CHAPTER 1

General Introduction and Methodology

Francis Crick and James Watson, walking along the Backs, Cambridge, England
Picture Source: https://paulingblog.wordpress.com/2009/04/30/the-watson-and-crick-structure-of-dna/

"I THINK THAT THE FORMATION OF DNA'S STRUCTURE BY WATSON AND CRICK MAY TURN OUT TO BE THE GREATEST DEVELOPMENTS IN THE FIELD OF MOLECULAR GENETICS IN RECENT YEARS."

— LINUS PAULING

1.1 An Overview of DNA:

1.1.1 History of DNA

DNA (deoxyribonucleic acid) stores biological and genetic information that is transmitted from one generation to other. DNA was first chemically investigated by the Swiss physician and biologist, Friedrich Miescher in 1869.[1] Before this discovery, very little was known about DNA and it was considered as uninteresting junk. Miescher isolated phosphorus containing microscopic substances from the pus of discarded surgical bandages. As it resided in the nuclei of cells, he called it "nuclein" and found that nuclein consist of an acidic and a basic portion. The acidic portion is the DNA and the remaining is the protein part. Many others including Miescher suspected that nuclein *i.e.* nucleic acid, was associated in some way with cell inheritance, however, its role and the structure was not elucidated with certainity until late 1940s.[2]

A series of experiments in the 1920s finally revealed that DNA is the genetic material. In 1928, Frederick Griffith discovered phenomenon of transformation in bacteria. Griffith concluded with the experiment that some chemicals were surviving heat treatment and retain genetic information.[3]

In 1944, Oswald T. Avery, Colin MacLeod and Maclyn McCarty for the first time reported that DNA is the bearer of genetic information.[4] From their study, Avery and coworkers discovered that the DNA extracted from the virulent (disease-causing) strain was responsible for carrying the inheritable genetic massage for virulency. The experiments conducted by Avery and colleagues were definitive, but many scientists were very reluctant to accept DNA as the genetic material rather presence of protein in the DNA extracts used, might be responsible for transformation. Later in 1952, Alfred D. Hershey and Martha Chase[5] provide independent evidence that DNA is the carrier of genetic information. They used radioactive phosphorus (^{32}P) and sulfur (^{35}S) tracers for their study and allowed the bacteriophage T2 containing radioactive traces to infect the host bacterial cell, *Escherichia coli.* They found that the phosphorus containing DNA of the viral coat entered into the host cell and furnished the genetic information for the viral replication. This experiment directly implies that DNA is the exclusive chromosomal component bearing the genetic information of the living cells.

At that time, though the function of this genetic material was well established but the structure was unknown. The most important evidence to DNA structure came from the work of Erwin Chargaff and coworkers in 1950.[6] They have collected different DNA samples from different animals, analyzed them and led to the following conclusions:

1. The base composition of DNA generally varies from one species to another.
2. DNA specimens isolated from different tissues of the same species have the same base composition.
3. The base composition of DNA in a given species does not change with an organism's age, nutritional state, or in different environment.
4. In all cellular DNAs, regardless of the species, the number of adenosine residues is equal to the number of thymidine residues (that is, A=T), and the number of guanosine residues is equal to the number of cytidine residues (G=C). From these relationships it follows that the sum of the purine residues equals the sum of the pyrimidine residues; that is, A+G=T+C.

These quantitative relationships are known as *'Chargaff's rules'*. These rules help in visualizing the three dimensional structure of DNA, how the genetic information is encoded in it and passed from one generation to other.[7]

The landmark in the DNA research was the discovery of its helical structure. Rosalind Frankllin, a young research associate in John Randalls laboratory at Kings College, London and Maurice Wilkins were the first scientists to discover the ribose-phosphate backbone of the DNA that lies outside the DNA strand. They used X-ray crystallographic technique to analyze DNA fibers and elucidate that DNA molecules were helical with two periodicities along their long axis, a primary one of 3.4 Å and a secondary one of 34 Å, in early 1950s.[1] The problem then was to formulate a three dimensional model of the DNA molecule which was solved, in 1953, by Watson and Crick. They realized that the relationship between the nitrogenous bases in DNA proposed in *Chargaff's rules* could be due to the complementary nature of adenine-thymine and guanine-cytosine bases and on this basis they discovered the hydrogen bonding between these base pairs, which is now known as Watson-Crick hydrogen bonding.[8] They postulated that DNA consists of two helical chains wound around the same axis to form a right handed double helix structure.

The phosphate backbone lied outside the helix and the bases held through hydrogen bonding are pointed towards the center of the helix.

1.1.2 Structure of DNA

DNA is a polymer of deoxyribose nucleotides. The nucleotides have three basic components *viz*, a nitrogenous base, a pentose ring and a phosphate, as shown in Figure 1.1 (a).[9] The molecules without the phosphate group are known as nucleoside. The nitrogenous bases are derived from two parent compounds, pyrimidine and purine Figure 1.1 (b).[10]

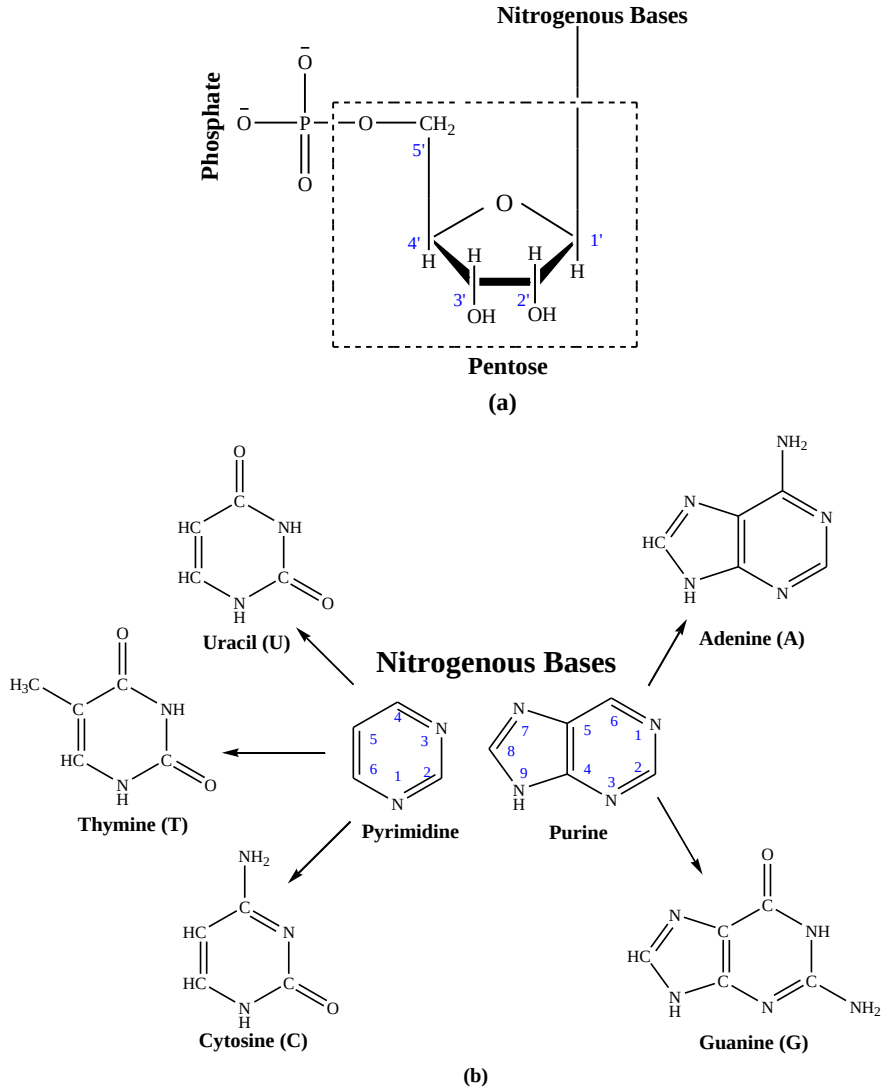


Figure 1.1 (a) Structure of deoxyribose nucleotides (b) The nitrogenous bases present in nature

Each DNA stand has a deoxy-ribose sugar backbond made up of linking phosphate units. These two stands run in opposite directions and are held together by hydrogen bonds between nucleotides of opposite stands resulting into a double DNA helical structure as shown in Figure 1.2. The base of a nucleotide is joined covalently (at N-1 of pyrimidines and N-9 of purines) in an *N*-*β*-glycosyl bond to the 1′ carbon of the pentose, and the phosphate is esterified to the 5′ carbon. The *N*-*β*-glycosyl bond is formed by removal of the elements of water (a hydroxyl group from the pentose and hydrogen from the base). The successive nucleotides of DNA are covalently linked through phosphate-group 'bridges', in which the 5-phosphate group of one nucleotide unit is joined to the 3′-hydroxyl group of the next nucleotide, creating a phosphodiester linkage. Thus, the covalent backbones of nucleic acids consist of alternating phosphate and pentose residues, and the nitrogenous bases may be regarded as side groups joined to the backbone at regular intervals. The backbones of DNA are hydrophilic. The hydroxyl groups of the sugar residues form hydrogen bonds with water. The phosphate groups, with a pK_a near zero, are completely ionized and negatively charged at pH equals to seven, and the negative charges are generally neutralized by ionic interactions with positive charges on proteins, metal ions, and polyamines. All the phosphodiester linkages have the same orientation along the chain, giving each linear nucleic acid strand a specific polarity and distinct 5′ and 3′ ends.

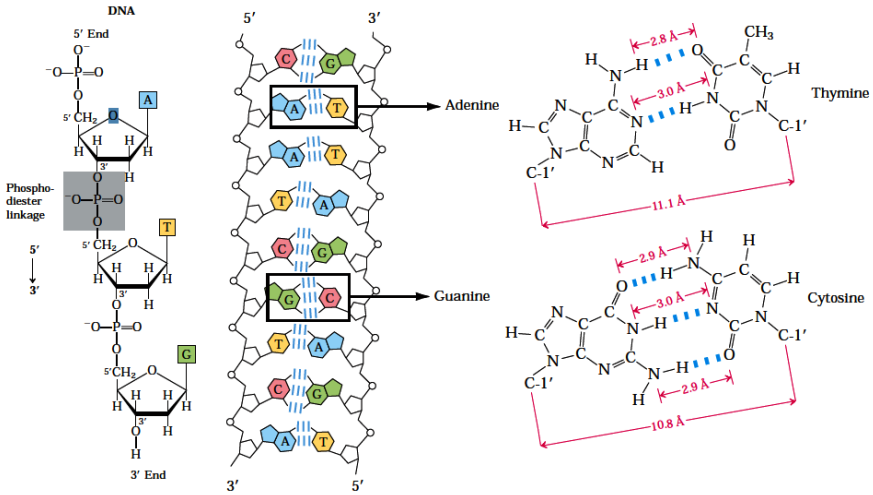


Figure 1.2 Hydrogen-bonding patterns in the base pairs defined by Watson and Crick (The picture is adapted from Reference 1)

DNA is consisting of two major purine bases, adenine (A) and guanine (G) and two major pyrimidines, cytosine (C) and thymine (T). The nucleic acid present in DNA is the 2′ deoxy-D-ribose and pentose sugar is the β-furanose (closed five-member ring). In Watson-Crick base pairing, the bases are hydrogen bonded with strict complimentary base pairs according to the *Chargaff rules*, described earlier.[6]

1.1.2.1 DNA is a double Helix

In 1953 Watson and Crick postulated a three dimensional model of DNA. It consists of two helical DNA chains wound around the same axis to form a right handed double helix. The hydrophilic backbones of alternating deoxyribose and phosphate groups are outside of the double helix, facing the surrounding water. Whereas, the furanose ring of each deoxyribose is in the C-2 endo conformation. The hydrophobic nature of purine and pyrimidine bases allow them to stack inside the DNA duplex very close together and perpendicular along the axis. The offset pairing of the two strands creates a major groove and minor groove on the surface of the duplex. Each nucleotide base of one strand is paired in the same plane with a base of the other strand. Watson and Crick further found that hydrogen bonding is possible only on the complementary bases *i.e.* A with T and G with C. It is significant to note that three hydrogen bonds is possible between G and C, symbolized G≡C, but only two can form between A and T, symbolized A=T. Thus, G-C interaction is stronger (by about 30%) than A-T, and A-T rich regions of DNA are more prone to thermal fluctuations.

When Watson and Crick constructed their three dimensional model, at the outset they had to decide whether 5′, 3′-phosphodiester bonds should run in the same or opposite directions *i.e.* whether the two stands of DNA should be parallel or antiparallel. They have realized that an antiparallel orientation produced the most convincing model which has been further established by experimental evidences (X-ray analysis).[5]

For this enormous contribution in recognizing the molecular structure of nucleic acid and its significance for information transfer in living material, the most prestigious Nobel Prize in Physiology or Medicine was awarded jointly to Francis Harry Compton Crick, James Dewey Watson and Maurice Hugh Frederick Wilkins in 1962.

1.1.2.2 DNA can occur in different three dimensional forms

Free rotation is possible in the phosphodeoxyribose (sugar phosphate) back bonds of DNA which give rise to a remarkable flexibility in the molecule. Moreover, thermal fluctuation can produce bending, stretching, and unpairing of the strands resulting into significant deviations from the Watson-Crick DNA structure in cellular DNAs. However, these structural variations generally do not affect the key properties of DNA defined by them *i.e.* stand complementary, antiparallel stands and base pairing.

In DNA, the structural variation is due to the three factors: the different possible conformations of the deoxyribose, rotation about the phosphodeoxyribose backbone and free rotation about the C-1´-N-glycosyl bond, as depiected in Figure 1.3 (a). Purine nucleotides are restricted to two stable conformations with respect to deoxyribose because of steric constraints, called *syn* and *anti*. However, the pyrimidines are generally *anti* because of steric interference between the sugar and the carbonyl oxygen at C-2 of the pyrimidine, as shown in Figure 1.3 (b).

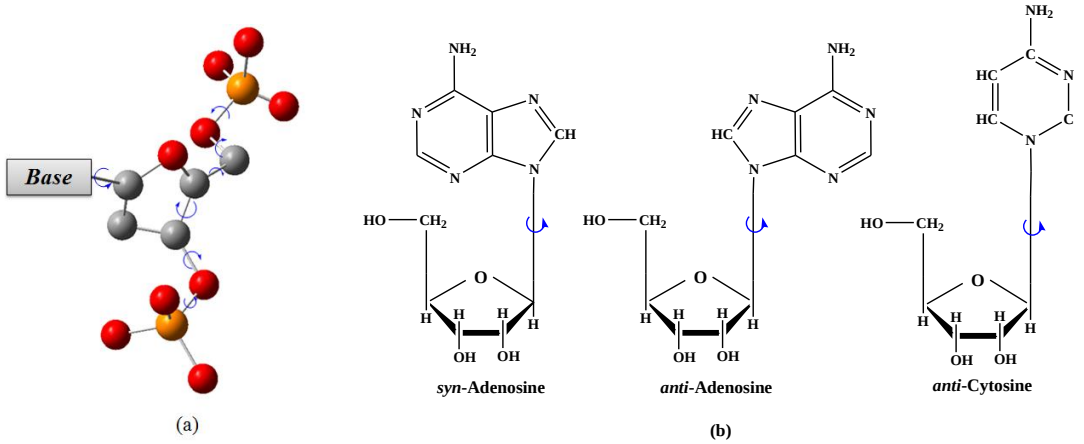


Figure 1.3 (a) The conformation of a nucleotide is affected by rotation about seven different bonds in DNA and (b) The two conformations with respect to the attached ribose for purine bases and pyrimidines occur only in the anti conformation

The model Watson-Crick has proposed is the B form DNA or B-DNA. This is the most stable structure for a random-sequence DNA molecule under physiological conditions and is therefore considered as standard point of reference in any study. There are also two different structural variants, the A and Z forms of DNA, Figure 1.4. The

DNA is still arranged in a right-handed double helix, but the helix is wider and the number of base pairs per helix may vary in each case. The basic differences between the three forms have been tabulated in Table 1.1.

Table 1.1 The variation in the three forms of DNA

Parameters	A-form	B-form	Z-form
Helical Sense	Right Handed	Right Handed	Right Handed
Diameter	≈26Å	≈20Å	≈18Å
Base pairs per helical turn	11	10.5	12
Helix rise per base pair	2.6Å	3.4Å	3.7Å
Base tilt normal to the helix axis	20˚	6˚	7˚
Sugar puncher conformation	C-3´ endo	C-2´ endo	C-2´ endo for pyrimidines; C-3´ endo for purines
Glycosyl bond conformation	Anti	Anti	Anti for pyrimidines; syn for purines

(The table is adopted form the Reference 1)

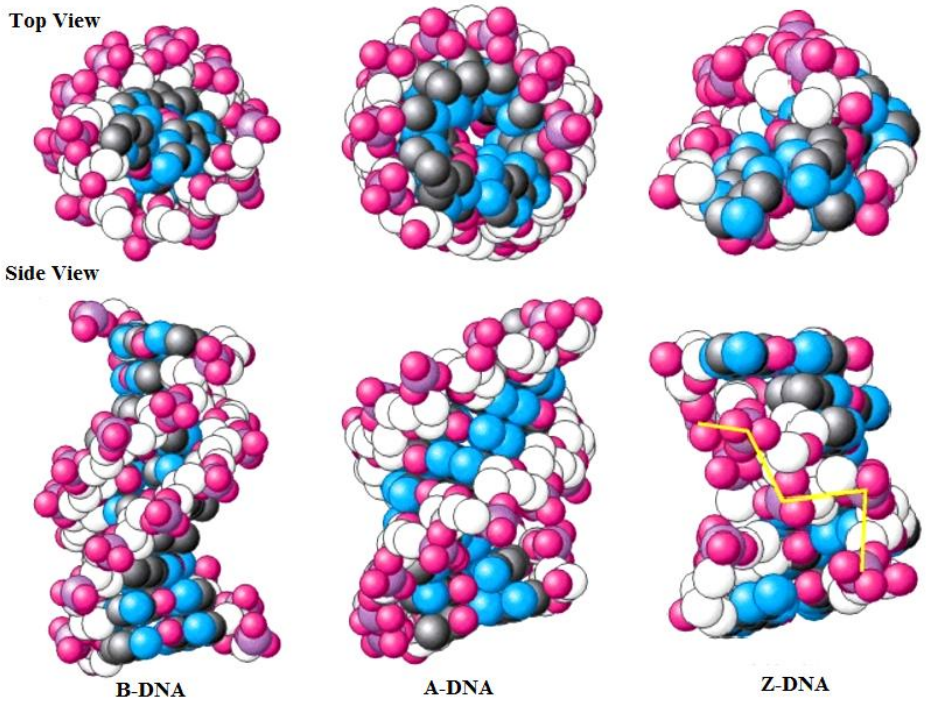


Figure 1.4 The comparision between the three forms of DNA

Picture source: http://www.ncbi.nlm.nih.gov/books/NBK22585/

1.1.3 DNA Damage

The replacement in a damaged protein or RNA molecule can be quickly done by using the information encoded in the DNA. However, the DNA molecule itself is irreplaceable. Damage in DNA molecule can occur via verity of processes, some are spontaneous and the others are catalyzed by the environmental agents. One should keep in mind that DNA is a chemical entity and hence it can participate in chemical reactions. The most serious outcome of this is the change in the base sequence, either replacement of one base pair with another or addition or deletion of one or more base pairs, known as unpaired DNA damage or a lesion.[11] This damage does not block replication machinery. The process through which mutation occurs is known as mutagenesis and is directly correlated with cancer. The mutations are majorly of two different types, spontaneous mutations and induced mutations.

1.1.3.1 Spontaneous Mutations

At molecular level, the spontaneous mutations[12] can be induced by

1. Tautomeric shift,
2. Replication mutation,
3. Spontaneous lesions and
4. Transposition.

1.1.3.1.1 Tautomeric Shift

These mutations arise due to the transient rearrangement of the DNA base pairs.[13] The tautomeric forms of the four common bases of DNA are shown below Figure 1.5:

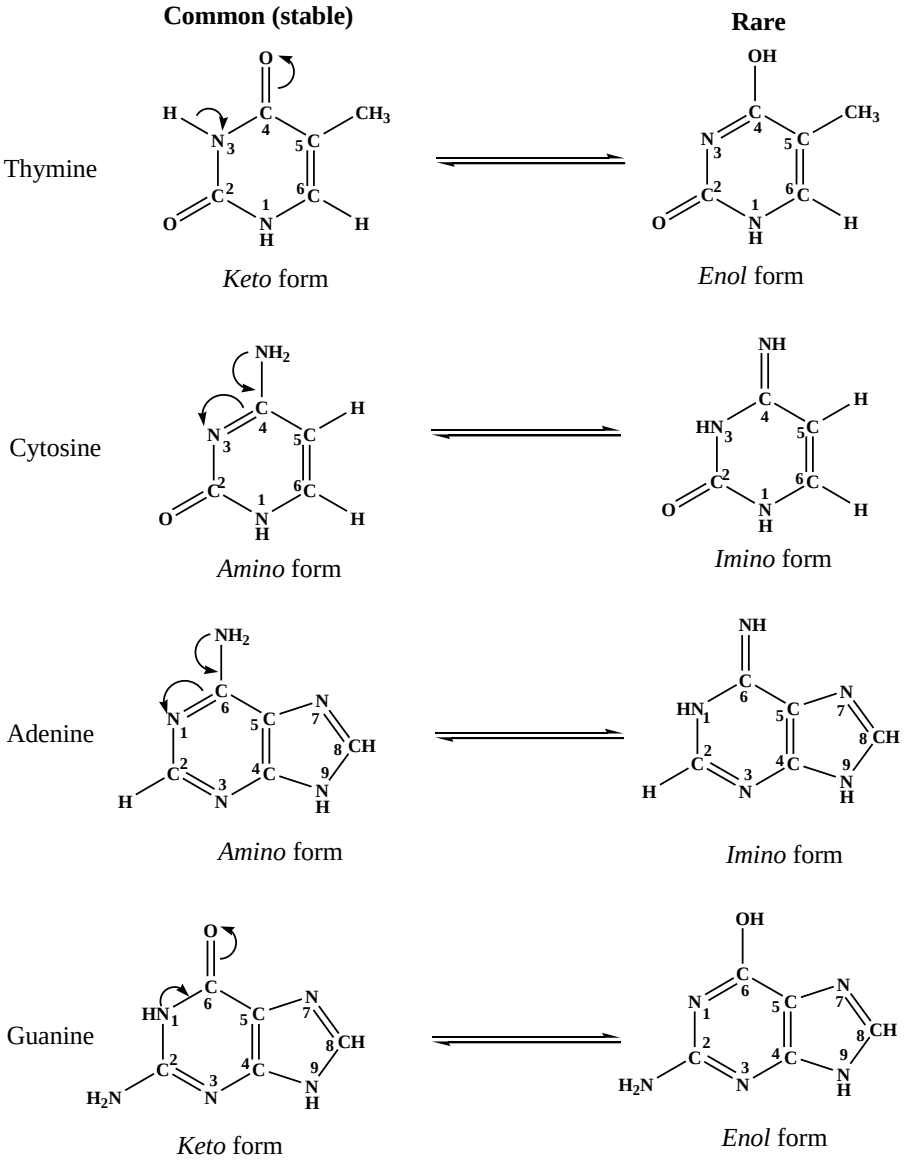


Figure 1.5 Tautomeric forms of the four common DNA bases

The more stable *keto* forms of thymine and guanine and *amino* forms of adenine and cytosine may undergo tautomeric shift (from 3 to 4 position in case of pyrimidines and 1 to 6 in case of purines) to less stable *enol* and *imino* forms, respectively. Though, the bases exist in its less stable structure for a very short period of time but if at that moment it was being replicated or being incorporate into a nascent DNA chain, it results in mutation, Figure 1.6.[16]

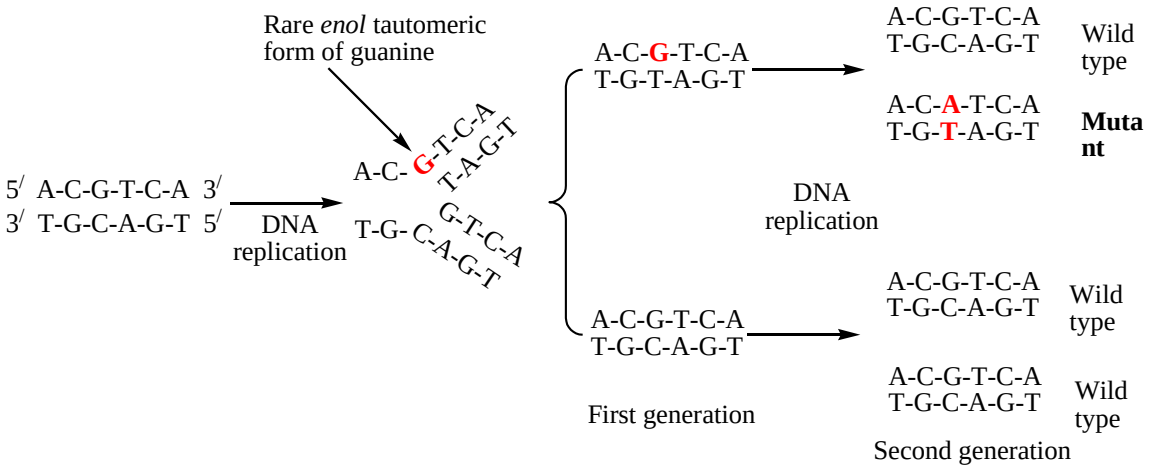


Figure 1.6 Mutation via tautomeric shift in the bases of DNA

1.1.3.1.2 Replication Mutation

They are of two types, either substitution mutation or frameshift mutation. The substitution mutation also known as point mutation or single site mutation, occurs due to the substitution of one base pair for another. It may be of transition mutation and transversion mutation. If one pyrimidine base is replaced by the other pyrimidine or one purine by another purine then it is known as transition mutation. On the other hand, if purine is replaced by pyrimidine or pyrimidine is replaced by a purine base then it is known as transversion.[14] The most common mutation under replication mutation is the transitions because an $A \rightarrow C$ and $G \rightarrow T$ mispairing does not necessarily destroy the DNA double helix as much transversion mispairing does.

Frameshift mutations[15] take place due to the insertions of extra nucleotides into the polynucleotide chain during replication or if some nucleotides are not being copied from

Term	Mutation			
Transition	G → A	A → G	C → T	T → C
Transversion	G → T	G → C	A → T	A → C
	T → A	T → G	C → A	C → G

the template. Addition or deletion of base pairs that occur within the protein codingportion resulting into a shift in the translational reading frame hampering the normal function of the protein. Since, this mutation causes a shift in the translational reading frame, shown in Figure 1.7, it is known as frameshift mutation.

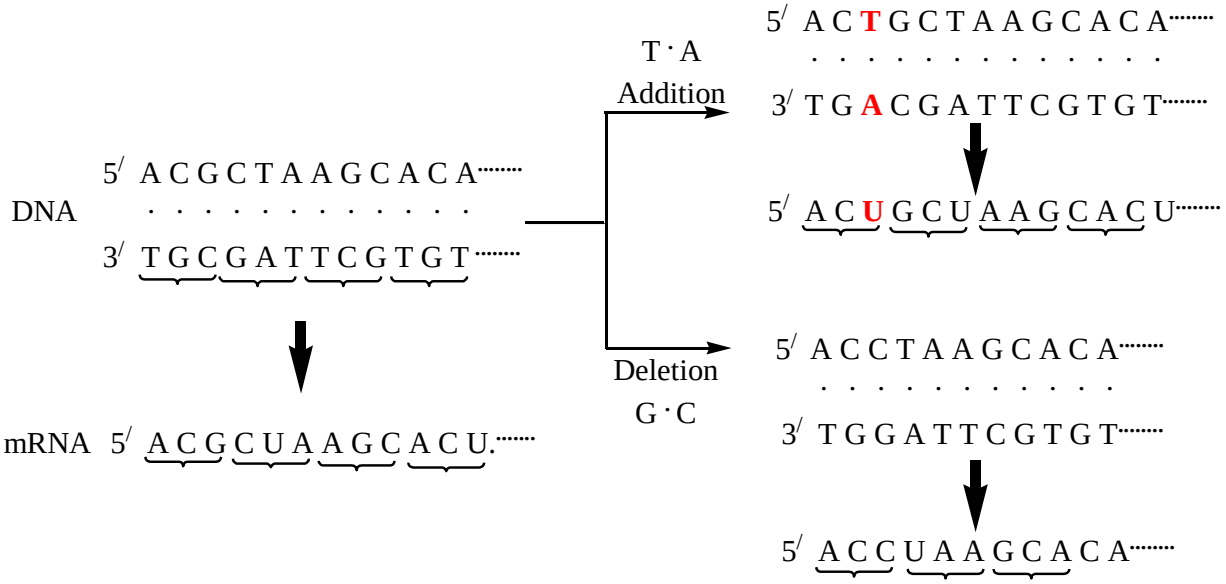


Figure 1.7 Frameshift mutation results from the addition or deletion of base pair

1.1.3.1.3 Spontaneous Lesions

This type of mutations occur due to the natural damages in DNA. The most appearing spontaneous lesions are depurination, depyrimidination, deamination and oxidative damage. As the name suggested, depurination and depyrimidination is the loss of purine or pyrimidine bases, respectively. They are more common in acidic conditions. Deamination refers to removal of an amino group from a molecular system. Deamination of cytosine results in the formation of uracil. Similarly, deamination of adenine and guanine results in hypoxantine and xanthine, respectively and 5-methylcytosine gives thymine (Figure 1.8).

Oxidative damage[16] is due to the influence of any reactive oxygen species which is going to interact with DNA and give rise to different by-products. Such reactive species are commonly peroxide radicals, hydrogen peroxide and hydroxyl radicals. For example, 8-oxo-7-hydrodeoxyguanosine (8-oxodG) and thymidine glycol are the products of oxidative damage. The 8-oxodG can mispairs with adenine and thymidine glycol inhibits DNA replication.

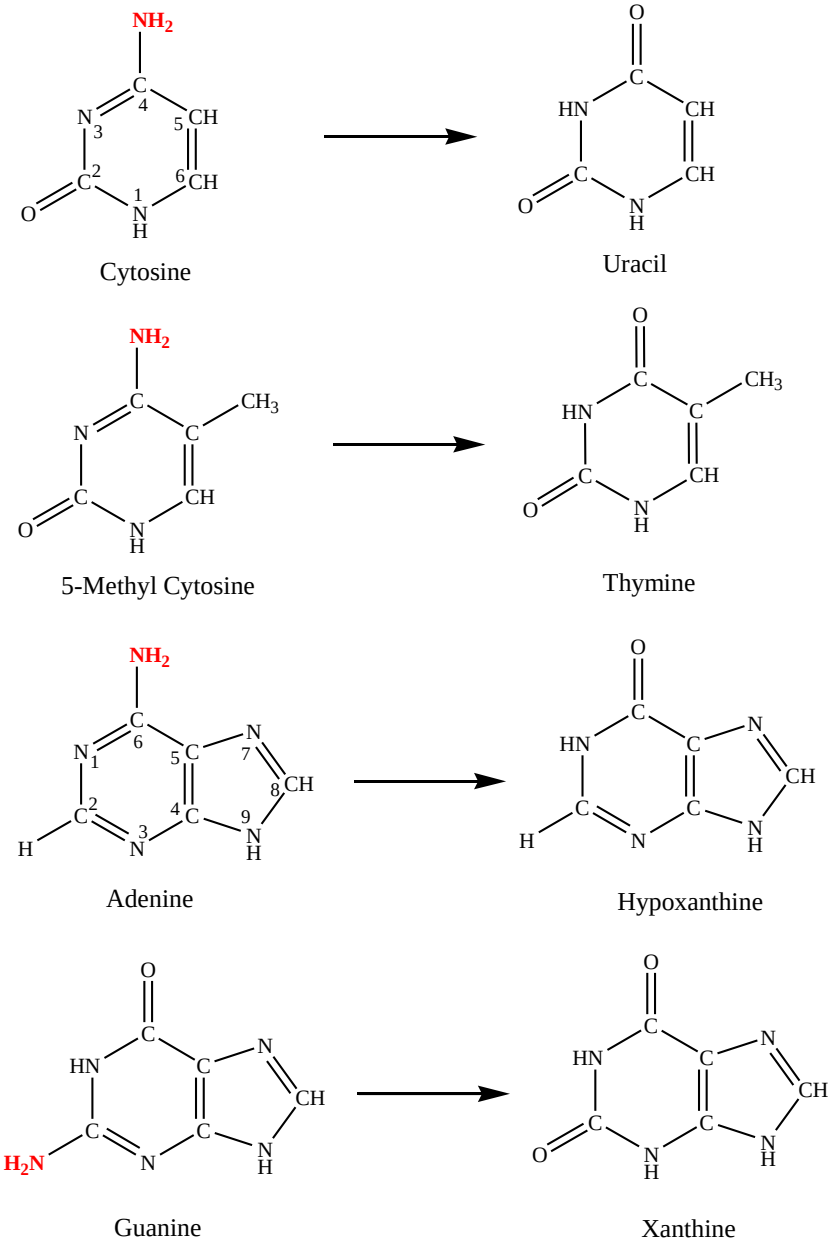


Figure 1.8 Deamination of DNA base pairs

1.1.3.1.4 Transposons

A transposon is a DNA sequence which is able to insert itself or a copy of itelf at a new location in the genome without having any sequence relationship with the target locus. Insertion of the transposable element into or near a functional gene can alter its expression.

1.1.3.2 Induced Mutation

This type of mutation at molecular level can be caused by chemicals or physical agents known as mutagens.[17] The chemical mutagens cause mutation in three different ways:

1. They can act as base analogs, which is structurally similar to one of the bases.
2. Some react directly causing structural changes in the nitrogenous bases.
3. Some act indirectly on DNA, they themselves do not affect the DNA structure but forcing the cells to synthesize some chemicals such as peroxides which cause mutations.

One of the examples of such mutation is hydroxylamine which causes GC-AT transitions. It preferentially hydroxylates the amino nitrogen of cytosine forming N-4-hydroxycytosine, which can mispair with adenine, as represented in Figure 1.9.

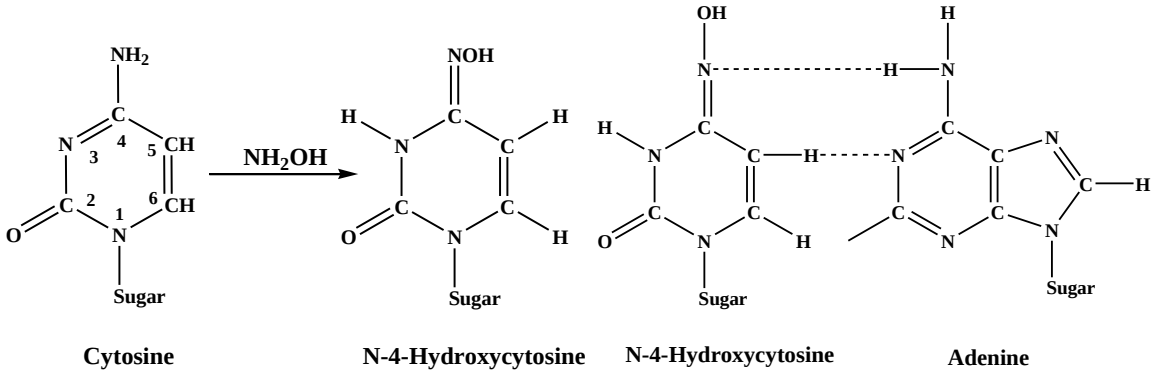


Figure 1.9 Hydroxylamine resulted from Chemical mutagens

The physical mutation includes ionizing (example, X-ray, β-ray) and non-ionizing (example, UV-rays) radiations. These types of radiation can either directly interact with DNA or it will form some reactive species which is responsible for DNA mutation.

UV-radiation is the potent physical mutagen and it will generate a number of photo-product in DNA. UV-radiation of 260 nm induces adjacent pyrimidine dimerization, especially if both are thymines, forming a cyclobutyl dimer. Another type of UV-induced photo-product is the cytosine hydrate. In this case, water is added across the 5, 6 double bond of cytosine to form the hydrated product, Figure 1.10.

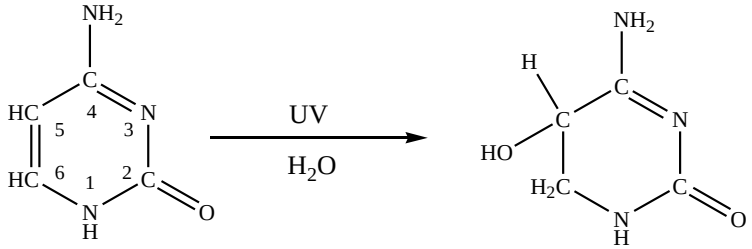


Figure 1.10 Mutations results from the UV-induced light

1.2 Cancer

Cancer (medical term: *malignant neoplasm*) is a class of diseases in which a group of cells display *uncontrolled growth* (division beyond the normal limits), *invasion* (intrusion on and destruction of adjacent tissues), and sometimes *metastasis* (spread to other locations in the body via lymphatic system or circulatory system). These three malignant properties of cancers differentiate them from benign tumors, which are self-limited, and do not invade or metastasize.[18] Most cancers form a tumor but some, like leukemia, do not. The branch of medicine concerned with the study, diagnosis, treatment, and prevention of cancer is *oncology*.

Cancer may affect people at all ages, even fetuses, but the risk for most varieties increases with age. Cancer is one of the leading causes of death worldwide. A report in 2012, found that an estimated 14.1 million new cases of cancer occurred worldwide and out of which around 8.2 million people died of cancer (Figure 1.11). More than half of cancer cases are in less developed regions, out of which six in ten cancer deaths worldwide occur in these regions.[19] The incidence of cancer is increasing rapidly in the developing countries. In India, cancer is one of the leading causes of mortality, which has affected nearly three million people. Annually, nearly 500,000 people die because of cancer in India.[20] According to WHO (World Health Organization) there will be 20 million new cancer cases by the end of 2025.[21,22]

Nearly all cancers are caused by abnormalities in the genetic material of the transformed cells. These abnormalities may be due to the effects of carcinogens, such as tobacco smoke, radiation, chemicals, or infectious agents. Other cancer-promoting genetic abnormalities may be randomly acquired through errors in DNA replication, or are inherited, and thus present in all cells from birth (congenital in nature).

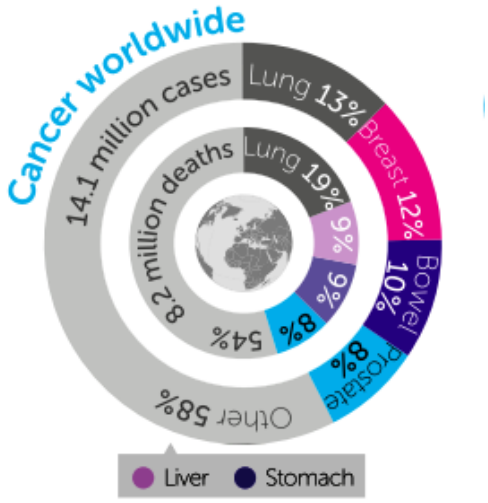


Figure 1.11 Distributation of Cancer Worlwide

(The picture is adapted from Reference 21)

Genetic abnormalities found in cancer typically affect two general classes of genes. Cancer-promoting oncogenes are typically activated in cancer cells, giving those cells new properties, such as hyperactive growth and division, protection against programmed cell death and/ or apoptosis, loss of respect for normal tissue boundaries, and the ability to become established in diverse tissue environments. Tumor suppressor genes are then inactivated in cancer cells, resulting in the loss of normal functions in those cells, such as accurate DNA replication, control over the cell cycle, orientation and adhesion within tissues, and interaction with protective cells of the immune system.

1.2.1 Classification of the Disease

Cancers are classified on the basis of their origin, histology and location. Examples of general categories include:

1. ***Carcinoma:*** Malignant tumors derived from epithelial cells (*i.e.*, tissue that lines organs and tubes). This group represents the most common type of cancers,

including the common forms of breast, prostate, lung and colon cancer. It accounts for 80 to 90 percent of all cancer cases. Carcinomas are divided into two major subtypes: adenocarcinoma, which develops in an organ or gland, and squamous cell carcinoma, which originates in the squamous epithelium.

2. ***Sarcoma:*** Malignant tumors derived from connective tissue or supportive tissue (e.g., bone, cartilage, muscle) or mesenchymal cells. Generally occurring in young adults, the most common sarcoma often develops as a painful mass on the bone. Examples of sarcomas are Osteosarcoma or osteogenic sarcoma (bone), Chondrosarcoma (cartilage), Leiomyosarcoma (smooth muscle), Rhabdomyosarcoma (skeletal muscle), Fibrosarcoma (fibrous tissue), Liposarcoma (adipose tissue) etc.
3. **Myeloma**: This group of cancer originates in the plasma cells of bone marrow. The plasma cells produce some of the proteins found in blood.
4. ***Lymphoma:*** Lymphomas develop in the glands or nodes of the lymphatic system, a network of vessels, nodes, and organs (specifically the spleen, tonsils, and thymus) that purify bodily fluids and produce infection-fighting white blood cells, or lymphocytes. They are considered as “solid tumors”
5. ***Leukemia:*** Leukemias ("liquid cancers" or "blood cancers") are cancers of the bone marrow (the site of blood cell production). Malignancies derived from hematopoietic (blood-forming) cells.
6. ***Germ Cell Tumor:*** Germ cell tumors are a varied group of benign and malignant neoplasms derived from primordial germ cells. They occur in a variety of sites, both gonadal (ovary and testis) and extragonadal (mediastinum, retroperitoneum, pineal gland and sacral area). Occurrence in young adults, mostly men and commonly found on midline location of body. They are classified as germinomatous or seminomatous and nongerminomatous or nonseminomatous germ cell tumors.
7. ***Blastic Tumor or Blastoma:*** A tumor (usually malignant) originates in embryonic tissue of organs. Many of these tumors are most common in children. Example of

this group of cancer are retinoblastoma (blastoma of the eye), nephroblastoma (blastoma of the kidney).

1.2.2 Causes of Cancer

Cancer is a diverse class of diseases which differ widely in their causes and biology. Anything which replicates (our cells) will probabilistically suffer from errors (in terms of mutations). Unless error correction and prevention is properly carried out, the errors will survive, and might be passed along to daughter cells. Normally, the body safeguards against cancer via numerous methods, such as apoptosis, helper molecules (some DNA polymerases), possibly senescence, etc. However, these error-correction methods often fail in small ways, especially in environments that make errors more likely to arise and propagate. For example, such environments can include the presence of disruptive substances called carcinogens, or periodic injury (physical, heat, etc.), or environments that cells did not evolve to withstand, such as hypoxia (see subsections). Cancer is thus a progressive disease, and these progressive errors slowly accumulate until a cell begins to act contrary to its function in the body.

1.2.3 Symptoms of Cancer

Symptoms of cancer metastasis depend on the location of the tumor. Roughly, cancer symptoms can be divided into three groups:

1. ***Local symptoms:*** unusual lumps or swelling (tumor), hemorrhage (bleeding), pain and/or ulceration. Compression of surrounding tissues may cause symptoms such as jaundice (yellowing the eyes and skin).
2. ***Symptoms of metastasis (spreading):*** enlarged lymph nodes, cough and hemoptysis, hepatomegaly (enlarged liver), bone pain, fracture of affected bones and neurological symptoms. Although advanced cancer may cause pain, it is often not the first symptom.
3. ***Systemic symptoms:*** weight loss, poor appetite, fatigue and cachexia (wasting), excessive sweating (night sweats), anemia and specific paraneoplastic phenomena, *i.e.* specific conditions that are due to an active cancer, such as thrombosis or hormonal changes.

1.2.4 Treatment of Cancer

There are several approaches for the treatment of the cancer available, such as surgery, chemotherapy, radiation therapy and/or combination of the treatment. The nature of the treatment is highly variable and dependent on a number of factors such as cancer type, its size, location and stage of the disease and the health condition of the patient.

If in case, the cancer is detected at an early stage, surgery is the most common and oldest technique to remove the tumor or cancerous cells. However, it is effective only if the tumor size is small and the affected area is reasonably well defined/localized. The treatment of radiation therapy involve the destruction of cancer cells by means of X-ray, laser beam or other radiopharmaceuticals. Surgery as well as radiation therapy is only effective to remove primary tumors but not in metastatic stage as these treatments are local.[23] Advanced cancer stages whose cell have already undergoes many mutations develop metastasis, accounts for more than 90% mortalities worldwide. Hence, at this stage advanced treatment such as chemotherapy is needed. Chemotherapy is very important in treating the blood cancer patients because the cancer cells are dispersed all over the body by means of blood or lymph and cannot be removed neither by surgery nor diminished by radiation. It uses drug molecules to treat the cancer cells in a non specific way. These drugs either taken orally or by injection (intravenous/intraperitoneal) are distributed throughout the body and selectively kill the cancer cells. Originally, nitrogen mustards were used as chemotherapeutic agents during the First World War time due to its anti-leukemic properties. Since then a number of chemotherapeutic agents have been developed and undergoing clinical trials.[24] However, the major problem associated with the chemotherapy is lack of tolerability in some patients and their enormous side effects including severe toxicity that damage vital organs such as kidney, liver etc.

1.3 Discovery and Use of Platinum Based Compounds

Cisplatin (cis-diamiminedichloroplatinum(II), CDDP), a square-planner platinum (II) was first chemically discovered in the year 1845[25], however its anticancer activity has its origin in 1960s, with the serendipitous discovery by Rosenberg of the inhibition of cell division by Pt complexes.[26] During the investigation of the effect of electric filed on the growth of the E.coli bacteria Rosenberg noticed that cells stopped dividing and displayed

strong filamentous growth. Later, it has been found that the cellular inhibition is due to the slow dissoprtion of the platinum into the ammonium chloride electrolyte forming complexes like *cis*-$[PtCl_2(NH_3)_2]$ and *cis*-$[PtCl_4(NH_3)_2]$.[27] From there onwards the interest towards the search of the anticancer platinum bases complexes are gradually increasing. The complexes having cis configuration are found to be more anticancer active. Cisplatin and its analogues such as carboplatin and oxaliplatin are extensively used anticancer drugs and is effective against a verity of tumors.[28,29] The success of platinum based drugs paved the way for the second to third generation platinum (II) based drugs such as carboplatin and oxaplatin while platinum (IV) complex satraplatin has recently undergone phase III trials and has been approved by the FDA, Figure 1.12. Sometimes some active agents *viz* Bleomycin, paclitaxel and 5-flurouracil have also been used along with platimun based combinatorial chemotherapy for improve outcome.[30] Platinum drugs continue to play a central role in the field of cancer chemotherapy. Due to the adverse side effects such as nephrotoxicity, neurotoxicity, and ototoxicity of the existing platinum drugs and due to the intrinsic and acquired resistance of these drugs, there is utmost need for a new platinum anticancer drug.[31] Therefore, many efforts have been directed towards the identification of compounds endowed with higher anticancer activity and lower toxicity than cis-DDP in the last thirty years.[32-38] Thousand platinum analogues have been synthesized for this purpose, but none of them have overcome the parent drug in efficacy, although, some of them display reduced toxicity and alternative modes of clinical delivery.[39]

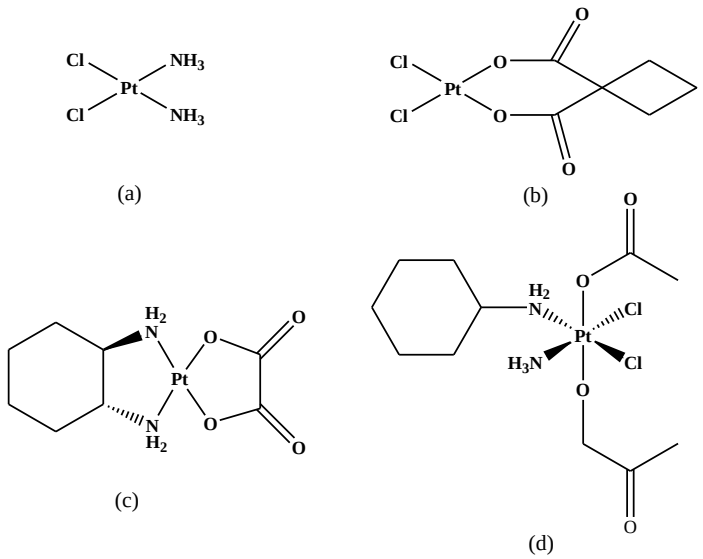


Figure 1.12 Molecular structure of cis platin analoges used in clinic (a) cisplatin, (b) carboplatin, (c) oxaliplatin and (d) satraplatin

1.3.1 Structure Activity Relationship (SAR) of Platinum Anticancer Agents

The anticancer activity of the platinum based drugs is determined by the ligand exchanged kinetics hence, governed by the nature of the ligand coordinating to the central platinum atom. Complexes containing either strong leaving groups or loosely coordinating groups are found to be inactive. It may be due to either very low or high reactivity of the complexes under physiological condition.[40,41] In addition, the trans effect of the ligand coordinating to the central platinum atom is also very important.

Moreover, the nature of the non-leaving groups are also responsible the reactivity of platinum complexes. For drugs of general formula $[PtCl_2(amine)_2]$ have found to be anticancer active. But the activity of the complexes decreases along the order $NH_3>RNH_2>R_2NH>R_3N$ (R is an alkyl substituent).[42] In addition the steric hindrance as well as the hydrogen bonding ability of the ligands present in the complexes are also found to be an important factor in determining the reactivity of the system.[43] The amino group will stabilize the Pt-Guanine bonding by acting as a hydrogen bond donor to the oxygen atom of the guanine and 5' phosphate group of the DNA. These interaction are very much crucial with respect to both kinetic (driving the platinum complex N7 position of guanine) as well as in thermodynamic point of view (by stabilizing the Pt-d(pGpG)

adduct).[44] All these observations resulted in a list of requirements for the structure of platinum complexes exhibiting anticancer activity, the so-called Structure Activity Relationships (SARs) of platinum complexes:

1. A cis geometry is required with the general formula *cis*-[$PtX_2(amine)_2$] for Pt(II), and for Pt(IV) the formula *cis*-[PtX_2Y_2(amine)2]. Monofunctional binding cationic complexes are inactive.
2. The X ligands (leaving groups) should be of intermediate strength (Cl^-, SO_4^{2-}, carboxylate ligands). For Pt (IV) complexes the Y ligands should have a trans orientation and can be Cl^-, OH^-, or $[O(CO)C_nH2_{n+1}]^-$.
3. The non-leaving group amine ligands should contain at least one NH moiety, necessary for hydrogen-bonding interactions with DNA (H-bonding to the O6 of guanine and to the 5´ phosphate group).

1.3.2 New Platinum Anticancer Agents/Trans-Platinum(II) Complexes

The design and synthesis of more efficient and less toxic platinum chemotherpetic agents constitutes a broad area of research with the aim of overcoming the limitations associated with the cisplatin.[45] Toxicity of the compounds can be decreased by exchanging the labile chloro ligands with comparatively more stable bidentate leaving groups thereby slowing down the hydrolytic activation of the drugs.[46] Cleare and Hoeschele demonstrated that the diaqua form of cisplatin is highly toxic and that toxicity decreases by decreasing the ability of the leaving ligand. More stable mutidentate leaving ligands can still give rise to similar activity profile and the activity only lost in the case of very strong coordinating ligands.[47] In recent years, an increasing trend has been observed describing the platinum complexes violating the previously established structure activity relationship (SARs) (Section 1.3.1). These complexes are known as "non-classical" platinum complexes.

It has been noticed that there are certain transplatin complexes which are showing higher antitumor activity both *in vitro* as well as *in vivo*.[48,49] Farrel et al. first reported that there are certain transplatin complexes having planner amino ligands showing a superior activity to cisplatin, especially in case of cisplatin resistant cell lines. The activity of such complexes could be increased by using bulky carrier ligand which reduces the rate of

replacement of the chloro ligands.[32] The bulky ligand would limit the axial access of the Pt-atom thereby inhibit the formation of five coordinated intermediate.[50] Examples of these antitumor trans-platinum complexes are the analogues of trans-platin in which one ammine group is replaced by ligands such as thiazole, piperidine, piperazine, 4-picoline and cyclohexylamine; the analogs with branched asymmetric aliphatic amines.

1.4 Mode of Action of Platinum Based Drugs

1.4.1 Hydrolysis Mechanism

Similar to that of the *cis*-platinum complex, hydrolysis is the key step for the intermolecular activation and anticancer mechanism of action of these types of complexes (Figure 1.13).

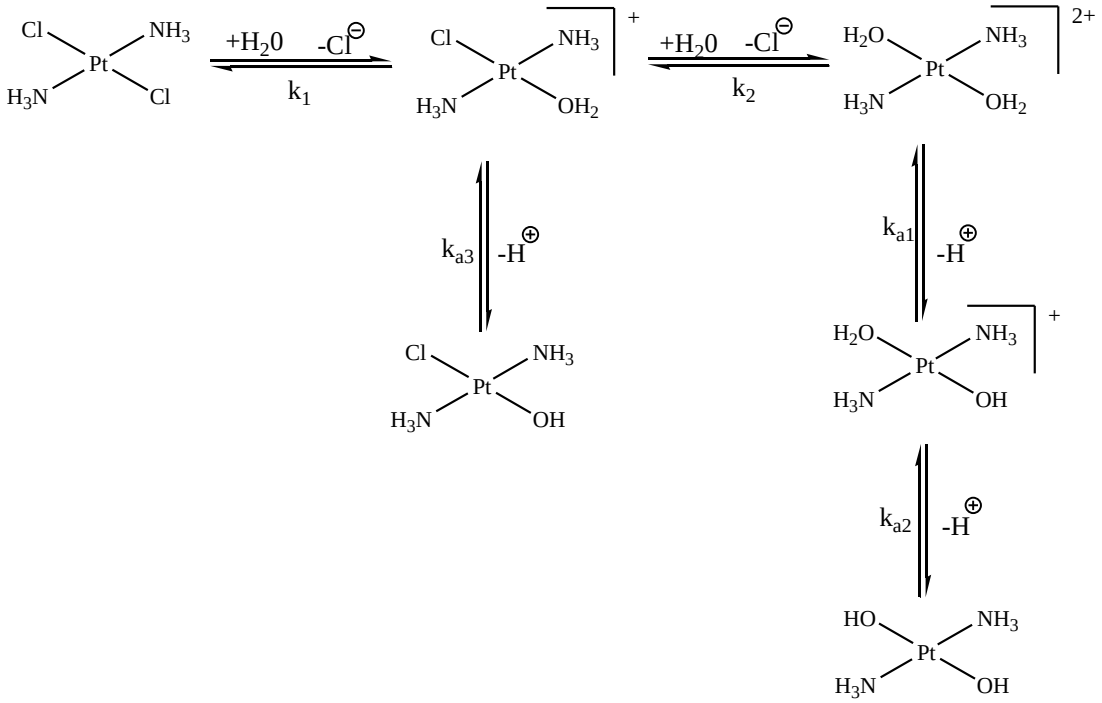


Figure 1.13 Hydrolysis of platinum based drugs under physiological condition

Neutral trans-platinum compounds need to be hydrolyzed to its monoaqua or diaqua species to react with its cellular target DNA. It has been observed that the abstraction of the first chloride from the trans-complex is relatively easy, this is because of the mutual trans effect of Cl^{-}.[51] However, removal of the second chloride ion is somewhat more difficult as it is positioned in trans to the weaker H_2O ligand, after

abstraction of Cl^-. For this, the monoaqua species are more likely to exit compared to its diaqua form under physiological condition.[52]

It has been observed that the equilibrium constant for the formation of the monoaqua specie for the trans complexes are also approximately an order of magnitude lowered compared to that of cisplatin complexes. The limited extent of hydrolysis implies that the reaction with nucleo-bases will be slower at intracellular chloride concentrations.[53] It has been also noted that the first pK_a of the diaqua species is uniformly lower than for the cis-isomer by an order of magnitude. The nature of the amino group attached to the central Pt-atom also have some impact on the pK_a values. In presence of planner amino ligands, the water molecule trans to it will be more acidic than NH_3, hence can be easily hydrolyzed. However, in the diaqua species, there exists an equilibrium between the aqua and hydroxo species, which allow it to react with the N-donor ligands such as N7 position of guanine.[54]

1.4.2 DNA Binding

It has been well established that the hydrolysis is the key activation step before reacting with its intercellular target DNA inside the cell.[55-57] Thus, the theoretical and experimental insight of the hydrolysis of the platinum based complexes has become a topic of interest in the current literature.[58-67] Platinum belongs to the B-type metal hence in ionic form it will preferentially react with N atom rather than O atom. In cellular DNA, under physiological pH condition, the N3 atom of thymine is protonated, the N3 positions of the purines are hindered sterically and other N atoms in the aromatic system without a σ lone pair are excluded for platinum coordination. The N1 of adenine and the N3 atom of cytosine are suitable positions for platinum binding.[68] All the potential binding positions of the DNA has been shown in the Figure 1.14 with respect to the platinum complex. Form the Figure 1.14, it has been observed that seven positions can be platinated, but it will preferably bind to the N7 positions of the purine bases[69,70] *i.e.* guanine (G) and adenine (A), which shows a strong kinetic preference in the nuclear DNA.[71] The tendency of the platinum complexes for the N7 of the purines results from the strong basicity at this position, ability to form hydrogen bond between the amino

group of the complex with O6 atom of guanine and the accessibility for the platinum complex.[44]

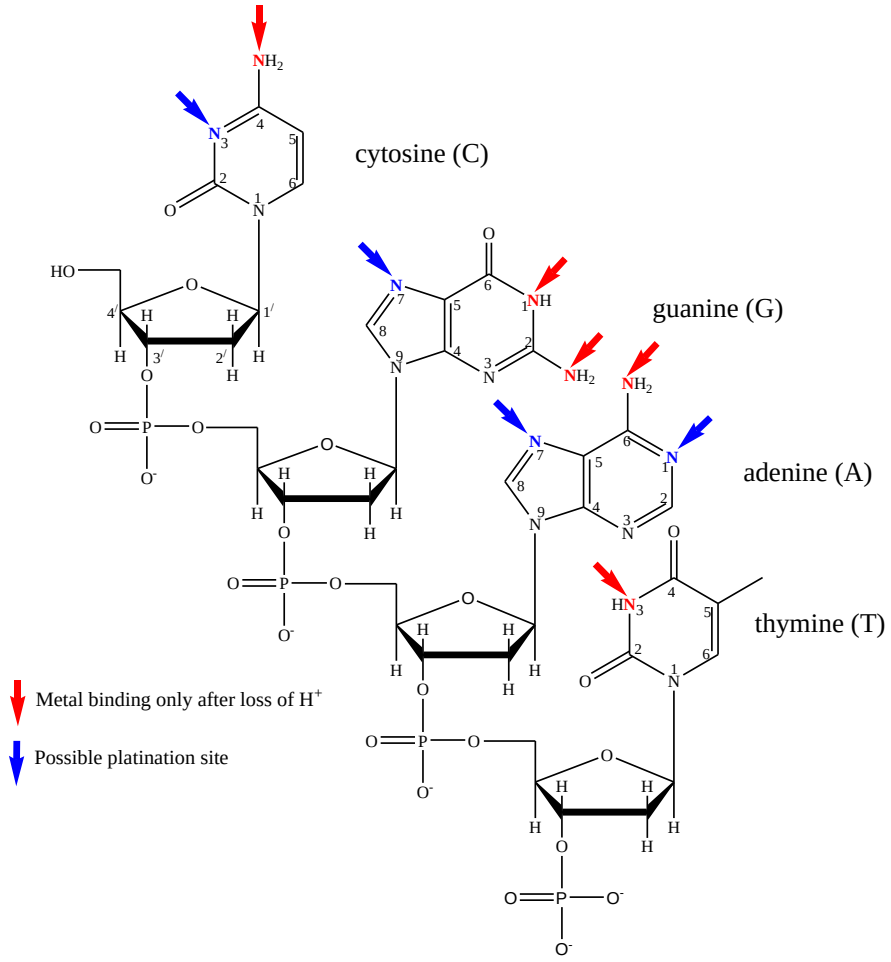


Figure 1.14 Possible binding site for the platinum based drugs on DNA

The reaction of platinum complexes with cellular DNA first produces monofunctional adduct, which subsequently coordinates to N7 atom of another purine base generating bifunctional adducts through intrastrand and interstrand cross-links (CLs), as shown in Figure 1.15. Cisplatin adducts in linear DNA include 90% 1,2-interstand d(GpG) cross-links on adjacent purine bases, followed by 1,2-intrastrand d(ApG) cross-links between an adjacent adenine and guanine, other lesions including 1,3-intrastrand d(GXG) and 1,4-intrastrand d(GXXG) cross-links between purines separated by one or two intervening bases.[72] Moreover, a small percentage of cisplatin also involved in interstrand cross-links, linking the two duplex strands of the DNA, or in monofunctional

adducts coordinated to a single purine and protein-DNA cross-link, in which cisplatin coordinates a protein molecule and a nucleobases.[73-78]

However, trans-platinum complexes cannot form such intrastrand cross-links due to steric hinderance of the two amino groups present in the complex. Instead it will only exhibit monofunctional adducts, 1,3-intrastrand CLs and ICLs between two G residues or between a G and a C residue, separated by at least one base. (Figure 1.15).[79] The amount of monofunctional adducts and protein-DNA cross-links formed by the transplatin complexes are much higher compared to its cis analoges. Another distict difference between the two is the interstand cross-links, the transplatin complex from interstrand CLs between complementary G and C residues, whereas cisplatin forms only G interstrand cross-links.[80] Thus, they will form a kinetically stable Pt-DNA adduct which is generally responsible for biologically activity of this class of compounds.[81]

These Pt-DNA adducts are considered to be responsible for the cytotoxic effect of these class of compounds. DNA containing Platinum adduct are distinctly different with respect to the normal B-type DNA, resulting in a loss of helix stability. But, the consequences of these CLs to cell or cell death are largely unknown. However, it has been believed that these CLs somehow perturb the cell cycle and inhibit in the G2-phase to allow DNA repair, and in the case of inadequate repair, the cells eventually undergo an abortive attempt at mitosis that results in cell death via an apoptotic mechanism.[82-84]

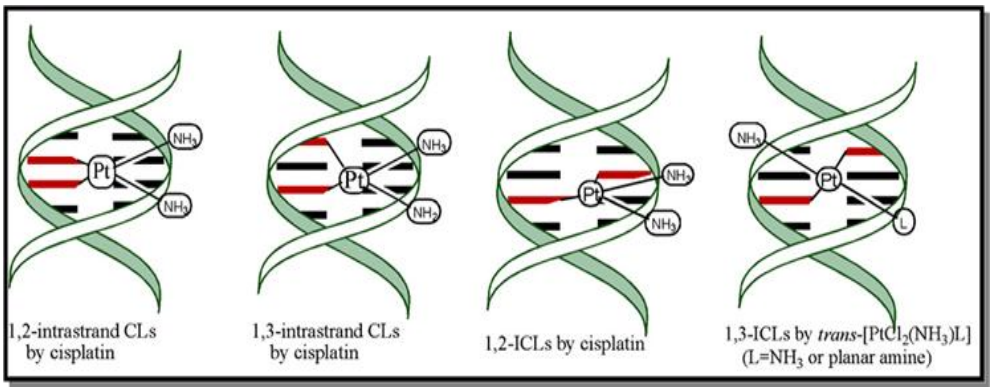


Figure 1.15 Schemetic representation of the modes of DNA-intra and inter stand CL by cisplatin and inter-stand CL by trans-$[PtCl_2(NH_3)L]$ ($L=NH_3$ or planner amino ligands) (The picture is adapted from Refference 82)

The first representative of this novel class of *trans*-platinum complexes is *trans*-[$PtCl_2$(dimethylamine)(isopropylamine)], shown in the Figure 1.16, which was initially characterized by elemental analysis, mass spectrum, and IR and NMR spectroscopic techniques. The compound is able to circumvent *cis*-DDP resistance in Pam212-*ras* tumor cells and induces cell death through apoptosis.[85]

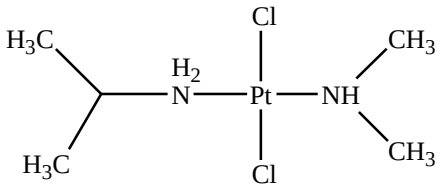


Figure 1.16 Structural formula of *trans*-[$PtCl_2$(dimethylamine)(isopropylamine)]

1.5 Gold anticancer agents

The potential anticancer activity of cis-platin and its platinum (II) analogues and their wide clinical success in current cancer treatments promoted a great deal of interest in the area of metal based antitumor agents. Among the several classes of metal based anticancer compounds such as ruthenium, palladium, titanium, cupper and so on, only a few investigations have focused on a variety of gold complexes. Gold compounds are being used over more than 70 years for the amelioration of the symptoms associated with the debilitating disease rheumatoid arthritis and the exploration of the anticancer potential of gold is a relatively recent aspect.[86] These complexes exhibiting anticancer properties similar to that of Pt (II) complexes as both Pt (II) and Au (III) are isoelectric in nature, having square planner coordination geometry i.e. isostructure. Investigation for anticancer potential of such gold complexes have been started in the late 1980's.[87]

However, unlike Pt (II) complexes, Au (III) complexes are very much photosensitive and easily reduce to metallic gold so they are quite unstable.[90] But during late 90's there was a revival of interest towards Au (III) based anticancer drugs derived from some novel compounds, endowed with improved stability and enhancing pharmacological activity.[88] These are the various organo-gold (III) complexes of this class, in order to increase the overall stability of Au (III) complexes multidentate ligands such as polyamide, cyclam, terpyridine and phenathroline were coordinated.[89] Their behavior have been investigated through various physico-chemical methods including

visible absorption spectroscopy, ESI mass spectrometry, and chloride selective potentiometric measurements and it has been found that they are quite stable which opened the way to extensive pharmacological testing *in vitro*.[90]

Gold (III) complexes are recognized as potent inhibitors of mitochondrial thioredoxin reductase.[91-93] The role of mitochondria in the mechanisms of cytotoxicity and the utilization of gold complexes in the antitumor action has been recently reviewed by McKeage et al.[94] Previously mitochondria are considered solely as powerhouse of cellular energy, but presently it is also known to play a key role in apoptosis or programmed cell death. The latter process, although essential for physiological functions, is also involved in several pathological conditions that span from degenerative diseases to cancer. Consequently, mitochondria are considered to be a more attractive general target for anticancer drugs.[95] Both gold(I) and gold(III) complexes are extremely efficient inhibitors of thioredoxin reductase showing IC_{50} ranging from 0.020 to 1.42 μM. The mitochondrial respiratory chain is seldomly affected by gold compounds while the other metal complexes exert their effect on different targets and thus exerted lower specificity. In particular zinc ion and zinc pyrithione, show a remarkable inhibitory effect that is associated with a rapid induction of membrane potential decrease that precedes swelling. Thus, other metal ions and metal complexes markedly inhibit the activity of thioredoxin reductase but to a lesser extent than that of gold complexes. It is concluded that gold compounds are highly specific inhibitors of mitochondrial thioredoxin reductase and this action influences other functions such as membrane permeability properties and ultimately lead to cell death.

1.6 Camptothecin

Camptothecin is the first topoisomerase I inhibitory drug isolated from *Camptotheca acuminata*[96,97] and is one of the prominent leading compounds in anticancer drug development. The discoveries of the naturally occurring compound 20-S camptothecin have lead to much interest in this type of pentacyclic system because of its marked activity in a number of leukemia and solid tumor systems.[98,99]

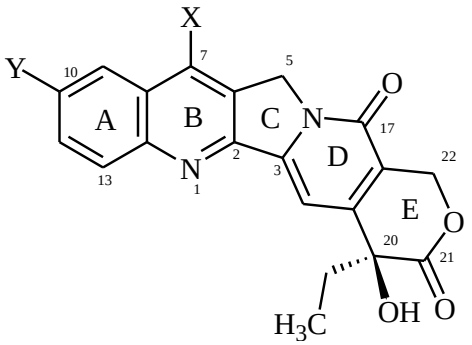
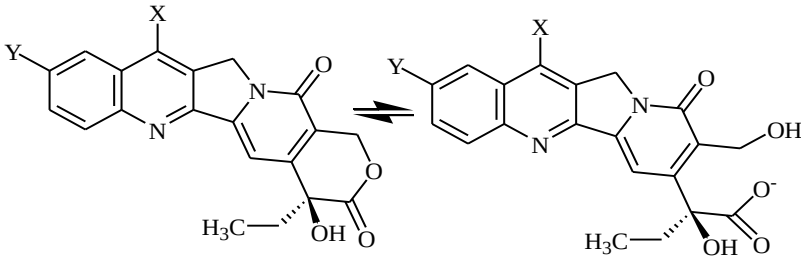


Figure 1.17 Structure of 20-S camptothecin

The compound (CPT) has a pentacyclic ring system with an asymmetrical center in ring E with 20-S configuration. The pentacyclic ring system includes a pyrrolo-quinoline moiety (rings A, B, and C), a conjugated pyridone (ring D), and a six-membered lactone (ring E) with an α-hydroxyl group as shown in Figure 1.17.[100,101] The camptothecin undergoes a reversible hydrolysis and there occurs a dynamic equilibrium between the close ring lactone and open ring carboxylic acid under physiological pH, Figure 1.18. The lactone form predominates at acidic pH whereas the inactive carboxylate form prevails at neutral and alkaline pH.



lactone form carboxylate form

Figure 1.18 Schematic of the interconversion between the lactone and carboxylic form of CPTs

Camptotecins are generally classified into two groups- the first group called water soluble camptothecins consists of Topotecan[102] and Irinotecan[103] (Figure 1.19.). Topotecan (Hycamptin) is currently used as a second line agent for the clinical treatment of the ovarian and small-cell lung cancers [104-107]and Irinotecan (Camptosar or CPT-11)[108,109] is presently used for the colon cancers.

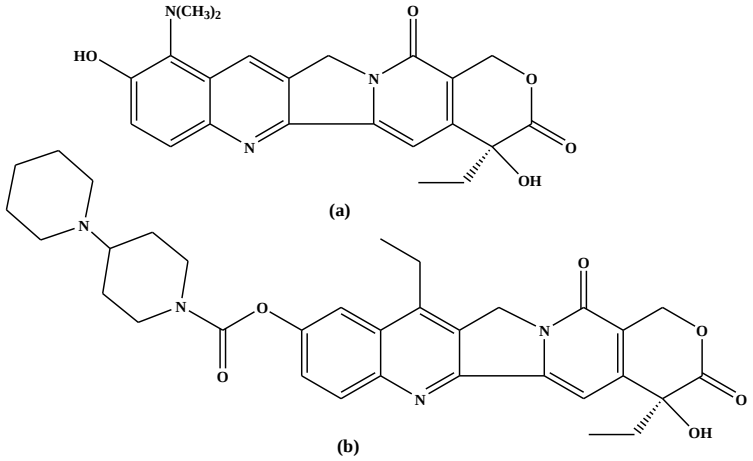


Figure 1.19 Structure of (a) Topotecan and (b) Irinotecan

The second group water insoluble, comprise the mother compound camptothecin[110] and synthetic derivative 9-nitro camptothecin, 9-amino camptothecin[111] which are in clinical trial, Figure 1.20. Natural camptothecin itself is insoluble in water. Therefore, camptothecin was evaluated clinically as the water soluble sodium carboxylate salt in the early stages. It results into the open lactone ring as stated earlier (Figure 1.18) which is actually only one by tenfold active as that of Camptothecin.

Camptothecin and some of its analogues are potent inhibitors of the enzyme DNA-topoisomerase I. In eukaryotic cell the DNA topoisomerase I (topo I) is an enzyme that acts to relax supercoils generated during transcription and DNA replication.[112] Because of the size of the eukaryotic chromosome, removal of these supercoils can only be accomplished locally by introducing breaks into the DNA helix. Topo I mediates DNA relaxation by creating a transient single-strand break in the DNA duplex. This transient nick allows the broken strand to rotate around its intact complement, effectively removing local supercoils, Figure 1.21 (a). Strand nicking results from the transesterification of an active-site tyrosine (Tyr-723) at a DNA phosphodiester bond forming a 3´-phosphotyrosine covalent enzyme-DNA complex. After DNA relaxation, the covalent intermediate is reversed when the released 5´ OH of the broken strand reattacks the phosphotyrosine intermediate in a second transesterification reaction, as shown in Figure 1.21 (b).[113]

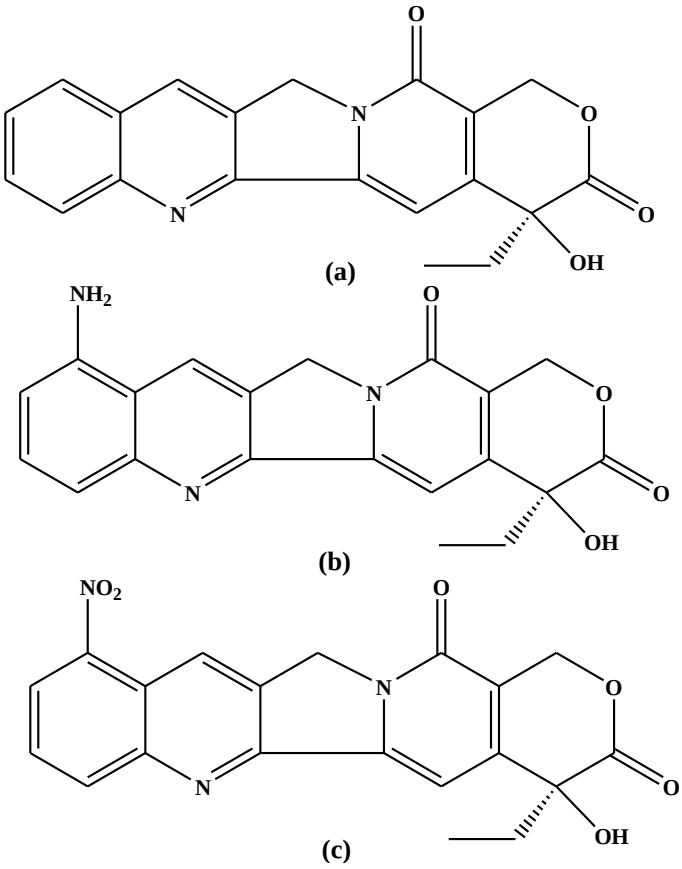


Figure 1.20 Structure of (a) Camptothecin, (b) 9-aminocamptotheci and (c) 9-nitrocamptothcin

CPTs bind the covalent 3´-phosphotyrosyl intermediate and specifically block DNA relegation[114], thus converting topo I into a DNA-damaging agent, Figure 1.21 (c).[115] Topo I is the sole intra-molecular target of CPT, and the cytotoxic effects of CPT poisoning are S-phase specific.[116] During DNA replication, the replication fork is thought to collide with the ‘‘trapped’’ topo I–DNA complexes, resulting in double-strand breaks and ultimately cell death.[117,118] It has also been suggested that topo I cleaves DNA at multiple sites. However, sites of cleavage stabilized by CPT exhibit a strong preference for guanine at +1 position, while thymidine remains the preferred nucleobase at the -1 position.[119] The exact mechanism by which CPT stabilizes the DNA topo I covalent binary complex is not fully understood because the drug acts as an uncompetitive inhibitor and binds only the transient binary complex.[120]

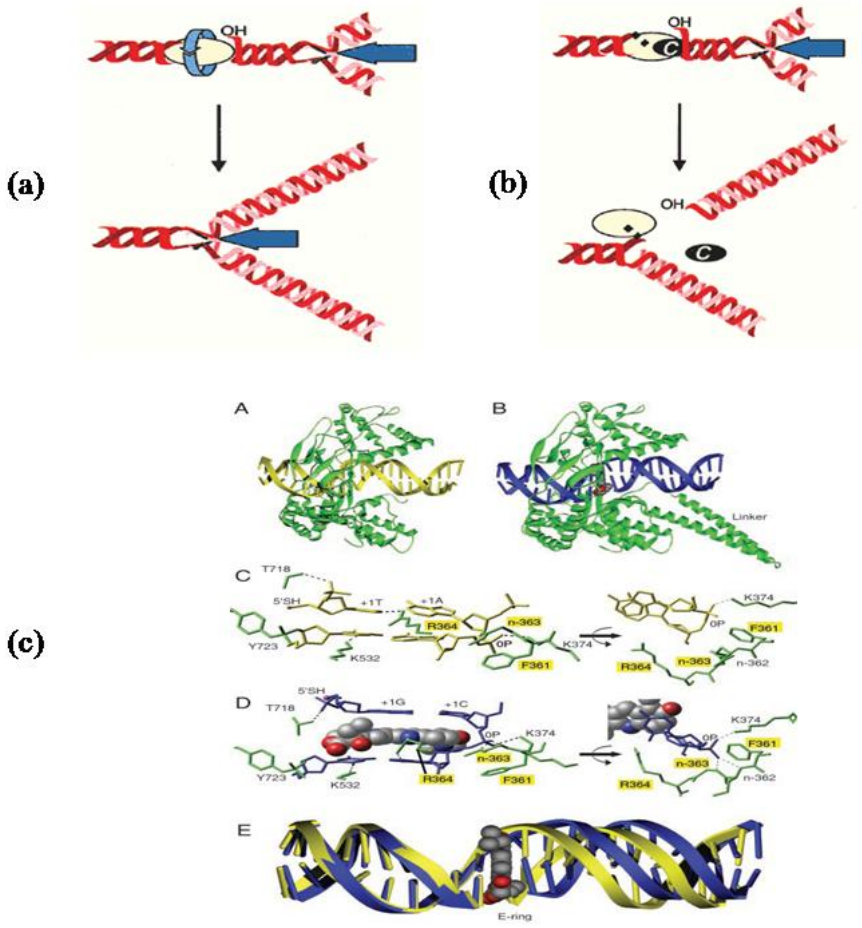


Figure 1.21 (a) Topoisomerases I introduces a nick in the DNA backbone allowing the rotation of one strand around another. (b) Camptothecin binds to the topoisomerase I-nicked DNA complex. This prevents the religation of the nicked strand and the release of the enzyme. (c) Structure of topo I-DNA complex without (*A*) and with (*B*) bound CPT. (*C*) Molecular diagram showing the nondrug-bound topo I–DNA complex. (*D*) CPTs intercalates between the +1 and -1 bases of the duplexDNA. (D) CPTs binds to the enzyme-substrate complex by intercalating in the DNA

1.7 Computational Chemistry Perspective

Computational chemistry is a bridge between the experimental method and theoretical models and defined as an implementation of chemistry, mathematics and computing expertise. It's a tool to investigate any chemical problem with the help of computer.[121] With the help of this method one can understand the basic concept of any

short of chemical problem which sometimes are unavailable directly from experimental support, say for example, the mechanism of any chemical reaction can be easily establish with the help of molecular modeling tools. However, the goals of chemistry are not changed by molecular modeling. It better be understood as a tool just like NMR or IR to achieve certain goals. Therefore, one should understand that molecular modeling forms a model of the real world and hence we study the model not the world and the model is valid as long as it reproduces the real world.[122]

Computational Chemistry helps us to investigate:

1. Determination of electronic structures
2. Geometry optimizations
3. Relative energies
4. Frequency calculations
5. Chemical Reactivity
6. Transition state and reaction pathways
7. Properties of molecules
8. Electron and charge distributions
9. Potential energy surfaces (PES)
10. Interaction of a substrate with host
11. Kinetic studies
12. Thermodynamic calculations
13. Calculations of many other molecular and bulk physical and chemical properties

Molecular mechanics, monte carlo, molecular dynamics, *Ab initio* method, semiempirical methods and density functional theory (DFT) are the most valuable tools in molecular modeling and understanding of these techniques is crucial to learn computational methods in chemistry. In this present thesis, we have employed DFT based quantum mechanics (QM) and hybrid quantum mechanics/molecular mechanics (QM/MM). In addition we also used Monte Carlo docking to find the active sites of bioactive molecules in protein and molecular mechanics based QSAR descriptors in determining structure activity relationship of various molecules.

1.8 Molecular Mechanics (MM)

Molecular mechanics employs a classical model of molecular structure comprising of sphere of different masses (atoms) held together by a variety of springs (bonds) of different lengths and stiffness, thus also referred as ball and spring model. It uses the parameters such as bond stretching and bending force constants and also interaction between non-bonded atoms to construct a potential energy expression that is a function of atomic positions.[123] If we know the normal spring lengths and the angles between them and how much energy it takes to stretch and bend the spring, we can calculate the energy of a given collection of balls and springs. The optimized geometry for a system can be thus obtained by simply changing the structure of the system. Here, the interactions between the atoms in a molecule are described by force field method. Force fields are assigned on the atoms depending on the atom's atomic number and the molecular environment. The various force fields available will vary with respect to three factors, the functional form of each energy term, cross terms included and the type of information used for fitting the parameters.[124-126] Some commonly used force fields include MM2, MM3, AMBER, CHARMM, CVFF, CFF91 and UFF.

The potential energy expression of a molecule in the force field can be written as:

$$E = \sum_{bonds} E_{stretch} + \sum_{angles} E_{bend} + \sum_{dihedrals} E_{torsion} + \sum_{pairs} E_{nonbond} \qquad (1.1)$$

in which,

$E_{stretch}$ = is the energy required to stretch or compressed a bond from its equilibrium bond length.

E_{bend} = is the energy required to bend a bond angle from equilibrium bond length.

$E_{torsion}$ = is the torsional energy due to twisting around bonds.

$E_{nonbond}$ = is the energy due to non-bonded interactions.

Molecular mechanics is very fast, the speed enables us to predict precise geometries and relative energies for a large bio-molecuar system such as DNA, protein, steriod molecules in a short span.

However, the weakness of this method stem from the fact that it ignores electrons, thus, unable to calculate electronic properties and processes which involves electronic effects such as bond breaking and bond formations. Also, MM cannot provide information about the shape and energies of molecular orbitals and cannot throw light on phenomena such as electronic spectra. Moreover, since the force field methods are zero dimensional, the probable error cannot be accessed within the method. It relies on experimental (or *ab-initio*) data for parameters.

1.9 Quantum Mechanics (QM)

The fundamental postulates of quantum mechanics proclaim that microscopic systems are described by wave functions that completely characterize all the physical properties of the system. There are numerous quantum mechanical operators corresponding to each observable physical properties, when applied to the wave function it will exhibit a particular eigen value or a range of values for that observable.

The electronic structure and total energy of a system can be obtained by solving Schrödinger eigen value equation. There are numerous representation of this equation, but the one concerned in the thesis is the time-independent, non-relativistic Schrödinger equation represented by:

$$\hat{H}\psi_i(\vec{x}_1,\vec{x}_2,.....,\vec{x}_N,\vec{R}_1,\vec{R}_2,.....\vec{R}_M) = E_i\psi_i(\vec{x}_1,\vec{x}_2,.....,\vec{x}_N,\vec{R}_1,\vec{R}_2,.....\vec{R}_M) \qquad (1.2)$$

where, $\hat{H}$ is the Hamiltonian operator for the molecular system consisting of M nuclei and N electrons in the absence of magnetic or electric field.

The Hamiltonian is a differential operator representing the total energy of the system and can be defined mathematically as:

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A=1}^{M}\frac{1}{M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \qquad (1.3)$$

Here, A and B run over M nuclei while i and j denote the N electron system. The first two terms describe the kinetic energy of the electron and nuclei, respectively and the Laplacian operator (∇_q^2) is defined as a sum of differential operators.

$$\nabla_q^2 = \frac{\partial^2}{\partial x_q^2} + \frac{\partial^2}{\partial y_q^2} + \frac{\partial^2}{\partial z_q^2} \tag{1.4}$$

and M_A is the mass of the nucleus. The remaining three terms define the potential part of the Hamiltonian and represent the attractive electrostatic interaction between the nuclei and electrons and the repulsive potential due to electron-electron and nucleus-nucleus interactions, respectively. The r_{pq} is the distance between the particles p and q, i.e. $r_{pq} = \left|\vec{r}_p - \vec{r}_q\right|$. $\psi_i(\vec{x}_1, \vec{x}_2, ..., \vec{x}_N, \vec{R}_1, \vec{R}_2, ...\vec{R}_M)$ stands for the wave function of the 'i' th state of the system, which depends on the 3N spatial coordinates and N spin coordinates of the electrons. The wave function ψ_i contains all the informations about the system and E_i is simply its eigen value described by the system ψ_i.

Schrödinger equation can be further simplified by considering the famous Born-Oppenheimer[127,128] or clamped-nuclei approximation. As there is a significant difference between the masses of the nuclei and the electrons, the electrons are considered to move in a field of fixed nuclei potential and hence the kinetic energy of the nuclei is considered to be zero. Thus, the complete Hamiltonian defined by equation (1.4) can be reduced to the so-called electronic Hamiltonian, defined as:

$$\hat{H}_{elec} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} = \hat{T} + \hat{V}_{Ne} + \hat{V}_{ee} \tag{1.5}$$

$$\hat{H}_{elec}\psi_{elec} = E_{elec}\psi_{elec} \tag{1.6}$$

The solution of the Schrödinger equation with $\hat{H}_{elec}$ is the electronic wave function (ψ_{elec}) and the electronic energy E_{elec}. ψ_{elec} depends on electron coordinates only, while the nuclear coordinates enter only parametrically and do not explicitly appear in ψ_{elec}. The total energy E_{total} is the sum of E_{elec} and the constant nuclear repulsion term.

$$E_{total} = E_{elec} + E_{nuc} \tag{1.7}$$

where, $E_{nuc} = \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}}$

1.9.1 Hartee-Fock Method (HF)

The HF theory is developed by Douglas Hartee (1928) and V. A. Fock (1930) to deal with the quantum many body problem.[129] An essential and unsolved problem in quantum mechanics is how to deal with indistinguishable, interacting particles (electrons) which determine the behavior of almost every object in nature. Since electrons are fermions, the solution of the electronic Schrödinger equation, must be antisymmetric with respect to permutation in any two electron system.[130] In Hartee-Fock method, an approximation to the many-electron wavefunction is done by considering the N-electron wavefunction as an anti-symmetric product of N one-electron wave functions. The product is referred to as the Slater determinant, ϕ_{SD} and has been used as an approximation to the exact wave function, ψ_{exact}.

$$\Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\vec{x}_1) & \chi_2(\vec{x}_1) & ..... & \chi_N(\vec{x}_1) \\ \chi_1(\vec{x}_2) & \chi_2(\vec{x}_2) & ..... & \chi_N(\vec{x}_2) \\ ..... & ..... & ..... & ..... \\ ..... & ..... & ..... & ..... \\ \chi_1(\vec{x}_N) & \chi_2(\vec{x}_N) & ..... & \chi_N(\vec{x}_N) \end{vmatrix} \tag{1.8}$$

Or, using the short-hand notation,

$$\Phi_{SD} = \frac{1}{\sqrt{N!}} \left| \chi_1(\vec{x}_1)\chi_2(\vec{x}_2).....\chi_N(\vec{x}_N) \right| \tag{1.9}$$

The one-electron functions $\chi_i(\vec{x}_i)$ are called spin orbitals which are composed of the spatial orbital $\phi_i(\vec{r})$ and one of the two spin functions $\alpha(s)$ or $\beta(s)$. Thus,

$$\chi_i(\vec{x}_i) = \phi_i(\vec{r})\sigma(s) \ , \tag{1.10}$$

Here, $\sigma = \alpha, \beta$

The physical interpretation of the spin orbitals is that, $\left|\chi(\vec{x})\right|^2 d\vec{x}$ represents the probability of finding the electron with spin σ within the volume element $d\vec{r}$. The pre-factor $\frac{1}{\sqrt{N!}}$ is given so that ϕ_{SD} fulfills the normalization condition. The anti-symmetry of the Slater determinant is confirmed form the fact that upon exchange of any of the two rows or columns, the determinant changes its sign. The variational principle[131] is taken

into account for the best Slater determinant which yields the lowest energy, keeping in mind the orthonormality of the spin orbitals.

$$E_{HF} = \min_{\Phi_{SD}\to N} E\left[\Phi_{SD}\right] \tag{1.11}$$

The final Hartee-Fock energy is given by the expression:

$$E_{HF} = \left\langle \Phi_{SD} \middle| H \middle| \Phi_{SD} \right\rangle = \sum_{i}^{N} \left(i \middle| \hat{h} \middle| i \right) + \frac{1}{2} \sum_{i}^{N} \sum_{j}^{N} \left(ii \mid jj \right) - \left(ij \mid ji \right) \tag{1.12}$$

where, $(i \mid \hat{h} \mid i) = \int \chi_i^*(\vec{x}_1) \left\{ -\frac{1}{2}\nabla^2 - \sum_{A}^{M} \frac{Z_A}{r_{1A}} \right\} \chi_i(\vec{x}_1) d\vec{x}_1$

Equation (1.12) gives the contribution to the kinetic energy and nucleus-electron attraction.

$$(ii \mid jj) = \iint \left|\chi_i(\vec{x}_1)\right|^2 \frac{1}{r_{12}} \left|\chi_j(\vec{x}_2)\right|^2 d\vec{x}_1 d\vec{x}_2 \tag{1.13}$$

$$(ij \mid ji) = \iint \chi_i(\vec{x}_1) \chi_j^*(\vec{x}_1) \frac{1}{r_{12}} \chi_j(\vec{x}_2) \chi_i^*(\vec{x}_2) d\vec{x}_1 d\vec{x}_2 \tag{1.14}$$

Equation (1.13) and (1.14) are known as *columb and exchange* intergrals, respectively and represents the interactions between two electrons. E_{HF} which is defined in equation (1.12) is a functional of spin orbitals.

$$E_{HF} = E\left[\{\chi_i\}\right] \tag{1.15}$$

The condition that the spin orbitals must remain orthonormal during energy minimization introduces the *Langrangian multipliers*, ε_i in the resulting Hartee-Fock equations which yield the best spin orbitals and attaining the lowest value of E_{HF}.

$$\hat{f}\chi_i = \varepsilon_i \chi_i \quad , \qquad (i = 1, 2, 3, \ldots\ldots, \mathrm{N}) \tag{1.16}$$

Lagrangian Multiplier, ε_i representing the orbital energy are the eigenvalues of the electron Fock operator $\hat{f}$.

$$\hat{f}_i = -\frac{1}{2}\nabla_i^2 - \sum_{A}^{M} \frac{Z_A}{r_{iA}} + V_{HF}(i) \tag{1.17}$$

The first two terms are the kinetic energy and the potential energy due to the electron-nucleus attraction and $V_{HF}(i)$ is the Hartee-Fock potential, defined as the

repulsive potential experienced by the "*i*" th electron due to the remaining (N-1) electrons.[132] The complicated two-electron repulsion operator $\frac{1}{r_{ij}}$ in the Hamiltonian is replaced and the electron-electron repulsion is taken into account in an average way. Thus, each electron is considered to be moving in the nuclear field and the average field of the other (N-1) electrons in Hartee-Fock theory. The two components of the $V_{HF}(i)$ are:

$$V_{HF}(\vec{x}_1) = \sum_{j}^{N} \left(\hat{J}_j(\vec{x}_1) - \hat{K}_j(\vec{x}_1) \right) \tag{1.18}$$

where, the columb operator is defined as:

$$\hat{J}_j(\vec{x}_1) = \int \left| \chi_j(\vec{x}_2) \right|^2 \frac{1}{r_{12}} d\vec{x}_2 \tag{1.19}$$

The columb operator represents the potential experienced by an electron at position $\vec{x}_1$ due to the average charge distribution of another electron in spin orbital, χ_j.

And, the exchange operator is:

$$\hat{K}_j(\vec{x}_1)\chi_i(\vec{x}_1) = \int \chi_j^*(\vec{x}_2) \frac{1}{r_{12}} \chi_i(\vec{x}_2) dx_2 \chi_j(\vec{x}_1) \tag{1.20}$$

Exchange operator refers to the exchange of electron within two spin orbitals. The exchange terms results from the anti-symmetry of the Slater determinant. The Slater determinant is an eigen function of a Hamiltonian operator defined as the sum of one-electron Fock operators:

$$\hat{H}_{HF}\Phi_{SD} = \hat{H}_{HF}^{0}\Phi_{SD} = \sum_{i}^{N} \hat{f}\Phi_{SD} = \sum_{i}^{N} \varepsilon_i \Phi_{SD} \tag{1.21}$$

The Slater determinant is thus the exact wave function of N non-interacting electrons moving in the field of the effective potential, V_{HF} and the Fock operator, $\hat{f}$ depends on the Hartee-Fock potential which in turn depends on spin orbitals.

1.9.1.1 The restricted and Unrestricted Hartee-Fock Methods

The restricted Hartee-Fock (RHF) method deals with closed shell system (even numbers of elctrons) with doubly occupied spatial orbitals i.e. the two spin orbitals, χ_p and χ_q are having the same value of energy and occupying the same spatial orbital

(ϕ_p)but with different α-and β- spin functions. On the other hand, Restricted Open Shell Hartee-Fock (ROHF) method is applicable for systems with unpaired electrons, occupying in the molecular orbitals in pairs as in RHF method, except for the unpaired electron (s). However, if the α- and β-spin electrons are allowed to occupy different spatial orbitals, this give rise to two sets of molecular orbitals one with the α and the other with an β electrons, the method is known as Unrestriced Hartee-Fock (UHF). In UHF, the effective potential experience by the α- and β-orbitals are different hence these orbitals differ in their spatial characteristics and also have different orbital energies.[133] The UHF scheme affords equations that are much simpler than their ROHF counterparts. The ROHF wave function is composed of a linear combination of few Slater determinants, where the expansion coefficient are determined by the symmetry of the state contrary to UHF method, which deals with a single-determinantal wave function.

1.9.2 Density functional theory

In the last few decades the exponential growth of density functional theory has been observed as it is a powerful quantum mechanical tool in studying the various chemical problems of interest.[134] The method is based on the pioneer work of Hohenberg-Kohn[135], followed by the Kohn-Sham[136] work, which led to the widespread use of DFT as a leading tool for calculating electronic structure. Figure 1.22 shows the increasing number of occurrence of DFT in journals form 1990 to 2012.

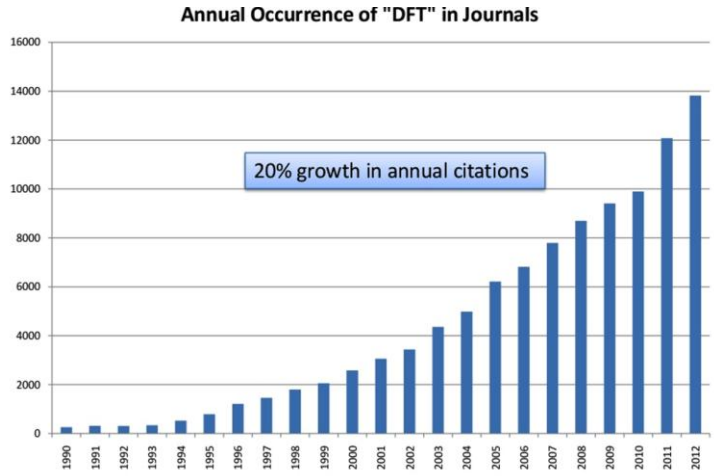


Figure 1.22 Annual occurrence of DFT in Journals Picture source: http://www.slideshare.net/Fitzgerald_G/cornell-computational-chemistry-seminar

DFT calculations are based on the Schrödinger equations, however, it does not calculate a wavefunction, but rather derives the electron distributions or electron density functionals directly.[137] According to this method, the ground state electronic energy of non-interacting electron system is determined completely by the electron density (ρ).

Mathematically,

$$\rho(\vec{r}) = N\int....\int\left|\Psi(\vec{x}_1\vec{x}_2\vec{x}_3.....\vec{x}_N\right|^2 ds_1 dx_2 dx_3.....dx_N \tag{1.22}$$

where, $\rho(\vec{r})$ determines the probability of finding any of the N electron within a volume element dr_1 with arbitrary spin (ds_1) while the other (N-1) electrons have arbitrary positions and arbitrary spin in the state represented by ψ .

Since, the electron density is a function of position only so three spatial variables is needed to describe an N electron system. However, the wave function is a function of 4N variables, three spatial and one spin variable for each electron. Thus, the complexity of the wave function increases with the increase of number of electrons while the electron density remains a function of only three variables. Apart from this, the electron density is an observable quantity and can measure experimentally unlike the wave function.

It is noteworthy that the 1998 Nobel Prize in chemistry was divided equally between Walter Kohn "for his development of the density-functional theory" and John A. Pople "for his role in developing practical wavefunction-based method", (Figure 1.23).[138]

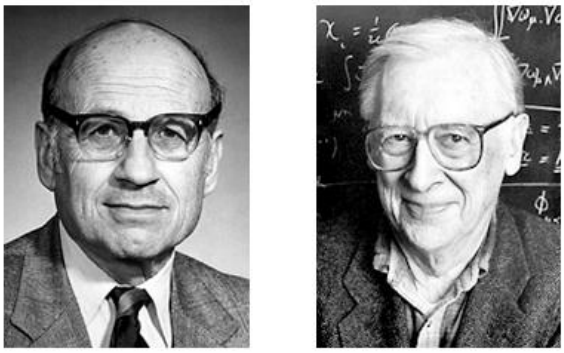


Figure 1. 23 Walter Kohn and John A. Pople received Nobel Prize in chemistry in 1998

Picture source: *http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1998/*

1.9.2.1 Hohenberg Kohn Theorem

In 1964, Hohenberg and Kohn[137] established this theorem which is considered to be the foundation of another theorem known as Kohn-Sham[138] approach, the main pillar for the DFT calculations. The significance of this theory is that it assures us that there is a principle way to calculate the molecular properties form the electron density itself.

The first Hohenberg-Kohn theorem[137], also known as the "*existence theorem*" states that all the properties of a molecule in a ground electronic state are determined by the ground state electron density function. In other words, if the ground state electron density [$\rho_0(x, y, z)$] is known we can in principle able to calculate any ground state property, say the ground state energy (E_0).

$$\rho_0(x, y, z) \rightarrow E_0 \tag{1.23}$$

So, it says that any ground state property of a molecule is a functional of ground state electron density, in this case it is the energy

$$E_0 = F[\rho_0] = E[\rho_0] \tag{1.24}$$

The main drawback of this theorem is that here, there is a functional F exists, but does not tell how to find it; hence, sometimes also known as existence theorem. The finding of the exact fucntional is the main problem with DFT. The second Hohenberg-Kohn theorem[137] says that any trial electron density function will give energy always higher than the true ground state energy. It will give the exact value of energy only if the trial function is exactly the true electron density function. This is in principle, the DFT analogue of the variation theorem (in context of ab initio method). The second theorem can thus be stated as

$$E_v[\rho_t] \geq E_0[\rho_0] \tag{1.25}$$

where, ρ_0 is the trial electron density, $E_0[\rho_0]$ is the ground state energy corresponding to the true electronic density, ρ_0 and $E_v[\rho_t]$ is the trial electronic energy under external or nuclear potential $v(r)$. However, the trial density must satisfy the condition $\int \rho_t(r)dr = n$, here n is the number of electrons in the system and $\rho_t(r) \geq 0$ for all *r*.

Thus the total electronic energy of a system can be expressed as:

$$E[\rho(r)] = T[\rho(r)] + E_{ee}[\rho(r)] + E_{Ne}[\rho(r)] \quad (1.26)$$

here, T is the kinetic energy of all the electron in the system, E_{ee} is the electron-electron interaction consists of the classical Coulomb term and non-classical terms and E_{Ne} is the potential energy due to the nucleus-electron attraction.

The potential energy due to the nucleus-electron attraction (E_{Ne}) can be expressed as

$$E_{Ne}[\rho(r)] = \int \rho(\vec{r}) V_{Ne} d\vec{r} \quad (1.27)$$

Thus, equation (1.27) can be further written as:

$$E[\rho(r)] = F_{HK}[\rho] + \int \rho(\vec{r}) V_{Ne} d\vec{r} \quad (1.28)$$

where, $F_{HK}[\rho] = T[\rho(r)] + E_{ee}[\rho(r)]$

Since, the functional forms of both the terms in $F_{HK}[\rho(r)]$ are unknown, therefore, the implementation of Hohenberg-Kohn theorem is not an easy task

1.9.2.2 The Kohn-Sham Theory

The key problem in DFT is that unfortunately we do not have the correct energy functional and certainly do not have the molecular electron density function. If we had the exact values of these two we could go directly from the electron density function to molecular energy. One year later, in 1965 Kohn and Sham put forwarded a new theory to alleviate these two problems.[138] This theory forms the foundation for the use of DFT in computational chemistry. The first Kohn-Sham theorem tells us that it is important to find a way to calculate molecular properties form the electron density. The second theorem proposed that a variational approach might yield a way to calculate the energy and electron density.

The two basic ideas behind the Kohn-Sham (KS) approach are:

1. To express the molecular energy as a sum of certain terms and among this one relatively small term containing an unknown functional. Thus, even moderately large errors in this term will not introduce large errors into the total energy.

2. To use an initial guess of the electron density in the KS equations to calculate an initial guess of the KS orbital and energy levels. This initial guess is then refined iteratively to get the orbitals and energy levels. The final KS orbitals are used to determine the electron density which is again utilized to calculate the energy of the system.

The Kohn Sham approach to DFT introduces a key idea to non-interacting reference system in which the ground state electron density distribution, ρ_r is exactly the same as that in the ground state system i.e. $\rho_r = \rho_0$. The ground state electronic energy of the molecular system is the sum of the electron kinetic energy, the nucleus-electron attraction potential energies and the electron-electron repulsion potential energies, defined by the equation:

$$E_0 = \langle T[\rho_0] \rangle + \langle V_{Ne} | \rho_0 | \rangle + \langle V_{ee} | \rho_0 | \rangle \tag{1.29}$$

The energy terms are expressed as the expectation values and each is a functional of the ground state electron density. Equation (1.30) can be written as:

$$E_0 = \langle T[\rho_0] \rangle + \int \rho_0(r) v(r) dr + \langle V_{ee} | \rho_0 | \rangle \tag{1.30}$$

Here, $v(r)$ is the external potential and the middle term in this equation is the classical electrostatic attraction potential energy. In order to solve equation (1.30), Kohn and Sham incorporated the idea of reference system of non-interacting electrons as the kinetic and potential energy functional terms $\langle T[\rho_0] \rangle$ and $\langle V_{ee} | \rho_0 | \rangle$ are unknown. The derivation of the real kinetic energy form the reference system can defined by $\Delta \langle T[\rho_0] \rangle$.

$$\Delta \langle T[\rho_0] \rangle \equiv \langle T[\rho_0] \rangle - \langle T[\rho_0] \rangle_{ref} \tag{1.31}$$

Similarly, the deviation of the real electron-electron repulsion energy from the classical electro-static repulsion energy is defined by $\Delta \langle V_{ee}[\rho_o] \rangle$.

$$\Delta \langle V_{ee}[\rho_o] \rangle = \langle V_{ee}[\rho_o] \rangle - \frac{1}{2} \iint \frac{\rho_0(r_1)\rho_0(r_2)}{r_{12}} dr_1 dr_2 \tag{1.32}$$

The last term in the equation (1.33) is the summation of all repulsion energies for a pair of infinitesimal volume elements $\rho_0(r_1)dr_1$ and $\rho_0(r_2)dr_2$. Thus, energy term is given by

$$E_0 = \int \rho_0(r)v(r)dr + \langle T[\rho_0]\rangle_{ref} + \frac{1}{2}\iint \frac{\rho_0(r_1)\rho_0(r_2)}{r_{12}}dr_1dr_2 + \Delta\langle T[\rho_0]\rangle + \Delta\langle V_{ee}[\rho_o]\rangle \quad (1.33)$$

Here, comes the concept of exchange-correlation energy (E_{XC}), the sum of the kinetic energy deviation from the reference system and electron-electron repulsion energy deviation from the classical system. E_{XC} can also be defined as the correlation to the kinetic energy term arising from the interacting nature of electrons and all non-classical correlations to the electron-electron repulsion energy.

$$E_{XC}[\rho_0] = \Delta\langle T[\rho_0]\rangle + \Delta\langle V_{ee}[\rho_o]\rangle \quad (1.34)$$

$\Delta\langle T\rangle$ and $\Delta\langle V_{ee}\rangle$ are the kinetic correlation energy and exchange energy, respectively. The correlation term refers to the interaction between electrons of opposite spin and the exchange term is connected with electrons of same spin. Using equation (1.34) in (1.33), gives

$$E_0 = \int \rho_0(r)v(r)dr + \langle T[\rho_0]\rangle_{ref} + \frac{1}{2}\iint \frac{\rho_0(r_1)\rho_0(r_2)}{r_{12}}dr_1dr_2 + E_{XC}[\rho_0] \quad (1.35)$$

More precisely, electronic energy of the ground state of a system comprising n electrons and N nuclei can be written as:

$$E_0 = -\sum_{X=1}^{N}\int \frac{Z_X}{r_{X_i}}\rho(r_1)dr_1 - \frac{1}{2}\sum_{i=1}^{n}\int \psi^*(r_1)\nabla_i^2\psi_i r_1 dr_1 + \frac{1}{2}\iint \frac{\rho_0(r_1)\rho_0(r_2)}{r_{12}}dr_1dr_2 + E_{XC}[\rho_0] \quad (1.36)$$

where, ψ_i (i =1, 2, 3,…., n) are the Kohn-Sham orbitals,

$$\int \rho_0(r)v(r)dr = -\sum_{X=1}^{N}\int \frac{Z_X}{r_{X_i}}\rho(r_1)dr_1$$

and

$$\langle T[\rho_0]\rangle_{ref} = -\frac{1}{2}\sum_{i=1}^{n}\int \psi^*(r_1)\nabla_i^2\psi_i r_1 dr_1$$

The biggest problem in DFT is to find the accurate exchange-correlation function (E_{XC}) part.

The ground state electron density can be written as a set of one-electron Kohn-Sham orbitals, mathematically,

$$\rho_0 = \sum_{i=1}^{n} |\psi_i(r)|^2 \tag{1.37}$$

The significance of the Kohn-Sham orbitals is that they allow us to calculate the density from the equation above. The Kohn-Sham orbitals are determined by solving the Kohn-Sham equation.

$$\hat{h}_i \psi_i(r_1) = \varepsilon_i \psi_i(r_1) \tag{1.38}$$

In equation (1.39), $\hat{h}_i$ represents the Kohn-Sham Hamiltonian and ε_i is the associated Kohn-Sham orbital energy. The Kohn-Sham Hamiltonian can be written as:

$$\hat{h}_i = -\frac{1}{2}\nabla_1^2 - \sum_{X=1}^{N} \frac{Z_X}{r_{X_i}} + \int \frac{\rho(r_2)}{r_{12}} dr_2 + V_{XC}(r_1) \tag{1.39}$$

V_{XC} in the above equation is the functional derivative of the exchange-correlation energy with respect to ρ and expressed as:

$$V_{XC}[\rho] = \frac{\partial E_{XC}[\rho]}{\delta_\rho} \tag{1.40}$$

Thus, V_{XC} can be readily obtained once E_{XC} is known. The Kohn-Sham equations will give the exact energy, if the exchange-correlation energy as a functional of density is known exactly. Several approximations to E_{XC} exits but there is still no systematic way to improve E_{XC}. Similar to Hartee-Fock method, the Kohn-Sham equations are solved in a self-consistent way.

1.9.2.3 Exchange-Correlation Energy Funcitonals

1.9.2.3.1 Local Density Approximation (LDA)

The simplest approximation to the exchange-correlation functional is given by the Local Density Approximation (LDA). The exchange-correlation energy in this approximation solely depends upon the value of the electronic density at each point in

space (and not, for example, derivatives of the density or the Kohn-Sham orbitals). For this LDA applies to a uniform electron gas and within its framework thus the exchange correlation energy at any point in space is a functional dependent on the electron density at that point only. The exchange correlation energy is given by the equation:

$$E_{XC}^{LDA}[\rho] = \int \rho(r) \in_{XC} (\rho) dr \tag{1.41}$$

Here, the integral is over all the space dr and $\in_{XC}$ is the exchange-correlation energy per electron in an uniform electron density, ρ. The functional derivative of the exchange-correlation gives,

$$v_{XC}^{LDA} = \frac{\partial E_{XC}^{LDA}}{\partial \rho} = \in_{XC} [\rho(r)] + \rho(r) \frac{\partial E_{XC}^{LDA}(\rho)}{\partial \rho} \tag{1.42}$$

$\in_{XC}$ consists of two terms, the exchange energy, $\in_X [\rho]$ and the correlation contribution, $\in_C [\rho]$

$$\in_{XC} = \in_X [\rho] + \in_C [\rho] \tag{1.43}$$

where, $\in_X = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/2} [\rho(r)]^{1/3}$

1.9.2.3.2 Generalized Gradient Approximation (GGA)

GGA tends for generalized gradient approximation which implies the use of both electron density and its gradient for calculating the exchange-correlation energy functionals E_{XC}, one of the most widely used approximation in density functional theory. It can be defined as any generic function of the local value of density and its squared gradient that is constructed to approximate the exchange-correlation energy per particle in an N electron system. The exchange correlation energy is defined as

$$E_{XC}^{GGA} = \int \rho(r) \in_{xc}^{GGA} \{\rho, \nabla\rho\} dr \tag{1.44}$$

where, $\nabla\rho$ denotes the gradient term.

1.9.3 Reactivity Descriptors

1.9.3.1 DFT-based Descripotrs

Several descriptors are used in computational chemistry in order to define the

reactivity of the system, known as reactivity descriptors. The prime objective for the formulation of these descriptors are essentially to quantify and analyze the conceptually important quantities such as chemical reactivity, selectivity and stability of the molecular systems form a general theoretical framework. There have been numerous works in this field brining out the usefulness of these descripotrs in generalizing the chemical reactivity problems within the framework of DFT.[139-143]

DFT provides a framework to discuss reactions in terms of change in number of electrons (N) or change in external potential $v(r)$ due to nuclei. The first derivative of $E(\rho)$ with respect to the number of electrons (N) under the constant external potential $v(r)$ is termed as the chemical potential, μ. In theoretical chemistry, the chemical potential (μ) is defined as the negative of the electronegativity (χ) by Iczkowski and Margrave[144] and expressed as

$$\mu = -\chi = \left(\frac{\partial \mathrm{E}}{\partial \mathrm{N}}\right)_{v(\vec{r})} \tag{1.45}$$

From this viewpoint, the electronegativity of a species is the drop in energy when an infinitesimal amount of electronic charge enters the system. It is a measure of how hospitable an atom or ion, or a molecule is to the ingress of electronic charge which fits. The chemical hardness (η) of an electronic system is defined as the second derivative of total energy (E) with respect to the number of electrons (N) at constant external potential, $v(\vec{r})$.[145]

$$\eta = \frac{1}{2}\left(\frac{\partial^2 \mathrm{E}}{\partial \mathrm{N}^2}\right)_{v(\vec{r})} = \frac{1}{2}\left(\frac{\partial \mu}{\partial N}\right)_{V(\vec{r})} = -\frac{1}{2}\left(\frac{\partial \chi}{\partial N}\right)_{V(\vec{r})} \tag{1.46}$$

The hardness of a species is then the amount by which its electronegativity, its ability to accept electrons, decreases when an infinitesimal amount of electronic charge is added to it.

Using the finite difference approximation, chemical hardness and chemical potential can be approximated as

$$\eta = \frac{\mathrm{IP} - \mathrm{EA}}{2} \tag{1.47}$$

$$\mu = -\left(\frac{\mathrm{IP} + \mathrm{EA}}{2}\right) \tag{1.48}$$

where, IP and EA are the first vertical ionization potential and electron affinity, respectively, of the chemical system.

Further, using Koopmans' theorem[146], the above parameter can be expressed as:

$$\mu = \frac{\mathrm{E}_{LUMO} + \mathrm{E}_{HOMO}}{2} \tag{1.49}$$

$$\eta = \frac{\mathrm{E}_{LUMO} - \mathrm{E}_{HOMO}}{2} \tag{1.50}$$

where, E_{LUMO} is the energy of the lowest unoccupied molecular orbital and E_{HOMO} is the energy of the highest occupied molecular orbital.

Parr et al.[147] introduced the global electrophilicity index (ω) in terms of chemical potential and hardness as:

$$\omega = \frac{\mu^2}{2\eta} \tag{1.51}$$

According to this definition, ω describes the electrophilic power of a molecule. When a molecule acts as a Lewis acid (an electron-pair acceptor), the incoming electron pairs are residing in its LUMO. The molecules with low-lying LUMOs are more capable to accommodating electrons than those with high LUMOs; thus, the LUMO descriptor should measure the electrophilic power of a molecule.

The Fukui function, a frontier MO reactivity index, is by far the most inportant local reactivity index and defined as

$$f(\vec{r}) = \left(\frac{\partial \rho(\vec{r})}{\partial N}\right)_{v(\vec{r})} = \left(\frac{\delta \mu}{\delta v(\vec{r})}\right)_N \tag{1.52}$$

where $\rho(\vec{r})$ is the electron density.

Mendez and Gazquez[148] and Yang and Mortier[149] introduced a procedure to obtained information about *f*(r). This method compute the values around each atomic site that signifies the atom in a molecule. With this approximation, the condensed FF becomes

$$f_k^+ = [q_k(N+1) - q_k(N)] \quad (1.53)$$

(for nucleophilic attack on the system)

$$f_k^- = [q_k(N) - q_k(N-1)] \quad (1.54)$$

(for electrophilic attack on the system

$$f_k^o = [q_k(N+1) - q_k(N-1)]/2 \quad (1.55)$$

(for radical attack on the system)

where, $q_k(N)$, $q_k(N+1)$ and $q_k(N-1)$ are the charges of the k^{th} atom for N, $(N+1)$ and $(N-1)$ electron systems, respectively.

1.9.3.2 Physiochemical Descripotrs

The physiochemical descriptors used in the studies obtained from the MM+ computations with HyperChem software.[150] These parameters were used to generate various QSAR equations and to calculate cytotoxity of the compounds. The descriptors used in these study are mainly, hydration energy, logP, surface area, molar refractivity and polarizability.

Hansch showed a correlation between biological activity of phenoxy acetic acid with octanol-water partition coefficient in terms of logP (octanol-water partition coefficient).[151]

Molar refractivity (MR) index was suggested by Viswanadhan et al. in combination with logP.[152] It belongs to molecular descriptor and was tested to estimate the atomic contribution to octanol-water partition coefficient.

Hydration energy was put forwarded by No et al.[153] for the study of the gas-phase proton transfer energies of several amino acids.

Stanton et al.[154,155]introduced the descriptor surface area in terms of the solvent-accessible surface area of each atom and atomic charges are calculated from the atomic electro negativity or with a quantum chemistry method.

1.9.4 Basis Sets

Basis set[156,157] is a set of mathematical functions use to represent atomic orbitals, the employed mathematical functions describe the radial and angular distribution of electron density. Basis set are used to create molecular orbitals by expanding as a linear

combination of atomic orbitals (LCAO). The basis set is classified into the following types:

1.9.4.1 Stater Type Orbitals (STO)

The STO was first introduced by J. C. Slater in 1930.[158] STO describes fairly well the radial electron distribution and closely resemble hydrogen-like atomic orbitals. However, it is difficult to handle as it does not converge. The STO do not have any radial nodes and exhibits an exponential dependence on the distance between the nucleus ad electrons and shows rapid convergence with increasing number of functions. STO is represented in the following functional form:

$$f^{STO}(r) = \left(\frac{\alpha^3}{\pi}\right)^{1/2} \exp(-\alpha r) \qquad (1.56)$$

1.9.4.2 Gaussain Type Orbitals (GTO)

In 1950, Frank Boys of Cambridge University proposed the use of Gaussain type Functions (GTF's) for the atomic orbitals in an LCAO wave function by replacing the exponential term, $\exp(-\alpha r)$ of STO with $\exp(-\alpha r^2)$ GTO.[159] The functional form of GTO can be thus written as:

$$f^{GTO}(r) = \left(\frac{2\alpha}{\pi}\right)^{1/2} \exp(-\alpha r^2) \qquad 1.57)$$

The main advantage of GTO is the converge criteria, that is why it is very suitable from the viewpoint of computational efficiency and are easier to handle. However, they describe the radial electron distribution less satisfactorily. Figure 1.24 shows the behavior of STOs and GTOs with respect to the radial decay of s-type functional.

Here, the dependence of r^2 in GTO makes it inferior to STO in two aspects: (a) at the nucleus GTO has zero slope but STO has a cusp or discontinuous derivatives and (b) GTO decays rapidly far away form the nucleus compared to STO.

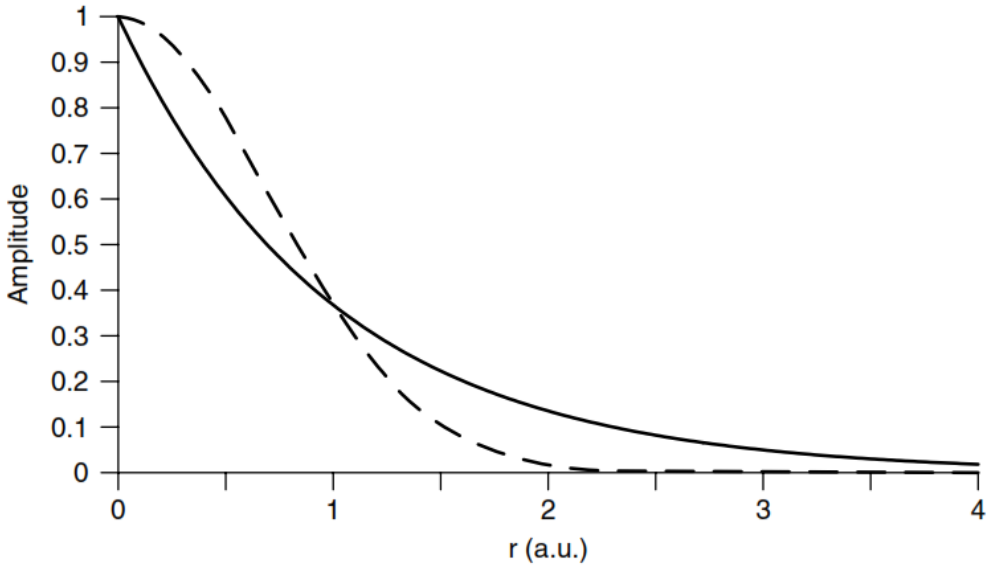


Figure 1.24 Behavior of e^{α}, where $\alpha = r$ (solid line, STO) and $\alpha = r^2$ (dashed line, GTO) The picture is adapted from Reference 155

1.9.4.3 Minimal Basis Set

The minimal basis set contains one basis function for each atomic orbital to describe a free atom. The most common of all the minimal basis sets are the STO-nG basis set devised by Pople and his group.[160] STO-nG involves a linear combination of "n" primitive gaussain to each STO. One example of this basis set is STO-3G, here, each basis function is a contraction of three primitive gaussain for the element H to Xe.

1.9.4.4 Split valence, Double and Triple-Zeta Basis Set

Split valence basis set is also introduced by the same group which uses a single contracted gaussain function for each core shell whereas the valence electrons are represented by two contracted gaussains.[161] The basis sets beyond minimal basis sets are termed as "extended" basis set, the most common improvement over the minimal basis set is the double-zeta (DZ) basis set which uses two contracted gaussain functions in place of each minimal basis functions. Similarly, triple-zeta (TZ) basis set uses three basis functions.

The valence double-zeta basis sets include 3-21G, 4-31G and 6-31G, where the first number is the number of primitive gaussaain in the core contracted gaussain functions

(CGF), the next is the two valence CGF and the last number is the primitives in the outer contracted gaussain. The smallest split-valence basis set is the 3-21G and it uses a three primitive expansion for the core orbitals and then splits the valence orbitals into a two basis functions, the inner function being a contraction of two gaussain and the outer function being just a single gaussain, as represented in Figure 1.25.

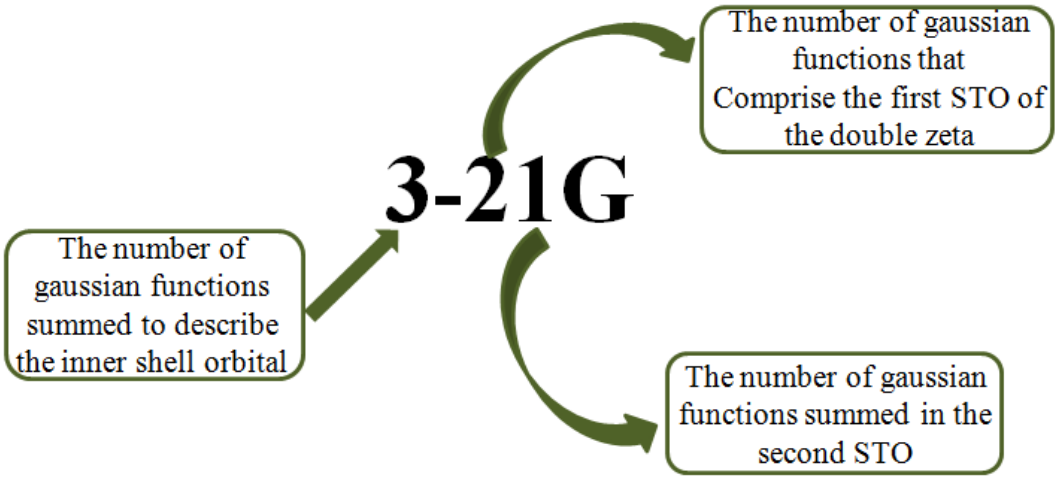


Figure 1.25 Example of split valence basis set

1.9.4.5 Polarized Basis Sets

This basis set includes polarization function into the basis set to improve the accuracy of the representation. As atoms are brought close together, there charge distribution causes a polarization effect which destroys the atomic orbitals. Thus, polarization functions improve the flexibility of the basis set by allowing atomic electron densities to be polarized in order to improve the accuracy of the calculation. For instance, 's' orbitals begin to have a little of the 'p' flavor and 'p' orbitals that of 'd' flavor. One single star (*) at the end of a basis set denotes that polarization has been taken into account in the 'p' orbitals to the heavy atom, excluding H-atom, example: 6-31G*, whereas two stars (**) indicate that, in addition to heavy atom polarization, p functions have been added to H, example: 6-31G**. In addition to polarization functions, diffuse functions (represented by '+') are used which considers a larger spatial extent, a single '+' corresponds to p-orbitals and '++'to both s and p-orbitals.

1.9.4.6 Numerical Basis Set

The numerical basis sets uses numerical basis functions which are generated by solving the KS equations numerically with the approximate exchange-correlation functional. The basis functions are given numerically as values on an atomic-centered spherical-polar mesh, rather than as analytical functions (i.e. Gaussain orbitals). Because of the better quality of the atomic orbitals, basis set superposition errors (BSSE) are minimized and an excellent descripotors of even weak bonds are possible. Example of this type of basis set are the double numeric polarization (DNP), triple numeric polarization (TNP) basis sets etc. Here, the 'N' emphasize the numerical nature of the orbitals. Double numerical (DN) basis set refers to the generation of an entire second set of functions resulting in the doubling the size of the basis sets. DMol3 program based on DFT utilizes Minimal, DN, DND, DNP, TNP numerical functions for representing the atomic orbitals.

1.9.4.7 Effective Core Potential (ECP)

Effective core potential (ECP) or pseudopotentails allow treatment of the core electrons as some average potential while the valence electrons are considered explicitly. ECPs take into account the relativistic effects which are prominent in heavy elements in the periodic table (third-row elements, *i.e.* the first transition metals).[162] The electrons near the very positive nucleus of a heavy element experience a large relative attraction than for lighter elements, which accelerates the electrons close to the speed of light. Thus, there is a significant difference in the kinetic energy of electrons in case of heavy atoms, if we go from core to the outer layer. In this situation, Einstein's theory of general relativity starts to have an effect on the shape of the atomic orbitals.[163,164] These consequences arise from the dependence of mass on velocity.[165] This dependence causes the masses of the inner electrons of heavy atoms to be significantly greater than mass of the rest electrons, since the Hamiltonian operator in the Schrödinger equation contains the electron mass, this change of mass should be taken into account. Relativistic effects in heavy-atom molecules affect geometries, energies, and other properties.[162] The net effect is that the inner core orbitals of a heavy element are contracted relative to the corresponding orbitals in a lighter atom. The ECP basis set includes valence electrons explicitly and replaces the

influence of the core electrons by a static, effective potential, commonly described by a polynomial function. ECP tremendously reduces the computational expense for heavier elements where the relativistic effects are significant for the core shell electrons. Example is the Los Alamos national Laboratory (LANL) ECP with double-zeta (DZ) valence basis set i.e. the LANL2DZ basis set.[166,167]

1.10 Regression Analysis

Regression analysis is a statistical tool for the investigation of relationships between different variables or molecular descriptor. They are qualitative or quantitative empirically defined relationships between molecular structure and observed properties. In some cases, this may seem to duplicate statistical mechanical or quantum mechanical results. However, structure-property relationships need not be based on any rigorous theoretical principles.

The first regression model has been developed by Adrien-Marie Legendre, a French mathematician, in 1805,[168] in order to describe the shape of the earth. Later in 1809, Carl Friedrich Gauss used the same method to calculate the orbit of celestial objects.[169,170] However, the very first computer-added approach in drug designing is developed in the early 1960 with pioneering work of Hansch, who used multiple linear regression (MLR) to build some predictive models inorder to describe the biological activity of certain compounds.[171,172]

The simplest case of structure-property relationships are qualitative rules of thumb. For example, the statement that branched polymers are generally more biodegradable than straight-chain polymers is a qualitative structure-property relationship. Such relationships are most often derived by using curve-fitting software to find the linear combination of molecular properties that best predicts the property for a set of known compounds. This prediction equation can be used for either the interpolation or extrapolation of test set results.

When the property being described is a physical property, such as the boiling point, this is referred to as a quantitative structure-property relationship (QSPR). When the property being described is a type of biological activity, such as drug activity, this is referred to as a quantitative structure-activity relationship (QSAR). For example, the

effect of electrophilicity, energy of lowest unoccupied molecular orbital, log P, and molar refractivity upon the logarithmic relative activity. Several physicochemical descriptors, such as hydrophobicity, topology, electronic parameters and steric effects, are usually used in QSAR studies in many disciplines, with many pertaining to drug design and environmental risk assessments.[173-178]

1.10.1 Development of a Regression Model

The development of a regressional model is very tidieous job, for this, the first step is to assamble a list of compounds for which the experimentally determined property is known. However, the choice of compounds are also very important, say for example if we want to constract a model by fitting hydrocarbon data then it will be reliable for predicting the properites of the hydrocarbon systems. In order to ensure that the method will be predictive, validation with some external data set is also some times needed.

The next step is to obtain geometries for the molecules. Crystal structure geometries can be used for this purpose; however, it is alwasys better to use theoretically optimized geometries so that the systematic errors druing computation will cancel out.

The next step is to compute the values of different molecular descriptors. Any numerical value that describes the molecule or its properties could be used for the development of the model. However, it is necessary to decide which descriptor will be used for the regression model. This is usually done by computing correlation coefficients. Correlation coefficients are measures of how closely two values (descriptor and property) are related to one another in a linear relationship. If a descriptor has a correlation coefficient of 1, that means that it can describe the property exactly. A correlation coefficient of zero means the descriptor has no relevance or the properties are fully independent of one another. The descriptors with the largest correlation coefficients are used in the curve fit to create a activity/property predicting equation in linear regression. However, if we want to develop a multiple regression model then it is necessary to check the autocorreation between the two descriptors involved in a single model. Only those descriptors which are having least

autocorelatin with one another can be used to build a multiple regression model.

A curve fit is then done to create a linear equation , such as

$$\text{Property}=c_0+c_1d_1+c_2d_2\ldots\ldots$$

where c_i are the fitted parameters or coefficients and d_i the independent descriptors.

1.10.2 Common Molecular Descriptors

Constitutional Descriptors

Molecular weight, Number of atoms of various elements, Number of bonds of various orders, Number of rings

Topological Descriptors

Weiner index, Randic indices, Kier and Hall indices, Information content, Connectivity index, Balaban index

Electrostatic Descriptors

Partial charges, Polarity indices, Topological electronic index, Multipoles

Charged partial surface areas, Polarizability, Anisotropy of polarizability

Geometrical Descriptors

Moments of inertia, Molecular volume, Molecular surface areas, Shadow indices, Taft steric constant, Length, width, and height parameters, Shape factor

Quantum Chemical Descriptors

Net atomic charges, Bond orders, HOMO and LUMO energies, FMO reactivity indices, Refractivity, Total energy, Ionization potential, Electron affinity, Energy of protonation, Orbital populations, Frontier orbital densities, Superdelocalizabilities, Sum of the squared atomic charge densities, Sum of the absolute values of charges, Absolute hardness

Statistical Mechanical Descriptors

Vibrational frequencies, Rotational enthalpy and entropy, Vibrational enthalpy and entropy, Translational enthalpy and entropy

1.10.3 Applications of Regression

1. Regression model (QSAR) can show way in identifying the features that makes

a molecule active or inactive.

2. It can be considered as the method of trying to build a model to explain why some keys works and others do not.

3. QSAR can predict quantities such as the binding affinity, pharmocokinetic parameters, toxicity and some environmental related parameters of a given molecule.

4. QSAR based analysis may be better regarded as an exercise to screen or filter drug candidates, before they are subjected to more intensive calculations such as docking or an experimental measurement of activity. Typically, this step will pick up a dozen of drug candidates from a library of millions of well-studied molecules.

1.11 Molecular Docking Simulation

Molecular docking study as it has gained enormous importance in the field of drug discovery. These approaches help us to recognize some potent drug candidates for some more rigorous calculations, whose target is already known. There are three basic parameters to evaluate for any docking calculation: (1) characterization of the binding cavity or active site of the target system. (2) Secondly, orientation of the ligand with respect to the binding cavity of the target system known as binding pose and finally (3) scoring, i.e. the evaluation of the strength of the interaction of the ligand-receptor system. It helps us for exploring the possible binding modes of a substrate to a given receptor, enzyme or other binding site.[179] In docking, computational stimulation is performed in order to ensure the best possible orientation of the substrate into the binding cavity of receptor of interest to form a stable complex so that the overall energy of the system is minimized. The scoring functions which in turn reflect the binding affinity were generally calculated by taking into account of all hydrogen bond that is being formed between the ligand with the amino acid of the receptor or target with respect to the different orientation of the ligands in to the active site of the specific target.

1.11.1 Docking Approaches

The first requirement for the docking calculation is the structure of the receptor of interest. It is usually available in some online protein data bank servers. These structures are determined using biophysical technique such as X-ray spectroscopy or using other analytical techniques. There are two popular approaches for molecular docking. The first

one is a matching technique that describes the receptor and the ligand as complementary surfaces.[180,181] Here, the receptor surface is described in terms of its solvent accessible surface area and whereas the ligand surface is described in terms of its matching surface descriptors.

In the second approach is the simulation method where the interaction energy of the ligand-receptor pair is calculated in terms of certain scoring functions.[182] Here, the ligand is trying to fit into the active cavity or binding site of the receptor after certain random moves in its confined space. The moves incorporate rigid body transformations such as translation and rotations, as well as internal changes to the lignads structure including torsion angle rotations. The ligand is hold into the active site by forming some hydrogen bond with the amino acid residues of the receptor.

The simulation method is comparatively more reliable as it incorporate ligand flexibility into the system. Moreover, the process is physically closer to what happens in the real system, when the protein and ligand approach each other after molecular recognition. But one disadvantage of this technique is the computational time. It will take longer time to evaluate the optimal binding pose and in order to score the binding poses.

1.12 Quantum Mechanics/Molecular Mechanices (QM/MM) Methods

The theoretical investigation of molecular and electronic structure enables the experimentalist to evaluate the relationship between the structure and its activity and help them to study binding mechanism at the molecular level. However, there are some limitations of both quantum and molecular mechanical methods. Molecular mechanics is unable to calculate or simulate the breaking or formation of bonds and has unable to emphasis the intermolecular interactions. Quantum mechanical methods can be implicating successfully in electronic structure calculations of small drug molecules. It will also able to describe the hydrogen, covalent as well as ionic bonding interaction fairly well. However, this method is computationally very expensive and hence their uses are usually restricted to small molecular systems. The development of hybrid QM/MM method allows us to study complicated system comprising of thousand of atoms. Applying QM/MM methodology to a system involves the division of the system into two

reasons. The region of interest is treated with a high level quantum mechanical method and the outer portion is treated with a low level molecular mechanical method.

An illustration of the QM/MM approach is, consider a metal ion immerge in a solvent, as shown in Figure 1.26. Here, the metal ion surrounded by the first solvent layer is defined by the QM (*'I' region*) and the remaining portion is taken care by the MM region (*'O' region*).

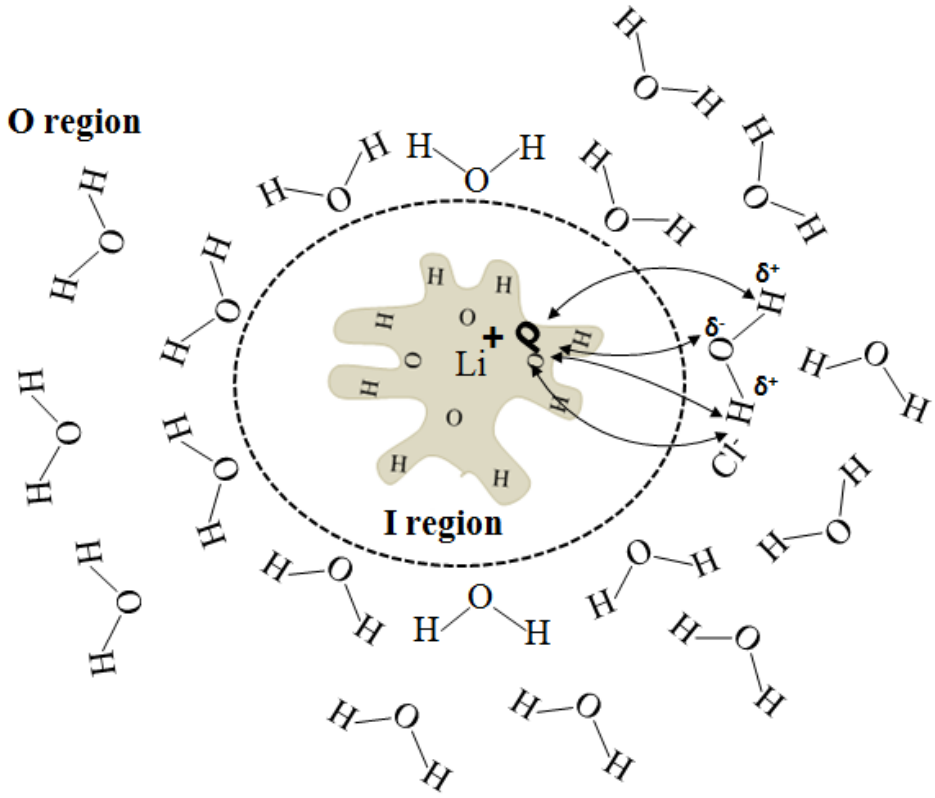


Figure 1.26 Graphical representation of Hybrid QM/MM method

We have used QMERA modules to perform the hybrid QM/MM calculations implemented in Material Studio. This is an effective method in handling the QM/MM computations as it combination of the accuracy of QM with the speed of force field method using ChemShell[183] environment. This method uses the DMol[3] for the QM region[184] whereas MM region is treated with the GLUP force field[185] engine.

The energy obtained for a real system is composed of three independent calculations.

$$E_{tot} = E_{QM} + E_{MM} + E_{QM/MM} \tag{1.58}$$

where,

E_{tot} = is the total potential energy expression of the QM/MM

E_{QM} = is the total QM energy (for the calculation in the 'I'th region)

E_{MM} = is the potential energy in the MM region (energy of the 'O' region)

$E_{QM/MM}$ = is the QM-MM interaction energy

1.13 Objectives of the Present Investigation

- Study of the effect of various DFT based descriptors and the physicochemical parameters on the cytotoxicity of camptothecin molecules.
- Development of some regression models to identify the particular descriptor that is responsible for the activity of the selected camptothecin molecules.
- Study of the inhibitory activity of the camptothecin molecules against their cellular target (1T8I).
- Kinetics and thermochemical study of hydrolysis mechanism of a novel anticancer agent *trans*-[$PtCl_2$(dimethylamine)(isopropylamine)].
- DFT study of energetic, binding affinity and stability of different possible adducts of transplatin complex using quantum mechanics/ molecular mechanics (QM/MM) method.
- Use of different computational strategies to identify potential candidates in the screening process of some selected gold (III) complexes.

References

1. Nelson, L.D. & Cox, M.M. *Lehninger Priciple of Biochemistry*, 4th ed., W.H. Freeman and Company, New York, 2000.

2. Dahm, R. *Dev. Biol.* **278** (2), 274--288, 2005.

3. Griffith, F. *J. Hyg. (Lond.)* **27** (2), 113--159, 1928.

4. Avery, O.T., et al. *J. Exp. Med.* **79** (2), 137--158, 1944.

5. Hershey A.D., & Chase, M. *J. Gen. Physiol.* **36** (1), 39--56, 1952.

6. Chargaff. E. *J. Cell. Physiol. Suppl.* **38**, 41--49, 1951.

7. Chargaff, E. *Experientia* **6**, 201--209, 1950.

8. Watson, J.D., & Crick, F.H. *Nature* **171** (4356), 737--738, 1953.

9. Ghosh, A., & Bansal, M. *Acta Crystallogr. D* **59** (4), 620--626, 2003.

10. Berg, J., Tymoczko, J. & Stryer, L. *Biochemistry*, 5th ed., W. H. Freeman and Company, New York, 2002.

11. Bertram, J.S. *Mol. Aspects Med.* **21** (6), 167--223, 2001.

12.Montelone, B. A. Mutation, Mutagens, and DNA Repair. http://www-personal.ksu.edu/~bethmont/mutdes.html#origins, 1998.

13. Gardner, E.J. et al. *Priniples of Genetics*. 8th ed., John Wiley and Sons, Singapore, 2005.

14. Griffiths, A.J.F. et al. *An Introduction to Genetic Analysis.* 7th ed., W. H. Freeman, New York, 2000.

15. Hogan, C. M. Mutation. http://www.eoearth.org/view/article/159530, 2010.

16. Hayashi, H., et al. *Curr. Biol.* **12** (4), 335--339, 2002.

17. Kozmin, S., et al. *Proc. Natl. Acad. Sci.* **102** (38), 13538--13543, 2005.

18. Jemal, A., et al. *CA Cancer J. Clin.***61** (2), 69--90, 2011.

19. Latest world cancer statistics Global cancer burden rises to 14.1 million new cases in 2012: Marked increase in breast cancers must be addressed. https://www.iarc.fr/en/media-centre/pr/2013/pdfs/pr223_E.pdf, 2013.

20.Cancer incidence to rise five-fold in India by 2025? http://timesofindia.indiatimes.com/life-style/health-fitness/health-news/Cancer-incidence-to-rise-five-fold-in-India-by-2025/articleshow/29823316.cms, 2014.

21. Ferlay, J., et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012, *Int. J. Cancer*, **136**, E359--E386, 2015

22. Alwan, A., et al. Global status report on noncommunicable diseases 2010, WHO Library Cataloguing-in-Publication Data, Italy, 2011.

23. Pardee, A.B., & Stein, G.S. *The biology and treatment of cancer: Understanding cancer*, Wiley-Blackwell, United Kingdom, 2009.

24. Sikira, K., & Timbs, O. *Exp. Rev. Anticaner. Ther.* **4** (1), 11--12, 2004.

25. Peyrone, M. *Ann. Chem. Pharm.* **51** (1), 1--29, 1844.

26. Rosenberg, B., et al. *Nature* **205**, 698--699, 1965.

27. Chu, G. *J. Biol. Chem.* **269** (2), 787--790, 1994.

28. Rosenberg, B. *Cancer* **55** (10), 2303--2316, 1985.

29. Rosenberg, B., et al. *Nature* **222**, 385--386, 1969.

30. Kelland, L. *Nat. Rev. Cancer* **7** (8), 573--584, 2007.

31. Klein, A.V., & Hambley, T.W. *Chem. Rev.* **109** (10), 4911--4920, 2009.

32. Farrell, N., et al. *J. Med. Chem.* **32** (10), 2240--2241, 1989.

33. Farrell, N., et al. *Cancer Res.* **52** (18), 5065--5072, 1992.

34. Kelland, L.R., et al. *Cancer Res.* **54** (21), 5618--5622, 1994.

35. Coluccia, M., et al. *J. Med. Chem.* **36** (4), 510--512, 1993.

36. Coluccia, M., et al. *Chem. Biol. Interact.* **98** (3), 251--66, 1995.

37. Goddard, P.M., et al. *Anticancer Res.* **16** (1), 33--38, 1996.

38. Kelland, L.R., et al. *J. Med. Chem.* **38** (16), 3016--3024, 1995.

39. Kelland, L.R. *Crit. Rev. Oncol. Hematol.* **15** (3), 191--219, 1993.

40. Harrap, K.R. *Cancer Treat. Rev.* **12**, 21--33, 1985.

41. Calvert, A.H., et al. *J. Clin. Oncol*. **7** (11),1748--1756, 1989.

42. Bloemink, M.J. & Reedijk, J. Cisplatin and derived anticancer drugs: mechanism and current status of DNA binding, in *Metal ions in biological systems*, A. Sigel et al, eds., Marcel Dekker Inc., New York, Basel, 1996, 641–675.

43. Reedijk, J. *Proc. Natl. Acad. Sci. U. S. A.* **100** (7), 3611--3616, 2003.

44. Reedijk, J. *Inorg. Chim. Acta*. **873**, 198--200, 1992.

45. Jamieson, E.R., & Lippard, S.J. *Chem. Rev.* **99** (9), 2467--2498, 1999.

46. Jakupec, M.A., et al. *Rev. Physiol. Biochem. Pharmacol.* **146**, 1--54, 2003.

47. Cleare, M.J., & Hoeschele, J.D. *Plat. Metal Rev.* **17**(1), 2--13, 1973.

48. Natile, G., & Coluccia, M. *Coord.Chem. Rev.* **216-217**, 383--410, 2001.

49. Farrell, N. *Met. Ions Biol. Syst.* **32**, 603--639, 1996.

50. Beusichem, M.V. & Farrell, N. *Inorg. Chem.* **31** (4), 634--639, 1992.

51. Appletion, T.G., et al. *Inorg. Chem.* **31** (14), 3077--3082, 1992.

52. Wimmer, S., et al. *J. Chem. Soc., Dalton Trans.* 403--412, 1989.

53. McGowan, G., et al. *Inorg. Chem.* **44** (21), 7459--7467, 2005.

54. Aris, S.M., & Farrell, N.P. *Eur. J. Inorg. Chem.* **2009** (10), 1293--1302, 2009.

55. Miller, S.E., et al. *Inorg. Chim. Acta* **190** (1), 135--144, 1991.

56. Miller, S.E., & House, D.A. *Inorg. Chim. Acta* **166** (2), 189--197, 1989.

57. Legendre, F., et al. *Chem. Eur. J.* **6** (11) 2002--2010, 2000.

58. Raber, J., et al. *Mol. Phys.* **102**, 2537--2544, 2004.

59. Arnesano, F., & Natile, G. *Coord. Chem. Rev.* **253**, 2070--2081, 2009.

60. Raber, J., et al. *J. Phys. Chem. B* **109** (24), 12195--12205, 2005.

61. Kozelka, J. *Inorg. Chim. Acta* **362** (3), 651--668, 2009.

62. Lau, J.K.C., & Deubel, D.V. *J. Chem. Theory Comput.* **2** (1), 103--106, 2006.

63. Costa, L.A.S.,et al. *J. Chem. Phys.* **118** (23), 10584--10592, 2003.

64. Chval, Z., & Sip, M. *J. Mol. Struct. (THEOCHEM)* **532** (1-3), 59--68, 2000.

65. Alberto, M.E., et al. *J. Phys. Chem. B* **113** (43), 14473--14479, 2009.

66. Pavelka, M., et al. *Chem. Eur. J.* **13**, 10108--10116, 2007.

67. Lucas, M.F.A., et al. *J. Phys. Chem. B* **113** (3), 831--838, 2009.

68. Reedijk, J. *Inorg. Chim. Acta.* **198-200**, 873--881, 1992.

69. Mansy, S., et al. *J. Am. Chem. Soc.* **95** (5), 1633--1640, 1973.

70. Ishida, S., et al. *Proc. Natl. Acad. Sci. U. S. A.* **99** (22), 14298--14302, 2002.

71. Blommaert, F.A., et al. *Biochemistry* **34** (26), 8474--8480, 1995.

72. Lippert, B., *Cisplatin: chemistry and biochemistry of a leading anticancer drug*, Wiley-VCH, Zurich, New York, 1999.

73. Fichtinger-Schepman, A. M. *J. Cancer Res.* **47** (11), 3000--3004, 1987.

74. Reed, E., et al. *Cancer Res.* **50**, 2256--2260, 1990.

75. Hambley, T.W. *J. Chem. Soc.-Dalton Trans.* **19**, 2711--2718, 2001.

76. Fichtinger-Schepman, A.M.J., et al. *Biochemistry* **24** (3), 707--713, 1985.

77. Plooy, A.C.M., et al. *Carcinogenesis* **6** (4), 561--566, 1985.

78. Eastman, A. *Biochemistry* **24** (19), 5027--5032, 1985.

79. Eastman, A., et al. *Chem. Biol. Interact.* **67** (1-2), 71--80, 1988.

80. Brabec, V., & Leng, M. *Proc. Natl. Acad. Sci. U. S. A.* **90** (11), 5345--5349, 1993.

81. Strandberg, M.C., et al. *Chemico-Biol. Interactions* **39** (2), 168--180, 1982.

82. Simonian, P.L., et al. *Blood* **90**, 1208--1216, 1997.

83. Sorenson, C.M., & Eastman, A. *Cancer Res.* **48** (16), 4484--4488, 1988.

84. Eastman, A. The mechanism of action of cisplatin: from adducts to apoptosis, in *Cisplatin. Chemistry and Biochemistry of a Leading Anticancer Drug*, B. Lippert eds., Wiley-VCH, Zürich, Switzerland, 2006, 111–134.

85. Montero, E. I., et. al. *J. Med. Chem.* **42** (20), 4264--4268, 1999.

86. Shaw, C. F.III *Chem. Rev*. **99** (9), 2589--2600, 1999.

87. Tiekink, E.R.T. *Inflammopharmacology* **16** (3) 138--142, 2008.

88. Abbate, F., et al. *Inorg. Chim. Acta* **311** (1-2), 1--5, 2000.

89 Gabbiani, C., et al. *Gold Bull.* **40** (1), 73--81, 2007.

90. Messori, L., et al. *J. Med. Chem.* **43** (19), 3541--3548, 2000.

91. Zoratti, M., & Szabo, I. *Biochim. Biophys. Acta* **1241** (2), 139--176, 1995.

92. Bernardi, P. *Physiol. Rev.* **79**, 1127--1155, 1999.

93. Rigobello, M.P., et al. *Br. J. Pharmacol.* **136** (8), 1162--1168, 2002.

94. McKeage, M.J., et al. *Coord. Chem. Rev.* **232** (1-2), 127--135, 2002.

95. Rigobello, M.P., et al. *J. Inorg. Biochem.* **98** (10), 1634--1641, 2004.

96. Wall, M.E., et al. *J. Am. Chem. Soc.* **88** (16), 3888--3890, 1966.

97. Wall, M.E. *Med. Res. Rev*. **18** (5), 299--314, 1988.

98. Wall, M.E. *Med. Pediatr. Oncol.* **11**, 480A--489A, 1983.

99. Suffness, M.I. & Cordell, G.A. Antitumor alkaloids, i*n The Alkaloids*, A. Brossi eds., Academic Press, New York, 1985, 198-212.

100. Gottlieb, J.A., & Luce, J.K. *Cancer Chemother. Rep.* **56**, 103--105, 1972.

101. Muggia, F.M., et al. *Cancer Chemother. Rep.* **56**, 515--521, 1972.

102. Kingsburg, W.D., et al. *J. Med. Chem.* **34**, 98--107, 1991.

103. Sawada, S., et al. *Chem. Pharm. Bull.* **39**, 3183--3188, 1991.

104. Markman, M. *Oncologist* **10**, 695--697, 2005.

105. Ahmad, T., & Gore, M. *Expert. Opin. Pharmacother.* **5**, 2333--2340, 2004.

106. Stewart, D.J. *Oncologist* **9**, 33--42, 2004.

107. Ormrod, D., & Spencer, C.M. *Drugs* **58** (3), 533--551, 1999.

108. Fuchs, C., et al. *Cancer Treat. Rev.* **32**, 491--503, 2006.

109. Pavillard, V., et al. *J. Anticancer. Res.* **24**, 579--585, 2004.

110. Gottlieb, J.A., & Luce, J.K. *Cancer Chemother. Rep.* **56**, 103--105, 1972.

111. Garcia, C.R., & Supko, J.G. *Clin. Cancer Res.* **8**, 641--661, 2002.

112. Wang, J.C. *Annu. Rev. Biochem.* **65**, 635--692, 1996.

113. Champoux, J.J. *Annu. Rev. Biochem.* **70**, 369--413, 2001.

114. Hertzberg, R.P., et al. *Biochemistry* **28** (11), 4629--4638, 1989.

115. Chen, A.Y., & Liu, L.F. *Annu. Rev. Pharmacol. Toxicol.* **34**, 191--218, 1994.

116. Bjornsti, M.A., et al. *Cancer. Res.* **49**, 6318--6323, 1989.

117. Legarza, K., & Yang, L.X. *Anticancer Res*. **26**, 3301--3305, 2006.

118. Hsiang, Y.H., et al. *Cancer Res.* **49**, 5077--5082, 1989.

119. Thomas, C.J., et al. *Bioorg. Med. Chem.* **12** (7), 1585--1604, 2004.

120. Wang, X., et al. *Biochemistry* **37** (26), 9399--9408, 1998.

121. Leach, A.R. *Molecular modeling: principle and application*, Addision Wesley Longam, England, 1996.

122. Koch, W. & Holthausen, M.C. *A chemist's guide to density functional theory*, 2nd ed., Wiley-VCH, Germany, 2001.

123. Rappi, A.K., et al. *J. Am. Chem. Soc.* **114** (25), 10024--10039, 1992.

124. Allinger, N.L. *Adv. Phys. Org. Chem.* **13**, 1--82, 1976.

125. Engler, E.M., et al. *J. Am. Chem. Soc.* **95** (24), 8005--8025, 1973.

126. Bartell, L.S. *J. Am. Chem. Soc.* **99** (10), 3279--3282, 1977.

127. Born, M., & Oppenheimer, J.R. *Ann. Physik*. **84** (20), 457--484, 1927.

128. Kolos, W., & Wolniewicz, L. *J. Chem. Phy.* **41** (12), 3663--3673, 1964.

129. Magnasco, V. *Methods of molecular quantum mechanics: an introduction to electronic molecular structure*, John Wiley and Sons Ltd, United Kingdom, 2009.

130. Hünenberg, P.H. *Adv. Polym. Sci*. **173**, 105--149, 2005.

131. Levine, I.N. *Quantum chemistry*, 5th ed., Prentice Hall, Engelwood Cliffs, NJ, USA, 2000.

132. Sazbo, A. & Ostlund, N.S. Modern *Quantum Chemistry: Introduction to advanced electronic structure theory*, McGraw-Hill Publishing Company, New York, 1989.

133. Wong, M.W., & Radom L. *J. Phy. Chem.* **99** (21), 8582--8588, 1995.

134. Argaman, N., & Makov, G. *Am. J. Phys.* **68** (1), 69--79, 2000.

135. Hohenberg, P., & Kohn, W. *Phys. Rev. B* **136** (3B), B864--B871, 1964.

136. Kohn, W & Sham, L. J. *Phys. Rev. A* **140** (4A), A1133--A1138, 1965.

137. Bader, R.F.W. *Atom in molecules*, Wiley, Oxford, New York, 1990.

138. Malakoff, D. *Science* **282**, 610, 1998.

139. Nalewajski, R.F. et al. Density functional theory, in *Topics in current chemistry*, R.F. Nalewajski eds., Springer, Berlin, 1996.

140. Pearson, R.G. *Chemical Harness: Application from molecules to solids*, Wiley-VCH Verlag GMBH, Weinheim, 1997.

141. Mondal, P., et al. *Phy. Chem. Commun.* **6**, 24--27, 2003.

142. Sarmah, P., et al. *J. Mol. Model.* **16** (3), 411--418, 2010.

143. Chandra, A.K., & Nguyen, M.T. *Faraday Discuss.* **135**, 191--201, 2007.

144. Iczkowski, R.P., & Margrave, J.L. *J. Am. Chem. Soc.***83** (17), 3547--3551, 1961.

145. Parr, R.G., & Pearson, R.G. *J. Am. Chem. Soc.***105**, 7512--7516, 1983.

146. Koopmans, T. A. *Physica* **1**(1-6), 104--113, 1933.

147. Parr, R. G., et. al. *J. Am. Chem. Soc.***121** (9), 1922--1924, 1999.

148. Mendez, F., & Gazquez, J. L. *J. Am. Chem. Soc.***116**, 9298--9301, 1994.

149. Yang, W., & Mortier, W.J. *J. Am. Chem. Soc*. **108** (19), 5708--5711, 1986.

150. HyperChem; Release 7, Hypercube; http://www.hyper.com/, 2002.

151. Hansch, C., et al. *J. Am. Chem. Soc.* **85** (18), 2817--2824, 1963.

152. Viswanadhan, V.N., et al. *J. Chem. Inf. Comput. Sci.* **29** (3), 163--172, 1989.

153. No, K.T., et al. *J. Phys. Chem.* **98** (42), 10742--10749, 1994.

154. Stanton, D.T., & Jurs, P.C. *Anal. Chem.* **62** (21), 2323--2329, 1990.

155. Stanton, D.T., et al. *J. Chem. Inf. Comput. Sci.* **32** (4), 306--316, 1992.

156. Bródka, A. *Chem. Phy. Lett.* **363** (5-6), 604--609, 2000.

157. Cramer, C.J. *Essentials of computational chemistry: theory and models*, John Wiley and Sons, New York, 2002.

158. Slater, J.C. *Phys. Rev.* **36** (9), 57--64, 1930.

159. Boys, S.F. *Proc. R. Soc. A* **200** (1063), 542--554, 1950.

160. Hehre, W.J., Radom, L., Schleyer, P.R. & Pople, J. *Ab initio molecular orbital theory*, Wiley, New York, 1986.

161. Ditchfield, R., et al. *J. Chem. Phys.* **54** (2), 724--728, 1971.

162. Lewars, E.G. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, 2nd ed., Springer, New York, 2010.

163. Balasubramanian, K. *Relativistic effects in chemistry Parts A, Theory & Techniques*, Wiley, New York, 1997.

164. Balasubramanian, K. *Relativistic effects in chemistry Part B, Applications*, Wiley, New York, 1997.

165. Okun, L The Concept of Mass.
http://scitation.aip.org/content/aip/magazine/physicstoday/article/42/6/10.1063/1.881171, 1989.

166. Hay, P.J., & Wadt, W.R. *J.Chem. Phys.* **82** (1), 299--310, 1985.

167. Hay, P.J., & Wadt, W.R. *J.Chem. Phys.* **82** (1), 270--283, 1985.

168. Legendre, A.M. *Nouvelles méthodes pour la détermination des orbites des comètes*, Firmin Didot, Paris, 1805.

169. Gauss, C.F. *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientum,* 1st ed., Hamburg: Frid. Perthes & I.H. Besser, 1809.

170. Gauss, C.F. *Theoria combinationis observationum erroribus minimis obnoxiae,* H. Dieterich, Gottingae, Paris, 1823.

171. Hansch, C. *J. Am. Chem. Soc.* **85** (18), 2817--2824, 1963.

172. Hansch, C., & Fujita, T. *J. Am. Chem. Soc.* **86** (8), 1616--1626, 1964.

173. Abhilash, T., & Mamta, T. *Bioinformatics India* **1**, 65, 2003.

174. Clare, B.W., & Supran, C.T. *Eur. J. Med. Chem.* **32** (4), 311--319, 1997.

175. Supran, C.T., & Clare, B.W. *Eur. J. Med. Chem.* **33** (6), 489--500, 1998.

176. Abhilash, T., & Mamta, T. *Res. J. Chem. Envioron.* **7**, 51, 2003.

177. Barua, N., et al. *Chem. Biol. Drug Des.* **79** (4), 553--559, 2012.

178. Supran, C.T., et al. *J. Med. Chem.* **43**, 1793--1806, 2000.

179. Seal, A., et al. *Bioinformation* **5**(10), 430--439, 2011.

180. Meng, E.C., et al. *J. Comput. Chem.* **13** (4), 505--524, 2004.

181. Morris, G.M., et al. *J. Comput. Chem.* **19** (14), 1639--1662, 1998.

182. Feig, M.I., et al. *J. Comput. Chem.* **25** (2), 265--284, 2004.

183. Sherwood, P., et al. *J. Mol. Struct. (THEOCHEM)* **632** (1-3), 1--28, 2003.

184. Delley, B.*J. Chem. Phys.* **92**, 508--517, 1990.

185. Gale, J. D. *Z. Kristallogr.* **220** (5-6), 552--554, 2005.