T.R.
BOLU ABANT İZZET BAYSAL UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
Department of Chemistry

DETERMINATION OF THE BIOLOGICAL ACTIVITY OF COMPLEXES FORMED WITH PT(II) AND 2-PYRIMIDINETHIOL USING QSAR METHOD BASED ON QUANTUM CHEMISTRY

DOCTOR OF PHILOSOPHY

HASAN CAN YAZICI

ACADEMIC SUPERVISOR
Prof. Dr. Ayşe MORKAN
ACADEMIC CO-SUPERVISOR
Prof. Dr. İzzet MORKAN

BOLU, ŞUBAT - 2022

APPROVAL OF THE THESIS

DETERMINATION OF THE BIOLOGICAL ACTIVITY OF COMPLEXES FORMED WITH PT(II) AND 2-PYRIMIDINETHIOL USING QSAR METHOD BASED ON QUANTUM CHEMISTRY submitted by **Hasan Can YAZICI** and defended before the Examining Committe Members listed below in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Department of Chemistry, Institute of Graduate Studies of Bolu Abant Izzet Baysal University in 2.02.2022** by

Examining Committee Members	**Signature**
Supervisor Prof. Dr. Ayşe MORKAN Bolu Abant İzzet Baysal University	.....
Member Prof. Dr. Fazilet Devrim ÖZDEMİRHAN Bolu Abant İzzet Baysal University	.....
Member Assoc. Prof. Dr. Bahadır ALTINTAŞ Bolu Abant İzzet Baysal University	.....
Member Assoc. Prof. Dr. Mecit AKSU Düzce University	.....
Member Assist. Prof. Dr. Aliye KAHYAOĞLU Düzce University	.....

Prof. Dr. İbrahim KÜRTÜL

Director of Institute of Graduate Studies

ETHICAL DECLARATION

In this thesis dissertation that was properly prepared according to the Thesis Writing Rules of Bolu Abant İzzet Baysal University of the Institute of Graduates Studies, I hereby declare that;

- All data, information, and documents presented in the thesis were obtained in accordance with the academic and ethical rules,
- All data, documents, assessments, and results were presented in accordance with the scientific ethical and moral rules,
- All works that were benefitted in the thesis were appropriately cited,
- No alteration was made in the data used,
- Study presented in this thesis is original,

  Otherwise, I declare that I accept the loss of all my rights in case any contradiction that may arise against me.

Based on the plagiarism report that was generated on the date of 03/02/2022 by using predetermined filtrations set by Directorate of Institute of Graduate Studies of the Turnitin program, a plagiarism detection software, the similarity index detected was 10 %.

_____

Hasan Can YAZICI

ABSTRACT

DETERMINATION OF THE BIOLOGICAL ACTIVITY OF COMPLEXES FORMED WITH PT(II) AND 2-PYRIMIDINETHIOL USING QSAR METHOD BASED ON QUANTUM CHEMISTRY
PHD THESIS
HASAN CAN YAZICI
BOLU ABANT IZZET BAYSAL UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF CHEMISTRY
(SUPERVISOR: PROF. DR. AYŞE MORKAN)
(CO-SUPERVISOR: PROF. DR. İZZET MORKAN)
BOLU, FEBRUARY 2022
(xiii + 133)

In the study, the subject of drug design was handled with theoretical and computerized chemistry calculations. Ligand-based and structure-based drug development methods were chosen as the main topic in this study. From studies on the A2780 human ovarian cancer gene, 154 platinum-based complexes were selected from the literature for ligand-based drug design. As a result of the calculations, HCTH407P was chosen as the DFT method to be used in geometric optimizations, and the basis set was ZORA relativistic full electron. Approximately 1400 descriptors were calculated taken from NMR, quantum mechanics, and PaDEL. Models were created for all groups using the multiple linear regression (MLR) method from the quantitative structure-activity relationship (QSAR) calculations together with the calculated descriptors. These models were validated with the four principles of the OECD and applied to 30 new drug candidate compounds formed together with platinum and the known biologically active 2-pyrimidinethiol ligand. As a result, 11 compounds were found to have low cytotoxicity and were identified as potential drug candidates. In structure-based drug development, the interactions of metal-containing molecules with DNA, which is one of the ultimate targets in cancer, were investigated. Here, the difficulties of integrating metal-containing compounds into this system were overcome. An effective core basis set (ECP), which deactivates core electrons on the metal atom, was used to make the calculations faster. As in ligand-based drug design, OLYP was chosen as the best DFT method, HayWadt for ECP, LanL2TZ for metal basis set, and def2-TZVP for other atoms.Since, DNA groove binding was examined, planar and +2 charged platinum and 2-pyrimidinethiol complexes consisting only of carrier ligands were formed. Molecular docking calculations were made with a total of 84 complexes composed of 21 biologically active functional compounds in 4 different groups with these complexes. The complexes with the best pose were selected from each group. The metal-containing complexes in the best exposures were subjected to 100ns simulations in the aqueous system. As a result of the simulations, the intermolecular interactions and binding affinities in the structures were examined in detail.

KEYWORDS: QSAR, Simulations, Molecular Docking, MLR, DFT, Basis Set, ECP, MM-GBSA, Drug Discovery, Descriptors, A2780, Cancer, LBDD, SBDD, IC_{50}, Cytotoxicity, Molecular Dynamics

ÖZET

PT(II) VE 2-PİRİMİDİNTİYOL İLE OLUŞTURULAN KOMPLEKSLERİN BİYOLOJİK AKTİVİTELERİNİN KUANTUM KİMYASINA DAYALI QSAR YÖNTEMİ İLE BELİRLENMESİ
DOKTORA TEZİ
HASAN CAN YAZICI
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
KİMYA ANABİLİM DALI
(TEZ DANIŞMANI: PROF. DR. AYŞE MORKAN)
(İKİNCİ DANIŞMAN: PROF. DR. İZZET MORKAN)
BOLU, ŞUBAT - 2022
(xiii + 133)

Çalışmada ilaç tasarımı konusunu teorik ve bilgisayarlı kimya hesaplamaları ile ele alındı. Temel olarak platin temelli ilaçların geliştirilmesi hedeflendi. Ligand temelli ve yapı temelli ilaç geliştirme yöntemleri bu çalışmada ana başlık olarak seçildi. A2780 insan ovaryum kanser geni üzerine yapılan çalışmalardan 154 adet platin içeren kompleks literatürden ligand temelli ilaç geliştirme için seçildi. Hesaplamaların sonucunda, geometrik optimizasyonlarda kullanılacak DFT metodu olarak HCTH407P ve temel set ise ZORA relativistik tam elektron olarak seçildi. NMR, kuantum mekanik ve PaDEL'den olmak üzere yaklaşık 1400 tanımlayıcı hesaplandı. Hesaplanan tanımlayıcılar ile kantitatif yapı-aktivite ilişkisi (QSAR) hesaplamalarından çoklu doğrusal regresyon (MLR) yöntemi kullanılarak tüm gruplar için modeller oluşturuldu. Bu modeller, OECD'nin dört prensibi ile doğrulandı. Platin ile biyolojik olarak aktifliği bilinen 2-pirimidintiyol ligandı arasında oluşturulan 30 adet yeni ilaç adayı bileşiklerde uygulandı. Bunun sonucunda, 11 adet bileşiğin sitotoksisitesi düşük bulundu ve potansiyel ilaç adayları olarak belirlendi. Yapı temelli ilaç geliştirmede ise, metal içeren moleküllerin kanserde nihai hedeflerden olan DNA ile etkileşimleri incelendi. Burada metal içeren bileşiklerin bu sisteme entegresinin zorluklarının üstesinden gelindi. Hesaplamaların daha hızlı olabilmesi için metal atomu üzerinde bağa katılmayan elektronları deaktif eden efektif çekirdek temel set (ECP) kullanıldı. Ligand temelli ilaç tasarımındaki gibi en iyi en iyi DFT metodu OLYP, ECP için HayWadt, metal temel set için LanL2TZ ve diğer atomlar için def2-TZVP seçildi. Daha sonra, DNA oluk bağlanması inceleneceği için düzlemsel ve sadece taşıyıcı ligandlardan oluşan +2 yüklü platin ile 2-merkaptoprimidin kompleksleri oluşturuldu. Bu kompleksler ile 4 ayrı grupta 21 adet biyolojik aktif fonksiyonel grup kullanılarak meydana getirilen toplamda 84 adet kompleks ile moleküler yerleştirme hesaplamaları yapıldı. Her bir gruptan en iyi poza sahip kompleksler seçildi. En iyi pozlardaki metal içeren kompleksler sulu sistemde 100ns'lik simülasyonlara tabi tutuldu. Simülasyonlar sonucunda, yapılardaki moleküller arası etkileşimler ve bağlanma afiniteleri detaylıca incelendi.

ANAHTAR KELİMELER: QSAR, Simülasyon, Moleküler Yerleştirme, MLR, DFT, Temel Set, ECP, MM-GBSA, İlaç Keşfi, Tanımlayıcılar A2780, Kanser, LBDD, SBDD, IC_{50}, Sitotoksik, Molekül Dinamik

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ABBREVIATIONS AND SYMBOLS

CADD : Computer Aided Drug Design

CBDCA : Carboplatin

CCDC : Cambridge Crystallographic Data Centre

CP : Cisplatin

DFT : Density Functional Theory

DNA : Deoksiribo Nucleic Acid

ECP : Effective Core Potential

GA-MLR : Genetic Algorithm-Multiple Linear Regression

GGA : Generalized Gradient Approximation

HOMO : Highest Occupied Molecular Orbital

LBDD : Ligand-Based Drug Design

LUMO : Lowest Unoccupied Molecular Orbital

MD : Molecular Dynamics

MLR : Multiple Linear Regression

MM-GBSA : Molecular Mechanics Generalized Born Surface Area

NBO : Natural Bond Orbital

OECD : Organisation for Economic Co-operation and Development

OX : Oxaliplatin

PCA : Principal Component Analysis

PDB : Protein Data Bank

QSAR : Quantitative Structure Activity Relationship

QSSR : Quantitative Structure Stability Relationship

QTAIM : Quantum Theory Atom in Molecules

SBDD : Structure-Based Drug Design

SCF : Self-Consistent Field

ZORA : Zeroth Order Regular Approximation

ACKNOWLEDGEMENTS

The author wishes to express his deepest gratitude to his supervisor Prof. Dr. Ayşe Morkan and co-supervisor Prof. Dr. İzzet Morkan for their guidance, advice, criticism, encouragements, and insight throughout the research.

I would like to express my heartfelt thanks to my dear wife Sülhiye Yazıcı for their unwavering support throughout my doctoral thesis.

Special thanks to David Case, who provided me with the AMBER 20 program free of charge, and to the support team who updated the Chemcraft program at my request.

I would also like to thank TÜBİTAK TRUBA and İTÜ UHEM (project 4007392020) for their contributions to my thesis, which made high-performance computers as free of charge.

1. INTRODUCTION

Platinum drugs are among the drug agents frequently used in cancer chemotherapy. Cancer occupies an important place among deaths worldwide (1). The cause of cancer is the uncontrolled proliferation of cells because of mutations in DNA replication while cell division takes place. However, differentiated cells spread to other organs (2).

The first platinum-derived cancer drug used in chemotherapy was Cisplatin (CP; [cis-diamminedichloroplatinum(II)]). CP was first synthesized by M. Peyrone in 1844, and its chemical structure was first revealed by Alfred Werner in 1893 (3). Cancer activity was discovered accidentally in 1965 by Barnett Rosenberg and his colleagues in a study to understand the effect of electrical fields on bacterial cells. CP is the first platinum complex compound approved by the FDA (Food and Drug Administration) in 1978 (4). It was used in the treatment of cancer in 1979 (5). CP is used as an active agent in many types of cancer. However, its use in treatment has been limited due to its high toxic effects such as neurotoxicity, nephrotoxicity (6). CP is used alone in the treatment of cancer, as well as in combination with other effective and less toxic drugs. It is successful in the early period in the treatment of many other types of cancer, especially in the ovary (7).

CP has a square plane structure. CP is a heavy metal complex containing two chloride molecules, two ammonia molecules and a platinum central atom surrounded in the cis position. There are labile or leaving and carrier groups in this structure. In the mechanism that provides the anticancer activity of CP, chlorine is the leaving group and nitrogen is the carrier group (8).

Drug trials using new analogues of CP are currently being carried out widely. In these studies, new platinum derivative compounds with cytotoxic effect, synthesized by using different functional groups instead of carrier and leaving groups, are subjected to in-vitro and in-vivo tests. Successful drug lead compounds participate in phase studies. The first condition for this success is to be able to detect compounds with low toxic effects and high anti-cancer activity. Because the indicator of success in cancer treatment is to stop the proliferation of tumor cells and destroy them without damaging the patient's healthy cells. Therefore, studies are underway to develop new chemotherapeutic agents that influence cancer types that develop CP resistance and whose side effects are minimized (9). In this content,

carboplatin (CBDCA; [cis-diammine(1,1-cyclobutanedicarboxylato)-platinum (II)]) and its main starting compound, CP, are platinum drugs commonly used in cancer treatment [10].A second-generation platinum-derived drug, CBDCA is much less neurotoxic and nephrotoxic than CP [11]. The relatively less toxicity of this compound is attributed to slower hydrolysis of the leaving group called cyclobutandicarboxylate compared to the chloride ligands in CP [12]. CP and CBDCA have been used extensively for many years in the treatment of ovarian cancer, testicular cancer, neck and head tumors, and various other solid tumors. Accordingly, many tumors have inherent resistance or develop acquired resistance to these cancer agents. The mutagenicity of CP in living cells (13) poses a problem because secondary malignancies have been associated with CP chemotherapy (14). Intensive efforts have been made to develop third-generation platinum anticancer drugs that are hoping to overcome these limitations. Oxaliplatin (OX; [trans-R,R-1,2-diaminocyclohexaneoxalatoplatinum (II)]) is a compound thought to achieve this success and has recently been approved for colorectal cancer and for the treatment of tumors resistant to CP and CBDCA. Although OX has some mutagenicity (15), it appears to be less mutagenic than CP (16). Nedaplatin, Laboplatin and Heptaplatin, which are other platinum analogues approved for use in far eastern countries, are used in the clinic (17).

2-pyrimidinethiol is an important compound with biological effect. It has been tested in some studies as a bacterial inhibitor and as an antiviral agent. It is also sometimes like biologically active structures such as 2-thiouracil found in t-RNA (18-20). It is predicted that new square planar complexes formed by the bonding of platinum metal with together nitrogen or sulfur atoms in the 2-pyrimidinethiol compound may be significant in terms of antitumor effect.

The discovery of new drug compounds is a demanding process that takes many years, requires high costs and can often result in failure. It is necessary for this process to contain many multi-disciplines together. Traditional approaches to drug discovery rely on the stepwise synthesis and screening of large numbers of compounds to identify a potential candidate. When this process is done specifically for diseases, the probability of success increases. Because each drug has a biological pathway specific to the disease. While the drug molecules act on this pathway to reach the target structure that causes the disease, they can interact with many structures such as enzymes, proteins, water, organic molecules. While these

interactions can have a positive effect, they often appear as a toxic effect. Experimental methods are frequently used to perform these operations. However, today there is an increasing effort to apply the increased computational capacity in biology and chemistry to facilitate drug discovery, design, development, and optimization (21-22). Computational calculations support the solution of problems in many fields of science with the development of technology.

Rational drug design methods occupy the minds of scientists and are constantly evolving. Rational drug design can be used to identify/design new drug candidates or to optimize the absorption, distribution, metabolism, excretion, and toxicity profile of identifying molecules from various sources. Developments in computational techniques and hardware have facilitated the application of in-silico methods in the discovery process. Rational drug design is implemented with two methods, ligand-based drug design (LBDD) and structure-based drug design (SBDD).

To predict the biological effect with the most accurate and least error, it is necessary to determine the most suitable and stable states of the initial geometries for the chemical structures. The physical structure should be selected from the least energy conformer if there is no single crystal structure for this. There are many methods for improving the structure. The most known of these methods is Density Functional Theory (DFT) (23-31). By using suitable DFT methods, the geometric structure is stabilized, and the existence of imaginary modes is determined by vibrational frequency calculations and the structure is verified. It is very valuable that the initial geometries of small ligands can be designed with minimum error based on all drug design methods.

The focus in LBDD is drug candidate small molecules and the cytotoxic effect of these molecules against the disease. The basis of LBDD is Quantitative Structure-Activity Relationship (QSAR) methods. QSAR studies are useful methods for screening chemicals that do not require experimental studies, especially in the early stages of drug discovery. If they are subjected to appropriate classification and validation analyzes (32), they become satisfactory tools for evaluating only those with the best results. In recent years, QSAR models have been used successfully to search for new molecular agents (33-38). Many other 2D and 3D chemical and physical properties of molecules whose 3D stable structure is

designed are revealed by quantum chemical calculations and these descriptors are associated with the biological activity of a living thing.

Another rational drug design method is SBDD. Because of the advances in determining biomolecular structures by spectroscopic methods, striking advances have been made in structural and molecular biology. These methods yielded more than 100,000 proteins as well as three-dimensional (3D) structures of DNA (39). Advances in bioinformatics, along with the discovery of many target proteins, have accelerated the pace of drug development. The presence of biologically important proteins and 3D structures of DNA facilitates the identification of binding gaps. These developments form the basis of SBDD (40). SBDD is more efficient and target-specific for the discovery and optimization of potential drug molecules. Among the related computational techniques, the most used methods are molecular docking and molecular dynamics (MD) simulations. These methods have numerous applications in the analysis of binding energies, ligand-protein interactions, and evaluation of conformational changes that occur during the insertion process (41). In docking studies, binding energies to the receptor with a 3D structure can be determined and the position of the ligand in the binding region of the receptor can be animated (42-44). In the simulations, the effect of the small ligand molecule placed in the target structure is simulated within a certain period (45-50). This can be useful for understanding the type of binding and for designing more compatible small molecule ligands for known target macromolecules.

In our study, it is aimed to design new complexes formed between platinum (II) and 2-pyrimidinethiol, which can be potential drug candidates and show high anticancer activity. It is planned to provide the structural characterization of all complexes first, and then to estimate the biological effects of the newly designed complexes by applying LBDD and SBDD methods.

First, many keywords are tested to minimize the structural errors caused by the algorithm errors in the geometric structures of the complexes and to detect the method constraints. It is aimed to find the most suitable structure characterization method and algorithm. After determining the parameters that make up the calculation infrastructure, it is important to determine the combinations of DFT method and basis set with the least standard deviation in terms of structural parameters by using a complex containing platinum, especially sulfur and nitrogen atoms. It is necessary to elucidate the single crystal structure of the complex and to

specify its structural parameters clearly. Single crystal 3D structures are stored in The Cambridge Crystallographic Data Centre (CCDC), the largest known crystal structure database. The desired single crystal structures from this database can be obtained free of charge in the desired formats.

The second part of the calculations is to calculate the antitumor activities of the new platinum complexes using the LBDD methods. Here, a disease-based method is adopted. One of the cancer types that is the basis of our study is the A2780 gene of the human ovarian cancer family. Researchers have used the A2780 gene many times in their in-vitro studies with CP-derived complexes. They perform chemical and physical characterization in their preliminary drug discovery studies and obtained the cytotoxic activity (IC_{50}) values of these complexes (51-74). Appropriate CP derivative complexes are expected to be categorized according to the conditions in which the researchers perform the experiment. It is necessary to reveal the most stable geometric structures in space by using the best combination of DFT and basis set, which we predicted to determine as in the previous section. Many chemical and physical descriptors, which are valuable to be revealed because of the calculations with these structures, are integrated into the QSAR statistical methods, and it is aimed to determine the descriptors and the regression equation that most accurately describes the cytotoxic activity. The determined regression equation is intended to be used to predict the activity of CP derivative complexes whose activity is unknown which is designed.

In the last part of the calculations, using SBDD, one of the computational-aided drug design (CADD) methods, it is aimed to design new complexes with planar structure containing platinum and 2-pyrimidinethiol with +2 charge, which cannot be leaving groups. After design, these complexes that bind to the target DNA, to predict the regions where they bind and interact, and finally to determine the changes that this interaction brings on the target. These complexes with suitable geometric structures taken from DFT are subjected to molecular docking and their regions in DNA are located docking poses with the best binding properties and energy are ranked. The best pose is selected and the complex structure containing this pose is designed together with the intracellular aqueous environment. By performing MD simulation, the effects of complex binding to the target are observed in the aqueous medium. Here, it is aimed to calculate the binding affinity of the drug candidate complex with the target structure DNA in a certain period.

DNA, the target receptor structure used in calculations with SBDD, can adopt a variety of different conformations that are affected by environmental factors such as hydration and ionic strength, and the primary structure of the polynucleotide (75). These conformations have been discovered as A-, B-, C-, D-, E-, and Z-DNA (76). However, under physiological conditions, the B-form of DNA predominates. The Watson-Crick model was based on the B form of DNA. This form is found in aqueous environments with low salt concentrations and is the conformation considered to be biologically important. In our calculations, it is planned to use B-form of DNA as a target.

In this study, the most important goal is to design new platinum derivative complexes that can be drug candidates by blending many disciplines with different computational methods.

2. AIM AND SCOPE OF THE STUDY

Many studies have been conducted in the literature on the importance of platinum-containing compounds. There are many theoretical, in-vitro, in-vivo and clinical studies, in which the biological effects of platinum-containing compounds are shown. The drug capacity of platinum-containing compounds remains interesting for researchers today. The literature examples explained in detail below, are shaped according to the calculation pattern to be made in this study.

2.1 Complexes Between 2-Pyrimidinethiol and Platinum

Complexes to be formed between platinum and 2-pyrimidinethiol are expected to be important drug candidates due to the biological activity of 2-pyrimidinethiol. The anti-tumor effect of complexes between platinum and 2-pyrimidinethiol by G. Cervantes et al. (77), the compound $[Pt_2Cl_2(Spym)_4]$, (Spym=2-pyrimidinethiol), was synthesized and studied as spectroscopic (the presence in the ^{195}Pt NMR spectrum). The interaction of this complex with DNA was studied by various techniques, including circular dichroism, electron microscopy, melting temperature determination, and atomic force microscopy. Frontal results showed a high activity against HL-60 and HeLa tumor lines for the Pt-Spym complex in contrast with the CP activity (figure 2.1-2.2). The best values for IC_{50} were obtained, while the values of LD_{50} were lower than those for CP. In vivo toxicity test with male mouse BDF gave a LD_{50} value for the Pt-Spym complex of 250 mg/kg. This value indicates that the toxicity is much lower than those of CP or CBDCA, which are the platinum complexes used in clinical trials.

Compound	$LC_{50}(\mu M)$	$LC_{50}(\mu g/ml)$
cisplatin	35.0	10.5
Pt-Spym	3.5	3.3

Figure 2.1. LC50 for cisplatin and Pt-Spym in the tumor line HL-60.

Compound	$LC_{50}(\mu M)$	$LC_{50}(\mu g/ml)$
cisplatin	37.0	11.1
Pt-Spym	3.5	3.3

Figure 2.2. LC50 for cisplatin and Pt-Spym in the tumor line HeLa.

2.2 Quantum Chemical Calculations of Platinum Containing Complexes

There are many methods and programs in quantum chemical calculations. Choosing the best and useful ones can take a long time. However, it is a very important phenomenon that the structures can be illuminated correctly in terms of physical and chemical aspects. There are many studies in the literature in which accurate and minimal error geometric structure optimizations for compounds are compared with experimental values. In these studies, different theoretical methods and basis set combinations were tested many times.

The study, which carried out detailed orbital and structural analyzes, was published in 2011 by Nour T. Abdel Ghani and Ahmed M. Mansour (78). A ligand and its Pd (II) and Pt (II) complexes have been synthesized as potential anticancer compounds. Geometry optimization, calculations including highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) molecular orbital definition and charge distribution were made using DFT. Natural bond orbital theory (NBO) was used to investigate the orbital interactions that provides main stabilization. Based on the results obtained from the physical and chemical techniques and ab-initio calculations of the metal complexes, about the way of bonding, it can be said that a neutral bidentate coordinated bond is formed between the nitrogen and the metal (figure 2.3). Thus, a planar geometry can be proposed for complexes.

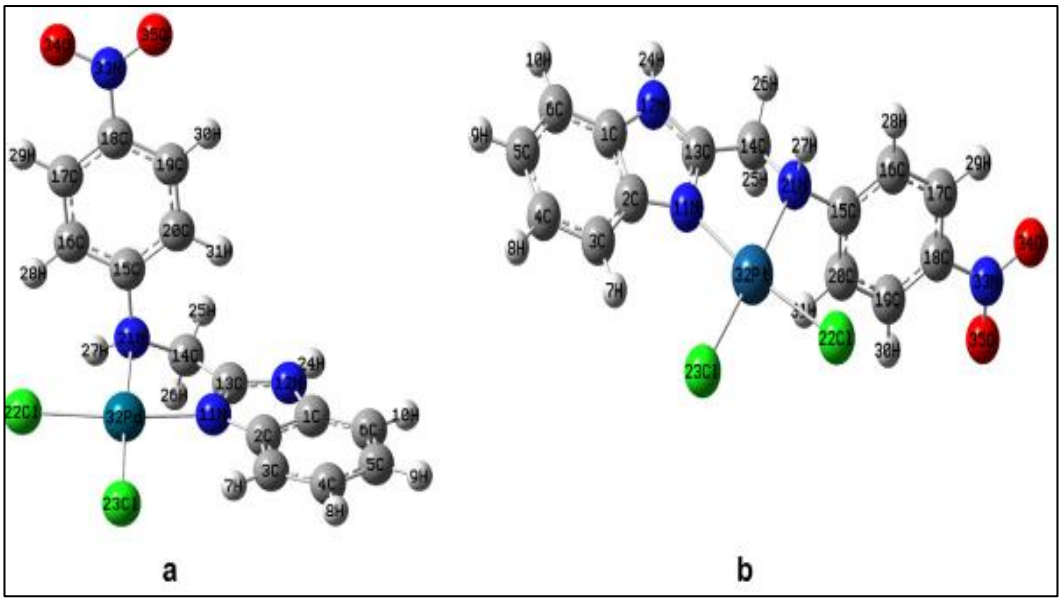


Figure 2.3. Optimized structures of (a) cis-PdLCl2 and (b) cis-PtLCl2.

In a study by Magdalena Malik and Danuta Michalska comparing DFT methods (79), various DFT methods have been examined for their abilities in the determination of raman and infrared spectra and geometric structure of CP. The estimated DFT methods content the range PBE0, B2PLYP, M06-2X, M06-L, M06, B3LYP, mPW1PW, CAM-B3LYP, wB97XD and LC-wPBE. The theoretical spectra and calculated bond lengths and angles have been checked against to the experimental values. The LC-wPBE method is success to other DFT methods in finding the geometry of CP according to results. Unhappily, the M06-2X, B3LYP and M06-L methods are insufficient in the assessment of the strength of two Pt-NH_3 bonds in CP. Both the functionals of mPW1PW and PBE0, in paperclip with the LanL2TZ(f) basis for Pt have counterpart results and look like that the best methods for estimating the Raman and infrared spectra of CP. DFT functionals of the long-range corrected (wB97XD, LC-wPBE and CAM-B3LYP) have shown successful performances in estimating the vibrational frequencies of Pt-ligand.

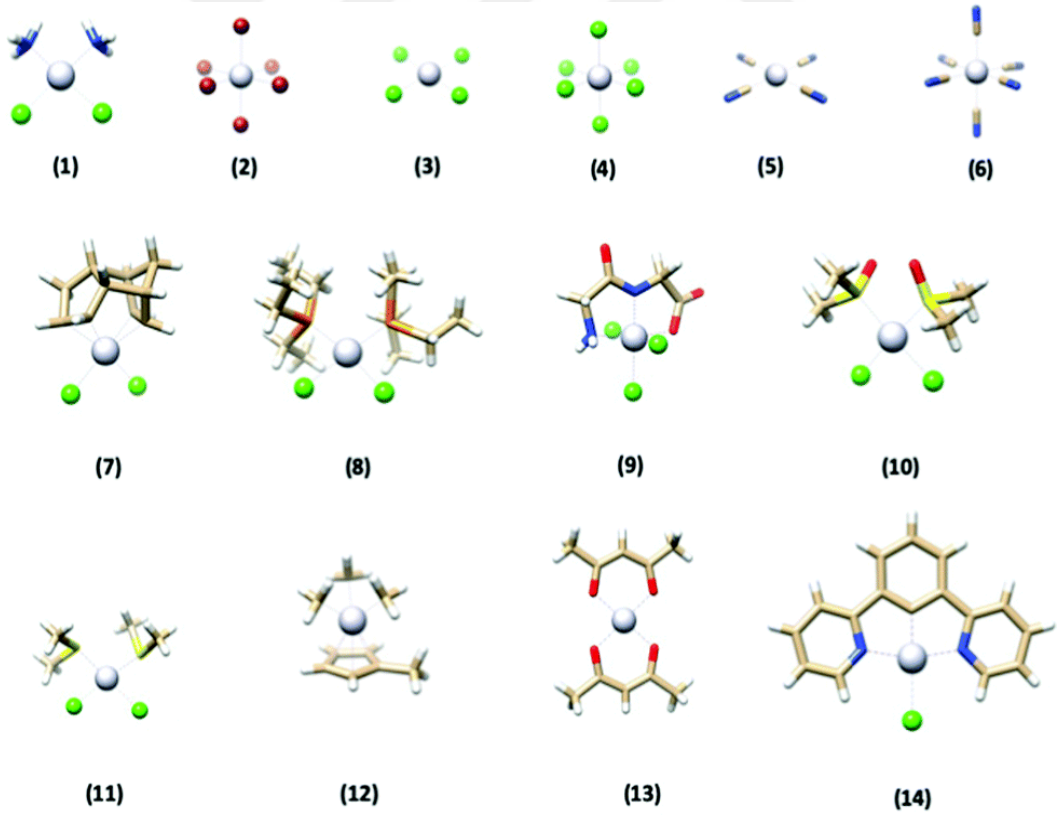


Figure 2.4. The structures of the 14 platinum-based complexes that comprised the training set used in this study. Gray = Pt, blue = N, red = O, tan = C, white = H, green = Cl, orange = P, yellow = S.

Louise M. Debefve and Christopher J. Pollock (80) tried to understand the nature of platinum in complex formation by using DFT methods on 14 platinum-containing compounds (figure 2.4) using the ORCA program. They selected five commonly used DFT methods, PBE, PBE0, TPSSh, B3LYP and BP86 for this sublect. Geometric optimizations were performed in the presence of various ligands. In addition, many calculation variations were applied in terms of basis sets, dispersion models, relativistic approximations, and solvation in these calculations. It turned out that the best performing method and basis set combination for the 14 platinum-containing complexes used here was the def2-TZVP basis set containing the ZORA relativistic approximation together with the PBE0 functional (these complexes are different in terms of their varying sizes, oxidation state, ligand diversity and number). Among the 80 methods tested for these 14 complexes with crystallographic data, the combinations of PBE0/ZORA/SVP/CPCM, PBE0/ZORA/TZVP/D3BJ/CPCM and PBE0/ZORA/TZVP/D3BJ/CPCM were found to be the best. In addition, the success of these tests was statistically tested by the researchers. As a result, the ZORA relativistic approximation and def2-TZVP basis set in platinum-containing complexes were found to be compatible with the experimental data. Many DFT methods and mentioned basis set combinations were used in the study.

2.3 Quantitative Structure Activity Relationship of Platinum Complexes

Complexes that can be drug candidates, whose structure-activity relationship is revealed by QSAR calculations, can shed light on further experimental studies. There must be some important physical and chemical descriptors for this process that can explain the biological activity. By using the best descriptors, theoretical activity values close to the known experimental cytotoxic activity specific to a particular disease type can be found. Appropriate descriptor selection and theoretical activity values are made by specific statistical analysis. QSAR is one of the LBDD methods that are still actively used today.

In this study published by Elena Monti et al. in 2004 (61), to find the appropriate descriptors for QSAR using statistical analysis was characterized 13 new synthesized or re-synthesized diamine-platinum (II) complexes and cytotoxic activity values against the A2780 and A2780/cp8 tumor cell lines which are effective in cisplatin resistance (IC_{50}). They detected and used for the cisplatin,

carboplatin, and oxaliplatin drugs routinely used in clinical treatments. QSAR study provides simple regression models to predict log ($1/IC_{50}$) values of diamine-platinum (II) complexes on these cells that are resistant to treatment for ovarian cancer. In total, 16 complexes were characterized by the best regression model using 197 molecular descriptors. They found the regression result that was best matched to the experimental data by using the principal component data analysis (PCA) with four compatible descriptors and determined their activity theoretically. The $Q^2_{LMO\text{-}50\%}$ value calculated by cross validation with four variable regression models provided 84.4% compatibility to experimental activity values (table 2.1).

Table 2.1. Molecular descriptors and statistical information for the best regressions of the log($1/IC_{50}$) on human ovarian cancer cells for the 16 Pt (II) complexes.

model	variables	$Q^2_{LOO}/Q^2_{LMO-50\%}$	R^2	s	F
1	CIC_2 $I(3)e$ $I(4)e$ $I(2)p$	0.971/0.920	0.988	0.111	221.7
2	$I(3)e$ $I(4)e$ $I(2)p$	0.950/0.914	0.972	0.161	137.8
3	nR09 $I(2)p$	0.897/0.845	0.932	0.241	88.5
4	$I(3)p$	0.680/0.637	0.761	0.434	44.5

Igor V. Tetko et al. (34) used Associative Neural Network and Multiple Linear Regression analysis of statistical methods in this study to confirm the QSAR models. After their QSAR study for 187 complexes they identified from the literature. 11 Pt (II) and 35 Pt (IV) compounds were preferred for this subject. The octanol/water partition coefficient of log P was chosen as experimental data because it is one of the most important physicochemical parameters for the development of new anticancer drugs with improved pharmacokinetic properties. The error of the consensus model, 0.65 for Pt (IV) and 0.37 for Pt (II) complexes, indicates that the predictions have good success. Less success for Pt (IV) complexes was attributed to experimental difficulties with log P measurements for some least soluble compounds. This model was formed using general descriptors such as molecular fragments, E-state indices and extended functional groups.

Hristo P. Varbanov (81) et al. found cytotoxic IC_{50} values for 53 new bis-, tris-, tetrakis-(carboxylato)platinum(IV) complexes (figure 2.5) that they synthesized on the system in their laboratory. They used the wb97x functional in the DFT for structural optimizations of the complexes and descriptors describing the activity. The cell lines of CH1 and SW480, which are cancer cell genes and

sensitive to CP, were used in cytotoxicity tests. They used multiple linear regression (MLR) and principal component analysis (PCA) methods to develop the QSAR models and analyze the data. They measured the robustness and predictability of the models in terms of R^2 and Q^2 (R^2 of cross-validated estimating, using the leave-two-out procedure-LTOP or leave one-out procedure-LOOP, RMS root mean squared error-RMSD, and average absolute error-AAR). The actual predictive ability of each model was checked by external validation by dividing the data set into training and prediction sets. In any case, the training set covers 75% of the data, while the remaining 25% is selected. As a result of the statistical calculations, the QSAR model for CH1 cells (R^2 = 85%, Q^2 = 82%) was constructed using the combination of MW, H_{don}, H_{acc} and E_{ea}s′ (table 2.2). Models consisting of Es, Hdon, Hacc and COOH (R^2 = 80%, Q^2 = 75%) and E_i, E_{ea}s, E_{ea}s′ and H_{don} (R^2 = 80%, Q^2 = 76%) were used for the SW480 cell line (table 2.3). According to these results, the proposed QSAR model is compatible with the descriptors and can successfully predict the biological activities of these complexes theoretically.

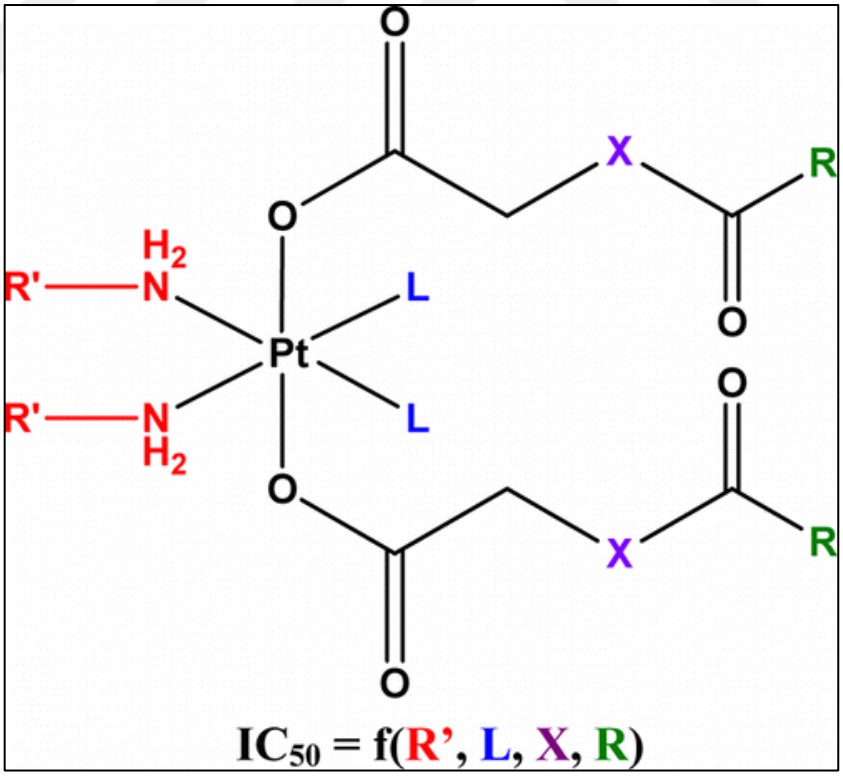


Figure 2.5. Octahedral platinum (IV) complexes in building system.

Table 2.2. The Best Regression Models in the CH1 Cell Line of the 53 Platinum (IV) Complexes for Cytotoxicity.

no. of variables	descriptors	R^2	Q^2(LTOP)	rms
1	H_{acc}	0.51	0.48	0.78
2	H_{don}, H_{acc}	0.70	0.67	0.62
3	q(Pt), $E_{ea}s'$, H_{don}	0.79	0.75	0.53
4	α, q(Pt), $E_{ea}s'$, H_{don}	0.86	0.82	0.45
4	MW, q(Pt), $E_{ea}s'$, H_{don}	0.85	0.81	0.47
4	MW, $E_{ea}s'$, H_{don}, H_{acc}	0.85	0.82	0.46
4	α, $E_{ea}s'$, H_{don}, H_{acc}	0.85	0.82	0.46
5	MW, α, q(Pt), $E_{ea}s'$, H_{don}	0.87	0.84	0.43
6	MW, α, q(Pt), $E_{ea}s'$, Es, H_{don}	0.88	0.84	0.43

Table 2.3. The Best Regression Models in the SW480 Cell Line of the 53 Platinum (IV) Complexes for Cytotoxicity.

no. of variables	descriptors	R^2	Q^2(LTOP)	rms
1	H_{acc}	0.63	0.60	0.75
2	q(Pt), H_{don}	0.73	0.70	0.65
3	Es, H_{don}, H_{acc}	0.77	0.73	0.61
4	Es, H_{don}, H_{acc}, COOH	0.80	0.75	0.59
4	E_i, $E_{ea}s$, $E_{ea}s'$, H_{don}	0.80	0.76	0.58
4	E_i, E_{ea}, $E_{ea}s$, H_{don}	0.82	0.79	0.54
5	E_{HOMO}, E_i, E_{ea}, $E_{ea}s$, H_{don}	0.84	0.80	0.52
5	q(Pt), H/Lgap, E_i, $E_{ea}s$, H_{don}	0.82	0.78	0.56
6	E_{HOMO}, E_i, E_{ea}, Es, $E_{ea}s$, H_{don}	0.85	0.81	0.51
7	E_{HOMO}, E_i, E_{ea}, Es, Es′, $E_{ea}s$, H_{don}	0.86	0.80	0.52

2.4 Structure Interactions Between DNA and Ligands

Platinum-based drug molecules show their effect by binding to DNA. DNA affinity is evaluated as an indicator of the effect of the drug on the cancer. Theoretical studies examining platinum-DNA interactions provide preliminary information to researchers in many respects.

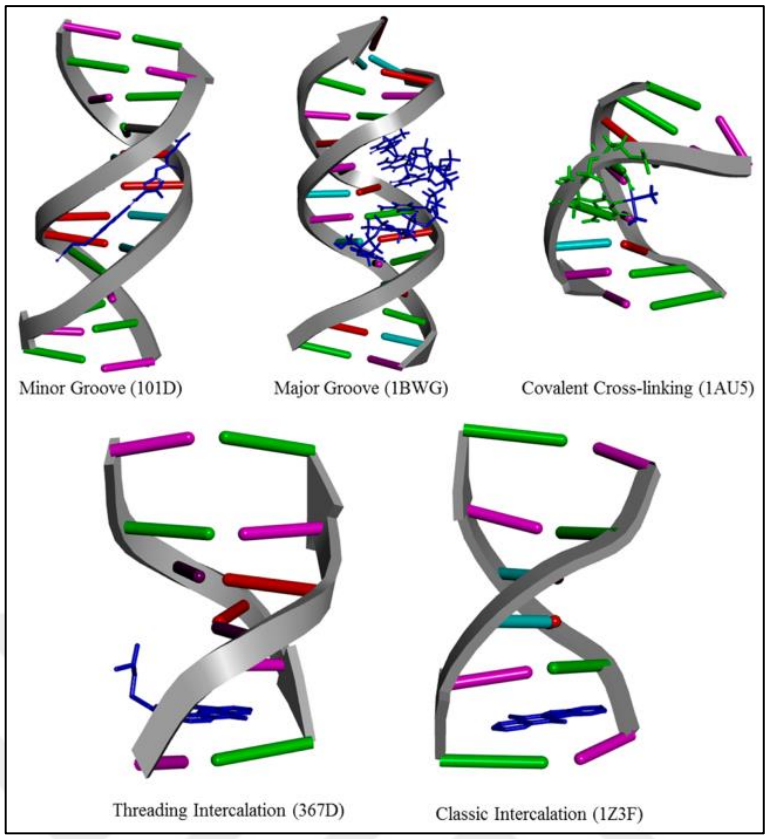


Figure 2.6. DNA-ligand binding modes. The ligand (stick content) is shown as blue. In 1AU5 the cross-linked bases are shown in a stick representation. Parentheses show PDB codes.

Yocheved Gilad and Hanoch Senderowitz (32) studied the complexes that figure out DNA-intercalator binding. They analyzed the binding sites of 63 high-resolution DNA-ligand intercalators (figure 2.6), which they downloaded from the Protein Data Bank (PDB) web site. As a result of this analysis, they found that ligands bind rather between G and C between the C and A base pairs (70% and 11%, respectively). For this test, they constructed decision trees to be able to generate different types of binding to known and unknown DNA regions. Here, they changed many parameters and settings in the AutoDock protocol. The percentages of success of bindings with RMSD < 2.00 Å were determined for all decision trees. In general, satisfactory results were obtained when ligands with known and unknown binding sites were docked into DNA constructs. The results of this study indicate that DNA intercalators can be used successfully in the docking procedure of the AutoDock program in drug design.

Ceyda icsel et al. (43) were examined synthesis, characterization, stability studies, interaction with DNA, DNA binding, molecular docking, antioxidant

properties, cellular uptake, and in vitro cytotoxic activity of new palladium (II) and platinum (II) 5,5-diethylbarbiturate complexes with 2,2'-dipyridylamine 2,2'-bipyridine and 2-phenylpyridine. The complexes were found to have a satisfactory binding ability by a non-covalent binding state and intercalation in accordance with molecular docking studies. PDB was scanned by the researchers and two forms of B-DNA, dodecamers and octamers, were downloaded and used for ligand docked (figure 2.7), respectively (d(CGCGATATCGCG)2 (1DNE) for groove binding and an octamer d(GAAGCTTC)2 (1DSC) for intercalation). In addition, complexes 1 and 2 displayed (figure 2.7) strong binding with supercoiled pUC19 plasmid DNA. Cellular uptakes were made to appreciate the subcellular localization of the current complexes. A moderate radical scavenging activity of 1 and 2 was confirmed by the DPPH and ABTS trials. Compounds of 1, 2 and 5 showed selectivity towards the HT-29 (column) cell line.

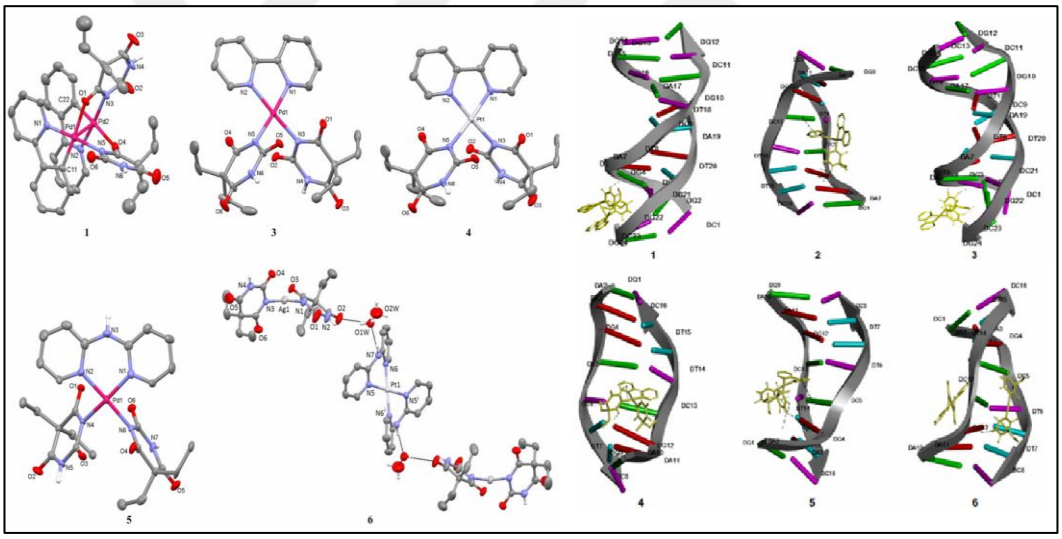


Figure 2.7. Platinum complexes and their DNA binding positions.

Hemant Kumar Srivastava and colleagues in their study (82) compared computational methods for ligands binding to Minor Groove DNA. Figure 2.8 shows some types of ligand-DNA interactions. Computational analyzes were performed for 57 non-covalent DNA minor grooves present in modeling DNA-ligand interactions of interest. In the study, firstly, a comparison of GLIDE, CDOCKER, GOLD and AUTODOCK docking programs was made. Next, quantitative structure stability relationship (QSSR) models were found and molecular dynamics simulations were developed. When the ligand docking

procedure was applied, the pose with the best score for the GOLD and GLIDE programs was found close to the lowest RMSD exposure (figure 2.9), and the deviation was detected at a lower value than the other docking protocols. Geometric optimizations were performed by applying the B3LYP functional, which is the DFT method for the structures, with the 6-31G* basis set, and the resulting quantum chemical descriptors were used to generate the best QSSR models. The best model was established for 46 ligands. Finally, the AMBER program was run for molecular dynamics simulations. Here, according to the results of the placement of the complexes in the water box of 10 Angstrom and in the initial minor groove regions, the changes after the production run after 5 ns were checked and whether there were fluctuations. The group of researchers thinks that the docking, molecular dynamics, and structure stability relationship approaches, which should further trigger the interaction, have successfully demonstrated, and validated their performance.

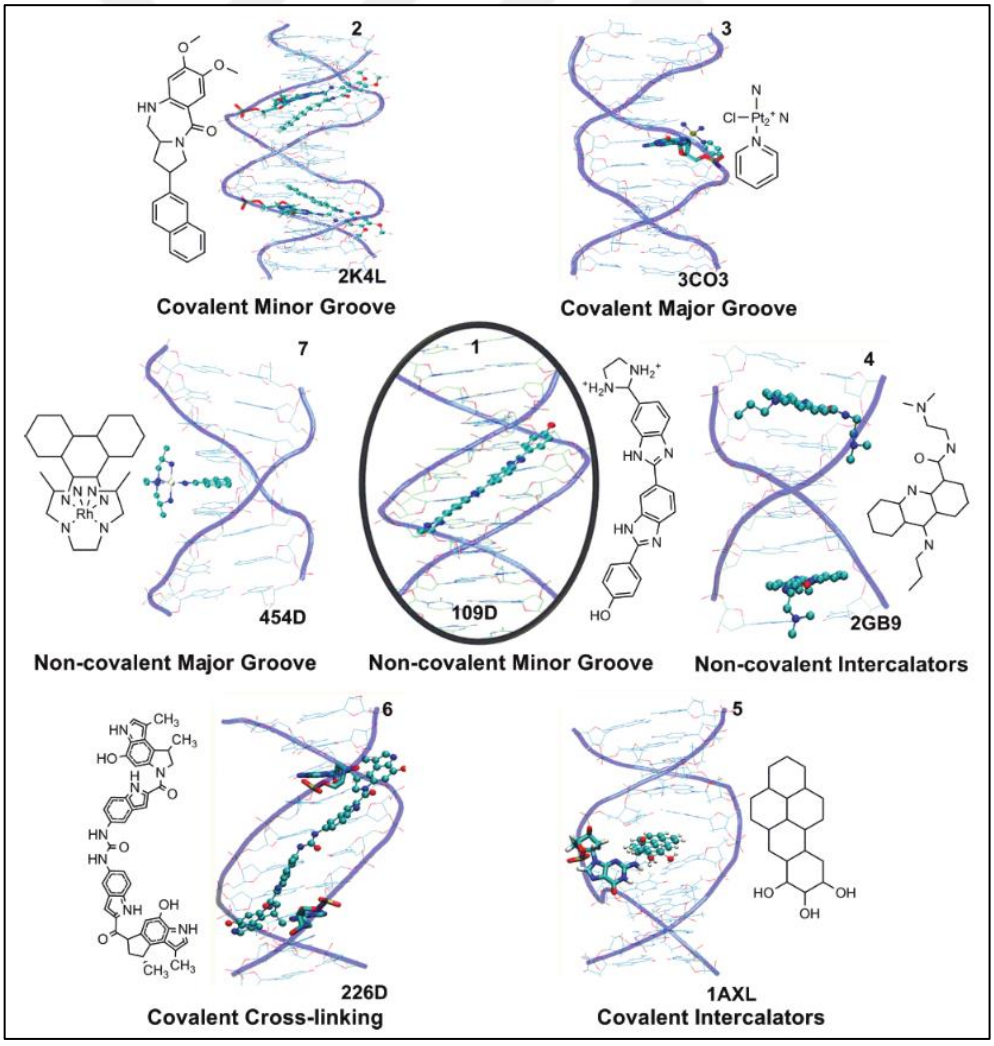


Figure 2.8. Some ligand-DNA interaction types and DNA binders

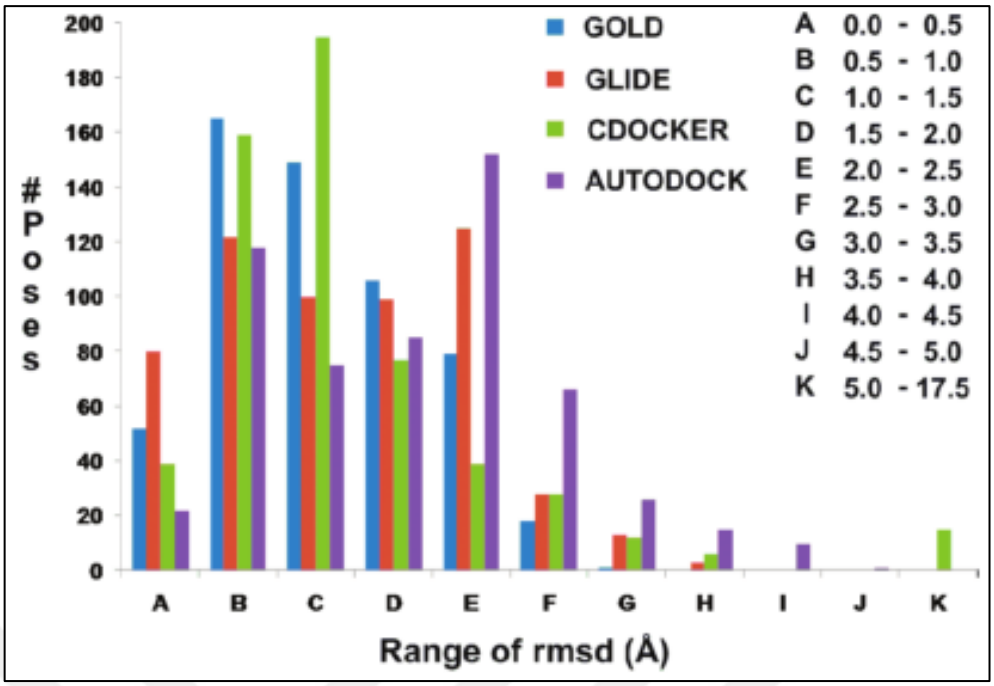


Figure 2.9. Collation of RMSD of formed ligand poses determined from the GLIDE, CDOCKER, GOLD and AUTODOCK docking from the experimental insertion.

3. MATERIALS AND METHODS

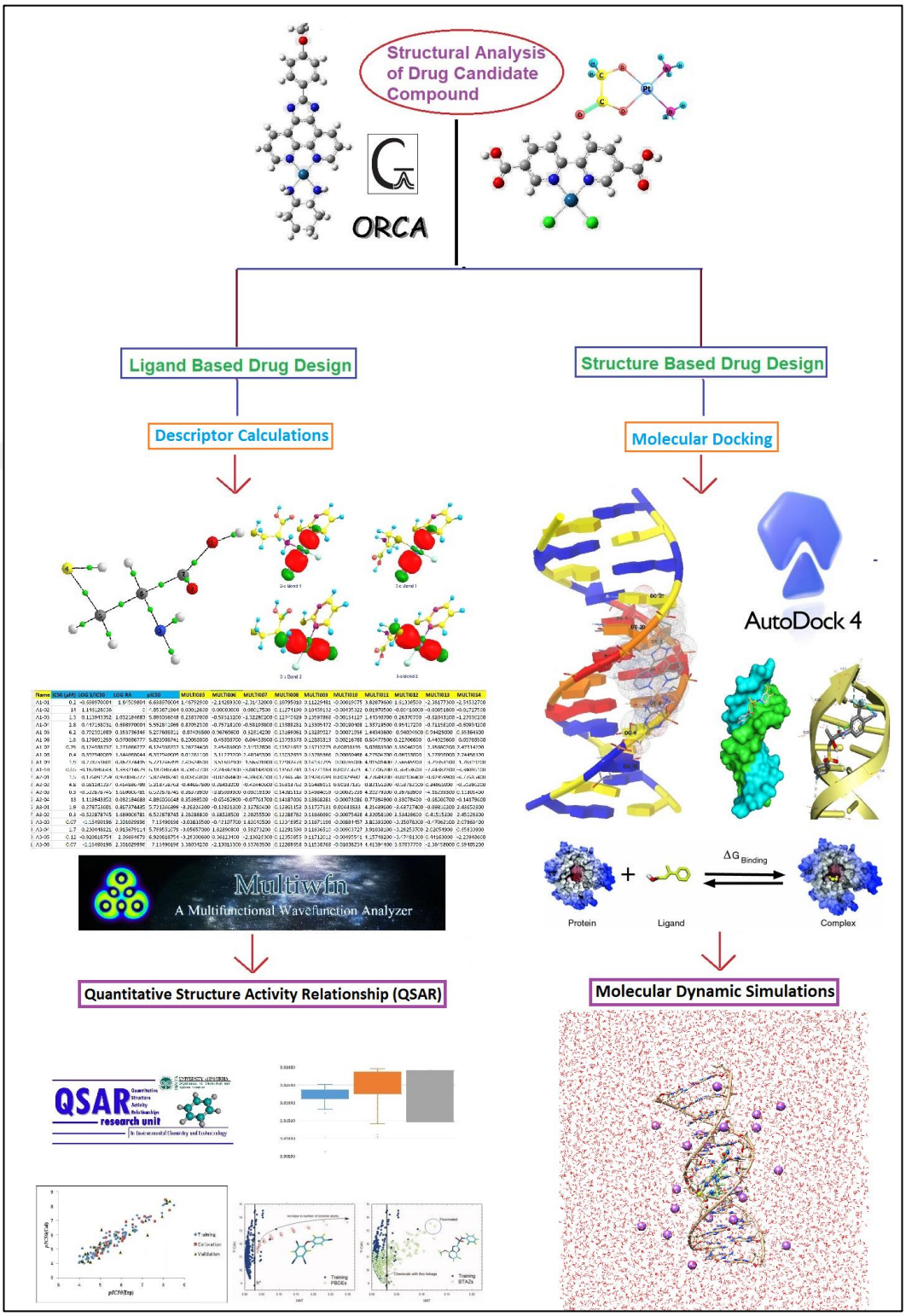


Figure 3.1. Simple flow diagram for Calculations

Many software and computer systems have been used in the calculations. According to the flow diagram in the figure 3.1, the calculations of the structural

parameters and geometries for the drug candidate compounds, with the Gaussian 03 and ORCA 4.2 programs; descriptor calculations with Multiwfn 3.7, ORCA 4.2 and PaDEL-Descriptor; QSAR analyses, QSARINS 2.2.4; molecular docking calculations with AutoDock 4.2.6, AutoDockTools 1.5.6 and python-based scripts; molecular dynamic simulations were made with the AMBER 20 program.

During the calculations, computers at the National High Performance Account Center (UHEM) affiliated to Istanbul Technical University, Tübitak ULAKBİM TRUBA and a home computer with intel i9 9900k processor were used. Calculations at UHEM were made depending on project number 4007392020. ORCA 4.2 software was installed on TRUBA and UHEM systems. input files of the calculations were sent as a job by providing remote access with the slurm script I created in these Linux-based systems. QSARINS 2.2.4, Gaussian 03, AutoDock 4.2.6, Multiwfn, PaDEL-Descriptor and AutoDockTools 1.5.6 programs and all other viewing programs (Chemcraft, Gaussview 3, BIOVIA Discovery Studio, and Chimera 1.15) were used on a home computer with an intel i9 9900k processor. In addition, all molecular simulations were made with the GeForce 1080ti graphics card and the GPU architecture integrated AMBER 20 program and its plugins. Apart from these, many necessary additional software and web-based supporting programs are used.

3.1 Determination of Initial Settings and Testing Critical Keywords Before Choosing Suitable DFT Method and Basis Set

Before finding the most stable chemical and physical structure for heavy metal-containing complexes, critical features of the software to be used were tested and additional parameters and settings to be used during the main quantum chemical calculations were determined. Initial calculations were based on Density Functional Theory (DFT). Before choosing the most suitable methods for the structures, preliminary studies were carried out with the ORCA 4.2 program.

ORCA 4.2 (83-85) is an ab initio quantum chemistry program package that consist of modern electronic structure methods bearing density functional theory, coupled cluster, multireference methods, many-body perturbation, and semi-empirical quantum chemistry methods. Its main field of application is transition metal complexes, larger molecules, and their spectroscopic properties. ORCA 4.2 has many useful keywords. The contributions of these keywords in solving the geometry of metal-containing complexes were examined in detail with various

variations. ORCA 4.2 software has some competencies as well as limitations and deficiencies. Therefore, some vital settings were chosen very well before starting the method and basis set selection also especially complexes including heavy metal. Number of grid points to be used on atoms to prevent geometrical structure errors, the self-consistent field (SCF) energy, geometry optimization algorithm, tolerance values that must be applied for geometric structure, SCF iterations, calculation times, the effect of NORI, RIJK, and RIJCOSX approximations on calculations, auxiliary basis set and initial guess effects were tested.

The other important state in the DFT calculation is the size of the integration grid point for numerical integration. 3D quadrature of numerical is necessary due to the complex and challenging analytical form for correlation function of exchange and energy. The values for these quantities are so sophisticated that there is no chance of the estimating analytical solutions to the required integrals and several numerical approach, usually 3D numerical integration, is essential. The numerical integration (86-91) is a main part in DFT calculation.

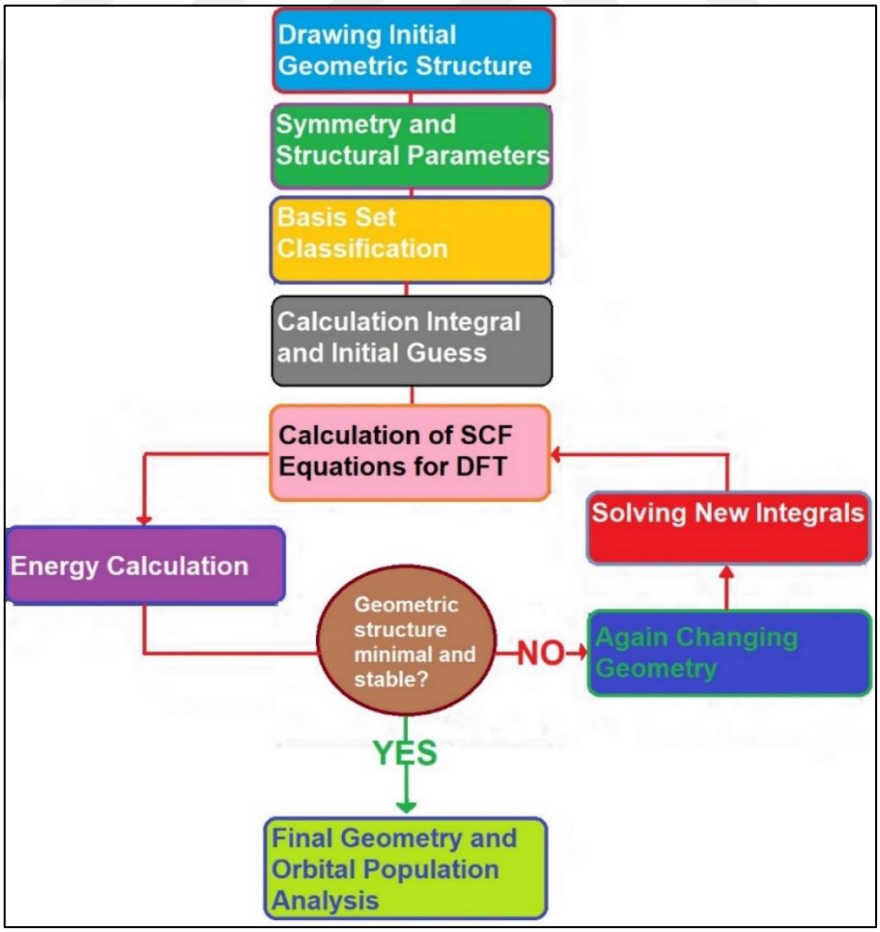


Figure 3.2. Simple geometric optimization flow chart.

Most approaches are based on the Resolution of Identity (RI) approach (89-95) (also called Density Fitting). Using these approaches speed up your ORCA 4.2 calculations significantly and is generally recommended. Calculations can often be done with or without these approximations. The use of the RI approach always requires an auxiliary basis set, and its choice depends on which integrals are approached and which basis set is used.

Geometric optimizations and DFT operation logic are theoretically constructed by software according to the diagram in figure 3.2. The Nedaplatin complex (96) for the initial calculations (figure 3.3), which had been experimentally characterized in the literature also used in the market as a drug, was chosen in the studies. Structural changes and deviations from experimental geometric structures were compared for the combination of many different parameters. The Nedaplatin structure is deposited at CCDC (https://www.ccdc.cam.ac.uk) and has the code 706387. The compound was obtained from here.

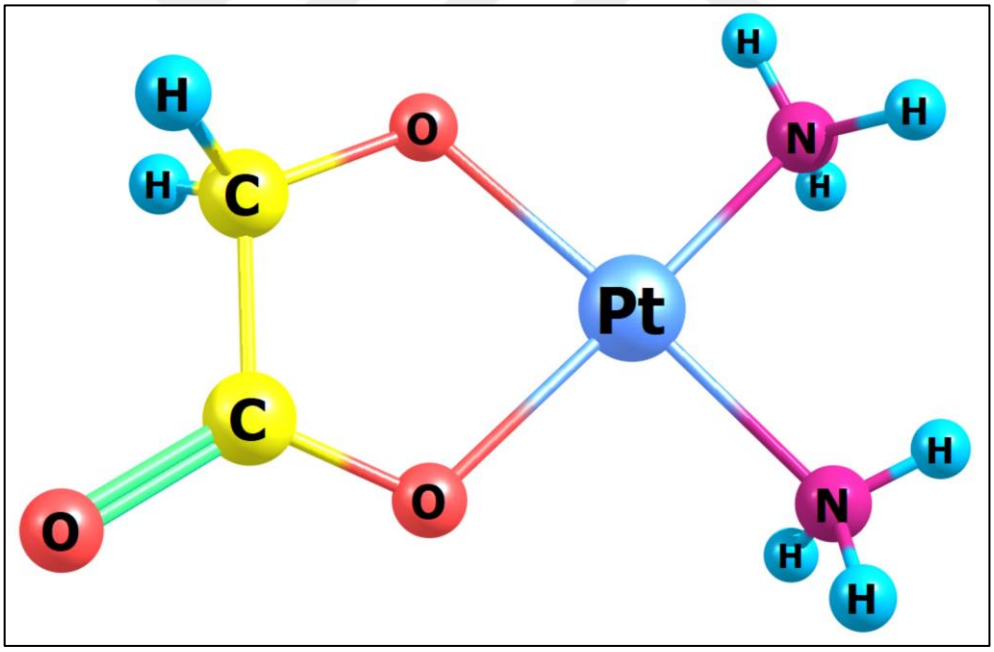


Figure 3.3. Nedaplatin structure.

Known DFT methods of M06L with high grid sensitivity and BP86 were used to determine the most accurate grid points, approximation, and other useful keywords for these tests. Platinum atom has grid sensitivity due to being a heavy metal. Therefore, the high grid values were chosen.

The Chemcraft program, which has a free trial version, was used to display the complex structure, and convert between formats. The Chemcraft program includes a set of graphical tools to facilitate working with quantum chemistry calculations and helps to prepare new work for computation and analyze the calculated results. ORCA uses “xyz” as the chemical compound format. The complex downloaded from CCDC in “mol” format was converted to “xyz” type. Next, the “xyz” coordinates are added to the input file of the ORCA in the calculation. Inputs were prepared by using the keywords in table 3.1 for different combinations. Many tests have been carried out. The resulting output files were opened with Chemcraft and examined one by one. The results were evaluated according to parameters such as total energy, calculation times and optimization step numbers, and the effects of keywords.

Table 3.1. Useful ORCA 4.2 keywords for DFT calculations, tested by preliminary work.

Definations	**ORCA Keywords**
Calculation Type	OPT (Geometric Optimization)
Shell Type	RKS (Closed-Shell DFT)
RI Methods	NORI (Not Used RI Approx.), RI & RIJCOSX
Split RI	NoSplit-RI-J & Split-RI-J
Auxiliary Basis	AutoAux (Automatic Auxiliary Basis) & DEF2-J
DFT Method	BP86 & M06L
Metal ECP Basis Set	def2-ECP
All Atom Basis Set	def2-TZVP
SCF Tolerance	TightSCF & VerytightSCF
Numerical Grid	GRIDn (n: 3,4,5,6,7)
Numerical Final Grid	FINALGRIDn (n: 3,4,5,6,7)
DFT Weight Scheme	Weight_Becke & Weight_AtomXC
Optimization Algorithm	RFO, QN & GDIIS
Initial Guess	MoRead, Hcore & Hueckel

3.2 Ligand-Based Drug Design

Ligand-based drug design (LBDD) was performed according to the flow diagram in figure 3.1. In the calculations, the determination of the most suitable

DFT method and basis set combinations, a structurally characterized complex with a platinum atom was selected from the literature for comparison (bond parameters). As a result of a series of calculations and certain statistical standard deviation analyzes, the appropriate DFT method and basis set combination was found. The literature review was performed and studies that included in-vitro cytotoxic tests for the human ovarian cancer gene A2780 were listed. The studies with platinum-containing complexes were excluded from these articles. These studies were separated according to their experimental methods and own groups (A-G). Subsequently, the platinum complexes in these studies were designed. Only those with experimental geometrical structures were obtained from CCDC. The categorized complexes, except the experimental ones, were drawn according to their 2D shapes as 3D. The suitable and stable geometric conformers were determined by conformational analysis. Geometric optimizations for the determined method and basis set combinations were performed for all complexes. Then, novel complexes containing 2-pyrimidinethiol, which is the subject of the study, were drawn and geometric optimizations were made by applying the same method and basis set combination. Then, many useful descriptor calculations were performed for complexes of newly designed and taken from the literature. QSAR analysis was performed for the complexes taken from the literature, the descriptors that best explained the cytotoxic activity were selected and the theoretical cytotoxic activity values were calculated. The best selected equations were applied to the newly designed complexes, each in its own calculation group, and the theoretical activity values were determined for novel compounds.

3.2.1 Determination of the Best DFT Method

It is very important that the complexes, which form the basis of the calculations to be made in the next process, have the most stable and correct geometric structure. DFT and basis set combinations were applied to the complexes and the selection of the best combination was determined by a series of statistical elimination methods to achieve this. After the determination of useful keywords and necessary algorithms, a reference complex with experimental crystal structure parameters was found for the DFT method and basis set tests. A structure containing sulfur, nitrogen and chlorine bonded to platinum atom was scanned for this reality. After short literature review, the complex in figure 3.4 with CCDC code of 160941, whose crystal structure was defined by D. Kovala-Demertzi et al. (97), was selected.

Metal-containing complexes are structurally accurately elucidated by quantum chemical calculations. There are many calculation methods to solve the structure. One of the most used calculation methods is DFT that were based on functions with many features related to many subgroups such as generalized gradient approximation (GGA), hybrid-GGA, hybrid-meta-GGA, meta-GGA and range-separated. These functions were used with known basis sets grouped in many different categories as Pople, Stutgart, Ahlrichs, etc. According to the features provided by programs, these methods and basis sets can be selected ready-made within the ORCA 4.2. Some DFT functions can be selected out of the box from ORCA, while others are imported from the LibXC (www.tddft.org/programs/libxc) system to ORCA. LibXC is a library of exchange-correlation and kinetic energy functionals for DFT. References and descriptions of all DFT functions in LibXC can be found at "www.tddft.org/programs/libxc/functionals" web address.

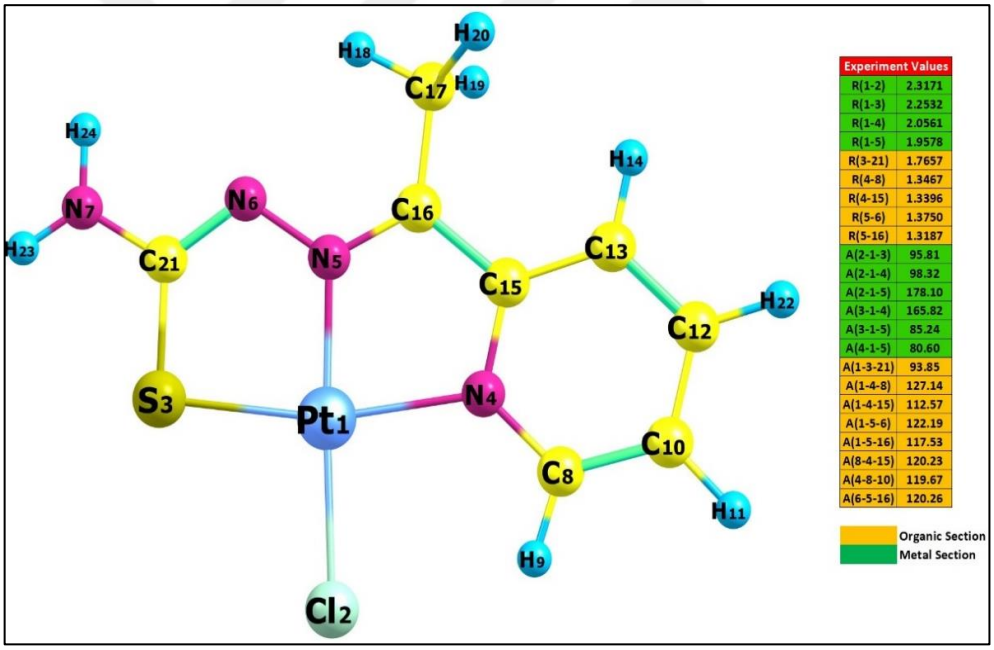


Figure 3.4. The complex selected; (2-acetyl pyridine thiosemicarbazonato)-chloro-platinum (II)

All electrons in a complex must be considered for future descriptor calculations for the determination of the quantum chemistry parameters The full electron approach (98-103) system with zero order regular approximation (ZORA) was used instead of the effective core potential (ECP) basis set.

All DFT methods to be used in comparison calculations were selected and categorized as in table 3.2 and 3.3. It was numbered in the best way for better understanding. So, GGA, hybrid-GGA, hybrid-meta-GGA, meta-GGA and range-separated were grouped as DFT1, DFT2, DFT3, DFT4 and DFT5, respectively. DFT functions that fall into these groups are listed. Some of the DFT functions were selected from the ORCA program, while others were selected from the LibXC library and included in the calculations. The functions in each group are numbered starting from one. ZORA relativistic approximation system was used for calculations as a basis set. In this system, SARC-ZORA-TZVP was used for metal basis set and ZORA-def2-TZVP was used for non-metal atoms.

Table 3.2. DFT functions and numbering system used for comparison taken ORCA. The purple, blue, green, pink, and red boxes contain the numbering of the DFT function as GGA, hybrid-GGA, hybrid-meta-GGA, meta-GGA and range-separated, respectively.

DFT1-01 B97	**DFT1-02** BLYP	**DFT1-03** BP86	**DFT1-04** MPWLYP
DFT1-05 MPWPW	**DFT1-06** OLYP	**DFT1-07** OPBE	**DFT1-08** PBE
DFT1-09 PW86PBE	**DFT1-10** PW91	**DFT1-11** PWP	**DFT1-12** REVPBE
DFT1-13 RPBE	**DFT1-14** RPW86PBE	**DFT1-15** XLYP	**DFT1-16** B97-D3BJ
DFT1-17 BLYP-D3BJ	**DFT1-18** MPWLYP-D3BJ	**DFT1-19** OLYP-D3BJ	**DFT1-20** PBE-D3BJ
DFT1-21 PW91-D3BJ	**DFT1-22** REVPBE-D3BJ	**DFT1-23** RPBE-D3BJ	**DFT1-24** RPW86PBE-D3BJ
DFT1-25 XLYP-D3BJ	**DFT2-01** B1PW91	**DFT2-02** B3LYP	**DFT2-03** B1LYP
DFT2-04 B3P86	**DFT2-05** B3PW91	**DFT2-06** BHANDHLYP	**DFT2-07** MPW1LYP
DFT2-08 MPW1PW	**DFT2-09** O3LYP	**DFT2-10** PBE0	**DFT2-11** PBEH-3C
DFT2-12 X3LYP	**DFT2-13** B3LYP-D3BJ	**DFT2-14** B1LYP-D3BJ	**DFT2-15** B3P86-D3BJ
DFT2-16 B3PW91-D3BJ	**DFT2-17** O3LYP-D3BJ	**DFT2-18** PW6B95-D3BJ	**DFT2-19** PBE0-D3BJ
DFT2-20 X3LYP-D3BJ	**DFT3-01** M06	**DFT3-02** M06-2X	**DFT3-03** TPSS0
DFT3-04 TPSSH	**DFT3-05** TPSS0-D3BJ	**DFT3-06** TPSSH-D3BJ	**DFT4-01** M06L
DFT4-02 revTPSS	**DFT4-03** TPSS	**DFT4-04** revTPSS-D3BJ	**DFT4-05** TPSS-D3BJ
DFT5-01 wB97	**DFT5-02** wB97X	**DFT5-03** CAM-B3LYP	**DFT5-04** LC-BLYP

Table 3.3. DFT functions and numbering system used for comparison taken LibXC. The purple, blue, green, pink, and red boxes contain the numbering of the DFT function as GGA, hybrid-GGA, hybrid-meta-GGA, meta-GGA and range-separated, respectively.

DFT1-26 HCTH407-D3BJ	**DFT1-27** B97-GGA1	**DFT1-28** EDF1	**DFT1-29** HCTH120
DFT1-30 HCTH147	**DFT1-31** HCTH407	**DFT1-32** HCTH407P	**DFT1-33** HCTH93
DFT1-34 HCTHP14	**DFT1-35** KT2	**DFT1-36** MOHLYP	**DFT1-37** MOHLYP2
DFT1-38 MPWLYP1W	**DFT1-39** PBE1W	**DFT1-40** PBELYP1W	**DFT1-41** TH1
DFT1-42 TH2	**DFT1-43** TH3	**DFT1-44** TH4	**DFT1-45** TH-FC
DFT1-46 TH-FCFO	**DFT1-47** TH-FCO	**DFT1-48** VV10	**DFT1-49** OPBE-D
DFT1-50 OBLYP-D	**DFT1-51** OPWLYP-D	**DFT2-21** B97-1-D3BJ	**DFT2-22** B97-2-D3BJ
DFT2-23 B1WC	**DFT2-24** B3LYP5	**DFT2-25** B3LYPS	**DFT2-26** B5050LYP
DFT2-27 B97-1P	**DFT2-28** B97-1	**DFT2-29** B97-2	**DFT2-30** B97-3
DFT2-31 B97-K	**DFT2-32** BHANDH	**DFT2-33** CAP0	**DFT2-34** EDF2
DFT2-35 HPBEINT	**DFT2-36** MB3LYP-RC04	**DFT2-37** MPW1K	**DFT2-38** MPW1PBE
DFT2-39 MPW3LYP	**DFT2-40** MPW3PW	**DFT2-41** MPWLYP1M	**DFT2-42** PBE0-13
DFT2-43 PBE50	**DFT2-44** REVB3LYP	**DFT2-45** SB98-1A	**DFT2-46** SB98-1B
DFT2-47 SB98-1C	**DFT2-48** SB98-2A	**DFT2-49** SB98-2B	**DFT2-50** SB98-2C
DFT3-07 MPW1B95-D3BJ	**DFT3-08** MPW1KCIS-D3BJ	**DFT3-09** MPWB1K-D3BJ	**DFT3-10** MPWKCIS1K-D3BJ
DFT3-11 PBE1KCIS-D3BJ	**DFT3-12** PW6B95-D3BJ	**DFT3-13** REVTPSSH-D3BJ	**DFT3-14** TPSS1KCIS-D3BJ
DFT3-15 B86B95	**DFT3-16** B88B95	**DFT3-17** BB1K	**DFT3-18** MPW1B95
DFT3-19 MPW1KCIS	**DFT3-20** MPWB1K	**DFT3-21** MPWKCIS1K	**DFT3-22** PBE1KCIS
DFT3-23 PW6B95	**DFT3-24** PW86B95	**DFT3-25** PWB6K	**DFT3-26** REVTPSSH
DFT3-27 TPSS1KCIS	**DFT3-28** X1B95	**DFT3-29** XB1K	

Geometric optimizations were made with the "compound" system in the ORCA 4.2 program choosing the best DFT method. This system allows many inputs to be made with a single calculation. “Compund” input example is in

Appendix 1. Also, the inputs are run with the slurm script. The slurm script sample and operating codes run in Tübitak TRUBA are in Appendix 2.

As a result of optimizations, calculated bond parameters were compared with experimental bond properties using some statistical methods and the method with the lowest standard deviation was chosen. Statistical analyzes were made by comparing the bond properties of the metal and the whole complex separately. MAPE (Mean Absolute Percentage Error), MAD (Mean Absolute Deviation), MSE (Mean Squared Error) and RMSE (Root Mean Square Error) were used for statistical analysis. Method selection was made with reference to MAD, which was one of the most used analyzes. In the following order (1-4) formulas are expressed:

$$MAD = \sum_{i=1}^{n} \frac{\left|Y_i - \hat{Y}_i\right|}{n} \quad (1)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \frac{\left|Y_i - \hat{Y}_i\right|}{Y_i} \quad (2)$$

$$MSE = \sum_{i=1}^{n} \frac{\left(Y_i - \hat{Y}_i\right)^2}{n} \quad (3)$$

$$RMSE = \sqrt{MSE} \quad (4)$$

3.2.2 Literature Search and Finding Appropriate Platinum Complexes Against A2780 Ovarian Cancer Cell Line

A comprehensive literature review was made. Studies with platinum complexes that have been tested in-vitro against the A2780 ovarian cancer cell gene and reliably cytotoxic (IC_{50}) values was stored. After the literature review was completed, all identified studies were reviewed. In-vitro methods in the studies; as the chemical used in the dyeing, the incubation time, the solvent that dissolves the complex, the type of medium used, the number of the tested complex and the number of bonds made by the platinum atom were categorized in table 3.4. Thus, it was ensured that the experimental error caused by different studies was minimized.

The A2780 human ovarian cancer cell has been extensively reviewed for in-vitro studies with platinum-containing complexes. All the articles with IC_{50} cytotoxic numerical values among the in-vitro studies on the A2780 cancer gene were selected and categorized by means of the ZOTERO program according to the criteria determined from hundreds of studies from these highly sensitive studies. As

a result, 19 articles listed in table 3.4, which may be suitable for the QSAR analysis to be done in the continuation of the studies, were selected.

Table 3.4. Selection criteria in articles selected from the literature.

No	Group	Method	Inc. Time (h)	Solvent	Type of Medium	Number of Complex	IC_{50} (µM) Control Value (CP)	Ref
1	A1	SRB	96	TCA	DMEM	10	0.2 ± 0.05	51
2	A2	SRB	96	TCA	DMEM	4	0.2 ± 0.07	52
3	A3	SRB	96	TCA	DMEM	6	0.3 ± 0.06	53
4	B1	MTT	72	DMSO	RPMI 1640	12	1.3 ± 0.2	54
5	B2	MTT	72	DMSO	RPMI 1640	4	2.5 ± 0.3	55
6	B3	MTT	72	DMSO	RPMI 1640	3	0.18 ± 0.07	56
7	B4	MTT	72	DMSO	RPMI 1640	6	2.64 ± 0.61	57
8	B5	MTT	72	DMSO	RPMI 1640	3	2.2 ± 1	58
9	C1	SRB	96	Water	DMEM	27	0.53	59
10	D1	MTT	48	DMSO	RPMI 1640	26	1.90 ± 0.20	60
11	E1	MTT	120	DMSO	RPMI 1640	16	1.37 ± 0.48	61
12	F1	MTT	96	DMSO	RPMI 1640	18	2.72 ± 0.58	62
13	G1	SRB	96	DMSO	RPMI 1640	11	0.55 ± 0.03	63
14	G2	SRB	96	DMSO	RPMI 1640	2	0.6 ± 0.1	64
15	G3	SRB	96	DMSO	RPMI 1640	1	0.55 ± 0.01	65
16	G4	SRB	96	DMSO	RPMI 1640	1	0.83 ± 0.02	66
17	G5	SRB	96	DMSO	RPMI 1640	2	0.44 ± 0.04	67
18	G6	SRB	96	DMSO	RPMI 1640	1	0.88 ± 0.01	68
19	G7	SRB	96	DMSO	RPMI 1640	1	0.88 ± 0.01	69

All articles in this list were grouped as per specific criteria, according to in-vitro experimental condition performed by the study groups. The following similarity criteria were used for this grouping as distinctive. Group numbers were given for A-G for these articles. The articles in the group were defined starting from number one. Another issue that is taken into consideration in the articles is whether it is done by the same working groups.

3.2.3 Initial Structural Analysis of Complexes Detected from The Literature

The most accurate determination of the initial structures of the complexes is the most important factor that determines the course of the later stages of the studies. From the literature, 154 platinum-containing compounds were identified, which were grouped and included in the calculations. Those with single crystal structures were downloaded from the CCDC site by finding their identifier codes. Those that do not have a single crystal structure were drawn in 3D with Spartan 18

and Chemcraft programs, which have limited time use licenses. The 2D drawings in the studies where the complexes were found were taken as reference for these designs. Two methods have been applied to identify the initial structures, the manual drawings of the complexes and the crystal structures taken from the CCDC.

In the first method in which the crystal structure of the complexes could not be found in the CCDC database, the complexes with 2D representations in the study were drawn by considering the whole structure through the Spartan 18 program. Later, the complexes were subjected to conformational analysis with the Spartan 18 program. In this analysis, the roughly drawn 3D structure was transferred to the program. Conformation distribution process was chosen. It is provided to keep the results up to 1000 kcal energy and up to 10000 structures for pre-settings. In addition, the turning groups and bonds were checked if missing. Molecular mechanics method was used for this scan. The conformers found after screening were sorted according to their energy. Geometric optimizations were performed for other conformers with the DFT B3LYP functional and LACVP basis set. The LACVP (104-105) series of basis sets are a combination of the successful 6-31G basis set with the LANL2DZ effective core basis set. Specifically, the atoms H-Ar are described with the 6-31G (or 6-31G*, 6-31+G** etc.) basis set while heavier atoms are modeled using the LANL2DZ basis set. After that, the structures were ranked according to their energies, and the lowest energy structure was selected for further calculations.

In the other method, the complexes with crystal geometries are selected from the CCDC database by the identifying codes given in the article and selected as the initial structure. The "mol" format compounds downloaded from CCDC were checked after they were opened with the Chemcraft program.

The coordinate system format used in ORCA of "xyz" was created for all complexes using Chemcraft and Spartan.

3.2.4 Geometric Optimizations of Complexes Taken from Literature

As a result of the calculations, using the best DFT method determined as described in the section 3.2.1, geometric optimizations were made for 154 complexes in different groups and the most stable structures were found. All geometric optimizations were carried out using the ORCA 4.2 program, which we installed on Tübitak's TRUBA and Istanbul Technical University's UHEM high performance computing systems. Geometric optimizations were made using the

same parameters for all complexes to ensure computational integrity. All calculations were performed using the "compound" scheme, exemplified in Appendix 1, using multiple inputs allowed by the ORCA 4.2 program. As useful keywords, optimization and SCF energy parameters with tight tolerance are applied to achieve better geometry and stable energy. In addition, a high grid frame was added to the complexes and a large-scale population analysis was requested for bonds, orbitals, and some charge methods. The coordinate system in "xyz" format, which was previously created by spartan and Chemcraft, was added to the inputs.

3.2.5 Designs and Geometric Optimizations of Novel Platinum 2-Pyrimidinethiol Complexes

After revealing the stable geometric structures of all available platinum-containing compounds detected in the literature, novel complexes containing platinum and 2-pyrimidinethiol, which could be potential drug candidates with unknown biological activity, were designed for QSAR analysis. In the design of new complexes, the skeletal structures of the complexes in the groups, which were already determined from the literature, were used as reference. By preserving the main skeleton structure, 2-pyrimidinethiol structure was added together with all possible conformers instead of changing functional groups in the structures. For example, when the 20 complex structures in group A are examined, chlorine atoms are definite present as -cis and -trans positions in all compounds. In some structures, NH3 is present together with chlorines. In the A3 group, there is no NH3 at all. While designing the new complexes, chlorine atoms are protected in -cis and -trans positions. 2-pyrimidinethiol compound was added by replacing other non-common functional groups. By applying this theory to all groups, structures were designed and formed. The novel designed complexes are put together with the suffix "X" next to the group name to represent that group according to which group it was created. For example, the design of complex 1, which belongs to group A, was designated AX-1. The same definitions were made for the new complexes in all groups.

The most stable and low-energy conformation determination was made for 30 new complexes, whose design and drawing were completed. Here, the possible orientations of the added 2-pyrimidinethiol compound were determined, since the main skeleton of complexes taken from literature was clear. Geometric optimizations of all conformations designed and drawn according to these

orientations were made with the ORCA 4.2 program. As a result of these optimizations, all conformers were ranked according to their energies. For the structures whose conformer analysis was completed, geometric optimizations were carried out with the ORCA 4.2 program by applying the input parameters and other settings described in the previous section.

3.2.6 Calculation of Useful Descriptors for QSAR

Many useful chemical and physical parameters of the complexes are used the most commonly properties to predict biological activity theoretically through a set of statistical methods. Valuable descriptors have been calculated through various programs and categories.

3.2.6.1 Descriptors Taken from The Final Structures of The Complexes

Some information of the output files belong the most stable geometric structures found as result of the calculations mentioned in all available and created complexes were taken to be used as descriptors. As a result of the calculations, the ORCA program creates an "out" output file. Normally, Chemcraft did not have the ability to display "compound" multi-input separately from the output file. As a result of contacting Chemcraft software support, this feature has been included in the program with a new software update. The results of all complexes opened from the output files were examined. 127 properties including dipole moment, bond properties of platinum atom, Loewdin and Mulliken reduced orbital charges and atomic charges of platinum atom, SCF energies and HOMO-LUMO orbital energies, were determined as descriptors for all complexes with the use of this feature of Chemcraft.

3.2.6.2 Descriptors Taken from The NMR Calculations of The Complexes

NMR parameters were found for each complex used as an descriptor in subsequent QSAR analyses. For this, the best DFT method was determined by using ZORA relativistic full electron basis set. ORCA 4.2 program was used for NMR calculations.

There is a restriction due to the ORCA program in NMR calculations. This restriction, NMR calculations cannot be made with the methods added from the LibXC library among the DFT methods listed in table 3.3. So, only calculations were made with the DFT methods included in the ORCA library.

In the calculations, cisplatin experimental NMR data were selected as comparisons. The reference compounds were selected when calculating the NMR shielding value. Nitromethane for nitrogen atoms, TMS for hydrogen atoms and hexaplatinated for platinum atom were selected as the reference compounds. As expressed in formula 5, isotropic shielding value was taken for the atoms in the reference compound and cisplatin to calculate the chemical shift (ppm) value. For example, σ_{MeNO2} is the isotropic shielding value for the atom in the reference compound.σ is the isotropic shielding value for the same atom in cisplatin. This formula can be applied to other atoms. Only the chemical shift value was calculated for the platinum atom common to all complexes.

$$\delta_{calc} = (\sigma_{MeNO2} - \sigma)/(1 - 10^{-6}\sigma_{MeNO2}) \qquad (5)$$

The DFT functions in table 3.2 were used to determine the best method for NMR calculations. The best DFT method selection was made for NMR. Both optimization and NMR calculations were made with the same DFT method and basis set. The same settings were chosen in the method, basis set and other parameters to ensure reliable results in the calculations of cisplatin and other reference complexes. The compound system in ORCA was also used here. SARC-ZORA-TZVP for metal atoms; ZORA-def2-TZVP basis sets for other atoms were integrated as ZORA relativistic full electron basis sets. Also, the RIJCOSX approach is also added to the inputs.

Mode	Pt	N	H
Value	-2101.00	-424.90	3.93

Figure 3.5. Pt, N and H-NMR experimental values in CP

Statistical deviation methods in previous formula 1-4 were used to detect errors caused by the methods, and as a result, the success of DFT methods was compared statistically using NMR shielding data taken from reference compounds and cisplatin complex.

After finding the best DFT method for NMR chemical shift value of CP, NMR calculations were performed for 154 complexes taken from the literature and 30 novel designed complexes. As a result of these calculations, 42 descriptors, including shielding values, chemical shift data, magnetic tensor shielding

contributions and diagonalized sT*s matrix of platinum atom, were determined from the ORCA output file for each complex.

3.2.6.3 Descriptors Taken from Quantum Theory Atom in Molecules

The Multiwfn 3.7 program was used for all quantum descriptor calculations. Multiwfn (106) is an extremely powerful program for realizing electronic wavefunction analysis, which is a key ingredient of quantum chemistry. Multiwfn is free, open-source, high-efficient, very user friendly and flexible. The program works with “cmd” screen. First, the wavefunction file, which is one of the outputs of the program for which the DFT calculations are made, is called to the Multiwfn 3.7. Then, any desired calculation is done by selecting the numbers to the left of the program types shown in figure 3.6.

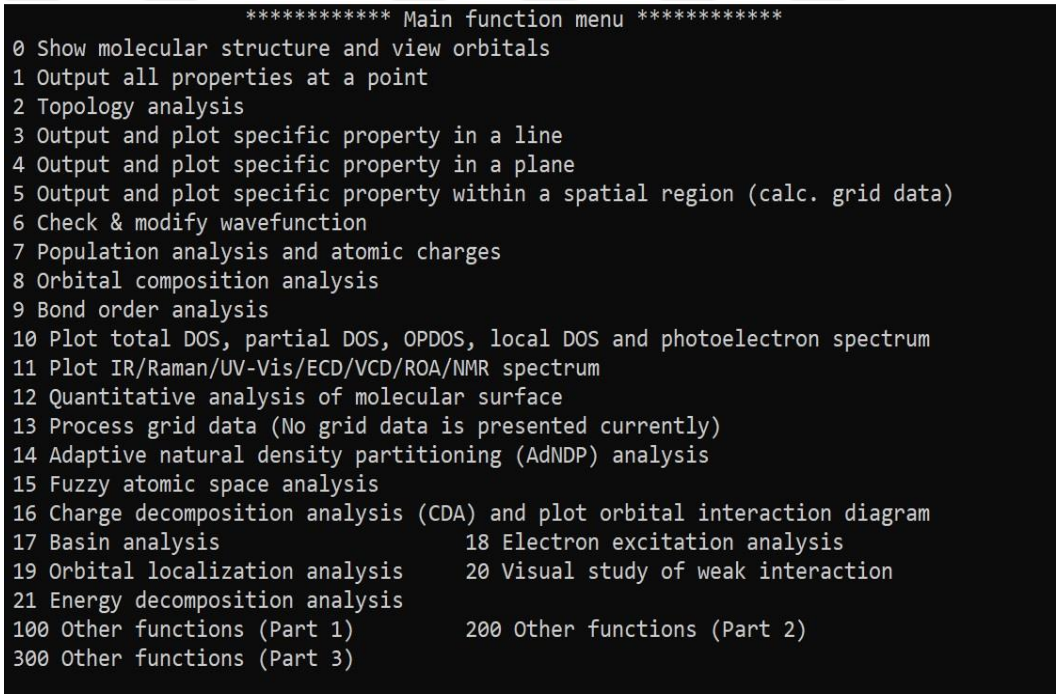


Figure 3.6. Multiwfn program main menu interface.

Figure 3.6 shows the interface of the Multiwfn 3.7 program. In order make calculations in the program, the wavefunction file containing all the structural features of the complex must be transferred to the Multiwfn 3.7 program as input. The code "orca_2mkl A1-01 -molden" was run in ORCA 4.2 program to create this Multiwfn 3.7 wavefunction input file format “molden”. The "gbw" structure, which contains all the features of the most stable structure created as result of geometric optimizations with the code orca_2mkl, has been converted to "molden" format for all complexes as Multiwfn 3.7 wavefunction input file.

Multiwfn 3.7 program offers users the opportunity to prepare useful scripts to perform many calculations with a single code. Multiwfn 3.7 sourced descriptor calculations were made for all complexes with the script example shown in Appendix 3 and the code to run this script. The numbers on the left of the script example were run the calculation types seen in Appendix 3, respectively.

434 parameters including population analysis (Hirshfeld, Voronoi deformation density, Mulliken, Löwdin, Modified Mulliken (including three methods: Ros & Schuit, Stout & Politzer, Bickelhaupt), Becke, Atomic dipole moment corrected Hirshfeld, CHELPG, Merz-Kollmann, Restrained Electrostatic Potential (RESP)), bond order/strength analysis (Mayer bond order, Wiberg bond order in Löwdin orthogonalized basis, Mulliken bond order, Laplacian bond order, intrinsic bond strength index), quantitative analysis of molecular surface, dipole moment, integral functions, energy index, polarity index and critical point density analysis of platinum atom, real space functions for whole system, basin analysis for attractors, analyzing real space functions in fuzzy atomic spaces and topology analysis for any real space function (such as electron density (AIM analysis), Laplacian, electrostatic potential and so on) were determined as descriptors for all complexes from Multiwfn.

3.2.6.4 Descriptors Taken from PaDEL-Descriptor

Finally, some 1D, 2D and 3D descriptors were calculated with the PaDEL-Descriptor 2.21 program. After final geometries of all complexes with the "mol2" format were transferred to the program, descriptors were found.

3.2.7 QSAR Analysis

The final part of the ligand-based drug design is quantitative structure analysis (QSAR). The QSARINS 2.2.4 (QSAR-Insubria) software developed at the University of Insubria (108-109) was used in the QSAR analysis for all complex groups.

All descriptors calculated in the previous section and cytotoxic data were divided into complex groups (A-G) and saved as "csv". Approximately 1400 descriptors were created in the "csv" files of each group. For model development, it is best to perform models using each of the currently calculated descriptors. However, the number of model combinations is so large that it is impossible to calculate all of them. Therefore, models were developed by reducing the number of descriptors. So, the "csv" file of each group was imported into the QSARINS

program, filtration was applied from the import data section for values and the descriptors that had more than 90% correlation and possessed more than 95% of constant values were removed. Random selection was made from the data setup section for the remaining descriptors. The complexes are divided into 70% training and 30% testing for this selection. Several random allocations were made until the best model was obtained. Then, genetic algorithm-multiple linear regression (GA-MLR) analysis was applied from the variable selection and model calculation part. Regression up to a maximum of six descriptors was accepted for this analysis. Also, the model per size, population size, generated model number, and mutation rate were set to 10, 10, 10000, and 20, respectively.

Finally, all model statistical parameters and selection descriptor names were listed in the section of view and select models. Many parameters were evaluated to select the best model in this section. Regression formulas consisting of descriptors selected in the best models determined for all groups were used to calculate the estimated cytotoxic activity values of the novel created complexes.

3.3 Structure-Based Drug Design

The main target of complexes containing platinum metal, which is a drug candidate, in the cell is the DNA macromolecule. In this section, the interactions of these complexes with DNA were simulated. The construct chosen for modeling was a 12-mer B-DNA.

The experimental structural data of the platinum-containing compound, which was used to compare the DFT methods in the previous section, were used for the best DFT and basis set selection. Differently from the previous section, effective core basis sets (ECP) were used instead of ZORA relativistic full electron basis sets for the selection of the basis set. The best basis set and DFT method were determined after statistical deviation calculations. Then, the identified functional groups taken from literature and the 2-pyrimidinethiol compound were added to each platinum-containing complex using Gaussview 3 software. The complexes were designed to interact with DNA in the form of non-covalent and groove binder. The conformations of each drawing compounds were made with Gaussian 03 program. After determining the best conformations for the complexes according to the lowest energy, all designed complexes were used in geometric optimization calculations with ORCA 4.2. After the most stable and best geometric structures of the complexes were found, the DNA structure was selected from protein data bank

(PDB) as code of 1DNE. Molecular docking operations were performed with AutoDock 4. Preparation of DNA and ligands for docking was performed with AutoDockTools 1.5.6. Docking was done with various scripts and using some python program. Docking calculations were performed quickly with the success of the scripts and the best positions and binding energies of all ligands in DNA were determined. The best of these pose was selected and prepared for simulations. These preparations were made using MCPB.py and Gaussian 03 programs. Simulations and minimization calculations were completed using AMBER 20. Ligand containing DNA constructs with the best poses were integrated into the aqueous medium. This medium was minimized and was heated at a constant volume. Afterwards, the simulation was run for 2 ns and the environment was brought to equilibrium. Finally, a simulation of 100 ns production was run and changes in structure and binding energies at various times were found. Viewing and monitoring of the simulations was done with Chimera 1.15. Binding energies were found by MM/GBSA methods with AMBER 20 program. All graphics were plotted with Xmgrace. The occupancy analyzes of the hydrogen bonds formed by the complexes with DNA as donor and acceptor were performed with the VMD 1.9.3 program. In addition, angular changes in the nucleobases with which the complexes interact were compared using the cpptraj module of AMBER 20. In this comparison, the results of molecular dynamics simulations of the B-DNA form of the de-complexed 1DNE were used.

3.3.1 Determination of The Best DFT Method and Basis Set Combination

As mentioned in the previous section 3.2.1, as a second calculation method, DFT method and basis set comparisons were made. In this second comparison, unlike section 3.2.1, effective core potential (ECP) basis set type was used instead of ZORA relativistic full electron basis set. This means that the full electron basis set is used in ligand-based drug design, especially in descriptor calculations, due to the application of quantum theory. Because all electrons must be included in the calculations in order that there are no errors in the geometrical structures of the complexes and in order achieve success in the quantum theory atom in molecule (QTAIM). On the other hand, the macromolecule is the key in molecular dynamics simulations. The most important factor affecting the calculations is the interaction between the ligand and the target. Therefore, it was aimed to shorten the calculation

time with a new comparison and the extent of the response of metal-containing complexes to the use of ECP in ORCA 4.2.

Table 3.5. Basis set list. 4 different metal atoms ECP system and 41 basis set combinations divided into different groups.

ECP Code 1	ECP Code 2	ECP Code 3	ECP Code 4	Metal Atom	Other Atom
DEF2-ECP-01	SDD-01	HayWadt-01	SK-MCDHF-RSC-01	cc-pVDZ-PP	cc-pVDZ
DEF2-ECP-02	SDD-02	HayWadt-02	SK-MCDHF-RSC-02		6-31G
DEF2-ECP-03	SDD-03	HayWadt-03	SK-MCDHF-RSC-03		6-31G(d)
DEF2-ECP-04	SDD-04	HayWadt-04	SK-MCDHF-RSC-04		def2-SVP
DEF2-ECP-05	SDD-05	HayWadt-05	SK-MCDHF-RSC-05	cc-pVTZ-PP	cc-pVTZ
DEF2-ECP-06	SDD-06	HayWadt-06	SK-MCDHF-RSC-06		6-311G
DEF2-ECP-07	SDD-07	HayWadt-07	SK-MCDHF-RSC-07		6-311G(d)
DEF2-ECP-08	SDD-08	HayWadt-08	SK-MCDHF-RSC-08		def2-TZVP
DEF2-ECP-09	SDD-09	HayWadt-09	SK-MCDHF-RSC-09		def2-TZVPP
DEF2-ECP-10	SDD-10	HayWadt-10	SK-MCDHF-RSC-10	cc-pVQZ-PP	def2-QZVP
DEF2-ECP-11	SDD-11	HayWadt-11	SK-MCDHF-RSC-11		def2-QZVPP
DEF2-ECP-12	SDD-12	HayWadt-12	SK-MCDHF-RSC-12		cc-pVQZ
DEF2-ECP-13	SDD-13	HayWadt-13	SK-MCDHF-RSC-13	def2-SVP	def2-SVP
DEF2-ECP-14	SDD-14	HayWadt-14	SK-MCDHF-RSC-14		cc-pVDZ
DEF2-ECP-15	SDD-15	HayWadt-15	SK-MCDHF-RSC-15		6-31G
DEF2-ECP-16	SDD-16	HayWadt-16	SK-MCDHF-RSC-16		6-31G(d)
DEF2-ECP-17	SDD-17	HayWadt-17	SK-MCDHF-RSC-17	def2-TZVP	def2-TZVP
DEF2-ECP-18	SDD-18	HayWadt-18	SK-MCDHF-RSC-18		def2-TZVPP
DEF2-ECP-19	SDD-19	HayWadt-19	SK-MCDHF-RSC-19		cc-pVTZ
DEF2-ECP-20	SDD-20	HayWadt-20	SK-MCDHF-RSC-20		6-311G
DEF2-ECP-21	SDD-21	HayWadt-21	SK-MCDHF-RSC-21		6-311G(d)
DEF2-ECP-22	SDD-22	HayWadt-22	SK-MCDHF-RSC-22	def2-TZVPP	def2-TZVPP
DEF2-ECP-23	SDD-23	HayWadt-23	SK-MCDHF-RSC-23		def2-TZVP
DEF2-ECP-24	SDD-24	HayWadt-24	SK-MCDHF-RSC-24		cc-pVTZ
DEF2-ECP-25	SDD-25	HayWadt-25	SK-MCDHF-RSC-25		6-311G
DEF2-ECP-26	SDD-26	HayWadt-26	SK-MCDHF-RSC-26		6-311G(d)
DEF2-ECP-27	SDD-27	HayWadt-27	SK-MCDHF-RSC-27	def2-QZVP	def2-QZVP
DEF2-ECP-28	SDD-28	HayWadt-28	SK-MCDHF-RSC-28		def2-QZVPP
DEF2-ECP-29	SDD-29	HayWadt-29	SK-MCDHF-RSC-29		cc-pVQZ
DEF2-ECP-30	SDD-30	HayWadt-30	SK-MCDHF-RSC-30	def2-QZVPP	def2-QZVPP
DEF2-ECP-31	SDD-31	HayWadt-31	SK-MCDHF-RSC-31		def2-QZVP
DEF2-ECP-32	SDD-32	HayWadt-32	SK-MCDHF-RSC-32		cc-pVQZ
DEF2-ECP-33	SDD-33	HayWadt-33	SK-MCDHF-RSC-33	LANL2DZ	def2-SVP
DEF2-ECP-34	SDD-34	HayWadt-34	SK-MCDHF-RSC-34		cc-pVDZ
DEF2-ECP-35	SDD-35	HayWadt-35	SK-MCDHF-RSC-35		6-31G
DEF2-ECP-36	SDD-36	HayWadt-36	SK-MCDHF-RSC-36		6-31G(d)
DEF2-ECP-37	SDD-37	HayWadt-37	SK-MCDHF-RSC-37	LANL2TZ	def2-TZVP
DEF2-ECP-38	SDD-38	HayWadt-38	SK-MCDHF-RSC-38		def2-TZVPP
DEF2-ECP-39	SDD-39	HayWadt-39	SK-MCDHF-RSC-39		cc-pVTZ
DEF2-ECP-40	SDD-40	HayWadt-40	SK-MCDHF-RSC-40		6-311G
DEF2-ECP-41	SDD-41	HayWadt-41	SK-MCDHF-RSC-41		6-311G(d)

The complex in figure 3.4 was used for both DFT and basis set comparisons. First, the best DFT method was selected from among the methods specified in table 3.2 and 3.3 as described in section 3.2.1. Differently, when choosing the method,

instead of ZORA relativistic full electron basis set, the def2-ECP (110-111) system was used for only the valance electrons of the platinum atom, and the def2-TZVP basis set was used for all atoms. Four ECP systems were used for the basis set for comparisons. These were Ahlrichs def2 basis set (def2-ECP) (110-111), Stuttgart-Dresden (SDD) (110,112), Hay and Wadt's family (LANL2) (113) and the Stuttgart-Koeln small-core multiconfiguration-Dirac-Hartree-Fock adjusted (SK-MCDHF-RSC) (114).

The established ECP system is detailed in table 3.5. In this system, all four ECP systems were used with basis sets in the same group. A total of 164 basis set combinations were designed for 41 basis sets in each group. One of the most important issues to be considered here is minimizing the errors originating from the basis set. In order achieve this, the basis set used especially for metal is double-zeta, and the basis set used for other atoms is determined as double-zeta. For example, when the triple-zeta basis set of LANL2TZ was chosen for the platinum atom, the basis sets used for other atoms were also chosen as 6-311G, which has triple-zeta character and was produced by a different research group.

3.3.2 Design and Conformational Analysis of Novel 2-Pyrimidinethiol and Platinum Containing Complexes

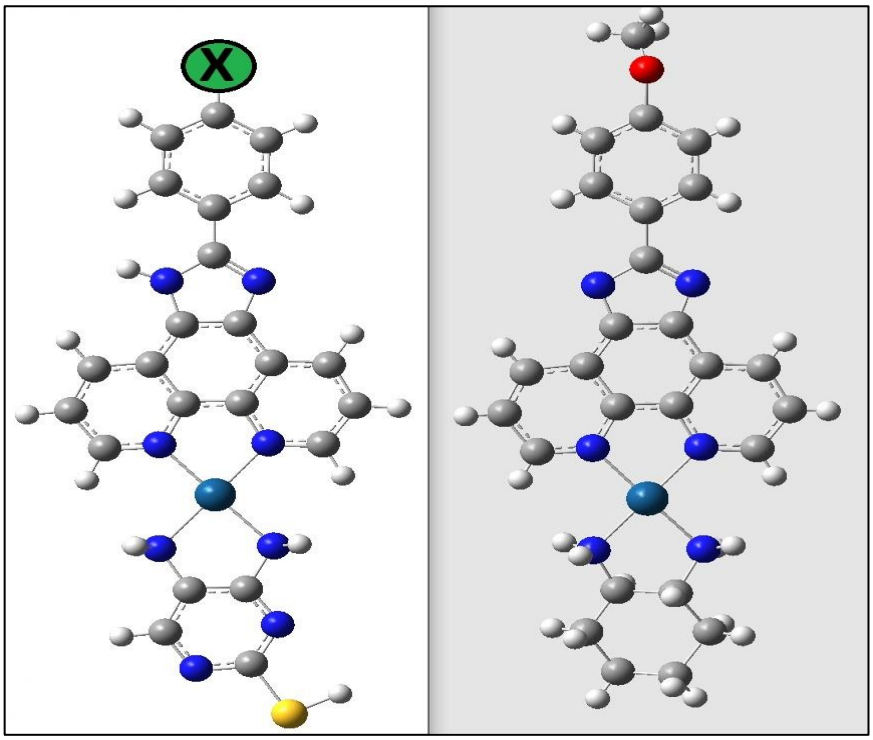


Figure 3.7 (continued)

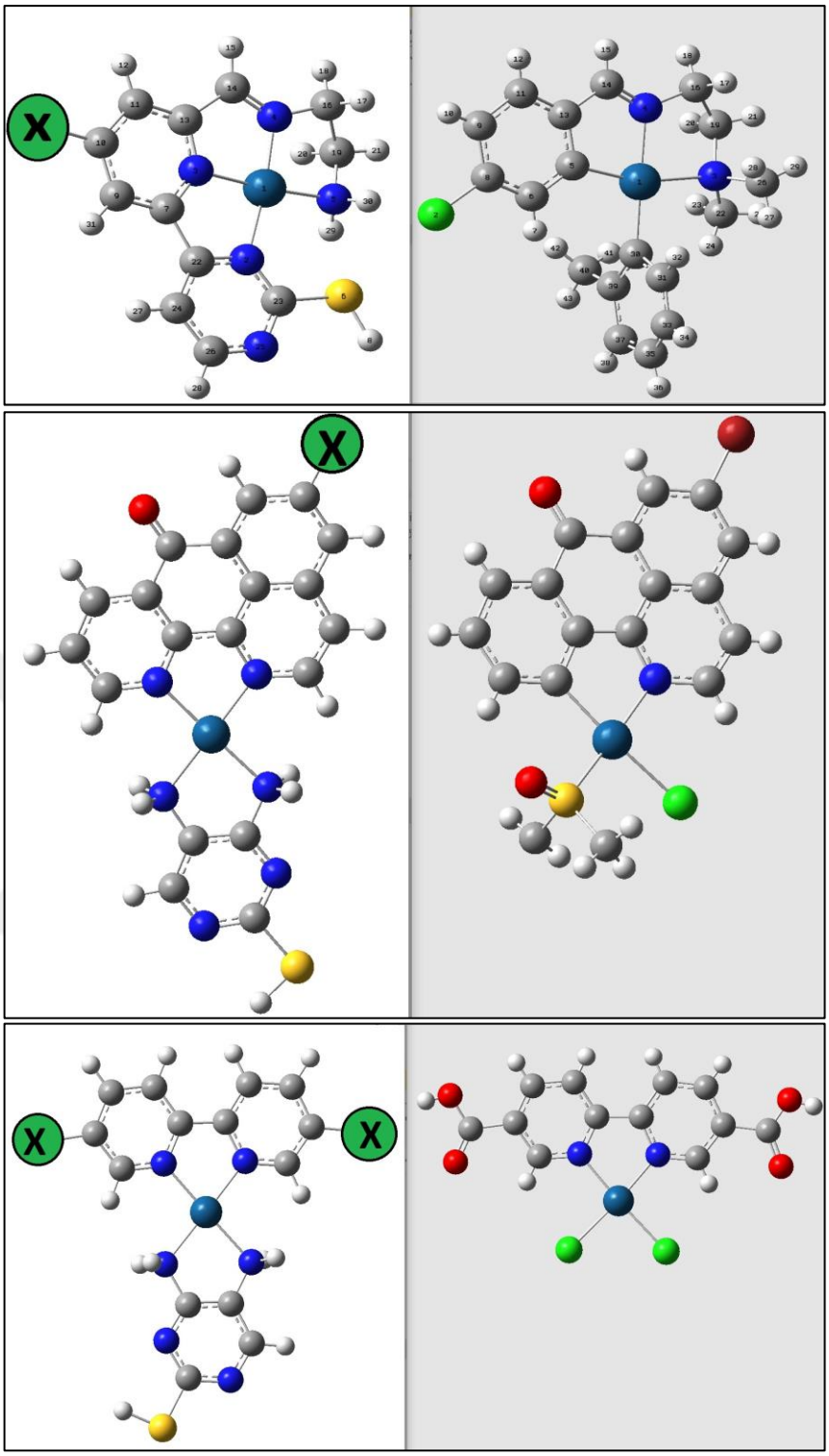


Figure 3.7. All novel platinum-2-pyrimidinethiol complexes. Four groups of complex systems are listed. The contents of the white boxes on the left are the designed structures, and the gray boxes on the right are the original structures from the CCDC. X is the functional group insertion position. Arrangement of atoms by color: Red, Oxygen; Navy Blue, Nitrogen; White, Hydrogen; Grey, Carbon; Yellow, Sulfur; Orange, Phosphorus; Blue, Platinum; Green, Chlorine; Ice blue, Fluorine; Claret red, Bromine.

A literature search was carried out on which of the platinum-containing complexes had groove binder interactions with DNA (115). On top of that, platinum-containing complexes with square planar geometry whose leaving group (Cl, O, etc.) could bind non-covalently with DNA were detected. Due to the necessity of forming the four bonds of the platinum atom in these complexes with the carrier ligand, the four sides of the platinum atom were coordinated with nitrogen atoms. Upon this, four different complex systems were designed. In choosing this complex layout, one side of the construct was selected from ligands previously studied and known in the literature. Many platinum-containing complexes with single crystals were screened in CCDC for this situation. The other side of the complexes was attached to the nitrogen atom(s) of the 2-pyrimidinethiol ligand. While creating the complex layout, the initial geometries taken from CCDC can be found in the access structure section with the codes 1472980, 1553814, 1541248 and 729232. This complex layout is as in figure 3.7, respectively.

In this system, some functional groups are bonded to the "X" port for the complexes in the white boxes on the left side of the images in figure 3.7. While determining these functional groups, the percentage of functional groups used in studies and their contribution to biological activity were considered. The functional groups were selected from the work by Peter Ertl (116). There is a list of functional groups showing biological activity in this study, and the organic functional groups shown in figure 3.8 are taken from this study. R_2 represents the -CH_3 group; R_1 is Hydrogen; R_{ar} represents the point at which the functional group binds to the complex for the functional groups in the image. According to this information, 21 functional groups were selected and a total of 84 molecules were designed in four complex groups.

After the designs and drawings were completed, conformations at all possible dihedral angles according to the position of the 2-pyrimidinethiol (especially for the -SH group) as well as the functional groups bonded to each of the complexes at the “X” point were manually generated before geometric optimizations. These conformers were created for 84 complexes by changing their dihedral angles using the Gaussview 3. Gaussian 03 program was used to find the best geometric structures of conformers drawn at different positions. In the geometric optimization calculations of conformers, B3LYP, which is one of the most used DFT functions, and LANL2DZ and LANL2 ECP for the metal atom, 6-

311G for the sulfur atom, and 6-31G for the other atoms in the basis set combination were selected.

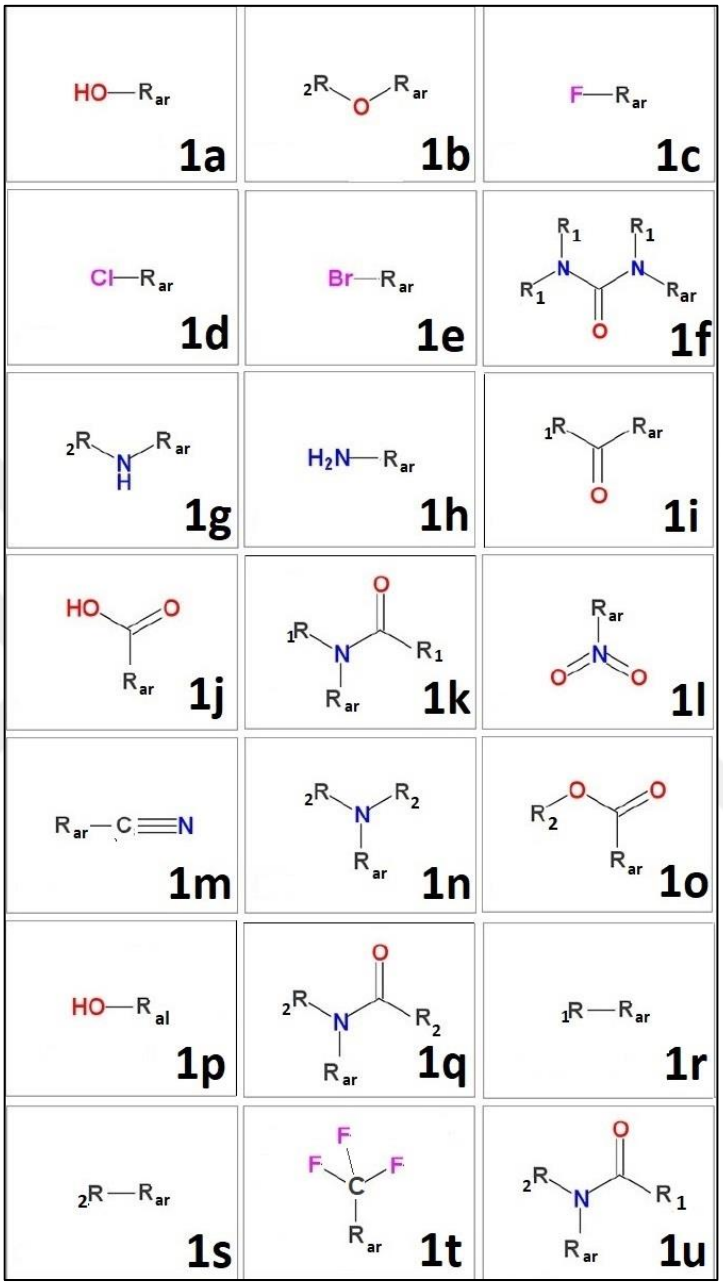


Figure 3.8. Functional groups that have disappeared from the literature. The functional groups were numbered 1a-1u. O: Oxygen, N: Nitrogen, F: Fluorine, Cl: Chlorine, Br: Bromine, R1: H, R2: CH3 R_{ar}: The point at which it attaches to the complex.

After optimizations in the conformation analysis, the best conformer was selected. In order select this, the conformers in all complex groups were ordered in energy, and the final geometry of the most energetic conformer was taken.

3.3.3 Geometric Optimizations of All Complexes

As a result of the conformational analysis, using the best DFT method and ECP basis set combination determined as described in the section 3.3.1, geometric optimizations were made for 84 complexes in different complex groups and the most stable structures were determined. All geometric optimizations were carried out using the ORCA 4.2 program, which we installed on Tübitak's TRUBA high performance computing systems. Geometric optimizations were made using the same parameters for all complexes to ensure computational integrity. All calculations were performed using the "compound" scheme, exemplified in Appendix 1, using multiple inputs allowed by the ORCA 4.2 program. As useful keywords, optimization and SCF energy parameters with tight tolerance are applied to achieve better geometry and stable energy. In addition, a high grid frame was added to the complexes and a large-scale population analysis was requested for bonds, orbitals, and some charge methods. The best detected metal ECP, basis set combination and DFT method were added to the system. The coordinate system in "xyz" format, which was taken Chemcraft, was added to the inputs.

3.3.4 Molecular Docking of All Complexes into The Target DNA

Docking calculations were made using AutoDock 4.2.6 program (117). It was done using some scripts that provide virtual screening in the molecular docking. Scripts downloaded from the website (www.autodock.scripps.edu) consist of a set of python packages that enable AutoDock 4.2.6 to be run. AutoDock 4.2.6 is a suite of automated docking tools. It is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure.

Before using the program, ligand, and receptor (target molecule) must be prepared. AutoDockTools 1.5.6 (ADT) program was used to prepare the necessary files. Simulation and docking processes in biomolecular systems using complexes containing metal atoms are challenging. Because calculations in macromolecular systems are usually made in the form of a quick scan on small organic compounds that do not contain metal. Since complexes containing platinum metal are used here, it is necessary to define the van der Waals parameters of the platinum atom. All Docking calculation flow is expressed in the image in figure 3.9, respectively. After these definitions were made, firstly 84 complexes taken from previous geometric optimizations and 1DNE coded 12-mer B-DNA structure shown in figure 3.10, which was downloaded from the protein data bank, were initially prepared. DNA

structure was opened as "pdb" format in ADT program and water molecules and existing ligands were deleted. All hydrogens and Gasteiger charge states in DNA were controlled and added in ADT. The charges of the complexes were calculated with the Chimera 1.15 program using the Gasteiger method. The point to be noted here is that the charge of the complexes is +2 since the platinum atom bonds with four nitrogen atoms in the complexes. The "pdbqt" files of the complexes were converted from "mol2" with the OpenBabel 3 program. Autogrid in AutoDock 4.2.6 was run according to the atom types in all complexes with ADT. The grid size was determined to include the entire complex as 90 90 120. Autogrid input file ("dpf" format) was prepared manually. All atoms in the complexes are added to the input one by one to create "map" files. After the “map” files of DNA and other atoms were created, scripts and packaged python programs were run in the Linux operating system so that the AutoDock 4.2.6 program could be used. These scripts were adapted by making some changes suitable for our own system.

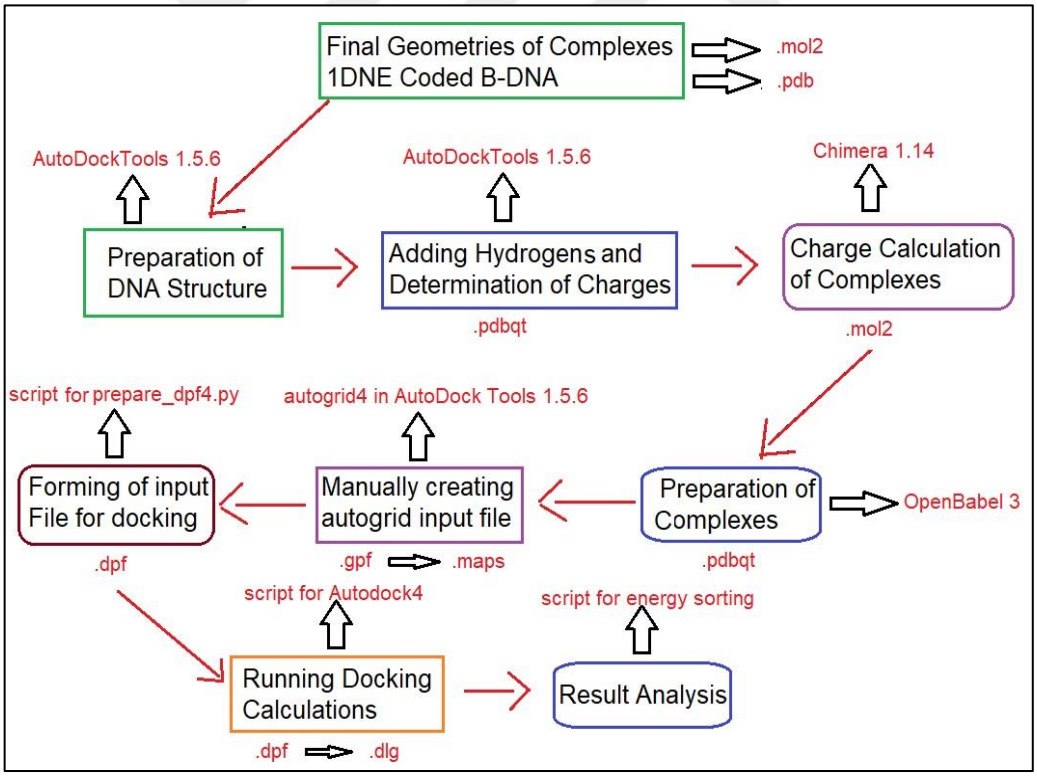


Figure 3.9. Molecular Docking flow diagram.

Input files running docking calculations were automatically created for all complexes one by one. This was provided by running the script with the

"prepare_dpf4.py" package program. Docking calculations were made according to the Lamarckian genetic algorithm. Settings and parameters in calculations were selected as population size 150, number of steps 200, root mean square tolerance (rms) 2.0, maximum number of energy evaluations 175000, genetic algorithm mutation rate 0.02, and genetic algorithm crossover rate 0.8. Other settings are left as default. After the input files were created with the above script, these inputs were automatically run with a script that directly called autodock and the best poses of all complexes on the target DNA were determined. The energy values of the poses with the best binding energy were extracted with a script that analyzes the results and ranks them.

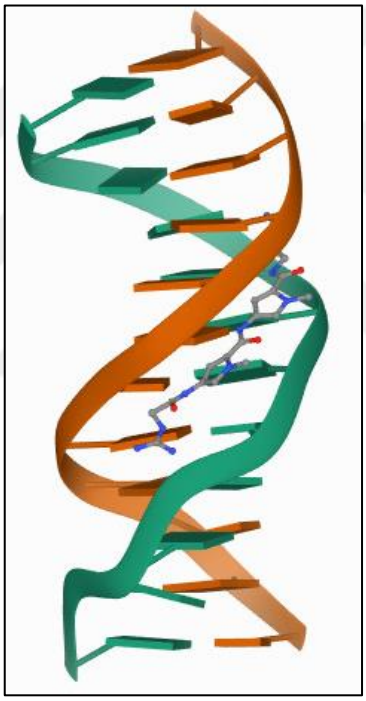


Figure 3.10. 1DNE coded B-DNA Structure

3.3.5 Molecular Dynamic Simulations

Molecular dynamics simulations were done with the AMBER 20 (118-121) program. AMBER 20 is a suite of biomolecular simulation programs. Molecular dynamics simulations involving metal-containing compounds pose a major problem. Because the bonds made by metal atoms undergo many changes according to the properties of the atoms to which they are attached, there are no ready-made force fields. These bonds and their types must be introduced through some program and stored in a parameter file that the simulation dynamics can define. Therefore,

force fields should be determined for the bond parameters (Bond length, bond angles and dihedral angles) found in metal-containing compounds. Based on this, the best poses of each of the four complex groups obtained from the previous docking calculations were taken from the ADT program in "pdbqt" format. The structure of these poses was saved in "pdb" format by converting them with OpenBabel 3. The complex in the build was saved in a separate "pdb" file. This compound was opened with the Gaussview 3, and the missing non-polar hydrogens were added manually and saved. Then, the ligand molecule was separated from the complex without the metal atom and the charge calculation was performed with the antechamber add-on, which is included as a package in AMBER 20. For the binding parameters in the ligand saved as "mol2", force filed calculations were made with the parmchk2 add-on. A separate "pdb" file was also created for the metal atom. Since the metal ion is known to have +2 charge, the "mol2" format for the metal atom was created with the metalpdb2mol2.py plugin that comes with AMBER 20. Hydrogens were added to the DNA-ligand structure without non-polar hydrogens using default settings on the H++ server (http://biophysics.cs.vt.edu). Then, the DNA structure, platinum ion and ligand molecule were combined into a single "pdb" file with the "cat" command. The atoms in the “pdb” file of the whole structure were renumbered with the pdb4amber plugin. In the next step, MCPB.py was used to generate force fields of metal-bonded compounds. MCPB.py (122), a python-based metal center parameter generator, was developed to generate force fields for simulation of metal complexes using the bonded model approach. An input ("in") was created for MCPB.py, where the ligand, metal ion, DNA structure and force fields were defined. All stages consisting of four formats (1-4) were run using the prepared input. When the format "1" was run, the necessary input files of Gaussian 03 were created. Energy, frequency (for force fields) and Merz-Singh-Kollman charge calculations of the complex were made using the inputs of Gaussian 03 program. By executing formats of 2,3, and 4 respectively using MCPB.py, the "tleap.in" file was created. The names and coordinates of atoms in all structures were ensured to be the same, and the metal atom was capitalized in all files using MCPB.py.

A classical AMBER parm99 (123-124) together with the parmbsc0 (125) refinement for DNA and gaff (126) force fields were defined for use. Na+ ions were added to neutralize the DNA structure. The structure was surrounded by a 20 Å

water box. The tleap plugin was run with the mentioned input file and the parameter ("prmtop") and coordinate (“rst”) files used during the simulation were created. After determining the necessary files for the next minimization and simulation processes, the protocol for all MD simulations was as follows: (1) The DNA and platinum containing complex contained in the main structure were constrained. only water molecules and Na+ ions were minimized with a cut-off value of 9 Å. Minimization was performed with 2500 steps with harmonic restraints of 500 kcal mol^{-1} $Å^{-2}$ on the DNA and platinum compound positions. (2) After initial minimization, 5000 steps of unrestricted and full system minimization were applied out before the heating process. A cut-off of 9 Å was used for unbound interactions. The SHAKE algorithm (127) was used to restrain hydrogen-containing bonds. (3) Each minimized structure was heated at a constant volume from 0 to 300 K. Heating was completed for 300 ps. The positions of DNA and platinum compound were restricted to a small value of 25 kcal $mol^{-1}Å^{-2}$. Constant volume was maintained during the heating. The structure was heated from 0 K to 300 K for a duration of 200 ps. It was held steady at 300 K for 200 ps to 300 ps. (4) An equilibration with 0.5 kcal $mol^{-1}Å^{-2}$ restrains on DNA and platinum complex was performed for 1 ns time under conditions at constant pressure and constant temperature, before a trajectory was done for a production simulation. An integration step of 1 fs was used for all equilibration simulation. The pressure and temperature were allowed to ripple at around 1 bar and 300 K, respectively, with a regardful coupling of 0.2 ps. (5) Production runs of 100 ns were carried out using the 9 Å cut-off and unrestrained on any structures. In each simulation, the time point at 1 ns after thermal equilibration was taken as the starting point for data collection. 25,000 structures from each simulation were recorded for post-processing, with uniform sampling from trajectory during production simulations. As a result of the production simulation, a trajectory file in "nc" format was created. This file was used together with the parameter file of the structure ("prmtop") with the Chimera 1.15 program, and the simulation was displayed. Input parameters of all simulations and minimizations are given in Appendix 4.

Hydrogen bond analysis was performed using the "Hydrogen Bonds" plugin in the analysis section of the VMD1.9.4 program. The "mc" trajectory and "prmtop" parameter files taken from the simulations were transferred to the program. DNA nucleobases were added to selection 1 and a platinum-containing compound was

added to selection 2. The donor and acceptor were selected on both sides to be investigated in all 100ns (25000 frames). Finally, hydrogen bond analyzes were performed by choosing the donor-acceptor distance of 3.5 Å and the angle cutoff of 120 degrees.

The conformational free energies for all production simulations were calculated, using MM-GBSA (128-131) energy analyzes, explanatory graphs were generated from many snapshots at different nanosecond durations. All free energy studies were done using the AMBER 20 program.

DNA parameters determined by AMBER's cpptraj module were created with Xmgrace program in graphic form by using trajectory (nc) and parameter file (prmtop) as input. Global base-pair parameters (tilt, stretch, stagger, buckle, propeller and opening) and global base-step parameters (shift, slide, rise, tilt, roll and twist) data were revealed with the cpptraj module. Also, comparative RMSD plots between complex-DNA and single B-DNA constructs were generated. Finally, snapshots of the structures at different nanoseconds during the simulations were created visually with the Chimera 1.15 program.

A script with the extension "sh" was prepared, where all the calculations were run in AMBER 20. By running this script on the terminal screen, simulations, minimizations, and the creation of graphics were performed automatically.

Molecular dynamic simulation calculations were made using “sander.cuda” program of AMBER 20 and GPU architecture of GeForce 1080ti graphics card in Linux operating system. GPU calculations significantly reduced simulation time.

4. RESULTS AND DISCUSSION

4.1 Determination of Initial Settings and Testing Critical Keywords Before Choosing Suitable DFT Method and Basis Set

Before the geometric optimization calculations related to metal-containing complexes in ORCA 4.2, the effect of the initial settings and critical keywords specified in table 3.1 calculations has been determined on the subsequent DFT. 72 tests were made using M06L and 161 trials were made with BP86 among the DFT functions. In these tests, errors caused by keywords were calculated by combining the different parameters given in table 3.1 with each other. Calculation times increased approximately 4 times in trials where RI or RIJCOSX approaches were not used. This significantly affected the computation time in systems with larger atoms. After the tests in which the GRID number was increased, the calculation times increased slightly. In the M06L method, the difference between the total energies was only 0.002%. In the BP86 function, this difference was calculated as 0.0001%. These differences show that keywords have little effect on total energy.

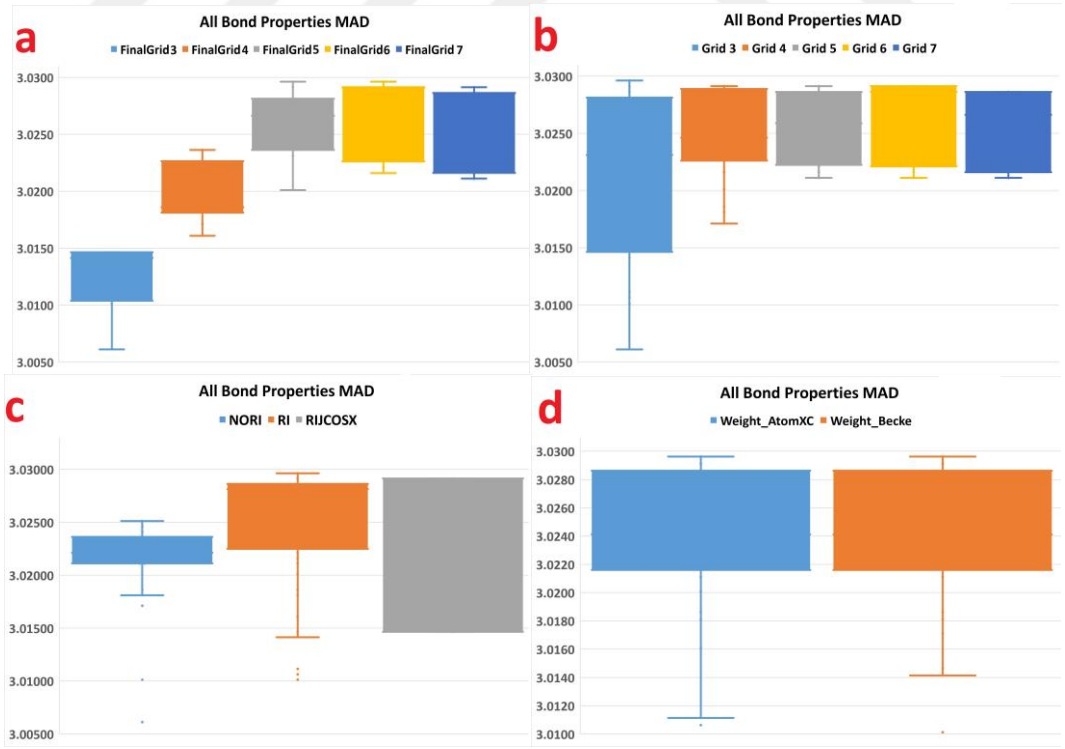


Figure 4.1. Comparison of all bond parameters with experimental data and mean absolute deviation (MAD) values using line-box plots over various settings and keywords a) FinalGrid b) Grid c) Weight d) RI Approximations.

The Nedaplatin complex was used to compare the structural parameters in the calculations. How the different settings and keywords used affect this deviation was evaluated with the graphs in figures 4.1, 4.2 and 4.3. In these graphs, mean absolute deviation (MAD) values were used in the comparison.

In figure 4.1, the results for the comparisons of all parameters in the complex (bond angle and bond length) using the BP86 functional were listed with a box plot. MAD values for this graph, which was the result of using different settings and parameters, were put as y-axis. As can be clearly seen in the graph, no visible change was detected when looking at all keywords. For example, the maximum variation between GRID3 and GRID7 in figure 4.1a is about 0.025. No significant change was observed for other parameters in figure 4.1.

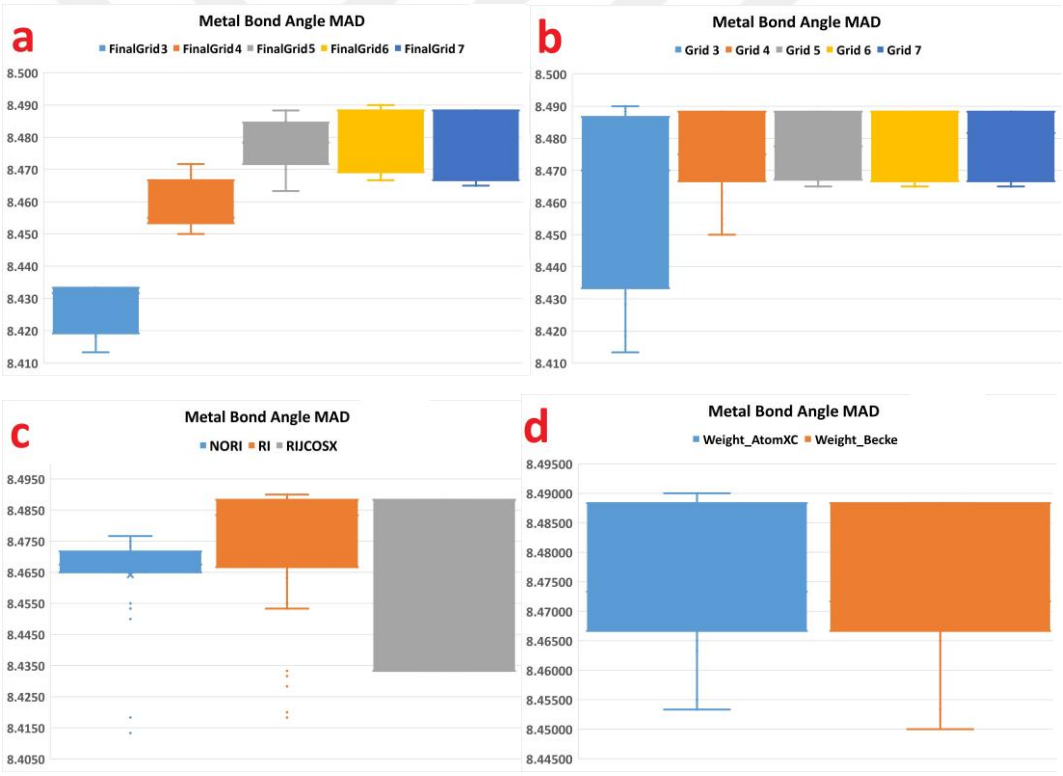


Figure 4.2. Comparison of bond angle for metal in center with experimental data and mean absolute deviation (MAD) values using line-box plots over various settings and keywords a) FinalGrid b) Grid c) Weight d) RI Approximations.

In Figure 4.2, only MAD statistical data to compare metal bond angles were listed in the box plot for the parameters specified in figure 4.1. Although the deviation value of GRID3 stands out in these graphs, it does not give confidence as the distribution and shows is in a very common range. Although the deviations in

GRID3 and FINALGRID3 (Figure 4.2 a,b) are negligibly smaller than the others, the range of distribution is unreliably large. There were negligible deviation differences in other parameters.

In Figure 4.3, the deviations in the bond lengths of the metal were added. Similar situations apply here as well. The differences in deviations seen in all tests are low.

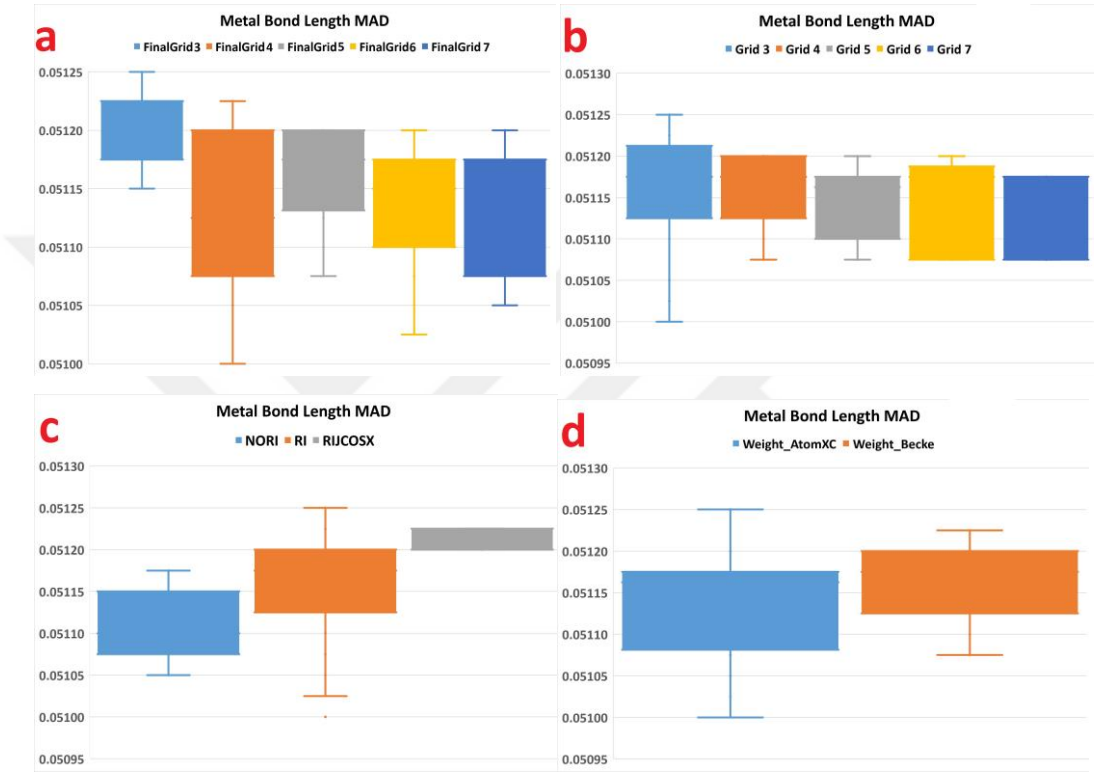


Figure 4.3. Comparison of metal containing bond length with experimental data and mean absolute deviation (MAD) values using line-box plots over various settings and keywords a) FinalGrid b) Grid c) Weight d) RI Approximations.

As a result, no significant difference was detected in the deviations between parameters and settings in all comparisons. All the settings and parameters for the next DFT calculations have been built and selected. First, the results for the reduction of the computation time were discussed. The RIJCOSX approximation method was chosen, which significantly reduces the computation time and does not create a significant deviation difference compared to the NORI keyword. As can be clearly seen in the graphs c in figures 4.1, 4.2 and 4.3, the deviation difference between the no approximation case (NORI) and the RIJCOSX approximation is negligible.

The next important issue is grid size. Since GRID is the number of integral points applied per atom and increasing GRID number slightly increases the computation time. The Nedaplatin complex was used in the comparison when selecting the initial settings. Nedaplatin is a simpler complex in terms of atomic number and complexity of structure to compare many platinum-containing complexes. Therefore, the GRID3 values pass with relatively negligible differences according to graphs. When the number of atoms in the complexes increases and the structure becomes more complex, the GRID3 integral number is insufficient for these structures. Therefore, it is inevitable that instabilities occur in geometric optimizations. This can cause SCF convergence (fluctuations in energy calculation) and optimization problems. Also, DFT functionals like the M06L have grid sensitivity. This may negatively affect the solution of the geometric structure. For all these reasons, GRID7 and FINALGRID7, which have grid integral values as high as possible, were chosen. Also, the RIJCOSX approximation method has its own grid keyword provided by the ORCA 4.2 program. GRIDX9 with high integral has been chosen for this keyword. The platinum atom is a large atom. Therefore, the radial grid (IntAcc) value around the atom has also been increased to make the permanent integral success in the grids knitted around this atom.

The deviation difference between the weight profiles used in DFT calculations is negligible as can be seen from the graphics. However, Weight_AtomXC has a more common range of bias distribution. Therefore, it was not preferred in the calculations. The basis for subsequent calculations is Weight_Becke, which is used by default in DFT calculations in ORCA 4.2. In addition, it was decided to perform optimization and SCF energy tolerances at a tight level to achieve better and stable geometric structures.

As a result of detailed initial comparison calculations, the success, and limitations of the ORCA 4.2 program in metal-containing complexes were tested. The best settings and keywords were determined for use in future calculations.

4.2 Ligand-Based Drug Design

An issue encountered in the literature was that the classical DFT method and basis sets formed the beginning of these calculations. To prevent this, the best DFT method from the ORCA 4.2 program was found with calculations. In addition, ZORA relativistic full electron basis set group was used instead of ECP, which freezes the valence orbitals and includes only the orbitals contributing to the bond,

as a basis set. By using this, it was tried to reduce the errors originating from the basis set. The first phase of ligand drug design was deciding which would be the best DFT method. It was carried out with the methods expressed in table 3.2 and 3.3 and the calculation methods mentioned in section 3.2.1 for this.

Geometric optimizations were performed with DFT calculations. Comparisons were made using 139 DFT methods using the complex with single crystal structure in figure 3.4. These DFT methods were compared with various statistical analyzes and graphs. During the geometric optimizations, the initial settings and keywords identified in section 4.1 were added to the inputs. In addition, ZORA relativistic full electron system was chosen as the basis set.

As a result of the calculations, the geometric parameters (bond length and bond angles) were transferred to the tables for all DFT methods and compared with the experimental bond parameters in figure 3.4 according to the mentioned statistical formulas.

The deviation formulas used in comparison were applied separately to bond lengths containing only metal, all platinum related bond lengths in the complex, bond angles containing only metal, all bond angles and all in the complex.

The methods were sorted according to the various statistical results in table 4.1-4.5. Among the 139 methods, the DFT functions that are in the first 50 ranks are listed.

In table 4.1, only the bonds made by the metal atom are compared with the experimental data. It has been determined that DFT1-31-32-33 functions have a remarkable success in all deviation groups compared to other functions. This DFT group is based on the "HCTH" function. Also, while the mean absolute deviation (MAD) value is very close for these three functions, the difference between the deviation value of DFT1-33 and DFT1-39 in the 50th order has approximately 54% deviation from the mean. This is a great value and indicates that the spread is over a wide area. Therefore, the HCTH group has come to the fore in metal-containing bonds.

In table 4.2, all metal-containing and metal-affected but metal-free bonds are considered statistically. As can be seen in the table, the success of the HCTH group decreased slightly when the lengths of the bonds in the organic parts were activated. However, when we analyzed the MAD analysis, the difference between the DFT1-01, which is the first DFT function in the ranking, and the 50th DFT

group, deviation from the mean is approximately 21%. This spread over a relatively lower range and shows that the MAD deviation values are close to each other. In addition, this value deviated from DFT1-33 by only 11%. This value is acceptable.

In table 4.3, only the bond angles with the metal atom in the center are compared with the experimental values. It is seen that the DFT1-37 function and the functions of the "HCTH" group have a remarkable success in all deviation groups compared to the other functions. While mean absolute deviation (MAD) value is very close between these functions, the difference between the deviation value of DFT1-37 in the first stage and DFT3-21 in the 50th order has a deviation from the mean by about 79%. This is a very high value and indicates that the spread is over a wide area. Therefore, HCTH group and MOHLYP2 (DFT1-37) DFT functions came to the fore in the angles of metal-containing bonds.

In table 4.4, metal-in-center and metal-containing bond angles were statistically compared. As seen in the table, DFT1-37, which had the least deviation in table 4.3, reached higher deviation when the non-metal atoms were in the center of the angles among the bond angles in the organic parts. The HCTH group is again successful. When we examine the MAD analysis, the difference between the DFT function DFT1-32, which is the first in the ranking, and the DFT1-23 group, which is in the 50th place, from the mean is approximately 28%. This spread over a relatively lower range and shows that the MAD deviation values are close to each other.

In table 4.5, all bond angles and bond lengths mentioned in the other tables are compared together statistically. As can be seen in the table, it is obvious that the HCTH group has significantly lower deviations. Again, the deviation from the mean for difference between the 50 DFT functions is approximately 28%. Although this difference seems relatively small, in the statistical analyzes (Metal bond length and meta bond angles) examining the bond properties of the platinum atom, which has the most important place in the calculations, the HCTH group took the lead with low deviations.

As a result of all these statistical analyzes, DFT1-32 included in the HCTH group transferred from LibXC to the ORCA program was selected to be used in the next DFT optimization calculations. The equivalent of the DFT1-32 expression is the HCTH407P function. This function is used with the ZORA relativistic full electron basis set.

Table 4.1. The statistical deviations of the bond lengths containing the metal atom from the experimental data by MAPE, MAD, MSE and RMSE analysis. The colors purple, blue, green, pink, and red represent DFT1, DFT2, DFT3, DFT4 and DFT5, respectively.

DFT	MAPE	DFT	MAD	DFT	MSE	DFT	RMSE
DFT1-33	0.29413362	DFT1-33	0.00663634	DFT1-33	0.00007278	DFT1-33	0.00853084
DFT1-32	0.29960100	DFT1-32	0.00671365	DFT1-32	0.00007363	DFT1-32	0.00858100
DFT1-31	0.30213126	DFT1-31	0.00677767	DFT1-06	0.00008450	DFT1-06	0.00919264
DFT1-06	0.33058849	DFT1-06	0.00741687	DFT1-31	0.00008996	DFT1-31	0.00948473
DFT1-30	0.34404556	DFT1-30	0.00773476	DFT1-30	0.00009302	DFT1-30	0.00964452
DFT1-29	0.34587364	DFT1-29	0.00774226	DFT1-29	0.00009582	DFT1-29	0.00978887
DFT1-28	0.40743929	DFT2-30	0.00903834	DFT2-46	0.00012593	DFT2-46	0.01122168
DFT2-30	0.41457453	DFT1-28	0.00904850	DFT2-27	0.00013292	DFT2-27	0.01152919
DFT2-29	0.42470867	DFT2-46	0.00944738	DFT1-28	0.00013498	DFT1-28	0.01161822
DFT2-46	0.43386239	DFT2-29	0.00945824	DFT2-29	0.00013651	DFT2-29	0.01168373
DFT3-15	0.44297498	DFT3-15	0.00982063	DFT3-15	0.00013708	DFT3-15	0.01170818
DFT2-27	0.44553312	DFT2-27	0.00985633	DFT2-45	0.00013985	DFT2-45	0.01182572
DFT1-01	0.46977814	DFT1-01	0.01028208	DFT3-23	0.00014250	DFT3-23	0.01193734
DFT3-23	0.47119791	DFT3-23	0.01041136	DFT3-19	0.00014711	DFT3-19	0.01212898
DFT2-45	0.47716138	DFT2-45	0.01045018	DFT2-30	0.00015188	DFT2-30	0.01232415
DFT2-47	0.48277415	DFT2-47	0.01053583	DFT2-26	0.00015606	DFT2-26	0.01249222
DFT3-22	0.48926711	DFT3-02	0.01064150	DFT5-03	0.00015751	DFT5-03	0.01255048
DFT3-12	0.49040296	DFT2-26	0.01070059	DFT3-24	0.00015791	DFT3-24	0.01256603
DFT2-18	0.49062868	DFT3-12	0.01082871	DFT3-12	0.00016040	DFT3-12	0.01266501
DFT3-19	0.49276562	DFT2-18	0.01083290	DFT2-18	0.00016045	DFT2-18	0.01266689
DFT2-26	0.49520930	DFT3-22	0.01083828	DFT3-22	0.00016056	DFT3-22	0.01267117
DFT2-28	0.50428993	DFT3-19	0.01089020	DFT3-27	0.00016251	DFT3-27	0.01274808
DFT5-03	0.50511764	DFT5-03	0.01096556	DFT2-34	0.00016544	DFT2-34	0.01286230
DFT2-50	0.50860357	DFT2-28	0.01102500	DFT2-21	0.00016978	DFT2-21	0.01302978
DFT2-48	0.51199135	DFT2-50	0.01113352	DFT2-05	0.00017716	DFT2-05	0.01331014
DFT1-22	0.52021858	DFT2-48	0.01114513	DFT2-49	0.00017912	DFT2-49	0.01338348
DFT2-49	0.52090165	DFT3-24	0.01123550	DFT1-16	0.00018063	DFT1-16	0.01343995
DFT3-02	0.52109090	DFT2-49	0.01136993	DFT3-04	0.00018522	DFT3-04	0.01360950
DFT3-24	0.52139374	DFT1-22	0.01137083	DFT1-01	0.00019011	DFT1-01	0.01378791
DFT2-05	0.52305792	DFT2-36	0.01138607	DFT2-50	0.00019109	DFT2-50	0.01382359
DFT2-36	0.52516182	DFT2-05	0.01157183	DFT2-04	0.00019222	DFT2-04	0.01386442
DFT3-16	0.52626176	DFT3-16	0.01159958	DFT1-08	0.00019258	DFT1-08	0.01387745
DFT3-21	0.52867462	DFT3-27	0.01165051	DFT2-47	0.00019439	DFT2-47	0.01394243
DFT3-27	0.52970896	DFT3-21	0.01166221	DFT4-01	0.00019753	DFT4-01	0.01405452
DFT2-01	0.53174153	DFT2-34	0.01169204	DFT1-05	0.00019806	DFT1-05	0.01407334
DFT2-34	0.53235410	DFT2-01	0.01177112	DFT2-36	0.00020074	DFT2-36	0.01416812
DFT2-40	0.54009615	DFT2-40	0.01196549	DFT2-28	0.00020126	DFT2-28	0.01418649
DFT3-25	0.54458300	DFT1-12	0.01196690	DFT2-40	0.00020204	DFT2-40	0.01421396
DFT3-03	0.54795409	DFT2-21	0.01201548	DFT5-02	0.00020370	DFT5-02	0.01427231
DFT2-21	0.54921499	DFT3-25	0.01202310	DFT3-02	0.00020400	DFT3-02	0.01428276
DFT2-04	0.54954927	DFT1-23	0.01209373	DFT1-22	0.00020561	DFT1-22	0.01433903
DFT1-23	0.55104623	DFT2-04	0.01213773	DFT2-01	0.00020580	DFT2-01	0.01434563
DFT1-05	0.55198050	DFT3-03	0.01213818	DFT4-03	0.00020602	DFT4-03	0.01435327
DFT1-12	0.55382901	DFT1-05	0.01214098	DFT2-48	0.00020626	DFT2-48	0.01436165
DFT1-08	0.55642717	DFT1-16	0.01225565	DFT1-10	0.00020629	DFT1-10	0.01436293
DFT3-04	0.55776442	DFT1-08	0.01227408	DFT1-20	0.00020717	DFT1-20	0.01439346
DFT1-16	0.56036390	DFT3-04	0.01230381	DFT4-02	0.00020774	DFT4-02	0.01441316
DFT2-08	0.56153456	DFT1-03	0.01233372	DFT1-23	0.00020880	DFT1-23	0.01444991
DFT1-03	0.56246374	DFT2-08	0.01240140	DFT1-03	0.00021011	DFT1-03	0.01449501
DFT4-03	0.56713706	DFT1-39	0.01242167	DFT4-05	0.00021150	DFT4-05	0.01454300

Table 4.2. The statistical deviations of the bond lengths and non-metal bonds from experimental data by MAPE, MAD, MSE and RMSE analysis. The colors purple, blue, green, pink, and red represent DFT1, DFT2, DFT3, DFT4 and DFT5, respectively.

DFT	MAPE	DFT	MAD	DFT	MSE	DFT	RMSE
DFT1-01	0.83667654	DFT1-01	0.01349435	DFT2-21	0.00025608	DFT2-21	0.01600241
DFT2-47	0.84609442	DFT2-47	0.01366369	DFT2-50	0.00025631	DFT2-50	0.01600975
DFT2-28	0.85349197	DFT2-28	0.01382744	DFT2-28	0.00025875	DFT2-28	0.01608569
DFT2-27	0.87053269	DFT2-27	0.01402988	DFT1-01	0.00025950	DFT1-01	0.01610905
DFT2-50	0.87670937	DFT2-50	0.01416060	DFT2-47	0.00026424	DFT2-47	0.01625548
DFT2-21	0.90033641	DFT2-21	0.01466768	DFT1-16	0.00026845	DFT1-16	0.01638434
DFT3-04	0.90653572	DFT2-46	0.01468516	DFT2-27	0.00027216	DFT2-27	0.01649741
DFT4-03	0.91110324	DFT1-06	0.01469074	DFT2-49	0.00027267	DFT2-49	0.01651282
DFT1-16	0.91220349	DFT2-48	0.01478523	DFT2-48	0.00027987	DFT2-48	0.01672917
DFT4-02	0.91450373	DFT4-03	0.01485081	DFT3-04	0.00029839	DFT3-04	0.01727394
DFT3-26	0.91944468	DFT1-16	0.01489138	DFT2-30	0.00030383	DFT2-30	0.01743062
DFT2-48	0.91969506	DFT2-49	0.01490443	DFT2-46	0.00031033	DFT2-46	0.01761606
DFT2-46	0.92168367	DFT1-22	0.01494106	DFT4-03	0.00031146	DFT4-03	0.01764832
DFT2-49	0.92571182	DFT3-04	0.01496627	DFT4-02	0.00031262	DFT4-02	0.01768106
DFT4-05	0.93236159	DFT1-28	0.01497220	DFT4-05	0.00031589	DFT4-05	0.01777327
DFT3-06	0.93656202	DFT4-02	0.01501243	DFT3-06	0.00031663	DFT3-06	0.01779412
DFT1-22	0.94024875	DFT1-12	0.01502940	DFT3-27	0.00031815	DFT3-27	0.01783662
DFT1-06	0.94418462	DFT1-33	0.01505513	DFT3-26	0.00031957	DFT3-26	0.01787651
DFT1-12	0.94512847	DFT2-30	0.01509811	DFT1-06	0.00032677	DFT1-06	0.01807671
DFT2-25	0.94533349	DFT1-49	0.01511948	DFT1-28	0.00032711	DFT1-28	0.01808629
DFT1-49	0.94597712	DFT1-29	0.01521974	DFT3-02	0.00033019	DFT3-02	0.01817122
DFT4-04	0.94810265	DFT3-26	0.01527664	DFT5-03	0.00033310	DFT5-03	0.01825112
DFT1-28	0.95082954	DFT4-05	0.01528500	DFT2-20	0.00033518	DFT2-20	0.01830790
DFT2-30	0.95872678	DFT1-30	0.01542470	DFT4-04	0.00033696	DFT4-04	0.01835661
DFT1-03	0.96142062	DFT1-39	0.01542885	DFT2-44	0.00033894	DFT2-44	0.01841024
DFT1-39	0.96505810	DFT1-03	0.01543171	DFT1-05	0.00034035	DFT1-05	0.01844862
DFT1-33	0.97272204	DFT2-25	0.01548440	DFT1-03	0.00034129	DFT1-03	0.01847409
DFT1-13	0.97408923	DFT3-06	0.01551272	DFT1-10	0.00034744	DFT1-10	0.01863975
DFT3-13	0.97521874	DFT4-04	0.01568931	DFT2-13	0.00034815	DFT2-13	0.01865867
DFT1-29	0.97571925	DFT1-13	0.01569913	DFT1-08	0.00034858	DFT1-08	0.01867029
DFT3-27	0.97793026	DFT1-05	0.01572067	DFT5-02	0.00034884	DFT5-02	0.01867729
DFT1-05	0.97977627	DFT3-27	0.01587528	DFT1-11	0.00034946	DFT1-11	0.01869378
DFT1-30	0.98781587	DFT1-32	0.01588910	DFT1-29	0.00034996	DFT1-29	0.01870732
DFT2-44	0.99113999	DFT2-44	0.01601783	DFT1-21	0.00035095	DFT1-21	0.01873376
DFT1-08	0.99949396	DFT1-08	0.01604299	DFT2-25	0.00035248	DFT2-25	0.01877456
DFT1-11	1.00079499	DFT3-02	0.01606598	DFT2-14	0.00035367	DFT2-14	0.01880610
DFT2-13	1.00120903	DFT1-10	0.01619344	DFT1-33	0.00035478	DFT1-33	0.01883554
DFT1-10	1.00391812	DFT2-20	0.01619426	DFT3-19	0.00035534	DFT3-19	0.01885049
DFT2-20	1.00533084	DFT1-23	0.01620336	DFT2-36	0.00035612	DFT2-36	0.01887122
DFT1-23	1.00675804	DFT1-11	0.01621327	DFT1-22	0.00035633	DFT1-22	0.01887682
DFT3-19	1.00928905	DFT3-19	0.01623853	DFT1-20	0.00035657	DFT1-20	0.01888318
DFT1-21	1.01147209	DFT2-13	0.01627762	DFT2-04	0.00035788	DFT2-04	0.01891765
DFT1-20	1.01667642	DFT1-31	0.01631295	DFT1-30	0.00035991	DFT1-30	0.01897134
DFT4-01	1.01893651	DFT3-13	0.01632012	DFT2-39	0.00036281	DFT2-39	0.01904767
DFT1-32	1.02516865	DFT1-21	0.01635110	DFT3-13	0.00036340	DFT3-13	0.01906315
DFT2-39	1.02588209	DFT1-20	0.01639772	DFT2-05	0.00036445	DFT2-05	0.01909069
DFT2-12	1.02713175	DFT2-39	0.01667249	DFT2-12	0.00036474	DFT2-12	0.01909807
DFT3-02	1.03012172	DFT2-29	0.01673713	DFT5-01	0.00036720	DFT5-01	0.01916251
DFT2-24	1.03112236	DFT2-36	0.01675646	DFT1-23	0.00036952	DFT1-23	0.01922286
DFT2-02	1.03117197	DFT5-03	0.01676229	DFT2-34	0.00037233	DFT2-34	0.01929582

Table 4.3. The statistical deviations of the bond angles where the metal atom is in the center from the experimental data by MAPE, MAD, MSE and RMSE analysis. The colors purple, blue, green, pink, and red represent DFT1, DFT2, DFT3, DFT4 and DFT5, respectively.

DFT	MAPE	DFT	MAD	DFT	MSE	DFT	RMSE
DFT1-37	0.24037330	DFT1-37	0.28233333	DFT1-37	0.11692777	DFT1-37	0.34194703
DFT1-32	0.32180604	DFT1-32	0.33180000	DFT1-36	0.17974770	DFT1-36	0.42396663
DFT1-31	0.34748659	DFT1-31	0.35880000	DFT1-32	0.18656283	DFT1-32	0.43192919
DFT1-36	0.34797545	DFT1-36	0.38180000	DFT1-31	0.20927674	DFT1-31	0.45746775
DFT1-33	0.37477850	DFT1-33	0.38643333	DFT1-06	0.21567240	DFT1-06	0.46440542
DFT1-06	0.38098646	DFT1-06	0.40880000	DFT1-33	0.23532149	DFT1-33	0.48509946
DFT1-30	0.46347418	DFT1-30	0.47551667	DFT1-27	0.33311138	DFT1-27	0.57715802
DFT1-27	0.46991539	DFT1-29	0.49653333	DFT1-30	0.39502272	DFT1-30	0.62850833
DFT1-29	0.48114154	DFT1-28	0.49891667	DFT1-29	0.44503032	DFT1-29	0.66710593
DFT1-28	0.48188966	DFT1-27	0.50340000	DFT1-28	0.46332169	DFT1-28	0.68067738
DFT1-13	0.50899762	DFT1-13	0.52950000	DFT1-13	0.47520599	DFT1-13	0.68935186
DFT1-12	0.55210933	DFT1-12	0.57728333	DFT1-12	0.54723106	DFT1-12	0.73975068
DFT1-49	0.59148032	DFT1-49	0.62330000	DFT1-49	0.66862788	DFT1-49	0.81769669
DFT2-27	0.59996474	DFT2-11	0.62550000	DFT2-11	0.68170311	DFT2-11	0.82565314
DFT2-11	0.60463686	DFT2-27	0.63241667	DFT2-27	0.69018403	DFT2-27	0.83077315
DFT2-46	0.60746527	DFT2-29	0.64075000	DFT1-07	0.69903140	DFT1-07	0.83608098
DFT2-29	0.60808071	DFT2-46	0.64085000	DFT2-29	0.77851053	DFT2-29	0.88233244
DFT2-30	0.62547212	DFT2-30	0.66201667	DFT2-46	0.78246193	DFT2-46	0.88456878
DFT1-01	0.63209342	DFT1-01	0.67160000	DFT2-09	0.83874924	DFT2-09	0.91583254
DFT2-47	0.63251309	DFT2-47	0.67231667	DFT1-01	0.85824800	DFT1-01	0.92641675
DFT1-39	0.63625787	DFT1-39	0.67428333	DFT1-34	0.86735713	DFT1-34	0.93132010
DFT1-07	0.66445989	DFT3-19	0.70911667	DFT1-39	0.86953674	DFT1-39	0.93248954
DFT3-19	0.66835620	DFT1-25	0.71623333	DFT2-47	0.87885180	DFT2-47	0.93747096
DFT1-25	0.67932323	DFT2-50	0.72556667	DFT2-30	0.88330233	DFT2-30	0.93984165
DFT2-50	0.68013172	DFT2-48	0.73165000	DFT1-25	0.92406712	DFT1-25	0.96128410
DFT2-48	0.68675895	DFT2-49	0.73803333	DFT1-19	0.93211579	DFT1-19	0.96546144
DFT2-49	0.69210084	DFT1-07	0.73868333	DFT2-17	0.93374937	DFT2-17	0.96630708
DFT2-28	0.70133287	DFT2-28	0.74991667	DFT1-05	0.95307020	DFT1-05	0.97625315
DFT2-05	0.70767383	DFT1-34	0.75188333	DFT1-02	0.97823146	DFT1-02	0.98905584
DFT1-34	0.71282175	DFT2-05	0.75298333	DFT3-19	0.97894490	DFT3-19	0.98941645
DFT3-22	0.71583856	DFT2-09	0.75781667	DFT2-22	0.98327023	DFT2-22	0.99159984
DFT2-09	0.71853031	DFT3-22	0.75878333	DFT3-08	0.98555316	DFT3-08	0.99275030
DFT1-05	0.72036173	DFT1-05	0.76626667	DFT2-50	0.98684549	DFT2-50	0.99340097
DFT1-02	0.72122305	DFT2-01	0.76981667	DFT1-38	1.02937143	DFT1-38	1.01457944
DFT2-01	0.72363707	DFT1-02	0.77160000	DFT1-08	1.03184608	DFT1-08	1.01579825
DFT1-50	0.73146394	DFT1-22	0.78223333	DFT2-49	1.03593631	DFT2-49	1.01780956
DFT1-22	0.73172682	DFT1-50	0.78355000	DFT2-33	1.03678201	DFT2-33	1.01822493
DFT1-51	0.73198218	DFT1-51	0.78426667	DFT1-22	1.04202687	DFT1-22	1.02079717
DFT2-40	0.73887792	DFT2-40	0.78793333	DFT3-11	1.04300400	DFT3-11	1.02127567
DFT1-38	0.74668234	DFT3-15	0.79523333	DFT1-15	1.04633290	DFT1-15	1.02290415
DFT2-21	0.75116433	DFT1-38	0.80031667	DFT1-23	1.05327004	DFT1-23	1.02628945
DFT3-27	0.75165999	DFT1-19	0.80083333	DFT2-21	1.05560926	DFT2-21	1.02742847
DFT3-15	0.75214651	DFT2-08	0.80203333	DFT2-05	1.05773433	DFT2-05	1.02846212
DFT1-04	0.75249879	DFT3-27	0.80606667	DFT1-03	1.06592192	DFT1-03	1.03243495
DFT2-08	0.75258047	DFT2-34	0.80661667	DFT2-28	1.06823303	DFT2-28	1.03355359
DFT2-34	0.75345308	DFT2-21	0.80693333	DFT1-40	1.08045588	DFT1-40	1.03944980
DFT1-19	0.75367247	DFT1-04	0.80711667	DFT1-04	1.08680044	DFT1-04	1.04249721
DFT3-11	0.75873724	DFT2-33	0.80808333	DFT1-10	1.09267499	DFT1-10	1.04531095
DFT2-33	0.76029015	DFT3-11	0.81111667	DFT2-01	1.09894185	DFT2-01	1.04830427
DFT1-03	0.76631071	DFT3-21	0.81285000	DFT1-50	1.10652075	DFT1-50	1.05191290

Table 4.4. The statistical deviations of the bond angles where the metal atom is in the center and other bond angles from the experimental data by MAPE, MAD, MSE and RMSE analysis. The colors purple, blue, green, pink, and red represent DFT1, DFT2, DFT3, DFT4 and DFT5, respectively.

DFT	MAPE	DFT	MAD	DFT	MSE	DFT	RMSE
DFT1-32	0.48497978	DFT1-32	0.55484286	DFT1-25	0.78910480	DFT1-25	0.88831571
DFT1-31	0.49806772	DFT1-31	0.56848571	DFT1-33	0.78945489	DFT1-33	0.88851274
DFT1-33	0.50302517	DFT1-33	0.57097143	DFT1-31	0.79276440	DFT1-31	0.89037318
DFT1-06	0.52293252	DFT1-06	0.60265714	DFT1-06	0.80586186	DFT1-06	0.89769809
DFT1-27	0.53495767	DFT1-28	0.60953571	DFT1-27	0.80895473	DFT1-27	0.89941911
DFT1-36	0.53806537	DFT1-27	0.61276429	DFT1-30	0.80985825	DFT1-30	0.89992124
DFT1-28	0.54127069	DFT1-30	0.61536429	DFT1-32	0.82028096	DFT1-32	0.90569363
DFT1-30	0.54685474	DFT1-36	0.62518571	DFT1-28	0.82158146	DFT1-28	0.90641131
DFT1-29	0.55762067	DFT1-29	0.62762857	DFT1-36	0.82520131	DFT1-36	0.90840591
DFT1-37	0.57233267	DFT1-13	0.65720714	DFT1-29	0.82768879	DFT1-29	0.90977403
DFT1-13	0.58230337	DFT1-25	0.66037143	DFT2-17	0.83483328	DFT2-17	0.91369211
DFT1-12	0.59410059	DFT2-27	0.66793571	DFT1-23	0.84058903	DFT1-23	0.91683643
DFT2-27	0.59515526	DFT1-12	0.66957857	DFT1-13	0.85148346	DFT1-13	0.92275861
DFT1-25	0.59620395	DFT2-46	0.67101429	DFT1-19	0.85182243	DFT1-19	0.92294227
DFT2-46	0.59622283	DFT2-29	0.67873571	DFT3-08	0.85335703	DFT3-08	0.92377326
DFT1-49	0.60303119	DFT1-37	0.67878571	DFT1-12	0.85898096	DFT1-12	0.92681226
DFT2-29	0.60346994	DFT1-49	0.67885000	DFT2-22	0.87218511	DFT2-22	0.93390851
DFT2-47	0.60850571	DFT2-47	0.68312143	DFT2-27	0.87242836	DFT2-27	0.93403873
DFT1-01	0.60950196	DFT1-01	0.68412143	DFT1-49	0.88711126	DFT1-49	0.94186584
DFT1-39	0.61583989	DFT2-30	0.69258571	DFT2-09	0.90017324	DFT2-09	0.94877460
DFT2-30	0.61741686	DFT1-39	0.69259286	DFT2-46	0.90695336	DFT2-46	0.95234099
DFT2-50	0.62732415	DFT2-50	0.70299286	DFT2-21	0.90750554	DFT2-21	0.95263085
DFT3-19	0.62939650	DFT2-48	0.70643571	DFT2-29	0.91604449	DFT2-29	0.95710213
DFT2-48	0.63002475	DFT3-19	0.70758571	DFT3-11	0.91632663	DFT3-11	0.95724951
DFT2-11	0.63135057	DFT2-11	0.70859286	DFT2-11	0.91986200	DFT2-11	0.95909436
DFT2-49	0.63259298	DFT2-49	0.70912143	DFT3-14	0.92185013	DFT3-14	0.96013026
DFT2-05	0.63995116	DFT2-05	0.71832143	DFT1-01	0.92723154	DFT1-01	0.96292863
DFT2-28	0.64100302	DFT2-09	0.71925714	DFT2-47	0.93680826	DFT2-47	0.96788856
DFT1-19	0.64191071	DFT2-28	0.71932143	DFT1-34	0.93771648	DFT1-34	0.96835762
DFT2-09	0.64216033	DFT1-19	0.72045000	DFT1-39	0.94851095	DFT1-39	0.97391527
DFT1-02	0.64694091	DFT2-21	0.72509286	DFT2-50	0.95499462	DFT2-50	0.97723826
DFT2-01	0.64709103	DFT1-02	0.72557857	DFT1-22	0.96069602	DFT1-22	0.98015102
DFT2-21	0.64805445	DFT2-01	0.72605000	DFT1-05	0.96215565	DFT1-05	0.98089533
DFT1-34	0.65037134	DFT2-33	0.73058571	DFT2-49	0.96328043	DFT2-49	0.98146851
DFT1-05	0.65117002	DFT1-34	0.73102857	DFT2-33	0.96560157	DFT2-33	0.98265028
DFT2-33	0.65220425	DFT1-05	0.73149286	DFT2-30	0.96981261	DFT2-30	0.98479064
DFT1-51	0.65225426	DFT1-51	0.73202857	DFT3-19	0.97505107	DFT3-19	0.98744675
DFT1-50	0.65282527	DFT1-50	0.73272143	DFT2-05	0.98570322	DFT2-05	0.99282588
DFT1-22	0.65285335	DFT3-22	0.73517857	DFT3-10	0.99026969	DFT3-10	0.99512295
DFT3-22	0.65517088	DFT1-22	0.73597857	DFT1-08	0.99708456	DFT1-08	0.99854121
DFT2-40	0.65587655	DFT2-40	0.73617143	DFT1-02	0.99879320	DFT1-02	0.99939642
DFT3-27	0.65997025	DFT2-17	0.73995000	DFT2-28	0.99911993	DFT2-28	0.99955987
DFT1-38	0.66146312	DFT3-27	0.74172857	DFT1-03	0.99987437	DFT1-03	0.99993718
DFT2-08	0.66201390	DFT1-38	0.74184286	DFT2-01	1.00446000	DFT2-01	1.00222752
DFT5-04	0.66276052	DFT2-08	0.74260000	DFT2-48	1.00769062	DFT2-48	1.00383795
DFT2-17	0.66358741	DFT1-04	0.74438571	DFT1-10	1.00779916	DFT1-10	1.00389200
DFT1-04	0.66364038	DFT5-04	0.74625714	DFT2-40	1.01744488	DFT2-40	1.00868473
DFT2-38	0.66595252	DFT2-38	0.74672857	DFT1-38	1.01769019	DFT1-38	1.00880632
DFT2-04	0.66615371	DFT2-04	0.74759286	DFT2-38	1.01836460	DFT2-38	1.00914052
DFT3-11	0.66883129	DFT1-23	0.74915000	DFT3-22	1.03236560	DFT3-22	1.01605394

Table 4.5. The statistical deviations of the all-bond properties from the experimental data by MAPE, MAD, MSE and RMSE analysis. The colors purple, blue, green, pink, and red represent DFT1, DFT2, DFT3, DFT4 and DFT5, respectively.

DFT	MAPE	DFT	MAD	DFT	MSE	DFT	RMSE
DFT1-33	0.63187403	DFT1-32	0.34394791	DFT1-25	0.48052547	DFT1-25	0.66489355
DFT1-06	0.63274868	DFT1-31	0.35241811	DFT1-33	0.48067659	DFT1-33	0.66499809
DFT1-32	0.64064939	DFT1-33	0.35343897	DFT1-31	0.48272593	DFT1-31	0.66641418
DFT1-01	0.64252465	DFT1-06	0.37258333	DFT1-06	0.49065247	DFT1-06	0.67186329
DFT2-47	0.64535719	DFT1-28	0.37688043	DFT1-27	0.49259958	DFT1-27	0.67319508
DFT1-28	0.64541022	DFT1-27	0.38039233	DFT1-30	0.49309803	DFT1-30	0.67353559
DFT2-27	0.64667871	DFT1-30	0.38060532	DFT1-32	0.49945954	DFT1-32	0.67786634
DFT1-31	0.65732451	DFT1-29	0.38799033	DFT1-28	0.50022106	DFT1-28	0.67838291
DFT1-30	0.66185237	DFT1-36	0.38822436	DFT1-36	0.50252922	DFT1-36	0.67994623
DFT1-29	0.66352650	DFT1-13	0.40618227	DFT1-29	0.50394751	DFT1-29	0.68090507
DFT2-46	0.66569091	DFT1-25	0.40906502	DFT2-17	0.50851870	DFT2-17	0.68398626
DFT2-28	0.66621880	DFT2-27	0.41205952	DFT1-23	0.51180748	DFT1-23	0.68619449
DFT2-50	0.66691690	DFT1-12	0.41345063	DFT1-13	0.51845919	DFT1-13	0.69063916
DFT1-12	0.67294258	DFT2-46	0.41418985	DFT1-19	0.51867184	DFT1-19	0.69078078
DFT1-13	0.67676201	DFT1-49	0.41912936	DFT3-08	0.51962709	DFT3-08	0.69141661
DFT1-49	0.67824923	DFT2-29	0.41969279	DFT1-12	0.52300796	DFT1-12	0.69366225
DFT2-48	0.68390408	DFT2-47	0.42115971	DFT2-27	0.53114985	DFT2-27	0.69904067
DFT2-21	0.68703160	DFT1-01	0.42170214	DFT2-22	0.53120326	DFT2-22	0.69907582
DFT2-49	0.68750833	DFT1-37	0.42564437	DFT1-49	0.54013411	DFT1-49	0.70492793
DFT2-30	0.69089508	DFT2-30	0.42748187	DFT2-09	0.54821944	DFT2-09	0.71018440
DFT1-39	0.69229125	DFT1-39	0.42761564	DFT2-46	0.55218000	DFT2-46	0.71274512
DFT1-22	0.70408742	DFT2-50	0.43344980	DFT2-21	0.55249488	DFT2-21	0.71294831
DFT3-04	0.71108797	DFT2-48	0.43578987	DFT2-29	0.55774359	DFT2-29	0.71632681
DFT1-16	0.71236995	DFT3-19	0.43705855	DFT3-11	0.55795453	DFT3-11	0.71646226
DFT3-19	0.71580610	DFT2-49	0.43747130	DFT2-11	0.56019533	DFT2-11	0.71789951
DFT2-25	0.71601451	DFT2-11	0.44035263	DFT3-14	0.56128488	DFT3-14	0.71859731
DFT1-05	0.71737467	DFT2-28	0.44325856	DFT1-01	0.56450335	DFT1-01	0.72065462
DFT4-03	0.71778330	DFT2-05	0.44385107	DFT2-47	0.57033452	DFT2-47	0.72436714
DFT2-29	0.71791111	DFT1-19	0.44568412	DFT1-34	0.57105960	DFT1-34	0.72482745
DFT1-25	0.71990347	DFT2-21	0.44710040	DFT1-39	0.57750347	DFT1-39	0.72890547
DFT1-27	0.72062600	DFT2-09	0.44716478	DFT2-50	0.58140137	DFT2-50	0.73136124
DFT1-03	0.72160827	DFT2-01	0.44889952	DFT1-22	0.58491092	DFT1-22	0.73356530
DFT3-27	0.72163823	DFT1-02	0.45014511	DFT1-05	0.58579314	DFT1-05	0.73411831
DFT1-36	0.73081683	DFT1-05	0.45140809	DFT2-49	0.58645131	DFT2-49	0.73453060
DFT2-05	0.73570663	DFT2-33	0.45224552	DFT2-33	0.58794340	DFT2-33	0.73546443
DFT3-26	0.73680317	DFT1-51	0.45321119	DFT2-30	0.59043961	DFT2-30	0.73702404
DFT3-06	0.73714075	DFT1-50	0.45365197	DFT3-19	0.59364840	DFT3-19	0.73902404
DFT1-23	0.73792651	DFT1-34	0.45381741	DFT2-05	0.60013588	DFT2-05	0.74305115
DFT1-08	0.73892458	DFT1-22	0.45383346	DFT3-10	0.60307612	DFT3-10	0.74486914
DFT1-10	0.73914013	DFT3-22	0.45436246	DFT1-08	0.60705743	DFT1-08	0.74732379
DFT2-44	0.74091971	DFT2-40	0.45490505	DFT1-02	0.60825588	DFT1-02	0.74806110
DFT4-05	0.74125138	DFT3-27	0.45769902	DFT2-28	0.60826120	DFT2-28	0.74806437
DFT4-02	0.74740957	DFT2-08	0.45915010	DFT1-03	0.60875273	DFT1-03	0.74836656
DFT2-20	0.74853199	DFT1-38	0.46011860	DFT2-01	0.61157032	DFT2-01	0.75009646
DFT2-04	0.74868503	DFT2-17	0.46103517	DFT2-48	0.61348641	DFT2-48	0.75127059
DFT1-21	0.74894116	DFT5-04	0.46128565	DFT1-10	0.61357892	DFT1-10	0.75132723
DFT2-40	0.75483426	DFT1-04	0.46140916	DFT2-40	0.61946684	DFT2-40	0.75492350
DFT1-20	0.75737147	DFT2-04	0.46166674	DFT1-38	0.61976075	DFT1-38	0.75510257
DFT1-11	0.75740882	DFT2-38	0.46189798	DFT2-38	0.62005117	DFT2-38	0.75527947
DFT2-24	0.75852040	DFT1-23	0.46234479	DFT3-22	0.62855717	DFT3-22	0.76044237

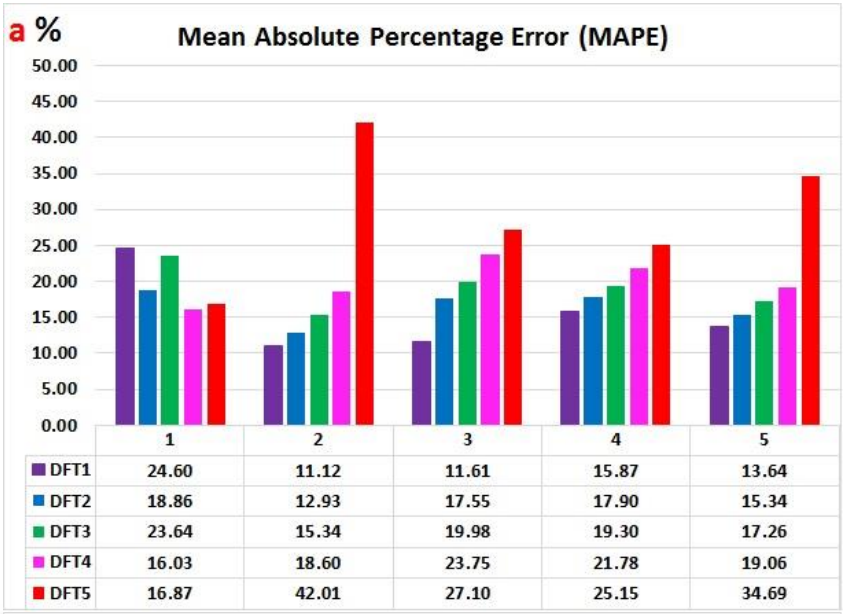
DFT functions were also evaluated according to the success of the groups in the figure 4.4 for these comparison calculations. 139 DFT functions were used throughout the calculations. The functions listed in table 3.2 and 3.3 contain the numbering of the DFT1-5 as GGA, hybrid-GGA, hybrid-meta-GGA, meta-GGA and range-separated, respectively.

The functions belonging to the TH group in DFT1 41-47 failed to calculate. These functions, which could not complete the geometric optimization, could not convergence the SCF energy. Therefore, GGA (DFT1), 44; hybrid-GGA (DFT2), 50; hybrid-meta-GGA (DFT3), 29; meta-GGA (DFT4), 5 and range-separated (DFT5), 4; from DFT groups were completed successfully.

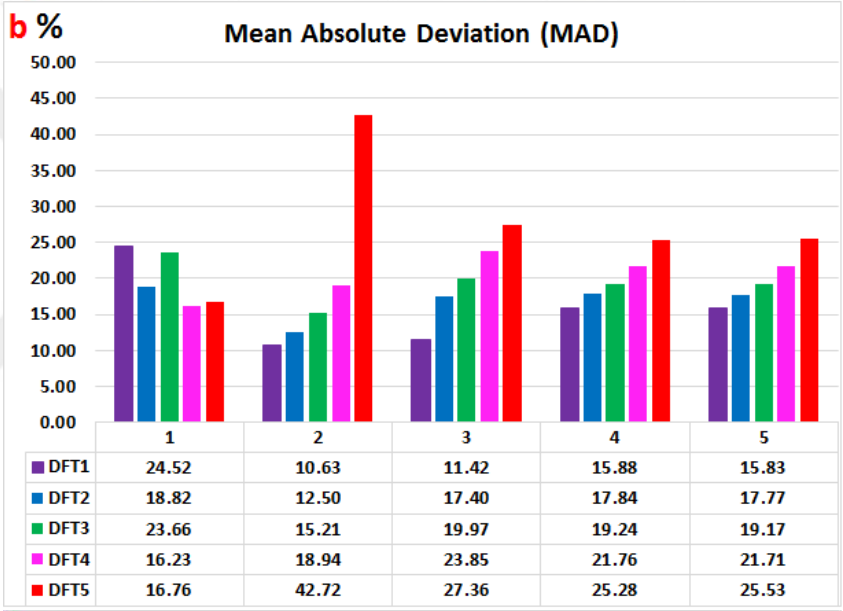
In figure 4.4, there are four column percentage comparison graphs between "a-d". These include MAPE, MAD, MSE, and RMSE, respectively. Numbers of 1-5 at the bottom of each graph represents metal-containing bond lengths, metal-containing and other bond lengths, metal-centered angles, metal-centered and metal-included angles, and all bond properties, and colors indicate DFT1-5, respectively. In addition, the values in the graphs are the total deviation value of each group as a percentage.

Similar results are observed in all statistical deviation parts of the graphs approximately. Considering graph 4.4b, according to the deviations in the lengths of the metal-containing bonds in the 1st group, the total deviation of the meta-GGA (DFT4) seems to be lower in general, although the HCTH group, which is one of the GGA functions, is successful. Interestingly, the most unsuccessful DFT subgroup was identified as GGA (DFT1). Some functions in DFT1 show high bias. In these cases, it is possible to cause instability in the geometry of the structure.

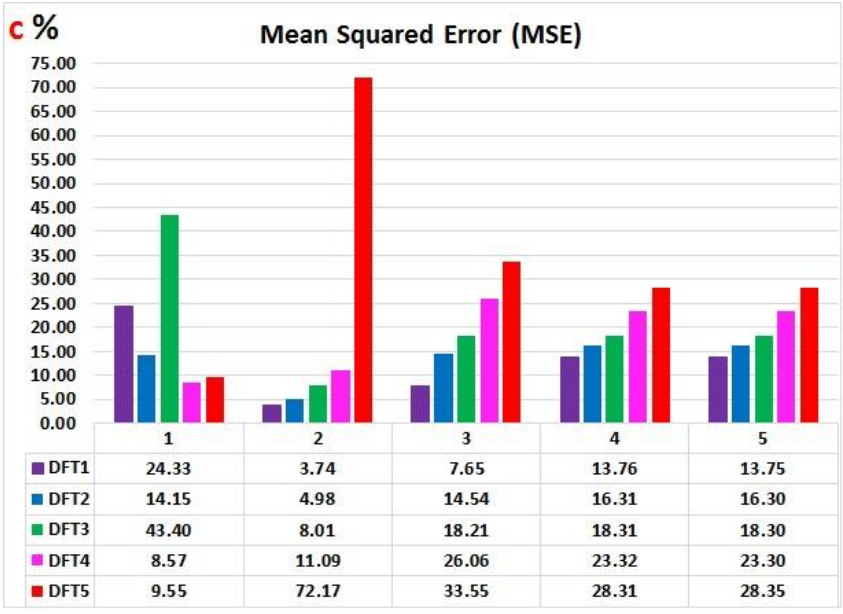
The increase in total deviations from DFT1 to DFT5 is obvious when we examine all the graphs in figure 4.4 and groups 2, 3, 4 and 5. There is a similar upward trend in all divergence parts. The success in GGA (DFT1) is in line with the previous results. Range-separated functionals (DFT5) are out of the trends with their high percentage of deviation. As a result, GGA group DFT functionals generally have lower deviation values in metal-containing complexes than other DFT groups.



	1	2	3	4	5
DFT1	24.60	11.12	11.61	15.87	13.64
DFT2	18.86	12.93	17.55	17.90	15.34
DFT3	23.64	15.34	19.98	19.30	17.26
DFT4	16.03	18.60	23.75	21.78	19.06
DFT5	16.87	42.01	27.10	25.15	34.69

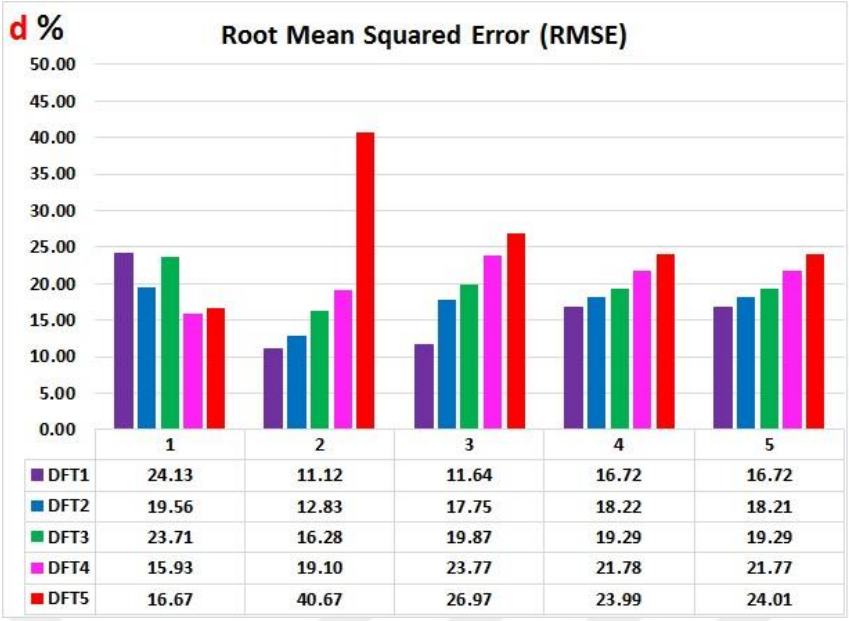


	1	2	3	4	5
DFT1	24.52	10.63	11.42	15.88	15.83
DFT2	18.82	12.50	17.40	17.84	17.77
DFT3	23.66	15.21	19.97	19.24	19.17
DFT4	16.23	18.94	23.85	21.76	21.71
DFT5	16.76	42.72	27.36	25.28	25.53



	1	2	3	4	5
DFT1	24.33	3.74	7.65	13.76	13.75
DFT2	14.15	4.98	14.54	16.31	16.30
DFT3	43.40	8.01	18.21	18.31	18.30
DFT4	8.57	11.09	26.06	23.32	23.30
DFT5	9.55	72.17	33.55	28.31	28.35

Figure 4.4 (Continued)



	1	2	3	4	5
DFT1	24.13	11.12	11.64	16.72	16.72
DFT2	19.56	12.83	17.75	18.22	18.21
DFT3	23.71	16.28	19.87	19.29	19.29
DFT4	15.93	19.10	23.77	21.78	21.77
DFT5	16.67	40.67	26.97	23.99	24.01

Figure 4.4. Comparative column chart. 1-5 at the bottom of each graph represents metal-containing bond lengths, metal-containing and other bond lengths, metal-centered angles, metal-centered and metal-included angles, and all bond properties, and colors indicate DFT1-5, respectively. a-MAPE, b-MAD, c-MSE, d-RMSE.

The best DFT method to be used in the next calculations was determined by statistical analyzes in the previous section. In the literature review, 154 compounds determined according to table 3.4 were divided into groups for ease of calculation. Single crystal states were investigated from the CCDC database for the complex in each group. It was saved without any conformational screening for those present with a single crystal structure. Structures that could not be found in CCDC, but only chemically synthesized structures were drawn in 3D format by looking at their equivalents and their own 2D structures in the article. The drawn structures were subjected to conformational analysis. For this analysis, Wavefunction Spartan 2018 program with a 6-month license period was used. The conformers of each complex were determined by molecular mechanical conformer scanning. These complexes were then ranked according to their energies. Those with less than 1000 kcal energy were kept. Geometric optimizations were made using the most known DFT method, B3LYP and LACVP basis set group, and ranked according to their energies for the remaining structures. The conformers with the lowest energy were selected and stored in the "xyz" format.

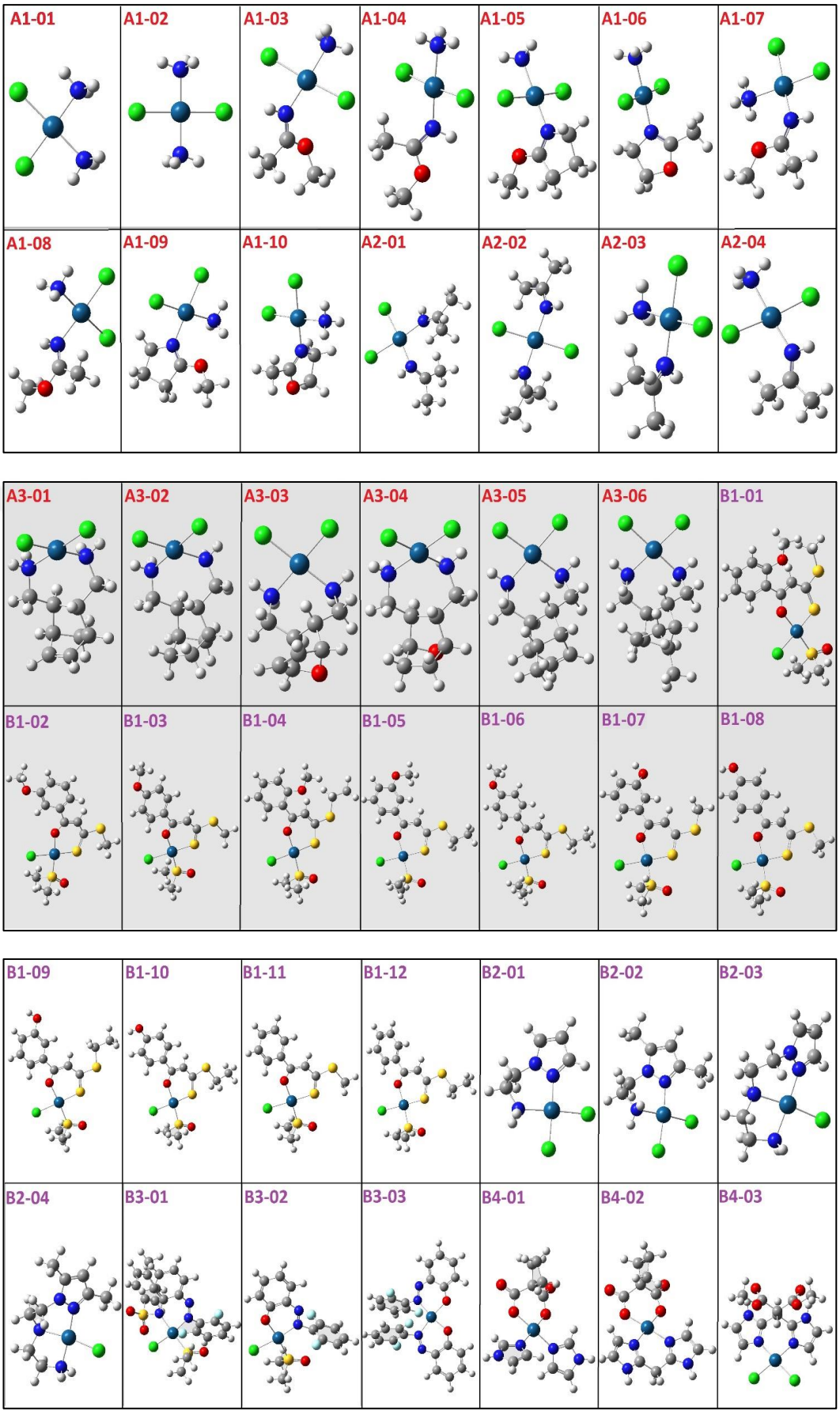


Figure 4.5 (continued)

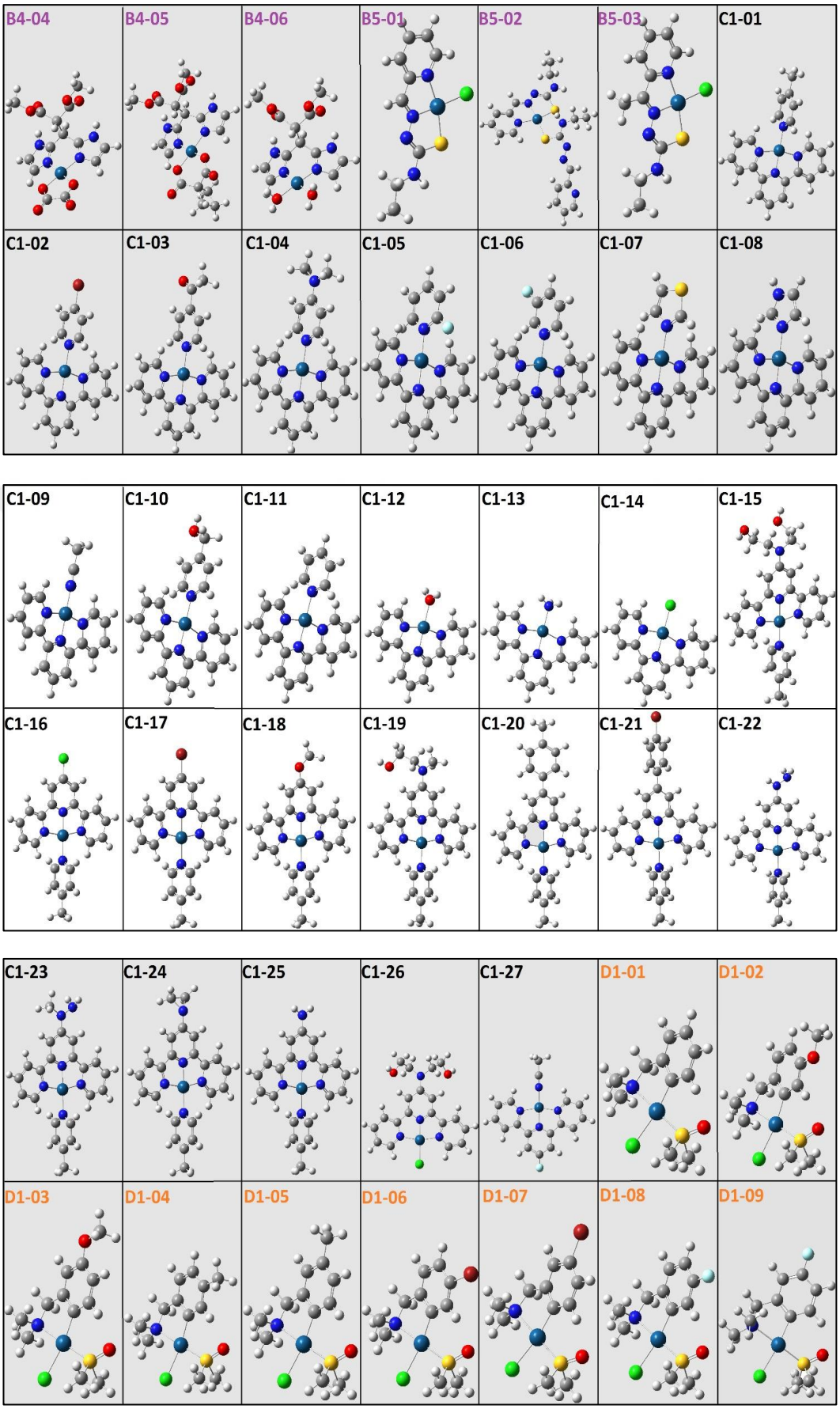


Figure 4.5 (continued)

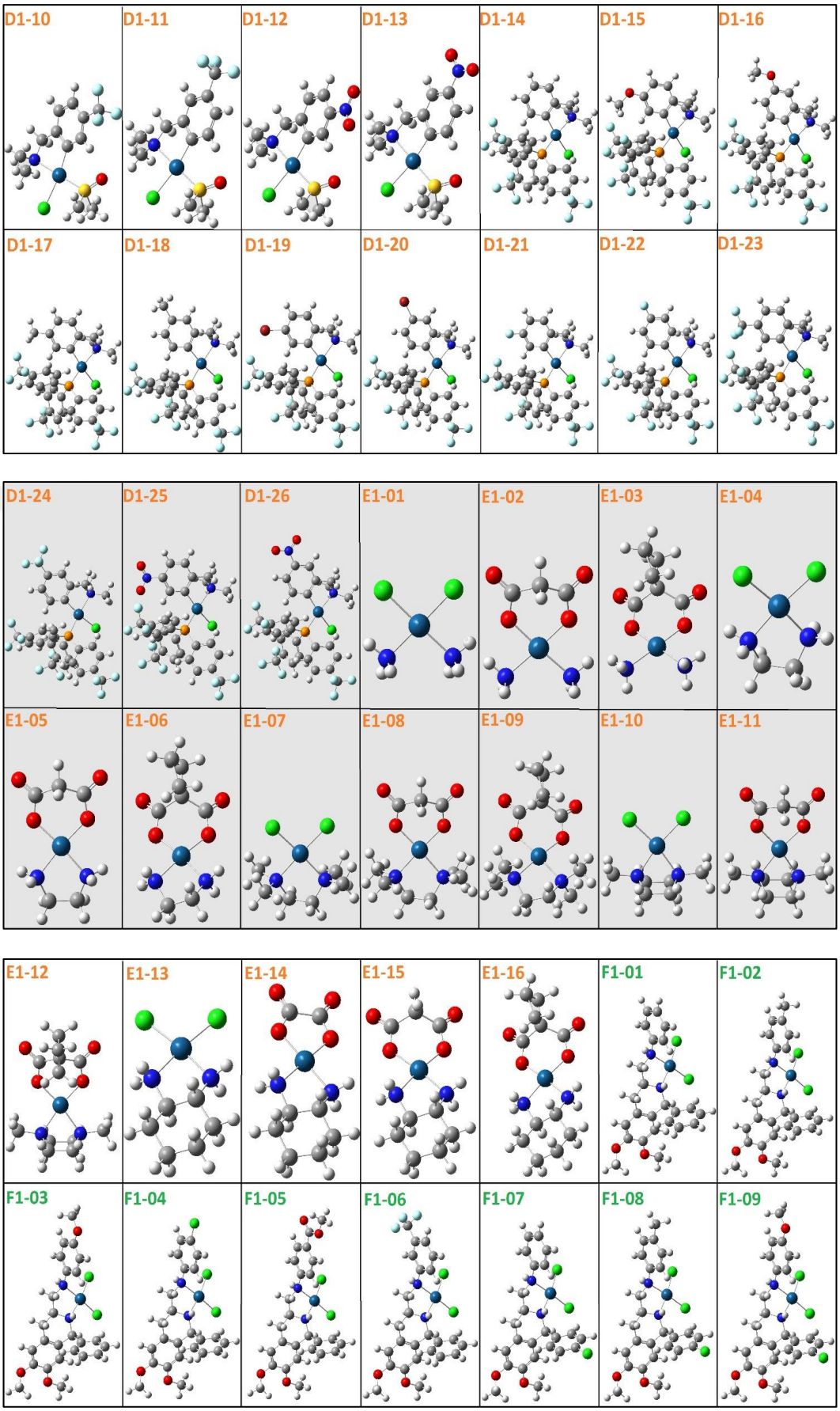


Figure 4.5 (continued)

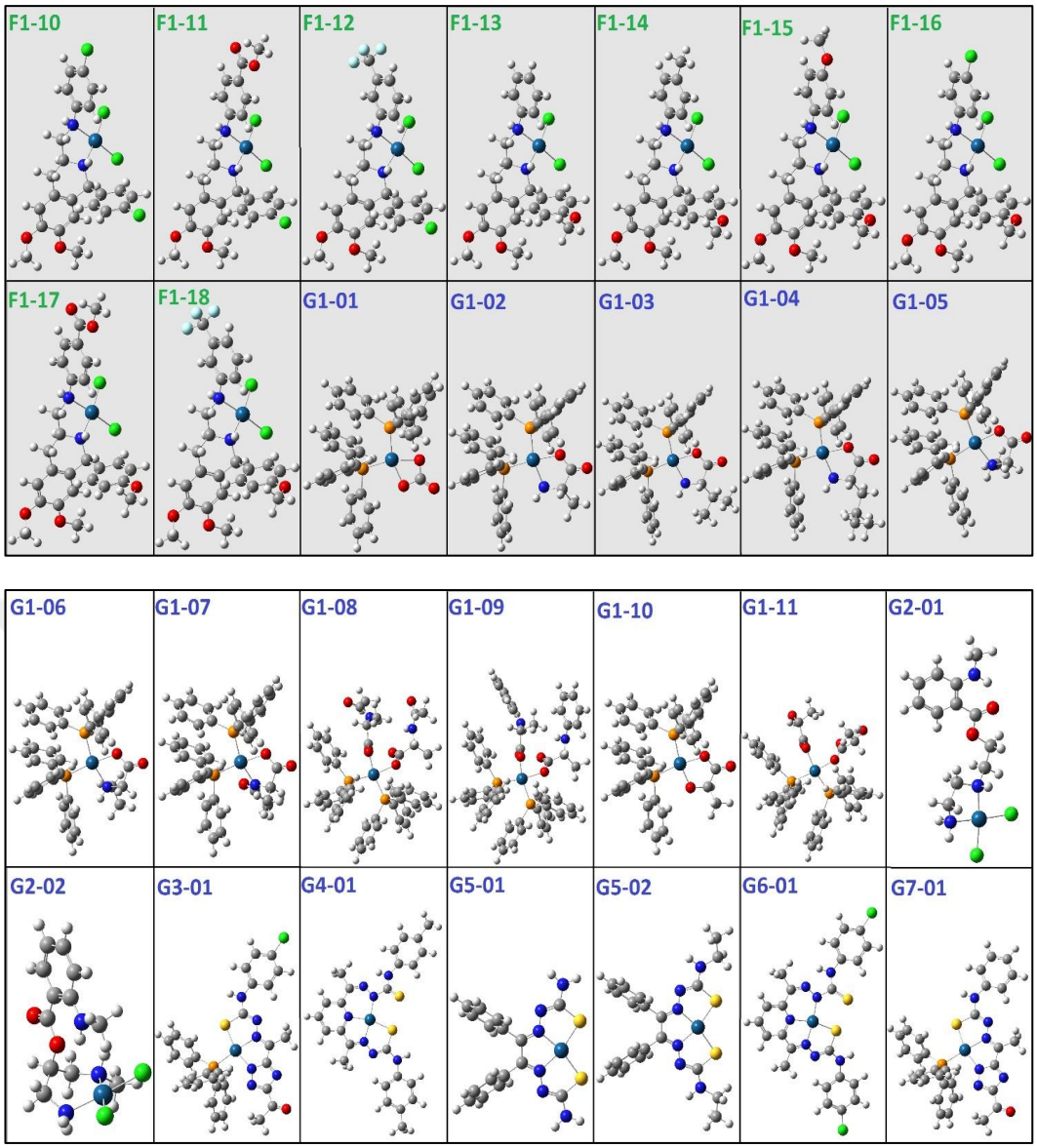


Figure 4.5. All 154 platinum complexes selected from literature. Arrangement of atoms by color: Red, Oxygen; Navy Blue, Nitrogen; White, Hydrogen; Grey, Carbon; Yellow, Sulfur; Orange, Phosphorus; Blue, Platinum; Green, Chlorine; Ice blue, Fluorine; Claret red; Bromine.

The most stable structures were revealed by geometric optimizations as described in section 3.2.4 after the best conformers were determined the complex. The stable geometries of the molecules in figure 4.5 were taken from the final step of the optimizations using the Chemcraft program and saved in "mol" format at the end of the calculations made with ORCA 4.2.

After revealing the most stable geometries of the platinum-containing complexes from the literature, new drug candidate complexes containing platinum

and 2-pyrimidinethiol were formed, like the skeletons of the complexes in each group.

After drawings for all new drug candidate compounds were made using Chemcraft, possible conformers of functional groups in space were designed one by one considering all the angles it could rotate. Five possible conformers for the EX-01 design is drawn as an example. The most stable optimizations with geometric minimum energy were found in ORCA 4.2 for these drawings. Then, the structures with the highest energy were eliminated from those that were geometrically similar structures. All conformers were compared in the own complex. The lowest energy (hartrees in the unit) conformer was selected. As seen in figure 4.6. EX-01_1 which is the lowest energy conformer for EX-01 was selected for further calculations. As seen here, rotatable such as 2-pyrimidinethiol and the -SH group in the ligand were manually prepared in Gaussview 3 and conformers were formed. The lowest energy conformer structures were designed by applying the same methods for the calculations in all other groups.

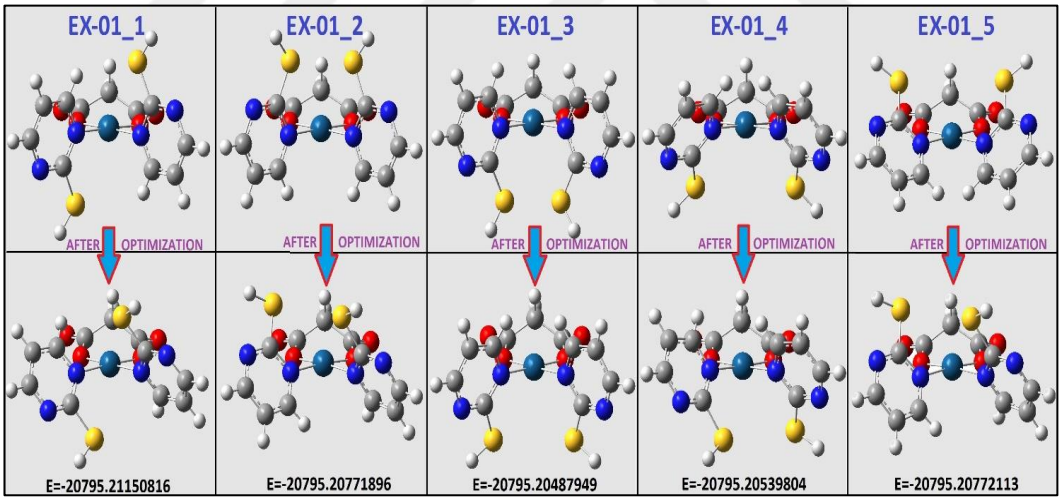


Figure 4.6. Initial drawings of all possible conformers and representation of the lowest energy conformers formed after optimization for EX-01 design. Arrangement of atoms by color: Red, Oxygen; Navy Blue, Nitrogen; White, Hydrogen; Grey, Carbon; Yellow, Sulfur.

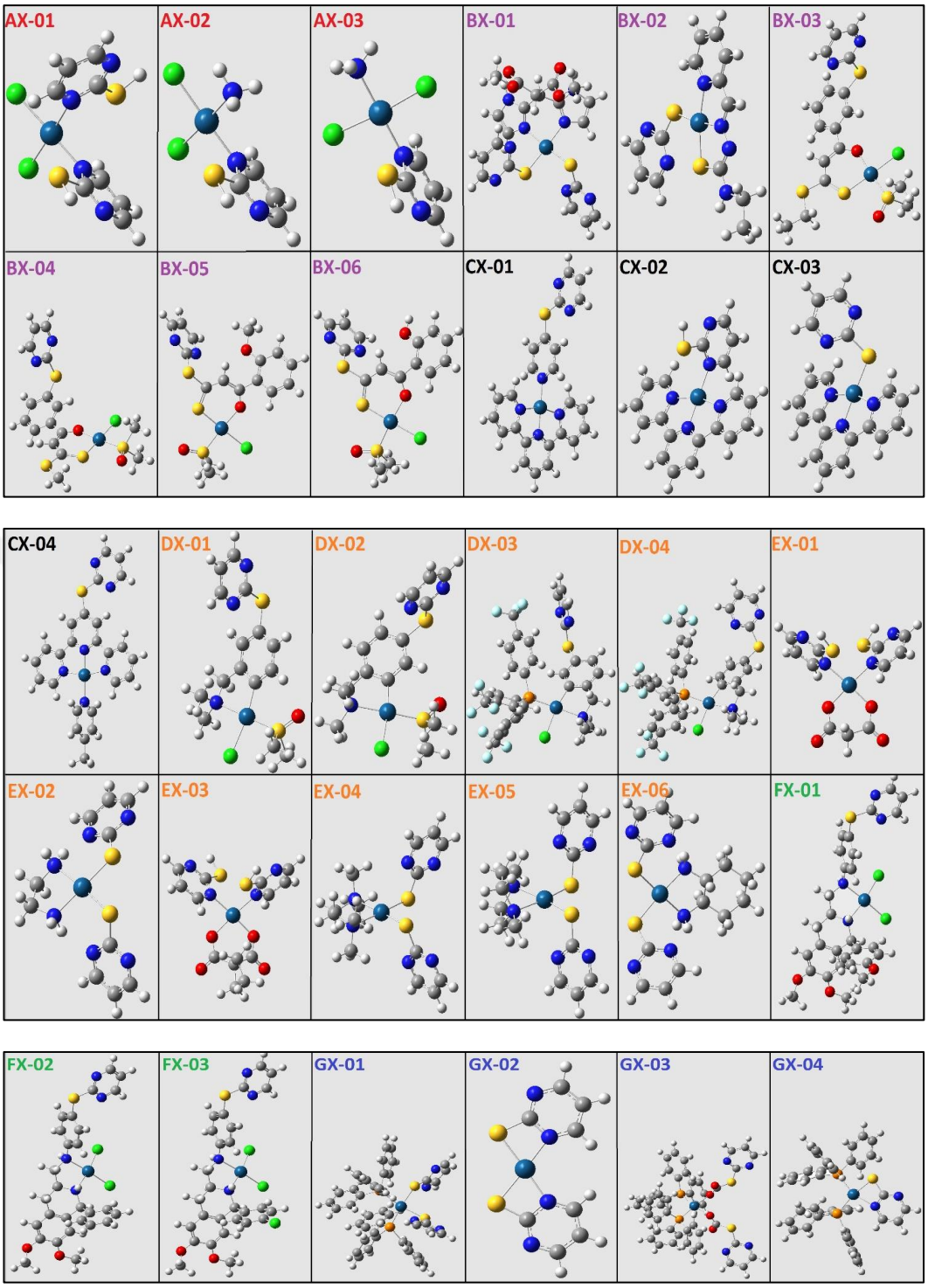


Figure 4.7. All novel platinum-2-pyrimidinethiol complexes. Arrangement of atoms by color: Red, Oxygen; Navy Blue, Nitrogen; White, Hydrogen; Grey, Carbon; Yellow, Sulfur; Orange, Phosphorus; Blue, Platinum; Green, Chlorine; Ice blue, Fluorine; Claret red, Bromine.

Geometric optimizations and the selection of the best complexes were performed for a totally 30 newly designed compounds. Final geometries taken from ORCA 4.2 output files are listed in figure 4.7.

The chemical, biological and physical descriptors of the structures were calculated for all complexes in the next step. Descriptors were calculated with NMR and geometric optimization in the ORCA 4.2 calculations for the first stage, then with quantum mechanics and finally with PaDEL-Descriptor.

As explained in section 3.2.6.1, 127 different descriptors were taken from the "out" file of geometric optimizations of complexes. It is very difficult to obtain these descriptorsone by one from the 154 complexes. To facilitate this, the "MID()" formula in excel was used. All data are transferred automatically with the "MID()" commands created for the values in figure 4.8 placed in a column in excel. The example image is the parameters of the A1-01 coded complex in the optimization result. Hirshfeld's chemical hardness (η), chemical potential (μ), electrophilicity index (ω) data from HOMO-LUMO values in figure 4.7 were calculated according to formulas 6,7 and 8, respectively. Also, the differences (e.g.: (LUMO+2)-(HOMO-1)) between all HOMO(-n) values and other LUMO(+n) values were calculated and added as descriptors.

$$\eta \approx \frac{\left(E_{LUMO} - E_{HOMO}\right)}{2} \quad (6)$$

$$\mu \approx \frac{\left(E_{HOMO} + E_{LUMO}\right)}{2} \quad (7)$$

$$\omega = \frac{\mu^{2}}{2\eta} \quad (8)$$

After the descriptors taken from the output files of the geometric optimizations, it was interesting that the theoretical Pt-NMR data in the study by Tsipis and Karapetsas (132) were successful in QSAR analysis. Therefore, in this section, the best DFT method was first determined in the NMR calculations made with the ORCA 4.2 program according to section 3.2.6.2. Using the mentioned general calculation process, 57 DFT methods have been tested. The top 30 ranks with the lowest statistical deviation were listed as seen in table 4.6. Especially, according to MAD and other statistical analysis results, DFT1-23 coded DFT method was successful in the calculation method. The DFT1-23 method is the RPBE function with D3BJ dispersion correction as seen in table 3.2.

```
The origin for moment calculation is the CENTER OF MASS  = ( 1.413938,  2.387978  3.816622)
-------------
DIPOLE MOMENT
-------------
                                X             Y             Z
Electronic contribution:      -0.94291       1.37771       1.48931
Nuclear contribution   :       2.61600      -3.81998      -4.12707
                        -----------------------------------------
Total Dipole Moment    :       1.67308      -2.44228      -2.63777
                        -----------------------------------------
Magnitude (a.u.)       :       3.96506
Magnitude (Debye)      :      10.07839
--------------------
Rotational spectrum
--------------------
Rotational constants in cm-1:     0.073591     0.056872     0.032418
Rotational constants in MHz :  2206.188406  1704.969040   971.858551
 Dipole components along the rotational axes:
x,y,z [a.u.] :    -0.000198    -3.965064    -0.000361
x,y,z [Debye]:    -0.000504   -10.078391    -0.000918

MAYER POPULATION ANALYSIS (Pt)
------------------------------
ATOM       NA         ZA         QA         VA         BVA        FA
0 Pt    77.7010    78.0000     0.2990     2.8901     2.8901     0.0000

LOEWDIN REDUCED ORBITAL CHARGES (Pt)
------------------------------------
0 Pts       :    10.573694  s :    10.573694
    pz      :     8.285923  p :    24.885917
    px      :     8.226152
    py      :     8.373842
    dz2     :     5.847153  d :    29.342752
    dxz     :     5.786893
    dyz     :     5.999155
    dx2y2   :     5.672459
    dxy     :     6.037091
    f0      :     2.021513  f :    14.245046
    f+1     :     2.046122
    f-1     :     2.037573
    f+2     :     2.043792
    f-2     :     2.019603
    f+3     :     2.024889
    f-3     :     2.051555

MULLIKEN REDUCED ORBITAL CHARGES (Pt)
-------------------------------------
0 Pts       :    10.630306  s :    10.630306
    pz      :     8.130595  p :    24.391627
    px      :     8.112759
    py      :     8.148273
    dz2     :     5.708901  d :    28.608264
    dxz     :     5.611566
    dyz     :     5.888624
    dx2y2   :     5.451301
    dxy     :     5.947873
    f0      :     2.007641  f :    14.070765
    f+1     :     2.012066
    f-1     :     2.010027
    f+2     :     2.012684
    f-2     :     2.006288
    f+3     :     2.008835
    f-3     :     2.013224

LOEWDIN ATOMIC CHARGE
---------------------
Pt:   -1.047410

MULLIKEN ATOMIC CHARGE
----------------------
Pt:    0.299039

ORBITAL ENERGIES
----------------
LOWEST:  -3149.784228
HOMO-5:  -0.262463
HOMO-4:  -0.223270
HOMO-3:  -0.201081
HOMO-2:  -0.196838
HOMO-1:  -0.192448
HOMO  :  -0.185257
LUMO  :  -0.081711
LUMO+1:  -0.064282
LUMO+2:  -0.017153
LUMO+3:  -0.004197
LUMO+4:   0.031858
LUMO+5:   0.043593

TOTAL SCF ENERGY
----------------
Total Energy       :       -20087.47691643 Eh
Nuclear Repulsion  :         1169.09301989 Eh
Electronic Energy  :       -21256.56993632 Eh
One Electron Energy:       -29639.12596385 Eh
Two Electron Energy:         8382.55602753 Eh
Potential Energy   :       -42386.20752061 Eh
Kinetic Energy     :        22298.73060418 Eh
E(XC Energy)       :     -443.075184859560 Eh

N(Total Electron)  :      132.000001186365 electron
Virial Ratio       :            1.90083500
```

Figure 4.8. An example of values from geometric optimization calculations in ORCA 4.2. The values refer to the A1-01 complex.

Experimental chemical shift values of cisplatin in figure 3.5 were used for comparison in NMR calculations. Chemical shift calculations of all atoms were

made with formula 5. The standard deviations from the experimental data for the atoms (Pt, N, and H), and the MAD statistical analysis for the platinum atom were performed with these calculations.

Table 4.6. The statistical deviations of the NMR valued from the experimental data by MAD analysis. The colors purple, blue, green, pink, and red represent DFT1, DFT2, DFT3, DFT4 and DFT5, respectively.

Isotropic Shielding	Cisplatin				Hexaplatinated	Nitromethane	TMS	Standard Deviation from Experimantal Value						RESULTS
	H_{ort}	N_{ort}	Cl_{ort}	Pt	Pt	N	H_{ort}	Pt-NMR	Pt (%SD)	N-NMR	N (%SD)	H_{ort}	H_{ort} (%SD)	MAD
DFT1-23	29.074	236.608	1001.584	1452.860	-968.497	-127.641	31.606	-2419.014	15.1363	-364.203	14.2851	2.532	35.5599	126.70306083
DFT2-17	29.302	245.583	1037.823	1548.428	-907.652	-126.214	31.703	-2453.853	16.7945	-371.750	12.5089	2.401	38.8933	135.84389428
DFT1-07	29.112	241.190	1009.973	2030.851	-461.036	-103.045	31.530	-2490.739	18.5502	-344.200	18.9928	2.418	38.4671	157.31696831
DFT1-19	29.305	239.418	1005.953	1514.232	-1019.729	-117.197	31.761	-2531.380	20.4845	-356.573	16.0808	2.456	37.5044	166.72696333
DFT2-09	29.323	245.687	1038.118	1489.741	-1164.686	-126.243	31.691	-2651.339	26.1941	-371.883	12.4776	2.368	39.7458	201.63949281
DFT1-11	28.883	236.815	988.820	1339.459	-1374.657	-131.467	31.436	-2710.390	29.0048	-368.233	13.3365	2.553	35.0467	222.47813005
DFT4-05	29.414	242.666	1016.236	1589.441	-1131.587	-126.201	31.970	-2717.952	29.3647	-368.820	13.1983	2.557	34.9428	224.80173250
DFT1-25	29.397	234.839	992.635	802.876	-1951.379	-147.546	31.762	-2748.891	30.8373	-382.329	10.0192	2.365	39.8242	230.67579156
DFT1-20	28.948	236.976	995.587	1411.178	-1324.371	-128.277	31.443	-2731.931	30.0300	-365.206	14.0490	2.496	36.4972	230.68653187
DFT1-21	29.002	236.960	994.036	1368.832	-1393.987	-130.954	31.516	-2758.973	31.3171	-367.866	13.4230	2.514	36.0412	238.80787392
DFT1-10	29.009	237.000	993.899	1358.315	-1414.905	-130.963	31.511	-2769.302	31.8087	-367.914	13.4116	2.503	36.3148	242.23818516
DFT4-03	29.428	242.755	1016.144	1559.660	-1219.345	-126.232	31.959	-2775.621	32.1095	-368.940	13.1702	2.531	35.6086	243.99335070
DFT1-08	28.963	237.054	995.555	1386.249	-1404.166	-128.302	31.434	-2786.502	32.6274	-365.309	14.0248	2.471	37.1206	248.85083235
DFT3-06	29.680	247.830	1051.617	1456.330	-1372.292	-130.866	32.042	-2824.746	34.4477	-378.646	10.8859	2.362	39.9090	257.18936680
DFT4-02	29.653	248.063	1006.782	2980.464	1573.411	-91.108	33.336	-1409.270	32.9238	-339.140	20.1835	3.683	6.2946	259.24563804
DFT4-04	29.647	248.034	1006.863	2988.939	1608.941	-91.096	33.343	-1382.222	34.2112	-339.099	20.1932	3.696	5.9553	268.27100683
DFT1-05	29.088	236.725	994.451	1361.513	-1502.907	-131.325	31.546	-2860.122	36.1314	-368.002	13.3910	2.458	37.4450	272.49714010
DFT3-04	29.692	247.908	1051.741	1427.942	-1448.828	-130.891	32.032	-2872.608	36.7258	-378.749	10.8615	2.340	40.4688	273.11636164
DFT1-03	29.013	236.398	989.337	1373.243	-1505.448	-131.668	31.487	-2874.364	36.8093	-368.017	13.3874	2.474	37.0464	277.23422431
DFT1-22	29.188	236.845	1001.941	1312.023	-1564.359	-128.066	31.547	-2871.889	36.6915	-364.864	14.1294	2.359	39.9854	277.49882879
DFT3-05	30.033	254.975	1098.459	1243.496	-1702.226	-137.004	32.139	-2940.716	39.9675	-391.925	7.7606	2.107	46.3956	291.50476022
DFT1-06	29.451	239.813	1008.668	1346.943	-1577.894	-117.536	31.694	-2920.229	38.9923	-357.307	15.9081	2.243	42.9244	296.16987106
DFT2-19	29.709	251.421	1087.241	1121.514	-1868.395	-156.891	31.750	-2984.333	42.0435	-408.247	3.9192	2.041	48.0603	300.62480195
DFT2-16	29.703	248.170	1070.572	1147.242	-1855.014	-141.415	31.813	-2996.697	42.6319	-389.530	8.3243	2.110	46.3024	310.96228440
DFT3-03	30.048	255.060	1099.159	1202.257	-1813.983	-137.046	32.128	-3010.778	43.3022	-392.052	7.7308	2.080	47.0678	314.82549820
DFT2-15	29.615	248.110	1066.920	1118.908	-1896.685	-141.760	31.749	-3009.884	43.2596	-389.815	8.2573	2.134	45.7023	315.25518562
DFT2-10	29.718	251.476	1087.586	1097.175	-1929.444	-139.731	31.743	-3020.791	43.7787	-391.152	7.9425	2.025	48.4801	318.48115965
DFT4-01	30.154	251.287	1058.621	1708.836	-1297.191	-98.755	32.450	-3002.133	42.8907	-350.007	17.6260	2.296	41.5778	325.88657810
DFT2-04	29.626	248.184	1067.186	1089.288	-1967.535	-141.777	31.741	-3050.820	45.2080	-389.906	8.2359	2.115	46.1836	328.87656665

To minimize program-related errors during NMR calculations, the inputs were prepared together with the geometric optimization of the studied DFT method. When table 4.6 is examined, the success of the GGA methods for the chemical shift value of the platinum atom is striking. The DFT method with the lowest standard deviation is DFT4-04 for the chemical shift value of hydrogen atom. It is seen that the method with the lowest deviation value among the chemical shift values of the nitrogen atom is DFT2-19. For the chemical shift values of hydrogen and nitrogen, GGA methods did not show sufficient success.

After choosing the DFT method with the lowest deviation for the chemical shift value of the platinum atom, NMR calculations were performed with the ORCA 4.2 program for the complexes taken from the literature and the newly designed platinum-containing complexes. In the calculations, ZORA relativistic approximation system basis set was used together with the DFT1-23 coded RPBE-D3BJ functional. SARC-ZORA-TZVP was used for metal basis set and ZORA-def2-TZVP was used for non-metal atoms as basis set combination. After the optimization and NMR calculations were made with the "compound" system in Appendix 1, the chemical shift values in all complexes were calculated using the formula 5 according to the chemical shift value of the platinum atom in the selected hexaplatin compound as a reference.

The descriptors shown in figure 4.9 are automatically retrieved with the "MID()" command, which is used to take descriptors from previous optimization calculations. Among the descriptors selected from the NMR output file are the following data as example of G1-02 complex code. As these data, descriptors such as NMR shielding, chemical shift, contributions to shielding parameters and diagonalized matrix of Pt atom in complex were distinguished among values.

After obtaining the NMR descriptors, various quantum mechanical calculations were performed for all complexes using the Multiwfn 3.7 program as described in section 3.2.6.3. We used scripts, an example of which is shown in Appendix 3 in these calculations. After the program was run with scripts, the duration of the calculations varies according to the difficulty of its complex structures. As the size of the complexes and the number of atoms increase, the computation time increases.

```
CHEMICAL SHIELDING SUMMARY (ppm)-Pt Atom
----------------------------------------
Isotropic     :  3078.004
Anisotropy    :  1594.540
Chemical Shift: -4042.586 (Predicted)

Diamagnetic contribution to the shielding tensor (ppm)-Pt Atom
--------------------------------------------------------------
 9659.044      18.249     -33.044
  -25.870    9610.931       6.707
  -24.017     -20.474    9614.870

Paramagnetic contribution to the shielding tensor (ppm)-Pt Atom
--------------------------------------------------------------
-7159.480     427.683   -1057.622
  288.361   -5564.681      87.445
 -903.192     109.426   -6926.672
Total shielding tensor (ppm)-Pt Atom
-----------------------------------
 2499.563     445.933   -1090.666
  262.491    4046.250      94.152
 -927.209      88.952    2688.198

Diagonalized sT*s matrix-Pt Atom
--------------------------------
sDSO      :    9613.699   9654.112   9617.033
sDSO iso  :    9628.282
sPSO      :   -8074.140  -6100.690  -5476.003
sPSO iso  :   -6550.278
Total     :    1539.559   3553.422   4141.030
Total iso :    3078.004
```

Figure 4.9. An example of values from geometric optimization calculations in ORCA 4.2. The values refer to the A1-01 complex.

As a result of multiple quantum mechanical calculations, output files consisting of hundreds of pages are created. To get the descriptors to be used from the output files in "txt" format, unnecessary lines need to be deleted. A macro created with visual basic is used in the tables for this aim. We added the statements that define the lines to be deleted to this macro and when we run the macro, these lines are deleted. It is ensured that only the lines containing the data we want to be retrieved remain. After the data is separated, only the necessary values are obtained with the "MID()" formula in the excel. This process has been applied for all output files from Multiwfn 3.7.

After quantum mechanical calculations with QTAIM theory, approximately 800 descriptors were calculated using the final geometries taken from geometric optimizations for each complex with the PaDEL-Descriptor program. Columns having blank cells were manually deleted and stored in "csv" format. To obtain the descriptors, the stable final structures of the complexes were saved in the "mol2"

format. This format was loaded into the program and 2D-3D Descriptors were calculated.

Finally, one of the most common problems in the literature was the standardization of the descriptors used for structures. We carried out a different study and calculated various quantum mechanical parameters that better explain the physical and chemical structures of the complexes with the Multiwfn 3.7 program and stored them as descriptors to do this more effectively. In addition, we transferred NMR data from ORCA 4.2, which were successful in some QSAR studies, as descriptors.

Human ovarian tumors are one of the most dangerous and deadly types of cancer. Many studies in the literature with the A2780 gene belonging to the ovarian cancer type are interesting. A rational drug that has a full effect on this species has not yet been developed. Based on this, many platinum-derived compounds tested in-vitro for the studies with the A2780 gene were taken from the literature to be used in QSAR studies. Platinum-derived compounds are not widely included in QSAR studies as they are metal-based and have less database information. In studies involving these compounds, theoretical cytotoxic activity values were tried to be estimated over a specific set of descriptors. In this study, we brought a different perspective to this issue. The compounds from the literature were categorized to minimize the error caused by the experimental method applied by different study groups.

After obtaining all descriptors, QSAR analysis was performed for all groups as described in section 3.2.7. QSAR analyzes were performed with the QSARINS 2.2.4 program. The descriptors that best describe the cytotoxic activities of the complexes have been theoretically determined. In the QSAR studies, the workflow in figure 4.10 was taken as the basis step by step. In the first step of the calculations, after the $\log(1/IC_{50})$ values of descriptors and cytotoxic activity (IC_{50}) values were transferred to the QSARINS program as "csv", correlated and fixed identifiers were eliminated. As a result of running the specified filtering algorithm, as seen in figure 4.11, the number of descriptors that each group would participate in the QSAR analysis was determined.

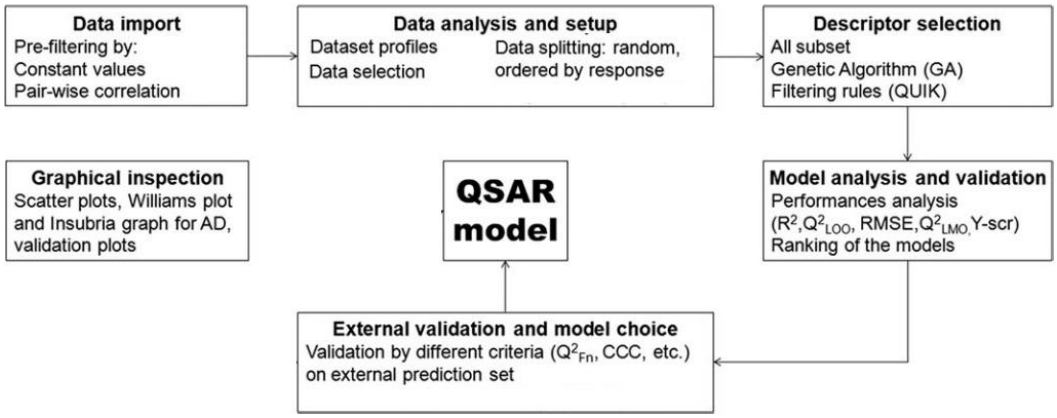


Figure 4.10. QSARINS Workflow taken from reference 32**.**

Many models were created for each group with random selection. These models were examined one by one. When performing the analyses, those with a maximum of six descriptors were allowed to build the model. Because increasing the number of descriptors can create overfitting. This creates an imaginary model. Therefore, as few descriptors as possible were selected for a good model. More than 20 random selection analyzes were performed for each group. Models with several different descriptor numbers from 1 to 6, were generated in each analysis. For the models in the analysis, firstly, by opening the view and select models' section of the program, internal validation was completed by using the 5000 iteration LMO and Y-scramble together with 30% prediction. Later, external validation was performed for all models.

The best models were selected for each group. The selected models and their statistical parameters are listed in table 4.7. This is important to check these model performances, that is, fitting, stability by internal and success to predict new chemicals with external validation (133). These steps must be done exactly right so that the model is not left to chance. For example, just a high value of validation R^2 or Q^2loo is not enough to validate the model. Because it may be an overfitting model. The model should be checked with the internal validation parameter Q^2 LMO and the external validation parameter R^2ext. The models should be examined according to Organization for Economic Co-operation and Development (OECD) principles in the final decision-making. These performances are measured by appropriate formulas/methods of various validation criteria (134-135).

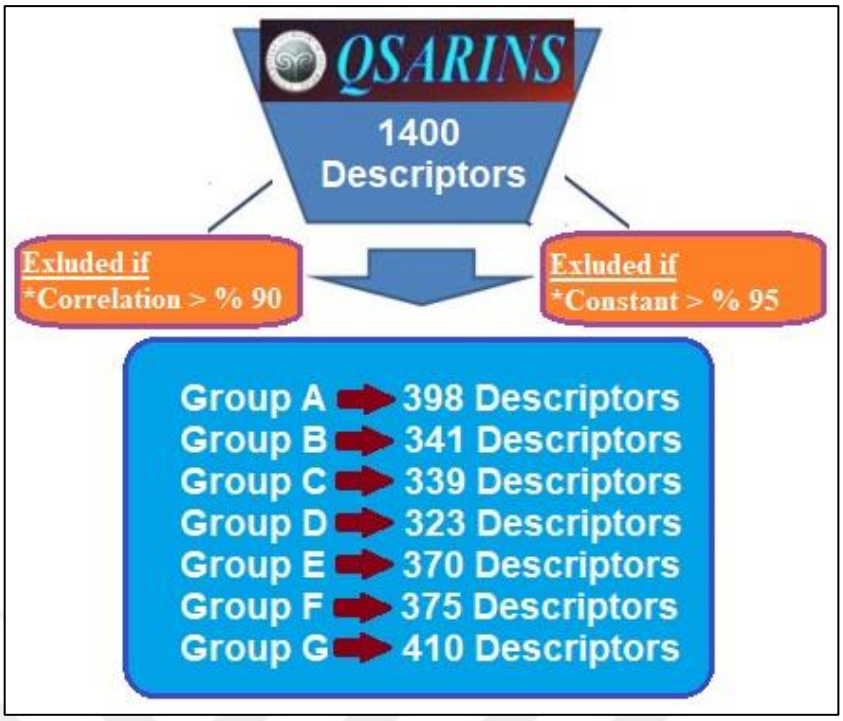


Figure 4.11. Diagram of descriptors filtering process.

OECD principles are a guide for the validation and regulatory purposes of QSAR models. These principles were agreed by OECD member countries at the 37thJoint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology in November 2004. The principles are intended to be read in conjunction with the associated explanatory notes which were also agreed at the 37thJoint meeting. To facilitate the consideration of a QSAR model for regulatory purposes, it should be associated with the following information: 1) a defined endpoint, 2) an unambiguous algorithm, 3) a defined domain of applicability, 4) appropriate measures of goodness of fit, robustness and predictivity, and 5) a mechanistic interpretation, if possible. It is necessary to validate the models according to OECD principles.

Principle 1 relates to the definition of an endpoint, where it refers to a pharmacological, biological, or physicochemical property that can be measured and thus modeled. The purpose of this principle is to provide transparency at the measurement point predicted by a particular model. To provide this, extensive literature studies were carried out. Complexes formed by similar groups and with similar experimental methods were included in the same groups.

According to Principle 2, QSAR models can be expressed as open algorithms and developed by one or more experts, given that the algorithm model

is the way to relate descriptors of chemical structure and activity (the endpoint of the model) through mathematical models or knowledge-based rules. The algorithm used was a mathematical model of Multiple Linear Regression detailed in the previous sections.

According to Principle 3, William graph was used to define the model application domain. Thus, successful predictions of the model have leverage data below the critical leverage with ± 3 standard deviations. The approach of leverage (h) and standardized residuals was used in the technical literature (138). Figure 4.12 shows the William graphs of the all model. The complexes that exceed the critical leverage (h) are excluded from the model. The "h*" values seen above the graphics in the figure are the limit values. Training compounds that fall outside the critical leverage are considered potential outlier compounds. According to the graphs, it is seen that only the B5-01 complex is left out in group B. There is cytotoxic data for compounds from many different studies. If one complex is left out of the model, when considered together with other parameters, the model does not fail. Although similar ones of all experimental methods are grouped at the selection stage, there is a margin of error. The aim is to minimize this with the chosen QSAR model.

In addition, Insubria graph was used for application domain. This was very useful to assess the reliability of the predictions of experimental compounds, to compare with the predictions of the database, and lacking response. Application domain graphics are added for all groups in Figure 4.13.

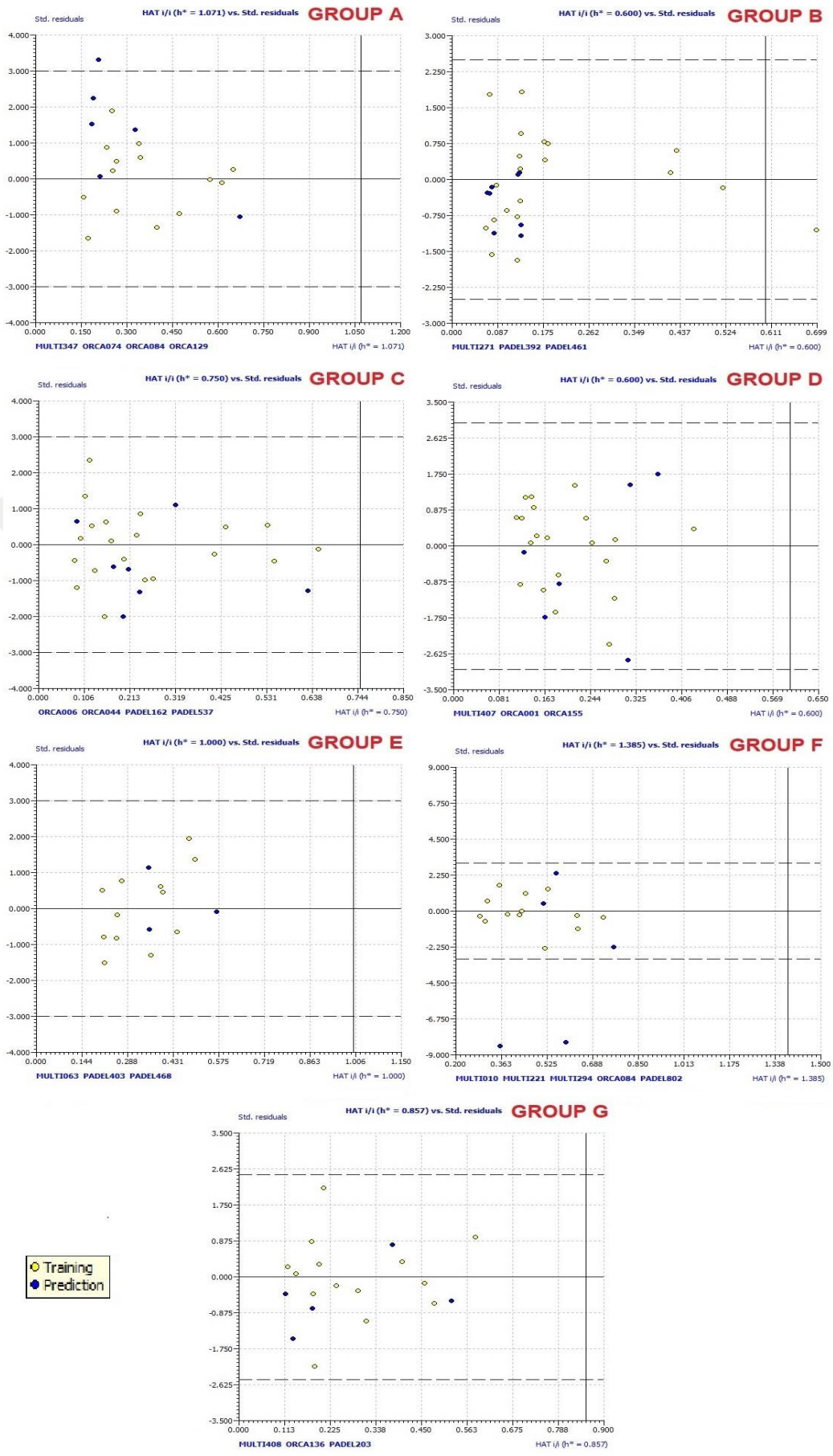


Figure 4.12. Williams graphs for all models.

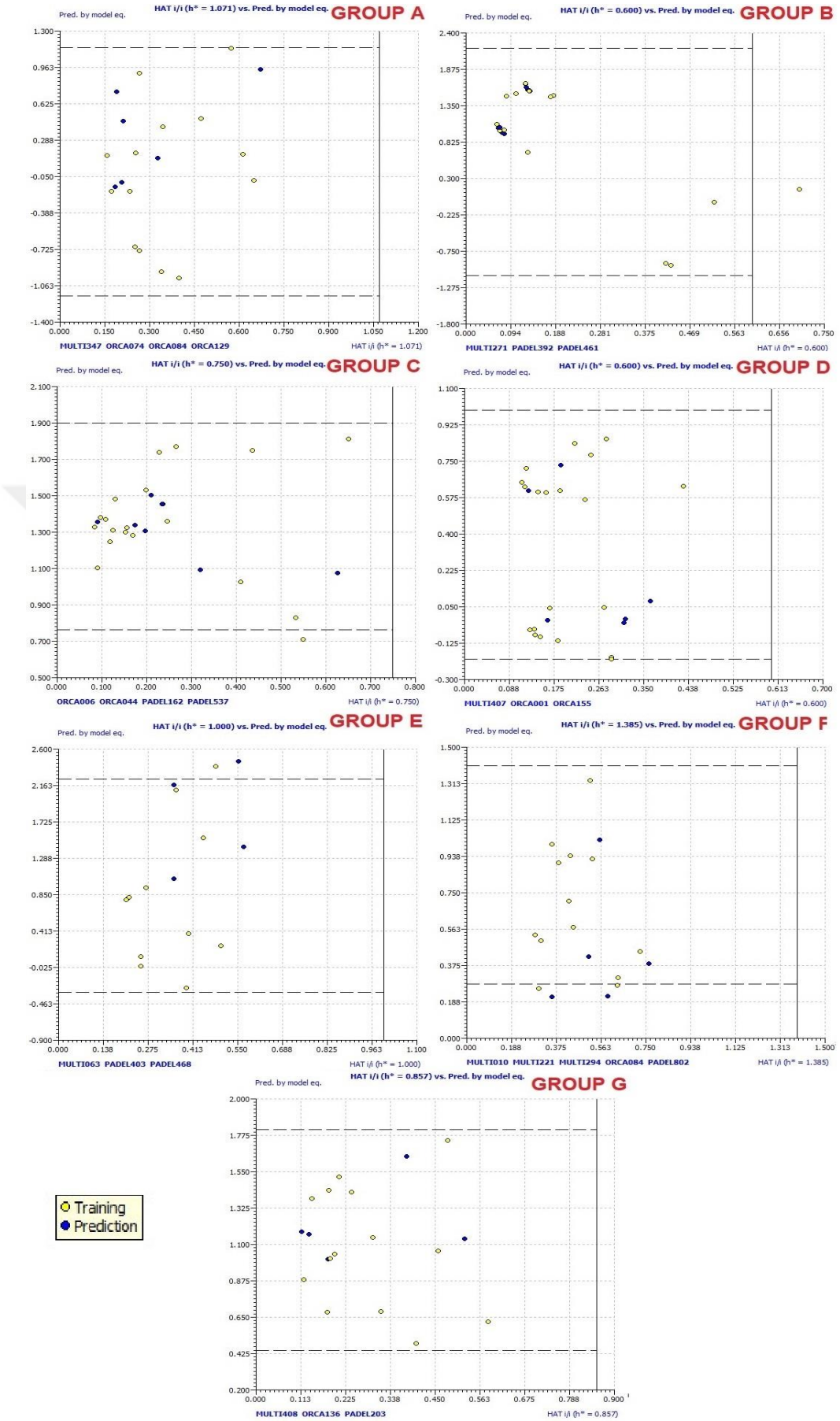


Figure 4.13. Application domain for all models.

Following the OECD Principle 4, the goodness of the fit was evaluated using the coefficient of determination of fitting, internal validation, and external validation in table 4.7. In addition, while evaluating the quality of the fit with the graphics in figure 4.14, statistical evaluations were made in detail for all models.

Table 4.7. Models selected for all groups and statistical parameters considered in model selection.

Statistical State	Parameter	A	B	C	D	E	F	G
Fitting Criteria	R^2	0.8895	0.8789	0.8015	0.9751	0.9880	0.9883	0.8986
	R^2adj	0.8404	0.8561	0.7486	0.9704	0.9835	0.9799	0.8681
	R^2-R^2adj	0.0491	0.0227	0.0529	0.0047	0.0045	0.0084	0.0304
	LOF	0.2687	0.1771	0.0558	0.0079	0.0307	0.0222	0.0468
	Kxx	0.2958	0.1676	0.3910	0.5745	0.2086	0.3112	0.2780
	ΔK	0.0913	0.1466	0.0228	0.1227	0.2581	0.0607	0.1974
	RMSE tr	0.2222	0.2946	0.1418	0.0620	0.0876	0.0344	0.1236
	MAE tr	0.1783	0.2426	0.1084	0.0510	0.0780	0.0263	0.0882
	RSS tr	0.6910	1.7353	0.4020	0.0769	0.0920	0.0154	0.2140
	CCC tr	0.9415	0.9355	0.8898	0.9874	0.9940	0.9941	0.9466
	s	0.2771	0.3293	0.1637	0.0693	0.1072	0.0469	0.1463
	F	18.1103	38.6901	15.1434	208.6210	219.7842	118.2120	29.5246
Internal Validation Criteria	Q^2loo	0.7807	0.8190	0.7118	0.9598	0.9680	0.9552	0.8151
	R^2-Q^2loo	0.1088	0.0598	0.0898	0.0152	0.0200	0.0331	0.0835
	RMSE cv	0.3129	0.3600	0.1709	0.0787	0.1430	0.0673	0.1669
	MAE cv	0.2588	0.3028	0.1378	0.0638	0.1229	0.0512	0.1248
	PRESS cv	1.3710	2.5924	0.5839	0.1240	0.2453	0.0589	0.3900
	CCC cv	0.8854	0.9058	0.8414	0.9797	0.9840	0.9776	0.8997
	Q^2LMO	0.6914	0.7098	0.6298	0.9550	0.9640	0.9444	0.7803
	R^2Yscr	0.3058	0.1571	0.2120	0.1546	0.2692	0.4230	0.2336
	RMSE AV Yscr	0.5519	0.7748	0.2815	0.3604	0.6778	0.2379	0.3379
	Q^2Yscr	-0.8110	-0.4336	-0.5107	-0.3360	-0.6870	-1.1394	-0.5665
External Validation Criteria	RMSE ext	0.4582	0.2108	0.1636	0.0979	0.1732	0.1833	0.1150
	MAE ext	0.3769	0.1629	0.1498	0.0864	0.1207	0.1404	0.1017
	PRESS ext	1.2596	0.3556	0.1873	0.0575	0.1200	0.1681	0.0661
	R^2ext	0.8460	0.7744	0.6790	0.9179	0.9350	0.7878	0.8555
	Q^2-F1	0.5224	0.8662	0.5449	0.9174	0.9705	0.3657	0.8082
	Q^2-F2	0.5204	0.5817	0.5396	0.9120	0.8908	0.1622	0.4527
	Q^2-F3	0.5299	0.9379	0.7359	0.9379	0.9531	0.6677	0.9122
	CCC ext	0.7046	0.7938	0.6932	0.9543	0.9503	0.7586	0.8256
	r^2m aver.	0.2683	0.6837	0.4018	0.8839	0.9053	0.3534	0.5190
	r^2m delta	0.1995	0.0239	0.3446	0.0469	0.0411	0.3656	0.2535
	k'	0.5852	0.8960	0.9318	0.9113	1.0542	0.8706	0.9544
	k	0.9007	1.1028	1.0623	1.0382	0.9429	1.0402	1.0413
	Clos'	0.7580	0.0133	0.6453	0.0001	0.0001	0.1295	0.0703
	Clos	0.3773	0.0227	0.0352	0.0043	0.0031	0.7791	0.3427

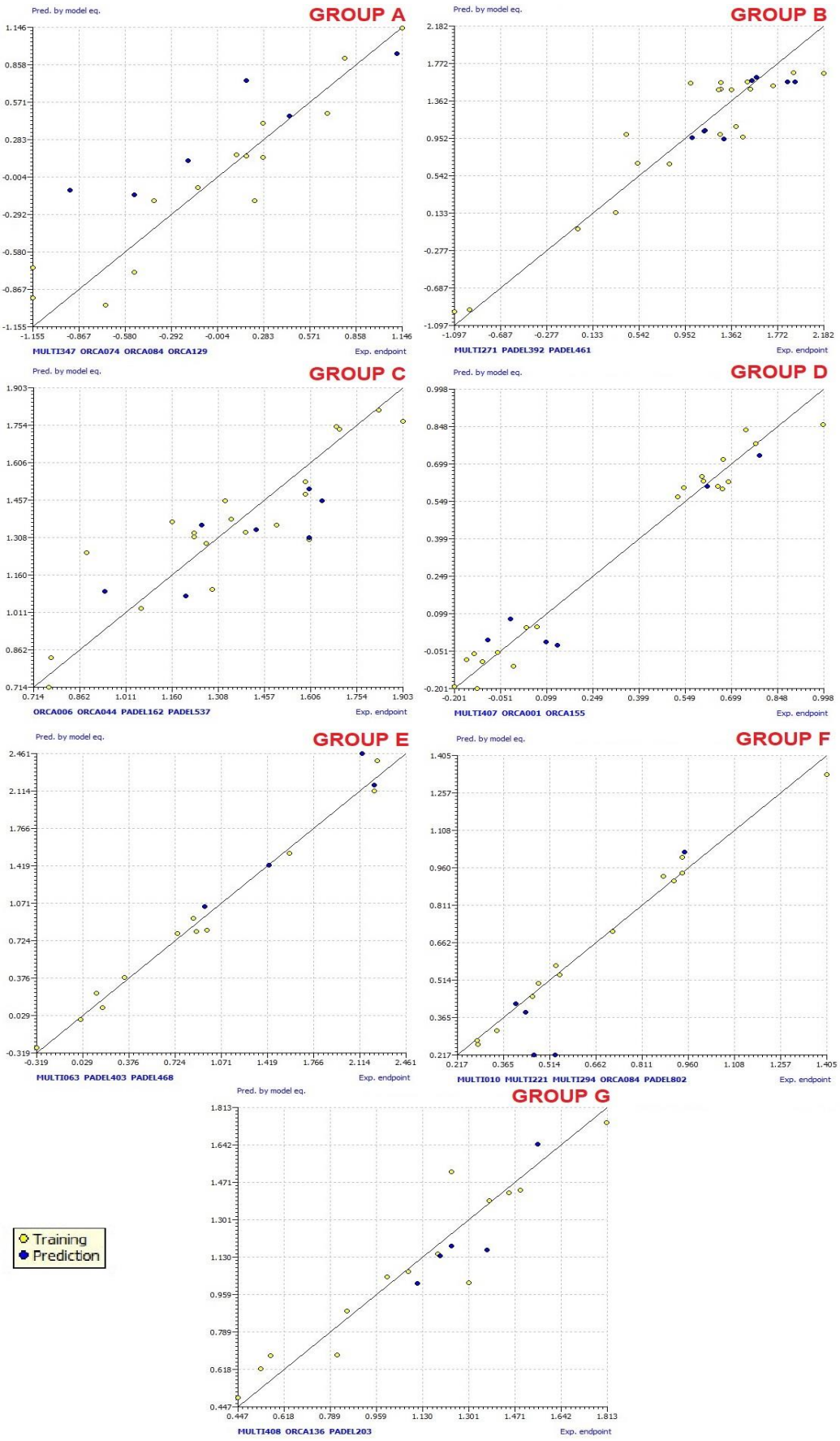


Figure 4.14. R^2 plots for the experimental endpoints of the complexes according to the predicted values in all groups.

After building a QSAR model, it is an important part to validate its performance. The following rules were used to select and validate a model (136-137): $R^2 \geq 0.6$, $Q^2loo \geq 0.5$, $Q^2LMO \geq 0.6$, $R^2 > Q^2loo$, $R^2ex \geq 0.6$, RMSEtr < RMSEcv, $\Delta K \geq 0.05$, CCC tr ≥ 0.80, Q^2-Fn ≥ 0.60, $r^2m \geq 0.5$, $0.9 \leq k \leq 1.1$, $0.9 \leq k' \leq 1.1$, with RMSE and MAE as low as possible. All QSAR models that did not reach the recommended lower limit values for the above statistical parameters were rejected outright.

The regression graphs of the experimental endpoints of the complexes according to the values predicted by the models (R^2) are found in figure 4.14. The comparative table of the values in these graphs for the complexes in all groups is available in table 4.8. In addition, the cytotoxic (IC_{50}) values of all the complexes found a result of academic studies were added into table 4.8. The complexes were grouped according to the articles in which they were found and some values such as LOG ($-1/IC_{50}$), LOG (RA) and pIC_{50} were calculated using cytotoxic IC_{50} values. These are parameters that will be useful when used in statistical analysis.

The best models in all groups have R^2 values greater than 0.8. This satisfies the first condition. In addition, the $R^2 > Q^2loo$ requirement is provided for all models. RMSE tr and MAE tr error values remained below 0.5 as desired. CCC tr is also higher than 0.8. The condition $\Delta K > 0.5$ was reached for all groups. It has been determined that successful fitting values have been achieved in all models.

When the internal validation values in table 4.7 are examined, Q^2loo being greater than 0.5 was provided for all groups. Only Group C has a relatively low value of 0.7118. The R^2-Q^2loo difference is also low enough. Only group A has a slightly high value that does not adversely affect the model. Values greater than 0.6 are acceptable for the Q^2 LMO. Only the value of Group C is close to the lower limit. But it is statistically successful. CCC cv value is greater than 0.8 for all groups. Yscr values should be far from the Q^2 and R^2 values to confirm that there is no correlation in the models. This was accurately ensured in all groups. RMSE cv and MAE cv error data is less than 0.5 for all groups and is at the desired level. Internal validation parameters also validate the models as well as the fitting values.

External validation must be done to calculate the activity of compounds of unknown activity as accurately as possible. The activities of the complexes separated as predictions are validated. The most important parameter here is R^2ext. While searching the models, firstly the models were sorted according to this R^2ext

value and values less than 0.6 were ignored. When we look at all groups, very successful results have been detected. Only Group C has a value of 0.6790 close to the limit. When Q^2-Fn values are examined, only Group F has a small number of values less than 0.6. However, it did not adversely affect the quality of this model considering other parameters. r^2m aver. value also usually has values greater than 0.5 and alone does not affect the model. RMSEext and MAEext error data are less than 0.5 for all groups and are at the desired level. $0.9 \leq k$ or $k' \leq 1.1$ was provided adequately in all groups. We can figure out from the comparative statistical analysis results that external validation can also successfully calculate compounds with unknown activity.

Table 4.8. IC_{50} cytotoxic values and other useful parameters of all complexes. Estimated cytotoxic values resulting from QSAR models and difference from experimental value.

ID	Complexs Name in Articles	A2780 Cytotoxic Values	LOG RA	pIC_{50}	QSAR Splitting Status	Log $(1/IC_{50})$	Estimated by Model Equation	Prediction Model Equation Difference
A1-01	cis-DDP	0.2 ± 0.07	1.8451	6.6990	Training	-0.6990	-0.9857	-0.2867
A1-02	trans-DDP	14 ± 4	0.0000	4.8539	Training	1.1461	1.1460	-0.0001
A1-03	1-trans-Z	1.3 ± 0.3	1.0322	5.8861	Training	0.1139	0.1701	0.0562
A1-04	2-trans-E	2.8 ± 0.7	0.6990	5.5528	Prediction	0.4472	0.4668	0.0196
A1-05	3-trans-Pyr	6.2 ± 2	0.3537	5.2076	Training	0.7924	0.9100	0.1177
A1-06	4-trans-Ox	1.5 ± 0.5	0.9700	5.8239	Prediction	0.1761	0.7376	0.5615
A1-07	5-cis-Z	0.75 ± 0.02	1.2711	6.1249	Training	-0.1249	-0.0805	0.0444
A1-08	6-cis-E	0.4 ± 0.07	1.5441	6.3979	Training	-0.3979	-0.1821	0.2159
A1-09	7-cis-Pyr	1.9 ± 0.3	0.8674	5.7212	Training	0.2788	0.4142	0.1355
A1-10	8-cis-Ox	0.65 ± 0.1	1.3332	6.1871	Prediction	-0.1871	0.1244	0.3115
A2-01	1	1.5 ± 0.3	0.9700	5.8239	Training	0.1761	0.1591	-0.0170
A2-02	2	4.8 ± 0.2	0.4649	5.3188	Training	0.6812	0.4898	-0.1914
A2-03	3	0.3 ± 0.07	1.6690	6.5229	Prediction	-0.5229	-0.1389	0.3840
A2-04	4	13 ± 1.5	0.0322	4.8861	Prediction	1.1139	0.9490	-0.1649
A3-01	7	1.9 ± 0.9	0.8674	5.7212	Training	0.2788	0.1497	-0.1291
A3-02	8	0.3 ± 0.1	1.6690	6.5229	Training	-0.5229	-0.7341	-0.2112
A3-03	9	0.07 ± 0.002	2.3010	7.1549	Training	-1.1549	-0.7002	0.4547
A3-04	10	1.7 ± 0.01	0.9157	5.7696	Training	0.2304	-0.1818	-0.4122
A3-05	11	0.12 ± 0.05	2.0669	6.9208	Prediction	-0.9208	-0.1009	0.8199
A3-06	12	0.07 ± 0.005	2.3010	7.1549	Training	-1.1549	-0.9314	0.2235
B1-01	Pt1	13.2 ± 4.3	1.0613	4.8794	Prediction	1.1206	1.0353	-0.0853
B1-02	Pt2	19.8 ± 1.6	0.8852	4.7033	Prediction	1.2967	0.9452	-0.3515
B1-03	Pt3	18.7 ± 3.1	0.9100	4.7282	Training	1.2718	1.4980	0.2261
B1-04	Pt4	33.9 ± 4.2	0.6516	4.4698	Training	1.5302	1.4969	-0.0333
B1-05	Pt5	25.6 ± 2.4	0.7736	4.5918	Training	1.4082	1.0877	-0.3205
B1-06	Pt6	23.1 ± 6.7	0.8182	4.6364	Training	1.3636	1.4879	0.1243
B1-07	Pt7	17.7 ± 1.5	0.9339	4.7520	Training	1.2480	1.4873	0.2393
B1-08	Pt8	29.1 ± 8.6	0.7180	4.5361	Training	1.4639	0.9704	-0.4935
B1-09	Pt9	10.3 ± 5.5	1.1690	4.9872	Prediction	1.0128	0.9642	-0.0487
B1-10	Pt10	2.7 ± 0.6	1.7505	5.5686	Training	0.4314	0.9968	0.5655
B1-11	Pt11	13.6 ± 3.5	1.0483	4.8665	Prediction	1.1335	1.0427	-0.0908
B1-12	Pt12	18.4 ± 4.1	0.9170	4.7352	Training	1.2648	1.0023	-0.2625

Table 4.8. (continued)

B2-01	**6**	10 ± 0.9	1.1818	5.0000	Training	1.0000	1.5619	0.5619
B2-02	**8**	32 ± 4.6	0.6767	4.4949	Training	1.5051	1.5765	0.0713
B2-03	**10**	152 ± 15	0.0000	3.8182	Training	2.1818	1.6655	-0.5163
B2-04	**12**	35 ± 8	0.6378	4.4559	Prediction	1.5441	1.5888	0.0448
B3-01	**1**	1.0 ± 0.3	2.1818	6.0000	Training	0.0000	-0.0378	-0.0378
B3-02	**2**	3.4 ± 0.7	1.6504	5.4685	Training	0.5315	0.6831	0.1516
B3-03	**3**	6.5 ± 0.1	1.3689	5.1871	Training	0.8129	0.6785	-0.1344
B4-01	**8**	38.64 ± 5.33	0.5948	4.4130	Prediction	1.5870	1.6232	0.0361
B4-02	**10**	85.56 ± 0.9	0.2496	4.0677	Prediction	1.9323	1.5750	-0.3572
B4-03	**12**	81.94 ± 1.29	0.2683	4.0865	Training	1.9135	1.6762	-0.2373
B4-04	**13**	54.33 ± 4.8	0.4468	4.2650	Training	1.7350	1.5342	-0.2009
B4-05	**14**	72.45 ± 5.8	0.3218	4.1400	Prediction	1.8600	1.5715	-0.2885
B4-06	**15**	18.7 ± 4.22	0.9100	4.7282	Training	1.2718	1.5684	0.2965
B5-01	**3**	2.16 ± 1.4	1.8474	5.6655	Training	0.3345	0.1446	-0.1898
B5-02	**5**	0.11 ± 0.08	3.1405	6.9586	Training	-0.9586	-0.9223	0.0363
B5-03	**6**	0.08 ± 0.1	3.2788	7.0969	Training	-1.0969	-0.9434	0.1535
C1-01	**2A0**	17	0.6726	4.7696	Training	1.2304	1.3125	0.0820
C1-02	**2A1**	5.8	1.1397	5.2366	Training	0.7634	0.7137	-0.0497
C1-03	**2A2**	5.9	1.1322	5.2291	Training	0.7709	0.8333	0.0624
C1-04	**2A3**	16	0.6990	4.7959	Prediction	1.2041	1.0778	-0.1263
C1-05	**2A4**	21.5	0.5707	4.6676	Training	1.3324	1.4556	0.1231
C1-06	**2A5**	19.5	0.6131	4.7100	Training	1.2900	1.1038	-0.1862
C1-07	**2A6**	18.6	0.6336	4.7305	Training	1.2695	1.2855	0.0160
C1-08	**2A7**	17	0.6726	4.7696	Training	1.2304	1.3270	0.0965
C1-09	**2A8**	50	0.2041	4.3010	Training	1.6990	1.7390	0.0400
C1-10	**2A9**	40	0.3010	4.3979	Training	1.6021	1.3025	-0.2996
C1-11	**2A10**	40	0.3010	4.3979	Prediction	1.6021	1.5039	-0.0982
C1-12	**2A11**	8.8	0.9586	5.0555	Prediction	0.9445	1.0959	0.1515
C1-13	**2A12**	11.5	0.8424	4.9393	Training	1.0607	1.0281	-0.0326
C1-14	**2A13**	49	0.2129	4.3098	Training	1.6902	1.7516	0.0614
C1-15	**4E**	31.5	0.4048	4.5017	Training	1.4983	1.3599	-0.1384
C1-16	**4I**	14.5	0.7417	4.8386	Training	1.1614	1.3712	0.2098
C1-17	**4M**	44	0.2596	4.3565	Prediction	1.6435	1.4570	-0.1865
C1-18	**4N**	22.5	0.5509	4.6478	Training	1.3522	1.3820	0.0299
C1-19	**4O**	40	0.3010	4.3979	Prediction	1.6021	1.3105	-0.2916
C1-20	**4P**	27	0.4717	4.5686	Prediction	1.4314	1.3400	-0.0913
C1-21	**4Q**	39	0.3120	4.4089	Training	1.5911	1.5327	-0.0584
C1-22	**4R**	18	0.6478	4.7447	Prediction	1.2553	1.3582	0.1030
C1-23	**4S**	7.7	1.0166	5.1135	Training	0.8865	1.2497	0.3632
C1-24	**4T**	39	0.3120	4.4089	Training	1.5911	1.4829	-0.1082
C1-25	**4X**	25	0.5051	4.6021	Training	1.3979	1.3299	-0.0680
C1-26	**D**	67	0.0770	4.1739	Training	1.8261	1.8152	-0.0109
C1-27	**Y2**	80	0.0000	4.0969	Training	1.9031	1.7707	-0.1324
D1-01	**1a**	0.98 ± 0.07	1.0070	6.0088	Training	-0.0088	-0.1097	-0.1009
D1-02	**2a**	1.36 ± 0.05	0.8647	5.8665	Prediction	0.1335	-0.0259	-0.1595
D1-03	**3a**	1.08 ± 0.01	0.9648	5.9666	Training	0.0334	0.0464	0.0129
D1-04	**4a**	0.63 ± 0.18	1.1989	6.2007	Training	-0.2007	-0.1912	0.0095
D1-05	**5a**	0.69 ± 0.01	1.1594	6.1612	Training	-0.1612	-0.0839	0.0773
D1-06	**6a**	0.81 ± 0.08	1.0898	6.0915	Prediction	-0.0915	-0.0056	0.0859
D1-07	**7a**	0.96 ± 0.04	1.0160	6.0177	Prediction	-0.0177	0.0791	0.0968
D1-08	**8a**	0.75 ± 0.19	1.1232	6.1249	Training	-0.1249	-0.1991	-0.0741
D1-09	**9a**	0.73 ± 0.10	1.1349	6.1367	Training	-0.1367	-0.0592	0.0775
D1-10	**10a**	0.87 ± 0.14	1.0587	6.0605	Training	-0.0605	-0.0548	0.0056
D1-11	**11a**	1.17 ± 0.20	0.9301	5.9318	Training	0.0682	0.0471	-0.0211
D1-12	**12a**	0.78 ± 0.05	1.1062	6.1079	Training	-0.1079	-0.0915	0.0165

Table 4.8. (continued)

D1-13	**13a**	1.25 ± 0.34	0.9013	5.9031	Prediction	0.0969	-0.0125	-0.1095
D1-14	**1b**	3.36 ± 0.07	0.4719	5.4737	Training	0.5263	0.5679	0.0416
D1-15	**2b**	3.51 ± 0.08	0.4530	5.4547	Training	0.5453	0.6054	0.0601
D1-16	**3b**	4.01 ± 0.04	0.3951	5.3969	Training	0.6031	0.6489	0.0457
D1-17	**4b**	4.67 ± 0.07	0.3289	5.3307	Training	0.6693	0.6012	-0.0681
D1-18	**5b**	4.89 ± 0.10	0.3090	5.3107	Training	0.6893	0.6288	-0.0605
D1-19	**6b**	4.72 ± 0.06	0.3243	5.3261	Training	0.6739	0.7188	0.0449
D1-20	**7b**	5.99 ± 0.43	0.2208	5.2226	Training	0.7774	0.7824	0.0050
D1-21	**8b**	4.52 ± 0.14	0.3431	5.3449	Training	0.6551	0.6111	-0.0441
D1-22	**9b**	4.17 ± 0.08	0.3781	5.3799	Prediction	0.6201	0.6102	-0.0099
D1-23	**10b**	4.06 ± 0.09	0.3897	5.3915	Training	0.6085	0.6304	0.0219
D1-24	**11b**	9.96 ± 0.25	0.0000	5.0017	Training	0.9983	0.8575	-0.1408
D1-25	**12b**	6.16 ± 0.12	0.2087	5.2104	Prediction	0.7896	0.7327	-0.0569
D1-26	**13b**	5.58 ± 0.08	0.2516	5.2534	Training	0.7466	0.8377	0.0910
E1-01	**1**	1.37 ± 0.48	2.1089	5.8633	Training	0.1367	0.2417	0.1050
E1-02	**2**	7.32 ± 0.39	1.3811	5.1355	Training	0.8645	0.9354	0.0709
E1-03	**3**	7.66 ± 0.72	1.3614	5.1158	Training	0.8842	0.8103	-0.0739
E1-04	**4**	5.55 ± 1.17	1.5013	5.2557	Training	0.7443	0.7943	0.0500
E1-05	**5**	8.84 ± 2.16	1.2991	5.0535	Prediction	0.9465	1.0457	0.0993
E1-06	**6**	9.21 ± 1.03	1.2813	5.0357	Training	0.9643	0.8216	-0.1426
E1-07	**7**	27.24 ± 1.74	0.8104	4.5648	Prediction	1.4352	1.4296	-0.0056
E1-08	**8**	176.03 ± 8.23	0.0000	3.7544	Training	2.2456	2.3969	0.1513
E1-09	**9**	135.90 ± 8.07	0.1124	3.8668	Prediction	2.1332	2.4614	0.3281
E1-10	**10**	38.86 ± 8.40	0.6561	4.4105	Training	1.5895	1.5386	-0.0509
E1-11	**11**	167.36 ± 27.13	0.0219	3.7763	Prediction	2.2237	2.1740	-0.0497
E1-12	**12**	167.52 ± 31.57	0.0215	3.7759	Training	2.2241	2.1143	-0.1098
E1-13	**13**	1.03 ± 0.22	2.2327	5.9872	Training	0.0128	-0.0029	-0.0157
E1-14	**14**	0.48 ± 0.16	2.5643	6.3188	Training	-0.3188	-0.2672	0.0516
E1-15	**15**	2.22 ± 0.42	1.8992	5.6536	Training	0.3464	0.3857	0.0393
E1-16	**16**	1.52 ± 0.11	2.0637	5.8182	Training	0.1818	0.1066	-0.0752
F1-01	**7a**	8.21 ± 0.31	0.4908	5.0857	Training	0.9143	0.9077	-0.0067
F1-02	**7b**	3.52 ± 0.41	0.8586	5.4535	Training	0.5465	0.5355	-0.0110
F1-03	**7c**	25.42 ± 2.05	0.0000	4.5948	Training	1.4052	1.3307	-0.0745
F1-04	**7d**	8.90 ± 0.18	0.4558	5.0506	Prediction	0.9494	1.0235	0.0741
F1-05	**7e**	7.61 ± 0.23	0.5238	5.1186	Training	0.8814	0.9268	0.0455
F1-06	**7f**	5.20 ± 0.38	0.6892	5.2840	Training	0.7160	0.7079	-0.0081
F1-07	**7g**	1.92 ± 0.05	1.1219	5.7167	Training	0.2833	0.2590	-0.0243
F1-08	**7h**	3.41 ± 0.75	0.8724	5.4672	Prediction	0.5328	0.2168	-0.3160
F1-09	**7i**	2.74 ± 0.33	0.9674	5.5622	Prediction	0.4378	0.3870	-0.0508
F1-10	**7j**	2.21 ± 0.09	1.0608	5.6556	Training	0.3444	0.3136	-0.0308
F1-11	**7k**	2.88 ± 0.12	0.9458	5.5406	Training	0.4594	0.4505	-0.0089
F1-12	**7l**	1.91 ± 0.43	1.1241	5.7190	Training	0.2810	0.2742	-0.0068
F1-13	**7m**	3.01 ± 0.38	0.9266	5.5214	Training	0.4786	0.5034	0.0249
F1-14	**7n**	8.72 ± 0.33	0.4647	5.0595	Training	0.9405	1.0014	0.0609
F1-15	**7o**	3.42 ± 0.25	0.8711	5.4660	Training	0.5340	0.5735	0.0395
F1-16	**7p**	2.91 ± 0.48	0.9413	5.5361	Prediction	0.4639	0.2192	-0.2447
F1-17	**7q**	8.72 ± 0.91	0.4647	5.0595	Training	0.9405	0.9409	0.0004
F1-18	**7r**	2.54 ± 0.31	1.0003	5.5952	Prediction	0.4048	0.4211	0.0162
G1-01	**1**	17.3 ± 1.3	0.5749	4.7620	Prediction	1.2380	1.1813	-0.0568
G1-02	**2**	17.3 ± 0.3	0.5749	4.7620	Training	1.2380	1.5198	0.2817
G1-03	**3**	31.1 ± 1.4	0.3202	4.5072	Training	1.4928	1.4385	-0.0542
G1-04	**4**	12.9 ± 0.5	0.7023	4.8894	Prediction	1.1106	1.0110	-0.0996
G1-05	**5**	23.4 ± 2.1	0.4437	4.6308	Prediction	1.3692	1.1658	-0.2034
G1-06	**6**	23.8 ± 2.2	0.4363	4.6234	Training	1.3766	1.3879	0.0113
G1-07	**7**	3.4 ± 0.6	1.2814	5.4685	Training	0.5315	0.6227	0.0912

Table 4.8. (continued)

G1-08	**9**	10.0 ± 0.1	0.8129	5.0000	Training	1.0000	1.0410	0.0410
G1-09	**10**	28.2 ± 1.5	0.3627	4.5498	Training	1.4502	1.4247	-0.0255
G1-10	**11**	15.4 ± 0.3	0.6254	4.8125	Training	1.1875	1.1466	-0.0409
G1-11	**12**	15.7 ± 0.4	0.6170	4.8041	Prediction	1.1959	1.1379	-0.0580
G2-01	**7a**	2.8 ± 0.3	1.3658	5.5528	Training	0.4472	0.4898	0.0427
G2-02	**7b**	3.7 ± 0.4	1.2447	5.4318	Training	0.5682	0.6837	0.1155
G3-01	**[Pt(HL$_2$)(PPh$_3$)]**	12 ± 1	0.7337	4.9208	Training	1.0792	1.0643	-0.0149
G4-01	**[PtL$_2$]**	20 ± 2	0.5119	4.6990	Training	1.3010	1.0156	-0.2854
G5-01	**5**	65 ± 3	0.0000	4.1871	Training	1.8129	1.7454	-0.0675
G5-02	**6**	36 ± 2	0.2566	4.4437	Prediction	1.5563	1.6472	0.0909
G6-01	**[PtL]**	7.12 ± 0.21	0.9604	5.1475	Training	0.8525	0.8865	0.0340
G7-01	**[Pt(HL)(PPh$_3$)]**	6.54 ± 1.16	0.9973	5.1844	Training	0.8156	0.6866	-0.1290

For the models detailed in Table 3.7 and Table 4.8 and verified according to OECD rules, the following prediction formulas between 9 and 15 have emerged. In these formulas, there are descriptors and coefficients that best explain the experimental activity. As can be understood from the descriptors in these formulas, the data obtained with the ORCA and Multiwfn programs are very successful in creating QSAR models.

Group A = 20.8605 + (-0.0343 * MULTI347) + (52.9199 * ORCA074) + (-77.8793 * ORCA084) + (-0.0010 * ORCA129**)** **(9)**

Group B = 1.9105 + (0.000015 * MULTI271) + (-0.0891 * PADEL392) + (-1.9682 * PADEL461**)** **(10)**

Group C = -433.2757 + (-0.0043 * ORCA006) + (30.5537 * ORCA044) + (-0.0161 * PADEL162) + (-46.0181 * PADEL537) **(11)**

Group D = 1.2952 + (0.0030*MULTI407) + (-0.2002*ORCA001) + (-0.0008*ORCA155) **(12)**

Group E = 2.4956 + (-4.6343 * MULTI063) + (0.3591 * PADEL403) + (-0.7036 * PADEL468) **(13)**

Group F = 23.1996 + (62.2195 * MULTI010) + (-0.0003 * MULTI221) + (-11.6578 * MULTI294) + (-2.6086 * ORCA084) + (-0.0840 * PADEL802) **(14)**

Group G = 0.9450 + (-0.0113 * MULTI408) + (-0.0177 * ORCA136) + (-3.0465 * PADEL203) **(15)**

In Table 4.9, there are descriptors and definitions specified in formulas 9-15 selected as models for all groups. According to this table, log(1/IC$_{50}$) and IC$_{50}$ estimated cytotoxic activity values obtained from the formulas are found. In addition, promising ones from these values are indicated with green color. Specifically, in all groups, at least one of the descriptors from ORCA or Multiwfn

participated in the activity estimation. The success of descriptors related to dipole moment, NMR shielding, and orbital analysis is remarkable. In addition, the use of ZORA full electron basis set in calculations makes the Multiwfn program that calculates QTAIM descriptors valuable. It has been determined that the diversity of 3D parameters directly affects the success of the models. The creation of overfitted models is prevented. Since the complexes in Group C have high cytotoxicity, it is not recommended to be investigated in future studies. Complexes designed in Group F have very low cytotoxicity. These complexes can be used as drug candidates in in-vitro studies. Group E and Group D contain promising compounds with low cytotoxicity.

Table 4.9. Estimated cytotoxic activity values from descriptors detected in QSAR models for 30 complexes that may be new drug candidates in all groups.

GROUP A		AX-01	AX-02	AX-03
MULTI347	Fuzzy Magnitude of the Traceless Quadrupole Moment Tensor	37.154573	14.455383	22.498538
ORCA074	LUMO+1	-0.113032	-0.099571	-0.088975
ORCA084	LUMO - (HOMO-5)	0.128245	0.133569	0.149402
ORCA129	Pt Anisotropy Shielding	-4367.420	-4703.419	3397.296
Prediction	$Log(1/IC_{50})$	7.9842	9.3966	0.3476
	IC_{50}	>100	>100	2.23

GROUP B		BX-01	BX-02	BX-03	BX-04	BX-05	BX-06
MULTI271	Generating Basins for Attractor Pt	26470.9188	18543.3799	10103.7073	13659.9332	17694.3419	9735.4547
PADEL392	nAtomP	6	11	11	11	12	12
PADEL461	nF8Ring	0	1	0	0	0	0
Prediction	$Log(1/IC_{50})$	1.7730	-0.7596	1.0820	1.1353	1.1067	0.9873
	IC_{50}	59.29	0.17	12.08	13.66	12.78	9.71

GROUP C		CX-01	CX-02	CX-03	CX-04
ORCA006	Dipole Moment Electonic Contribution (Z)	-40.905350	1.495290	-1.372170	2.397140
ORCA044	Loewdin Reduced Total "f "Orbital Charge of Pt Atom	14.242227	14.238687	14.240730	14.239574
PADEL162	AATSC8v	12.731119	9.317838	1.625832	2.144338
PADEL537	JGI8	0.008426	0.000000	0.007442	0.007760
Prediction	$Log(1/IC_{50})$	1.4602	1.6124	1.4685	1.3941
	IC_{50}	29.85	40.97	29.41	24.78

GROUP D		DX-01	DX-02	DX-03	DX-04
MULTI407	Hirshfeld ODI % in HOMO-2	49.33	20.63	32.55	26.16
ORCA001	X of The Origin for Moment Calculation	1.069563	2.190986	4.949096	4.686205
ORCA155	Total Shielding Tensor (Z1)	1794.591	1818.742	-909.062	-554.688
Prediction	$Log(1/IC_{50})$	-0.2066	-0.5365	1.1293	0.8793
	IC_{50}	0.62	0.29	13.47	7.57

Table 4.9. (continued)

GROUP E		EX-01	EX-02	EX-03	EX-04	EX-05	EX-06
MULTI063	Hirshfeld Atomic Dipole Moment for Total	0.301546	1.149436	0.482018	0.023421	0.424558	1.168015
PADEL403	MDEC-12	0.000000	0.000000	0.000000	6.419940	4.625084	0.000000
PADEL468	nF9Ring	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
Prediction	Log(1/IC_{50})	1.0981	-2.8312	0.2618	4.6925	2.1889	-3.6209
	IC_{50}	**12.54**	**0.001475**	**1.83**	**>100**	**>100**	**0.000239**

GROUP F		FX-01	FX-02	FX-03
MULTI010	VDD Total Charge	0.012713	0.012456	0.003213
MULTI221	Integrate RDG with Promolecular Approximation for Pt Atom in Whole Space	76473.5464	78404.5366	80386.3610
MULTI294	Atomic quadrupole moments Q_2,1 for Pt	0.175412	0.142581	0.148059
ORCA084	LUMO - (HOMO-5)	0.123764	0.123898	0.124507
PADEL802	RDF80p	15.285886	13.859722	16.058118
Prediction	Log(1/IC_{50})	-2.6033	-2.6963	-4.1161
	IC_{50}	**0.002493**	**0.002012**	**0.000077**

GROUP G		GX-01	GX-02	GX-03	GX-04
MULTI408	Hirshfeld ODI % in HOMO-1	17.09	18.95	27.00	34.91
ORCA136	Diamagnetic Contribution to the Shielding Tensor (Y3)	0.246	6.437	1.330	-1.520
PADEL203	MATS7v	-0.140803	-0.020113	-0.123315	-0.163208
Prediction	Log(1/IC_{50})	1.1765	0.6782	0.9920	1.0746
	IC_{50}	**15.01**	**4.77**	**9.82**	**11.88**

Many descriptors, including quantum mechanics, stable geometric structure, NMR and some 2D (PaDEL), were calculated one by one for platinum-containing complexes, which were determined from the literature and categorized according to their experimental procedures in QSAR analyses. These descriptors were also used to construct models that could theoretically predict activity for complexes. These models were evaluated by adhering to many statistical analyzes and OECD guidelines. Successful models were selected. These models were used to predict activity for 30 new drug candidate complexes. Interesting results were obtained that can be used in in-vitro and subsequent in-vivo studies. As a result of the activity estimation, the compounds marked with green in table 4.9 were predicted to be at a level that could be evaluated experimentally in future studies. In addition, novel complexes using 2-pyrimidinethiol as a functional group have been evaluated to reduce cytotoxicity.

As a result of QSAR analysis, we used MLR, which is one of the most common methods for QSAR studies, with the QSARINS program and separated the transferred descriptors at a rate of 70 (training):30 (test) with random splitting, which is one of the screening methods mentioned in the literature. Constant and highly correlated descriptors were eliminated, and many models were created. In choosing the best model, we adhered to many criteria. We have followed all OECD guidelines and tested models with graphs and analysis. In addition, the quality of all models was checked on 37 statistical parameters. Unlike other studies, we created successful models with high contributions from quantum mechanics, orbital properties, and NMR parameters. For example, the differences between low and high orbitals (such as HOMO-(LUMO+5)) from HOMO and LUMO, which are not encountered in other studies, were successful in creating a model. The basis of our study was to discover potential drug candidate compounds that may occur between 2-pyrimidinethiol and platinum. While forming these compounds, it was tried to make similar complexes that were divided into groups. Adhering to similar studies in the literature while creating the QSAR models, the main structure of the complexes was preserved, and 30 new platinum derivative complexes were formed by replacing some groups with 2-pyrimidinethiol. Estimated cytotoxic activity values were found for this complex by using the model formulas in their groups and successful results were obtained for 11 complexes.

These complexes can be synthesized in further studies and experimentally cytotoxic activities can be found because of in-vitro tests with the A2780 gene.

4.3 Structure-Based Drug Design

The best combination of basis set and DFT method was determined using the methods mentioned in section 3.3.1 initially. The most important parameter in structure-based drug design is the most accurate interaction between the ligand and the macromolecule. When using virtual screening methods for these interactions, it is necessary to calculate the geometric structures as quickly as possible, since many compounds are tested simultaneously. Therefore, instead of the ZORA full electron basis set group selected for metal and other atoms in ligand-based drug design, metal atom ECP basis set combinations that could perform faster SCF energy and geometric iteration calculations were used. In addition, an important comparison study was carried out by statistically evaluating the effects of these selected basis sets on geometry.

Geometric optimizations were performed with DFT calculations. Comparisons were made using 139 DFT methods using the complex with single crystal structure in figure 3.4. These DFT methods were compared with various statistical analyzes and graphs. During the geometric optimizations, the initial settings and keywords identified in section 4.1 were added to the inputs. The def2-ECP system was used for only the valance electrons of the platinum atom, and the def2-TZVP basis set was used for all atoms. As a result of the calculations, the geometric parameters (bond length and bond angles) were transferred to the tables for all DFT methods and compared with the experimental bond parameters in figure 3.4 according to the mentioned statistical formulas.

The MADn (n=1,2,3,4,5) statistical analysis in comparison was applied separately to bond lengths containing only metal, all platinum related bond lengths in the complex, bond angles containing metal in bond center, all bond angles including platinum atom, and all bond properties in the complex, respectively.

The methods were sorted according to the MAD statistical results in table 4.10. Among the 139 methods, the DFT functions that are in the first 50 ranks are listed.

In table 4.10-MAD1, only the bonds made by the metal atom are compared with the experimental data. It has been determined that DFT1-27 functions have a remarkable success in all deviation groups compared to other functions. This DFT group is the "B97-GGA1" function. The difference between the deviation value of DFT1-27 and DFT3-29 in the 50th order has approximately 31% deviation from the mean. The DFT2-21 functional has little success. Because DFT2-21 has a difference of about 10% from the average with the 50th method in MAD2. The deviation of the bond angles where the metal is in the center from the mean is around 55% in MAD3. This is a high difference. It has a widespread. In MAD4, on the other hand, the deviation of the bond angles where the metal is not in the center between the DFT functionals in the first row and the 50th row from the mean is around 19%. This is a slightly higher value. MAD5 represents the deviations for all mentioned bond properties. Here, the deviation difference is also around 19 percent. It is a relatively high value.

Table 4.10. The statistical deviations of the all-bond properties from the experimental value by MAD analysis. The colors purple, blue, green, pink, and red represent DFT1, DFT2, DFT3, DFT4 and DFT5, respectively.

DFT	MAD1	DFT	MAD2	DFT	MAD3	DFT	MAD4	DFT	MAD5
DFT1-27	0.00776000	DFT2-21	0.01459587	DFT1-32	0.30295000	DFT1-06	0.57938571	DFT1-06	0.35873041
DFT1-31	0.00871558	DFT4-02	0.01474427	DFT1-31	0.31978333	DFT1-33	0.58412857	DFT1-33	0.36193911
DFT1-06	0.00906557	DFT3-26	0.01476254	DFT1-37	0.32701667	DFT1-32	0.58507143	DFT1-32	0.36310867
DFT2-29	0.00924030	DFT2-27	0.01481821	DFT1-33	0.34775000	DFT1-36	0.59008571	DFT1-31	0.36767389
DFT1-33	0.00939057	DFT1-16	0.01497377	DFT1-36	0.34858333	DFT1-31	0.59300000	DFT1-36	0.36784875
DFT3-15	0.00959030	DFT4-03	0.01506649	DFT1-06	0.37191667	DFT1-27	0.60805714	DFT1-27	0.37654683
DFT1-19	0.01033500	DFT3-04	0.01511809	DFT1-27	0.43358333	DFT1-28	0.63214286	DFT1-28	0.39116954
DFT3-16	0.01044030	DFT4-04	0.01513315	DFT1-30	0.47978333	DFT1-30	0.64857143	DFT1-30	0.40136085
DFT3-23	0.01049031	DFT2-50	0.01516266	DFT1-28	0.50978333	DFT1-13	0.65974286	DFT1-13	0.40874811
DFT3-17	0.01054030	DFT1-01	0.01517377	DFT1-29	0.51825000	DFT1-29	0.66862857	DFT1-29	0.41369128
DFT3-22	0.01061531	DFT2-28	0.01517377	DFT1-13	0.53941667	DFT1-12	0.67474286	DFT1-12	0.41743072
DFT3-25	0.01066530	DFT2-47	0.01535155	DFT1-12	0.58441667	DFT1-37	0.67478571	DFT1-43	0.42163507
DFT3-21	0.01069030	DFT3-13	0.01539587	DFT1-46	0.60128333	DFT1-43	0.67850000	DFT1-37	0.42409982
DFT1-30	0.01076558	DFT4-05	0.01546649	DFT1-45	0.60295000	DFT1-41	0.68505714	DFT1-41	0.42480464
DFT3-12	0.01086530	DFT1-06	0.01548883	DFT1-43	0.60461667	DFT1-49	0.69061429	DFT2-27	0.42666800
DFT3-19	0.01086531	DFT1-03	0.01561105	DFT1-41	0.61491667	DFT2-27	0.69142857	DFT1-49	0.42736116
DFT2-18	0.01089030	DFT3-06	0.01565143	DFT1-44	0.63461667	DFT1-44	0.71135714	DFT1-44	0.43997420
DFT3-18	0.01091531	DFT2-49	0.01574044	DFT1-49	0.63645000	DFT1-47	0.71507143	DFT1-01	0.44292887
DFT3-20	0.01106000	DFT1-05	0.01583327	DFT2-27	0.63811667	DFT2-46	0.71720000	DFT2-46	0.44296800
DFT1-32	0.01109057	DFT1-23	0.01584427	DFT1-47	0.64128333	DFT1-01	0.71791429	DFT1-47	0.44649159
DFT1-23	0.01114030	DFT3-27	0.01601809	DFT2-11	0.65478333	DFT2-29	0.72291429	DFT2-29	0.44656795
DFT2-11	0.01131530	DFT1-08	0.01613315	DFT1-07	0.66025000	DFT2-47	0.72492857	DFT2-47	0.44726800
DFT2-01	0.01136530	DFT1-22	0.01628883	DFT2-29	0.68491667	DFT1-46	0.72601429	DFT1-39	0.45217420
DFT3-07	0.01139031	DFT1-33	0.01631105	DFT2-46	0.68991667	DFT1-45	0.72672857	DFT1-19	0.45381640
DFT2-05	0.01141531	DFT1-28	0.01632216	DFT1-01	0.70825000	DFT1-39	0.73148571	DFT1-46	0.45507918
DFT2-27	0.01141558	DFT1-10	0.01633315	DFT1-39	0.70991667	DFT1-19	0.73484286	DFT1-45	0.45551397
DFT3-26	0.01146530	DFT2-48	0.01634044	DFT2-47	0.72295000	DFT1-25	0.73777143	DFT2-50	0.45683756
DFT4-02	0.01161530	DFT3-19	0.01636254	DFT1-34	0.75978333	DFT1-07	0.73794286	DFT1-25	0.45723072
DFT2-21	0.01169031	DFT2-46	0.01638488	DFT1-42	0.76128333	DFT2-50	0.74077143	DFT2-21	0.45822447
DFT3-27	0.01171531	DFT1-11	0.01638871	DFT2-09	0.76275000	DFT2-21	0.74341429	DFT1-07	0.45879860
DFT2-08	0.01179030	DFT1-27	0.01641969	DFT2-50	0.76491667	DFT2-09	0.74841429	DFT2-09	0.46383225
DFT3-11	0.01179031	DFT1-20	0.01645538	DFT1-05	0.77478333	DFT2-11	0.74942857	DFT1-05	0.46402172
DFT2-37	0.01179031	DFT1-21	0.01646649	DFT3-19	0.77825000	DFT1-42	0.75125714	DFT2-28	0.46423321
DFT2-33	0.01183500	DFT2-30	0.01657457	DFT2-30	0.77961667	DFT1-05	0.75214286	DFT2-11	0.46424621
DFT2-40	0.01184030	DFT1-19	0.01666413	DFT1-25	0.78758333	DFT1-22	0.75275714	DFT1-22	0.46457389
DFT1-28	0.01184058	DFT2-29	0.01669587	DFT2-28	0.78991667	DFT2-33	0.75285714	DFT2-33	0.46528007
DFT2-10	0.01191531	DFT1-30	0.01681105	DFT2-05	0.79491667	DFT2-28	0.75291429	DFT2-49	0.46662887
DFT2-42	0.01191531	DFT2-05	0.01691809	DFT2-49	0.79658333	DFT2-49	0.75648571	DFT1-42	0.46737918
DFT2-38	0.01194031	DFT2-04	0.01696254	DFT1-22	0.80311667	DFT2-05	0.76077143	DFT2-05	0.46969838
DFT2-34	0.01194058	DFT1-29	0.01712216	DFT2-01	0.80658333	DFT3-19	0.76291429	DFT3-19	0.47078534
DFT1-29	0.01196558	DFT1-31	0.01716661	DFT2-33	0.80811667	DFT1-03	0.76341429	DFT1-03	0.47079563
DFT2-04	0.01206530	DFT1-12	0.01716740	DFT2-21	0.80941667	DFT2-30	0.76350000	DFT2-30	0.47122483
DFT1-05	0.01211558	DFT3-15	0.01729587	DFT1-19	0.81608333	DFT2-01	0.76434286	DFT2-01	0.47217230
DFT3-03	0.01214031	DFT3-14	0.01736254	DFT3-11	0.81608333	DFT1-34	0.76500000	DFT1-34	0.47353814
DFT2-30	0.01219237	DFT2-40	0.01744032	DFT1-03	0.81775000	DFT1-23	0.77055714	DFT1-23	0.47523471
DFT1-08	0.01221530	DFT2-15	0.01750698	DFT1-08	0.83275000	DFT3-11	0.77484286	DFT3-11	0.47887230
DFT3-10	0.01221530	DFT3-22	0.01752921	DFT2-40	0.83325000	DFT2-17	0.77555714	DFT1-08	0.47969558
DFT1-16	0.01226558	DFT2-34	0.01757377	DFT3-22	0.83658333	DFT1-08	0.77770000	DFT2-40	0.48120708
DFT4-04	0.01231530	DFT2-01	0.01768476	DFT1-10	0.84311667	DFT2-40	0.77934286	DFT2-17	0.48154053
DFT3-09	0.01233263	DFT1-39	0.01768963	DFT1-26	0.84345000	DFT3-08	0.77984286	DFT3-08	0.48176360

In the MAD5 analysis, in which all bond properties were compared with experimental data according to table 4.10, the DFT1-06 functional stood out. On the other hand, DFT1-27-31-06 methods are successful in the analysis of MAD1 containing platinum-containing bonds. The bonds between non-metal atoms are also included in the MAD2 analysis. It is interesting that the DFT2-21 of hybrid-GGA has the least errors. However, in this group, there was no big difference between the deviation of the other methods. DFT2-21 has low deviation differences only 0.0009 from DFT1-06, 0.002 from DFT1-27 and 0.003 from DFT1-21. Such low differences are not exactly a measure of success. In the MAD3 and MAD4 analyzes where the deviation of the angles from the experimental data is examined, the superior success of the GGA methods (DFT1) is striking. The difference of DFT1-32 from DFT2-27 in MAD3 is 0.34. The deviation difference is around 111%. This value is a high value. It is notable as a measure of success. There is a negligible difference of 0.07 between DFT1-32 and DFT1-06, which is one of the interesting methods. In the MAD4 analysis, DFT1-06, which has a ranking that can be considered successful in the deflection of the angles where the metal is in the center, got the lowest deviation value in the deflection analysis of the angles where the atoms without metal are in the center.

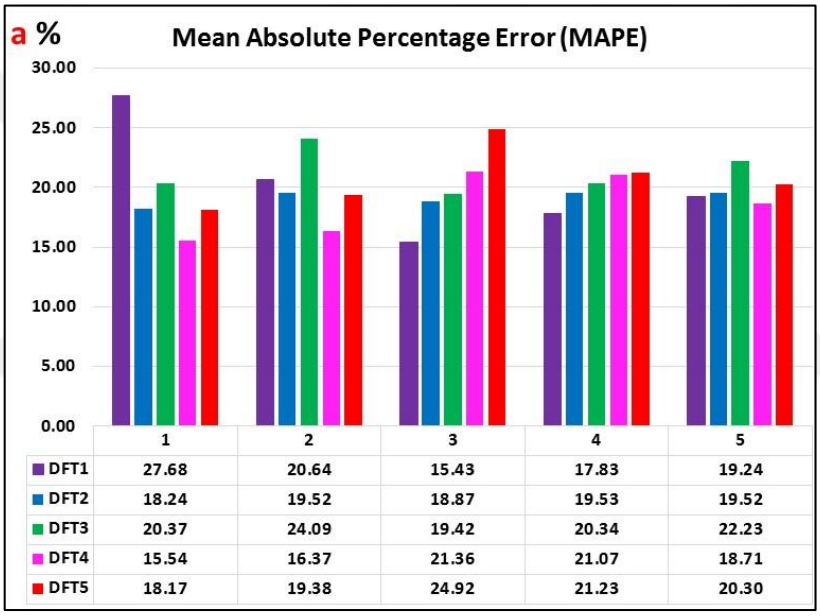
As a result, the success of DFT1-06-27-31-32-33-36 methods, which is one of the GGA methods, is remarkable. There are not very high differences between the deviation values of these DFT functionals. According to all the analyzes mentioned above, DFT1-06 was chosen for use in the next calculations. DFT1-06 corresponds to the OLYP functional.

In figure 4.15, there are four column percentage comparison graphs between “a-d”. These include MAPE, MAD, MSE, and RMSE, respectively. Numbers of 1-5 at the bottom of each graph represents metal-containing bond lengths, metal-containing and other bond lengths, metal-centered angles, metal-centered and metal-included angles, and all bond properties, and colors indicate DFT1-5, respectively. In addition, the values in the graphs are the total deviation value of each group as a percentage.

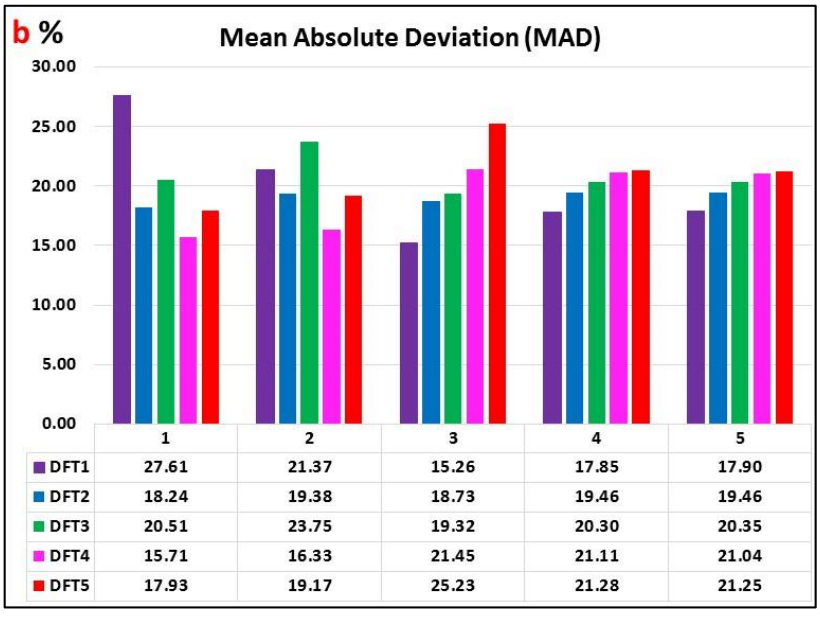
Similar results are observed in all statistical deviation parts of the graphs approximately. Considering graph 4.15a, according to the deviations in the lengths of the metal-containing bonds in the 1st group, the total deviation of the meta-GGA (DFT4) seems to be lower in general, although the GGA functions are successful.

Interestingly, the most unsuccessful DFT subgroup was identified as GGA (DFT1) in bond length comparisons. Some functions in DFT1 show high bias. In these cases, it is possible to cause instability in the geometry of the structure.

The increase in total deviations from DFT1 to DFT5 is obvious when we examine all the graphs and groups of 2, 3, 4 and 5 in figure 4.15. There is a similar upward trend in all divergence parts. The success in GGA (DFT1) is in line with the previous results. On the other hand, range-separated functionals (DFT5) are out of the trends with their high percentage of deviation. As a result, GGA group DFT functionals generally have lower deviation values in metal-containing complexes than other DFT groups.



	1	2	3	4	5
DFT1	27.68	20.64	15.43	17.83	19.24
DFT2	18.24	19.52	18.87	19.53	19.52
DFT3	20.37	24.09	19.42	20.34	22.23
DFT4	15.54	16.37	21.36	21.07	18.71
DFT5	18.17	19.38	24.92	21.23	20.30



	1	2	3	4	5
DFT1	27.61	21.37	15.26	17.85	17.90
DFT2	18.24	19.38	18.73	19.46	19.46
DFT3	20.51	23.75	19.32	20.30	20.35
DFT4	15.71	16.33	21.45	21.11	21.04
DFT5	17.93	19.17	25.23	21.28	21.25

Figure 4.15. (continued)

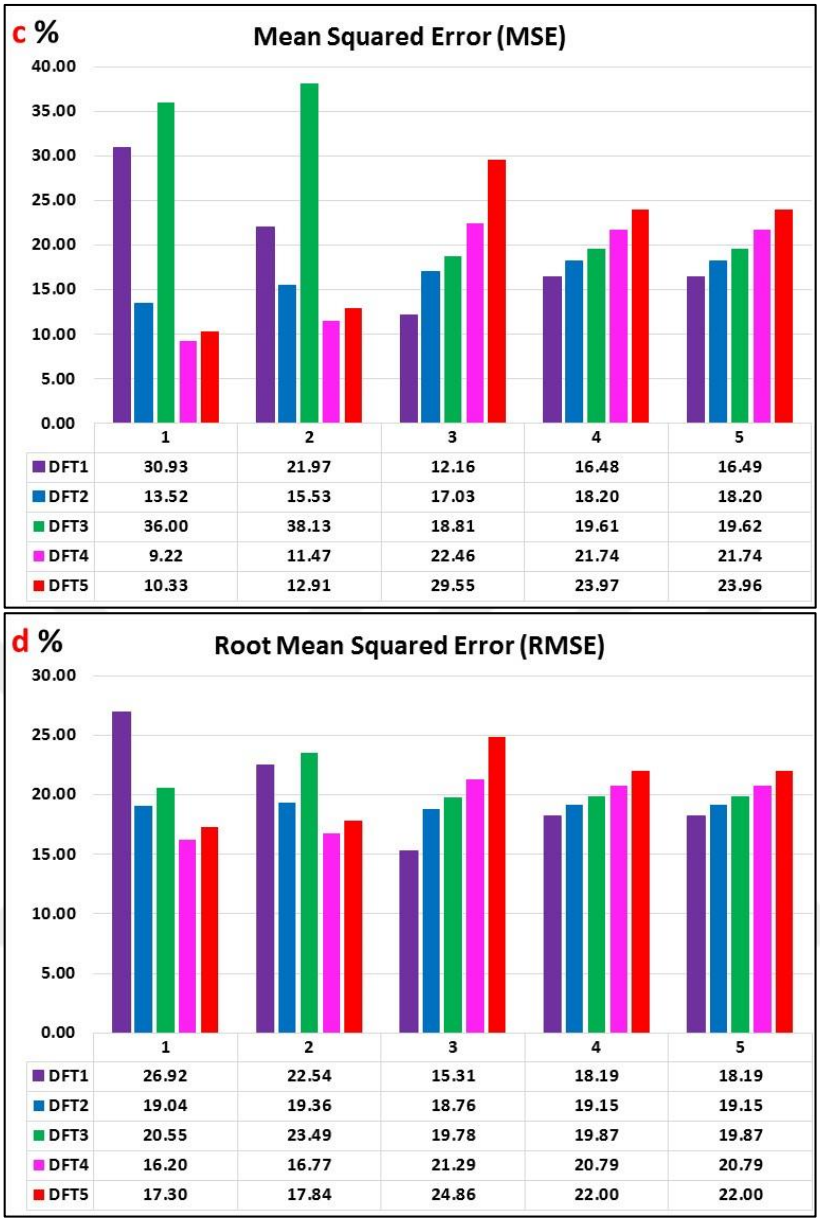


Figure 4.15. Comparative column chart. 1-5 at the bottom of each graph represents metal-containing bond lengths, metal-containing and other bond lengths, metal-centered angles, metal-centered and metal-included angles, and all bond properties, and colors indicate DFT1-5, respectively. a-MAPE, b-MAD, c-MSE, d-RMSE.

After choosing OLYP as the best DFT method, a total of 164 basis set combinations under four ECP groups were tested together with the OLYP function for basis set selection. Basis sets were grouped as double, triple, and quadruple-zeta as in table 3.5. Throughout the calculations, the initial settings from section 4.1 were integrated into the geometric optimization input files. Calculations made according to section 3.3.1 and statistical analysis results are given in table 4.11. The

MADn (n=1,2,3,4,5) statistical analysis in comparison was applied separately to bond lengths containing only metal, bond lengths related to platinum in the complex, bond angles for metal in the bond center, all bond angles including platinum, and all bond properties in the complex, respectively.

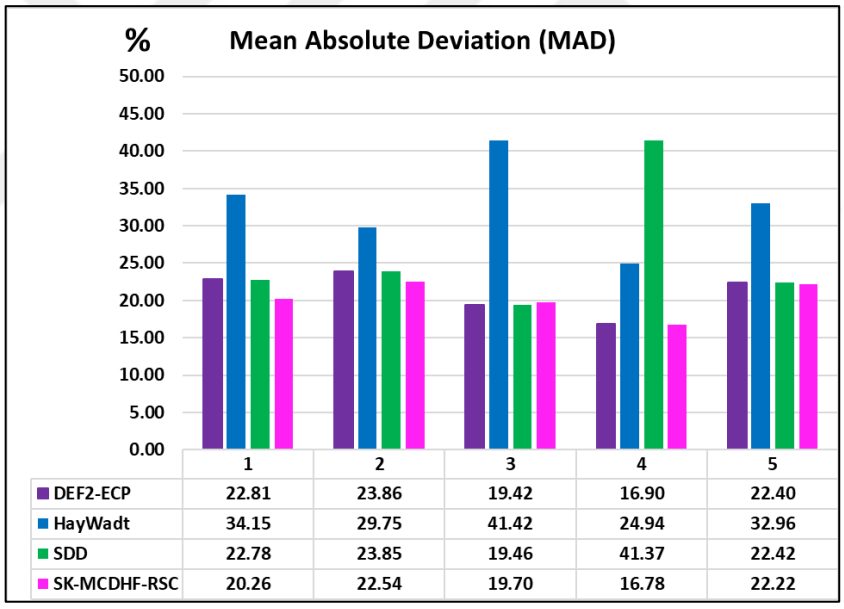
In Table 4.11, there are 35 basis set results for each MADn error analysis. According to these results, when the metal-containing and non-metallic bond lengths are compared, it is seen that the DEF2-ECP-12 basis set group has the lowest error. However, especially in MAD2 analysis, there is no high difference between the other groups. DEF2-ECP-12 has a negligible difference of only 0.0008 with the 35th combination. The DEF2-ECP-12 group, which showed the least success of deviation in metal-containing bonds, could not enter the ranking in table 4.11 in terms of MAD3 and MAD4 analysis. Although it has good values for bond lengths, it could not show the desired success in angle values. Deviation was examined for all the bond properties mentioned in MAD5. Interestingly, HayWadt-37 and 38 basis set combinations took the first place. These two groups are also successful in MAD4 (comparison of non-central metal atom) error analysis. However, there are no high differences in bias between groups in MAD4. This information is not enough to choose the best group. Therefore, HayWadt-37 is more successful than HayWadt-38 in the MAD1 (0.002) analysis, where the spread between the 1st and 35th groups is relatively high.

As a result of all these evaluations, the HayWadt-37 basis set combination was chosen to be used in the geometric optimization of the designed complexes. This combination includes LANL2TZ for metal atom, HayWadt for metal ECP, and def2-TZP basis set for non-metal atoms. Another advantage of this combination is that it is fast in calculations.

Table 4.11. The statistical deviations of the all-bond properties from the experimental value by MADn (n=1,2,3,4,5) analysis.

BASIS SET GROUP	MAD 1	BASIS SET GROUP	MAD 2	BASIS SET GROUP	MAD 3	BASIS SET GROUP	MAD 4	BASIS SET GROUP	MAD 5
DEF2-ECP-12	0.005885	DEF2-ECP-12	0.013664	DEF2-ECP-40	0.248417	HayWadt-38	0.557243	HayWadt-38	0.345039
DEF2-ECP-10	0.005910	SDD-12	0.013697	SDD-40	0.248417	HayWadt-37	0.557957	HayWadt-37	0.345448
DEF2-ECP-11	0.005910	SK-MCDHF-RSC-07	0.013712	SK-MCDHF-RSC-40	0.256617	HayWadt-41	0.561729	SDD-27	0.349374
SDD-12	0.005935	SK-MCDHF-RSC-26	0.013712	HayWadt-41	0.259083	DEF2-ECP-27	0.564371	SDD-28	0.349374
SDD-10	0.005935	HayWadt-03	0.013854	DEF2-ECP-16	0.292283	DEF2-ECP-28	0.564371	DEF2-ECP-27	0.349387
SDD-11	0.005935	DEF2-ECP-05	0.013886	SDD-16	0.293950	SDD-27	0.564371	DEF2-ECP-28	0.349387
SDD-09	0.006135	SDD-05	0.013920	SK-MCDHF-RSC-16	0.305250	SDD-28	0.564371	DEF2-ECP-30	0.349796
SK-MCDHF-RSC-30	0.006135	SK-MCDHF-RSC-19	0.013955	DEF2-ECP-21	0.315250	DEF2-ECP-30	0.565086	DEF2-ECP-31	0.349796
SK-MCDHF-RSC-31	0.006135	SK-MCDHF-RSC-32	0.014042	SDD-21	0.315250	DEF2-ECP-31	0.565086	HayWadt-41	0.349918
SDD-08	0.006141	DEF2-ECP-10	0.014064	DEF2-ECP-14	0.324783	SDD-30	0.565800	SDD-30	0.350217
DEF2-ECP-09	0.006191	DEF2-ECP-11	0.014064	SDD-14	0.324783	SDD-31	0.565800	SDD-31	0.350217
SK-MCDHF-RSC-27	0.006210	SK-MCDHF-RSC-29	0.014075	SK-MCDHF-RSC-14	0.326083	SDD-18	0.579386	SDD-18	0.358739
SK-MCDHF-RSC-28	0.006210	SDD-10	0.014075	HayWadt-36	0.342283	DEF2-ECP-18	0.580100	DEF2-ECP-17	0.359165
DEF2-ECP-08	0.006216	SDD-11	0.014075	DEF2-ECP-27	0.346917	DEF2-ECP-17	0.580100	SDD-17	0.359165
SK-MCDHF-RSC-23	0.006610	SK-MCDHF-RSC-30	0.014120	DEF2-ECP-28	0.346917	SDD-17	0.580100	DEF2-ECP-18	0.359178
SK-MCDHF-RSC-22	0.006635	SK-MCDHF-RSC-31	0.014120	SDD-27	0.346917	DEF2-ECP-22	0.585086	DEF2-ECP-22	0.362091
SK-MCDHF-RSC-10	0.006735	SK-MCDHF-RSC-24	0.014153	SDD-28	0.346917	DEF2-ECP-32	0.585800	DEF2-ECP-32	0.362213
SK-MCDHF-RSC-11	0.006735	SK-MCDHF-RSC-27	0.014164	SK-MCDHF-RSC-21	0.350250	SDD-32	0.586514	SDD-32	0.362639
SK-MCDHF-RSC-32	0.006885	SK-MCDHF-RSC-28	0.014164	DEF2-ECP-30	0.350250	SDD-22	0.586514	SDD-22	0.362943
SK-MCDHF-RSC-29	0.006985	SDD-09	0.014220	DEF2-ECP-31	0.350250	DEF2-ECP-23	0.586514	DEF2-ECP-10	0.362947
SK-MCDHF-RSC-17	0.007141	SDD-08	0.014233	HayWadt-37	0.350250	DEF2-ECP-10	0.587229	DEF2-ECP-11	0.362947
SK-MCDHF-RSC-18	0.007191	SDD-24	0.014233	SDD-30	0.350250	DEF2-ECP-11	0.587229	DEF2-ECP-23	0.362956
DEF2-ECP-05	0.007285	DEF2-ECP-09	0.014244	SDD-31	0.350250	SDD-23	0.587229	SDD-23	0.363374
SDD-05	0.007385	DEF2-ECP-24	0.014267	HayWadt-38	0.351917	DEF2-ECP-29	0.588657	DEF2-ECP-29	0.363965
SK-MCDHF-RSC-19	0.007390	DEF2-ECP-08	0.014267	DEF2-ECP-26	0.356917	SK-MCDHF-RSC-40	0.589257	SDD-10	0.364256
HayWadt-37	0.007591	SK-MCDHF-RSC-23	0.014331	DEF2-ECP-32	0.360250	SDD-10	0.589371	SDD-11	0.364256
SDD-32	0.007616	SK-MCDHF-RSC-22	0.014342	SDD-32	0.360250	SDD-11	0.589371	SK-MCDHF-RSC-27	0.365160
SDD-29	0.007641	SK-MCDHF-RSC-10	0.014375	HayWadt-39	0.360250	SK-MCDHF-RSC-37	0.589471	SK-MCDHF-RSC-28	0.365160
DEF2-ECP-32	0.007666	SK-MCDHF-RSC-11	0.014375	HayWadt-34	0.361583	SDD-29	0.590800	SDD-29	0.365256
DEF2-ECP-29	0.007716	SDD-32	0.014389	DEF2-ECP-29	0.361917	SK-MCDHF-RSC-27	0.590800	SK-MCDHF-RSC-37	0.365670
HayWadt-38	0.007716	HayWadt-39	0.014401	SDD-26	0.361917	SK-MCDHF-RSC-28	0.590800	SK-MCDHF-RSC-30	0.367316
SK-MCDHF-RSC-08	0.007760	DEF2-ECP-32	0.014411	SDD-29	0.365250	DEF2-ECP-40	0.592171	SK-MCDHF-RSC-31	0.367316
SK-MCDHF-RSC-12	0.007760	SDD-29	0.014411	SDD-18	0.368583	SDD-40	0.592171	SK-MCDHF-RSC-10	0.367416
SDD-30	0.007766	DEF2-ECP-29	0.014444	DEF2-ECP-18	0.370250	SK-MCDHF-RSC-38	0.592329	SK-MCDHF-RSC-11	0.367416
SDD-31	0.007766	SK-MCDHF-RSC-12	0.014464	HayWadt-16	0.370250	SK-MCDHF-RSC-10	0.594371	SK-MCDHF-RSC-38	0.367439

In Figure 4.16, the ECP types used on the platinum atom were compared. In this graph, it is seen that DEF2-ECP and SK-MCDHF-RSC groups generally have a high rate and similar success. In Figure 4.17, the parameters in the basis set combinations were evaluated by MAD analysis in terms of metal-containing bond lengths and metal-centered bond angles. When Figure 4.16 and 4.17 are examined together, the high error of LANL2DZ caused some failure for HayWadt metal ECP. In Figure 4.17a, it was determined that def2-QZVP had the lowest error, especially in bond lengths containing metal. Also, in the same graph, it is seen that LANL2TZ is more successful at angles where metal is in the center. It is understood that increasing the basis set size has chance the success in bond lengths but decreases it slightly in angles. Interestingly, def2-SVP, which showed a relatively high deviation in bond lengths, was the basis set with the lowest deviation in metal-centered bond angles.



	1	2	3	4	5
DEF2-ECP	22.81	23.86	19.42	16.90	22.40
HayWadt	34.15	29.75	41.42	24.94	32.96
SDD	22.78	23.85	19.46	41.37	22.42
SK-MCDHF-RSC	20.26	22.54	19.70	16.78	22.22

Figure 4.16. Comparative column chart. 1-5 at the bottom of each graph represents metal-containing bond lengths, metal related bond lengths, metal-centered angles, metal-included angles, and all bond properties, and colors denote metal ECPs.

In the comparison in Figure 4.17b, basis set sizes were evaluated. A result confirming the assessments identified in graph 4.17a emerged in graph 4.17b. The Quadruple-zeta is the largest sized basis set and increasing the size of the basis set reduces the deviation in metal-containing bond lengths. It is clear from the graph

that the double-zeta basis sets show very high deviations. However, there is a different situation for metal-centered bond angles. It has been observed that basis set errors occur on the angles by increasing the size of the bass set. Especially in bond angles, triple-zeta basis set showed lower deviations.

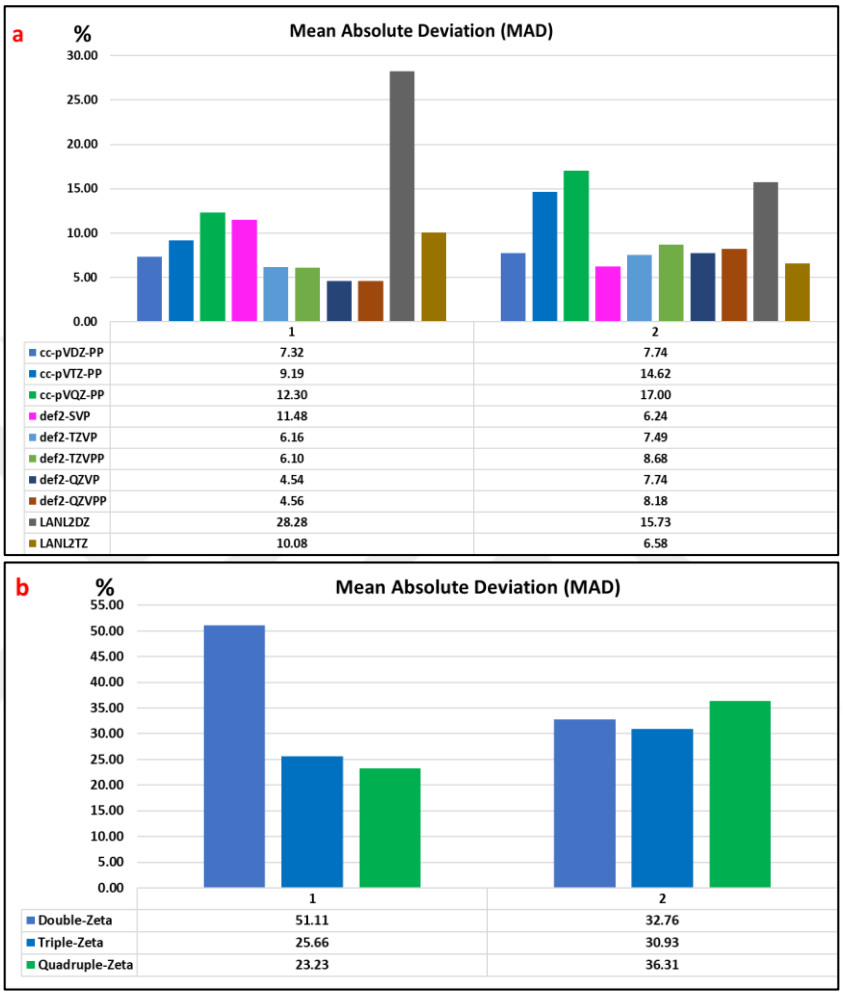


Figure 4.17. Comparative column chart. 1-2 at the bottom of each graph represents metal-containing bond lengths and metal-centered angles of MAD analysis. Comparison of the metal ECP (graph a) and the basis sets used for all metal atoms (graph b).

As described in Section 3.3.2, the best conformers were determined using the Gaussian 03 program for 84 platinum and 2-pyrimidinethiol-containing complexes that are expected to be drug candidates and designed. Geometric optimizations of the best conformers were performed using the selected DFT method, OLYP, and a basis set combination (HayWadt-37) with ORCA 4.2 in TÜBİTAK TRUBA high computing system. As a result of the optimizations, the final geometries of the complexes were saved in "mol2" format.

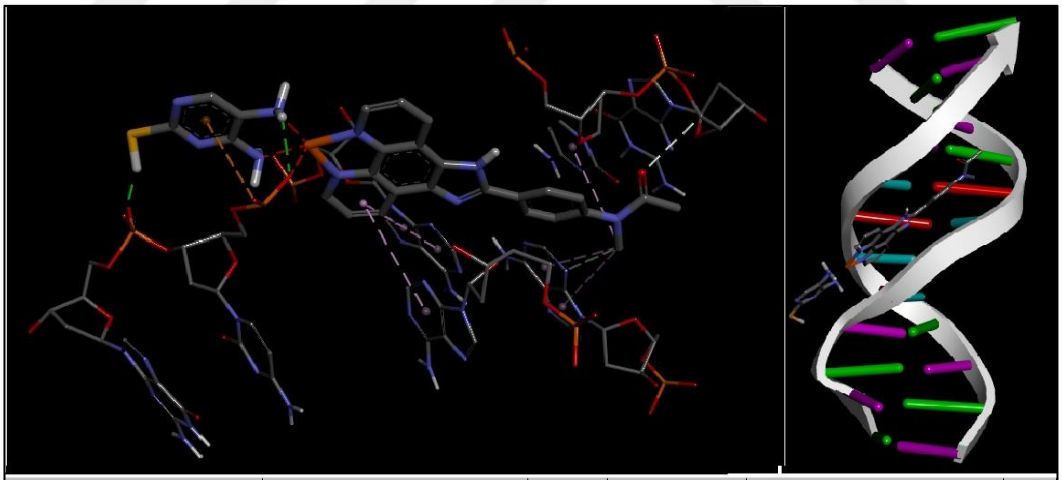
Molecular docking is a well-established computational technique to predict the interaction between molecules and DNA and to find the best orientation and binding of the molecule that will form a new complex with overall minimum energy. Molecular docking calculations were carried out with the methods described in section 3.3.4 for 84 DNA intercalator complexes, the best geometries of which were revealed and divided into 4 groups. In these calculations, those with an RMSD value above 2.0 Å were never included as the first criterion. In the second criterion, the binding energy to DNA, which is the target of the complexes, was taken. Inhibition constant was added for the third criterion and the sum of intermolecular interaction energies was chosen as the fourth criterion. All these criteria are given in table 4.12.

Table 4.12. Molecular docking study including the binding energies of 84 complexes with DNA, the inhibition constant of binding, and intermolecular interactions.

NO	Binding Energy (kcal/mol)				Inhibition Constant, Ki (nm)				vdW+Hbond+desolv Energy (kcal/mol)			
	SBDD1	SBDD2	SBDD3	SBDD4	SBDD1	SBDD2	SBDD3	SBDD4	SBDD1	SBDD2	SBDD3	SBDD4
1a	-14.01	-9.87	-11.81	-9.65	0.0541	58.62	2.20	83.85	-10.05	-7.14	-8.72	-6.79
1b	-14.36	-9.90	-11.69	-9.74	0.0298	55.56	2.69	72.87	-10.40	-7.25	-9.25	-7.56
1c	-14.00	-9.43	-11.58	-9.79	0.0543	122.02	3.24	66.15	-9.69	-6.51	-8.49	-6.93
1d	-14.70	-10.24	-11.94	-10.23	0.0168	30.99	1.76	31.49	-10.43	-7.33	-9.25	-6.86
1e	-14.96	-10.48	-12.19	-10.61	0.0108	20.84	1.15	16.67	-10.67	-7.59	-9.40	-6.74
1f	-14.57	-11.28	-12.51	-11.02	0.0208	5.40	0.67	8.36	-10.64	-8.57	-9.45	-7.79
1g	-14.37	-10.41	-12.01	-10.28	0.0295	23.35	1.58	29.18	-10.43	-7.73	-9.07	-6.75
1h	-13.94	-9.96	-11.63	-9.81	0.0609	49.90	2.99	64.18	-10.01	-7.27	-8.66	-7.61
1i	-14.54	-10.05	-11.86	-10.50	0.0220	43.18	2.03	20.20	-10.49	-7.37	-8.70	-7.23
1j	-14.87	-10.34	-12.12	-10.48	0.0126	26.38	1.30	20.73	-11.04	-7.86	-9.87	-7.49
1k	-14.55	-10.45	-12.22	-10.75	0.0216	21.80	1.10	13.13	-10.66	-7.81	-9.22	-7.19
1l	-15.07	-10.76	-12.29	-10.64	0.0090	12.88	0.99	15.96	-11.03	-8.10	-7.69	-8.04
1m	-14.53	-10.17	-11.75	-9.79	0.0224	35.08	2.42	66.65	-10.49	-7.57	-8.64	-6.44
1n	-14.75	-10.09	-12.13	-10.10	0.0155	40.41	1.28	39.25	-10.84	-7.42	-9.59	-7.57
1o	-14.84	-10.47	-12.27	-10.02	0.0132	21.02	1.01	45.04	-11.20	-8.12	-10.07	-7.18
1p	-14.58	-10.09	-11.96	-10.21	0.0205	40.42	1.71	32.86	-10.79	-7.72	-9.77	-7.14
1q	-15.19	-10.68	-12.84	-11.34	0.0074	14.74	0.39	4.88	-11.48	-7.42	-10.55	-8.28
1r	-14.03	-9.49	-11.71	-10.27	0.0524	111.31	2.62	29.52	-9.78	-6.21	-8.64	-7.25
1s	-14.53	-10.08	-11.85	-10.11	0.0222	41.15	2.07	38.75	-10.29	-7.17	-8.78	-6.77
1t	-14.44	-9.74	-11.60	-9.48	0.0262	72.84	3.16	112.27	-10.43	-7.10	-9.24	-6.18
1u	-14.86	-10.14	-12.43	-10.43	0.0128	36.86	0.78	22.56	-11.07	-7.57	-9.56	-7.42

The RMSD ≤ 2.0 Å criterion was used in the calculations for a good solution of the structure as the root mean square deviation (RMSD). According to the table, it was determined that the complexes containing the "1q" functional group generally interact with DNA with high binding energies. In particular, the binding affinities of the complexes in the SBDD1 group are higher than in the other groups. Accordingly, the inhibition constants are also lower. Intermolecular interactions are also more. The SBDD1 group complexes are longer in structure than the others.

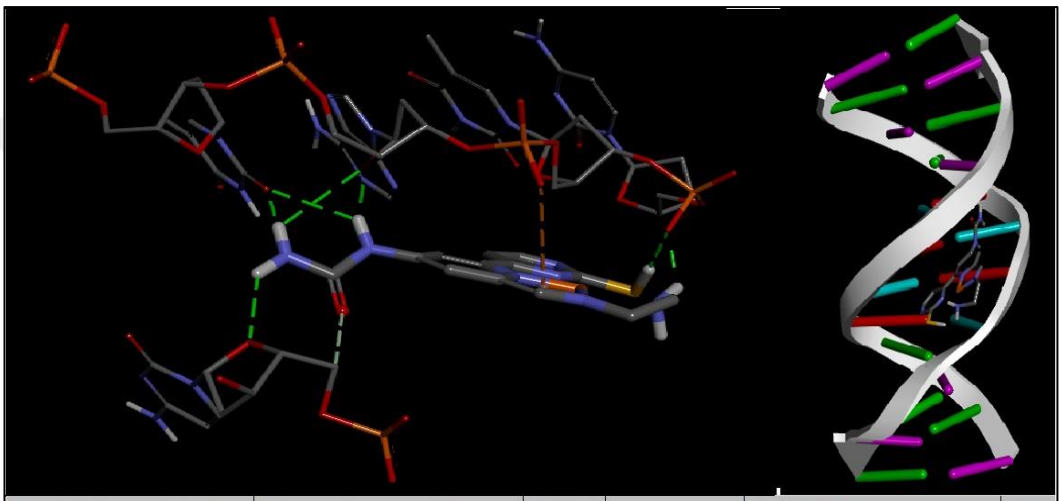
Figure 4.18 lists the intermolecular interactions in the binding of the SBDD1-1q complex with DNA. Here, it was observed that the "SH" group in the 2-pyrimidinethiol ligand, which is the subject of the study, forms a strong hydrogen bond with the oxygen in the phosphate bound to Guanine (DG10 residue). In addition, it was determined that hydrogen bonded with one of the slightly weaker platinum-bound donor nitrogen atoms and the acceptor “OP1” in the phosphate group bound to cytosine (DC10). As hydrophobic interactions occur with alkyl groups in this binding, the complex remained in the DNA groove with good binding affinity. These interactions arise from the pi orbitals of residues in DNA.



From Interaction Group 1		To Interaction Group 2		Distance	Category	Interaction Types	Color
LIG:SBDD1-1q:H41	H-Donor	A:DC9:OP1	H-Acceptor	2.3265	Hydrogen Bond	Conventional Hydrogen Bond	
LIG:SBDD1-1q:H9	H-Donor	A:DG10:OP2	H-Acceptor	1.9792	Hydrogen Bond	Conventional Hydrogen Bond	
B:DG22:C5'	H-Donor	LIG:SBDD1-1q:O53	H-Acceptor	3.4655	Hydrogen Bond	Carbon Hydrogen Bond	
LIG:SBDD1-1q:C58	H-Donor	A:DA5:N3	H-Acceptor	3.1515	Hydrogen Bond	Carbon Hydrogen Bond	
A:DC9:OP2	Negative	LIG:SBDD1-1q	Pi-Orbitals	4.7392	Electrostatic	Pi-Anion	
A:DA5	Pi-Orbitals	LIG:SBDD1-1q:C58	Alkyl	5.3283	Hydrophobic	Pi-Alkyl	
A:DA5	Pi-Orbitals	LIG:SBDD1-1q:C58	Alkyl	4.5022	Hydrophobic	Pi-Alkyl	
A:DA7	Pi-Orbitals	LIG:SBDD1-1q	Alkyl	5.1495	Hydrophobic	Pi-Alkyl	
B:DA19	Pi-Orbitals	LIG:SBDD1-1q	Alkyl	5.2937	Hydrophobic	Pi-Alkyl	
B:DA19	Pi-Orbitals	LIG:SBDD1-1q	Alkyl	5.1643	Hydrophobic	Pi-Alkyl	
B:DC21	Pi-Orbitals	LIG:SBDD1-1q:C58	Alkyl	5.4277	Hydrophobic	Pi-Alkyl	

Figure 4.18. Docking interaction properties of SBDD1-1q with target DNA.

Figure 4.19 lists the intermolecular interactions in the binding of the SBDD2-1f complex with DNA. The binding energies of the complexes in this group are lower than those of the SBDD1 group. The reason for this is clearly obvious. Because the donor "SH" group in the complex made a shorter and weaker hydrogen bond with the (DC9) acceptor "OP1" in cytosine than SBDD1-1q. Also, hydrophobic "pi" interactions with alkyl groups are not found. Meanwhile, the inclusion of the "1f" functional group, which is rich in polar hydrogens, in the structure enabled the formation of more than one hydrogen bond with DNA.

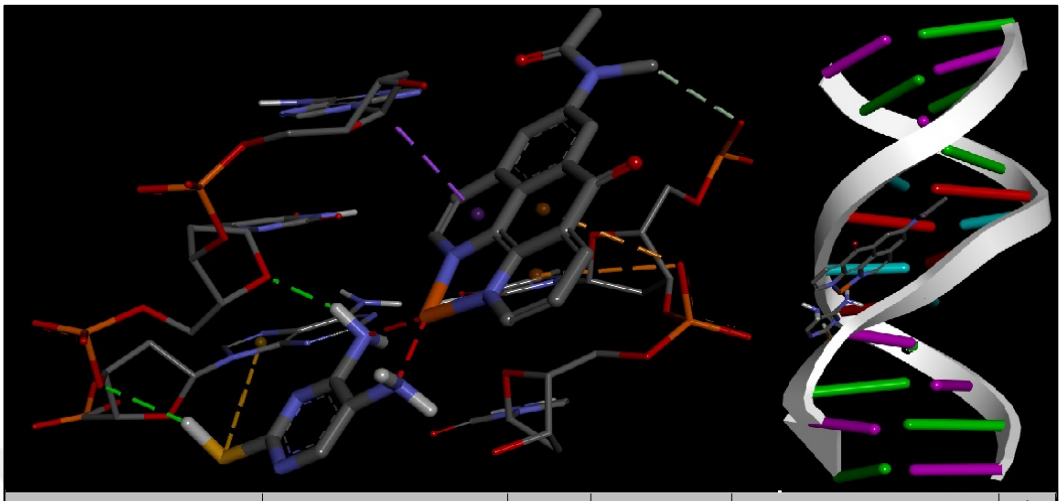


From Interaction Group 1		To Interaction Group 2		Distance	Category	Interaction Types	Color
LIG:SBDD2-1f:H29	H-Donor	A:DC9:OP1	H-Acceptor	1.9708	Hydrogen Bond	Conventional Hydrogen Bond	
LIG:SBDD2-1f:H33	H-Donor	A:DT6:O2	H-Acceptor	2.8041	Hydrogen Bond	Conventional Hydrogen Bond	
LIG:SBDD2-1f:H33	H-Donor	A:DA7:O4'	H-Acceptor	1.8807	Hydrogen Bond	Conventional Hydrogen Bond	
LIG:SBDD2-1f:H37	H-Donor	A:DT6:O2	H-Acceptor	1.9500	Hydrogen Bond	Conventional Hydrogen Bond	
LIG:SBDD2-1f:H37	H-Donor	A:DA7:O4'	H-Acceptor	3.0034	Hydrogen Bond	Conventional Hydrogen Bond	
LIG:SBDD2-1f:H38	H-Donor	B:DC21:O4'	H-Acceptor	2.5403	Hydrogen Bond	Conventional Hydrogen Bond	
LIG:SBDD2-1f:H8	H-Donor	A:DC9:OP1	H-Acceptor	2.0921	Hydrogen Bond	Conventional Hydrogen Bond	
B:DC21:C5'	H-Donor	LIG:SBDD2-1f:O35	H-Acceptor	2.7422	Hydrogen Bond	Carbon Hydrogen Bond	
A:DT8:OP1	Negative	LIG:SBDD2-1f	Pi-Orbitals	4.1280	Electrostatic	Pi-Anion	

Figure 4.19. Docking interaction properties of SBDD2-1f with target DNA.

Figure 4.20 lists the intermolecular interactions in the binding of the SBDD3-1q complex with DNA. Although the binding energies of the complexes in this group are lower than those of the SBDD1 group, they are higher than the other groups. In this context, there is a non-strong long distance hydrogen bond between the donor "SH" group in 2-pyrimidinethiol and the acceptor "OP2" in guanine (DG10). However, the "pi-anion" interaction that occurred between cytosine (DC9) in DNA and the rings of the complex ensured the attachment of the complex in the target DNA groove. In addition, one hydrophobic interaction occurred. As the

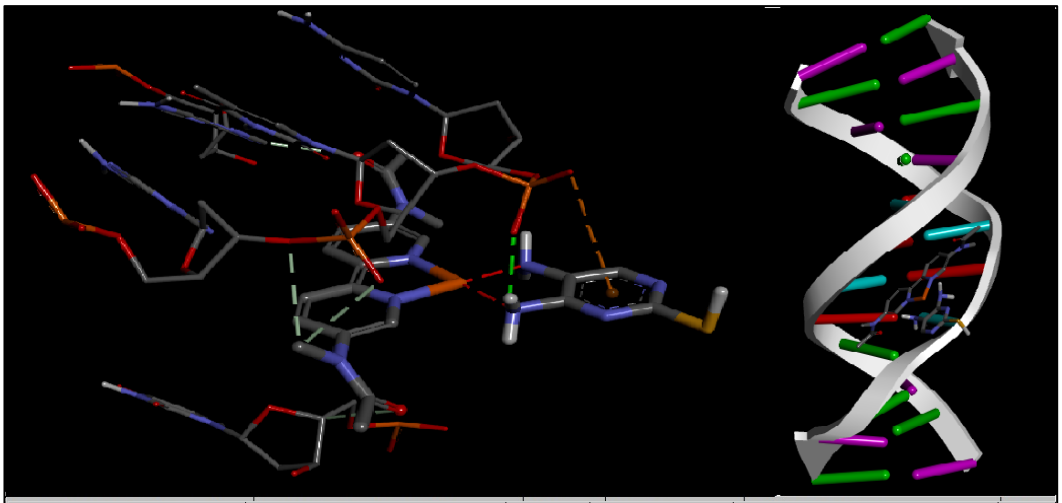
number of aromatic rings with double bonded carbons in the complexes increased in the SBDD3 group, the "pi" interactions with DNA increased.



From Interaction Group 1		To Interaction Group 2		Distance	Category	Interaction Types	Color
LIG:SBDD3-1q:H36	H-Donor	B:DT18:O4'	H-Acceptor	2.0361	Hydrogen Bond	Conventional Hydrogen Bond	
LIG:SBDD3-1q:H9	H-Donor	B:DA17:O3'	H-Acceptor	2.5906	Hydrogen Bond	Conventional Hydrogen Bond	
LIG:SBDD3-1q:C49	H-Donor	A:DT8:OP1	H-Acceptor	3.2406	Hydrogen Bond	Carbon Hydrogen Bond	
A:DC9:OP1	Negative	LIG:SBDD3-1q	Pi-Orbitals	4.1614	Electrostatic	Pi-Anion	
A:DC9:OP1	Negative	LIG:SBDD3-1q	Pi-Orbitals	3.6606	Electrostatic	Pi-Anion	
B:DA19:C1'	C-H	LIG:SBDD3-1q	Pi-Orbitals	3.6567	Hydrophobic	Pi-Sigma	
LIG:SBDD3-1q:S6	Sulfur	B:DA17	Pi-Orbitals	5.3644	Other	Pi-Sulfur	

Figure 4.20. Docking interaction properties of SBDD3-1q with target DNA.

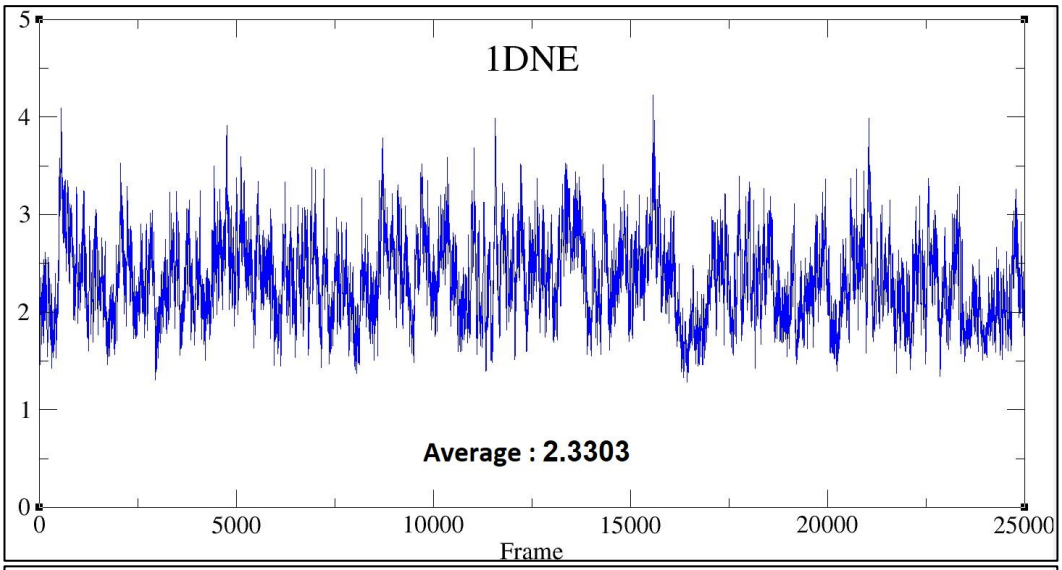
Figure 4.21 lists the intermolecular interactions in the binding of the SBDD4-1q complex with DNA. The binding energies of the complexes here are lower than those of the SBBD1 and SBDD3 groups. In other groups, the hydrogen bond formed between the "SH" group in 2-pyrimidinethiol, and DNA was not formed in SBDD4-1q. When the complex is examined, it is seen that the ligand on the side entering the DNA groove is long and planar, and the interactions that provide the bond with DNA are in this region. Electrostatic pi-anion interaction appears to occur between the ring of the 2-pyrimidinethiol ligand outside the groove and the "OP2" atom of the cytosine residue (DC9). In addition, carbon-hydrogen bonds are more common, and their distances are too far to be called weak.

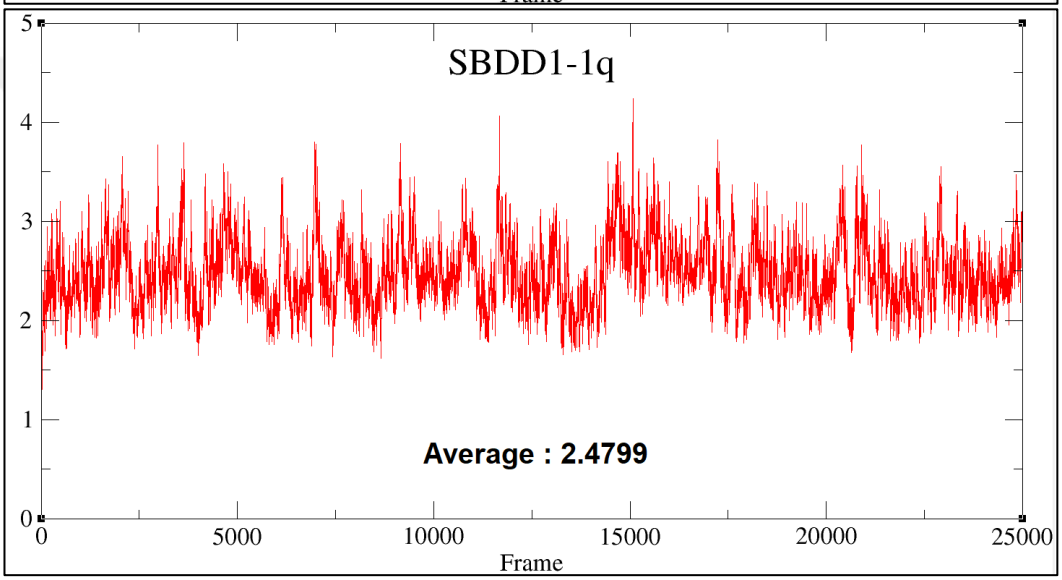


From Interaction Group 1		To Interaction Group 2		Distance	Category	Interaction Types	Color
LIG:SBDD4-1q:H	H-Donor	A:DC9:OP1	H-Acceptor	2.3006	Hydrogen Bond	Conventional Hydrogen Bond	
B:DA17:C2	H-Donor	LIG:SBDD4-1q:O	H-Acceptor	2.7649	Hydrogen Bond	Carbon Hydrogen Bond	
B:DT20:C4'	H-Donor	LIG:SBDD4-1q:O	H-Acceptor	3.1286	Hydrogen Bond	Carbon Hydrogen Bond	
LIG:SBDD4-1q:C	H-Donor	A:DA7:O3'	H-Acceptor	3.3376	Hydrogen Bond	Carbon Hydrogen Bond	
LIG:SBDD4-1q:C	H-Donor	A:DT8:OP1	H-Acceptor	3.2059	Hydrogen Bond	Carbon Hydrogen Bond	
A:DC9:OP2	Negative	LIG:SBDD4-1q	Pi-Orbitals	4.6475	Electrostatic	Pi-Anion	

Figure 4.21. Docking interaction properties of SBDD4-1q with target DNA.

The best intercalator pose in each group whose Molecular Docking studies were completed and analyzed was taken in "pdb" format for use in molecular dynamics simulations. 100-ns unconstrained molecular dynamics simulation was performed for four platinum-DNA intercalator complexes (SBDD1-1q, SBDD2-1f, SBDD3-1q and SBDD4-1q) and ligand-removed 12-mer B-DNA (1DNE). The simulations were completed in five stages using the methods described in section 3.3.5. The RMSD values were examined for all backbone atoms referenced to their starting structures corresponding to the binding patterns of platinum-containing complexes with DNA from the groove. Graphs provide information about whether the system has reached equilibrium. Small RMSD values indicate that the system has reached equilibrium and is stable. Large and fluctuating RMSD values indicate large conformational changes in the system. Figure 4.22 shows the RMSD graphs of the simulations performed for five structures. In the graphs, it is seen that the simulations performed in all groups reached equilibrium and were stable. Also, small RMSD values are taken.


1DNE
Average : 2.3303
Frame


SBDD1-1q
Average : 2.4799
Frame

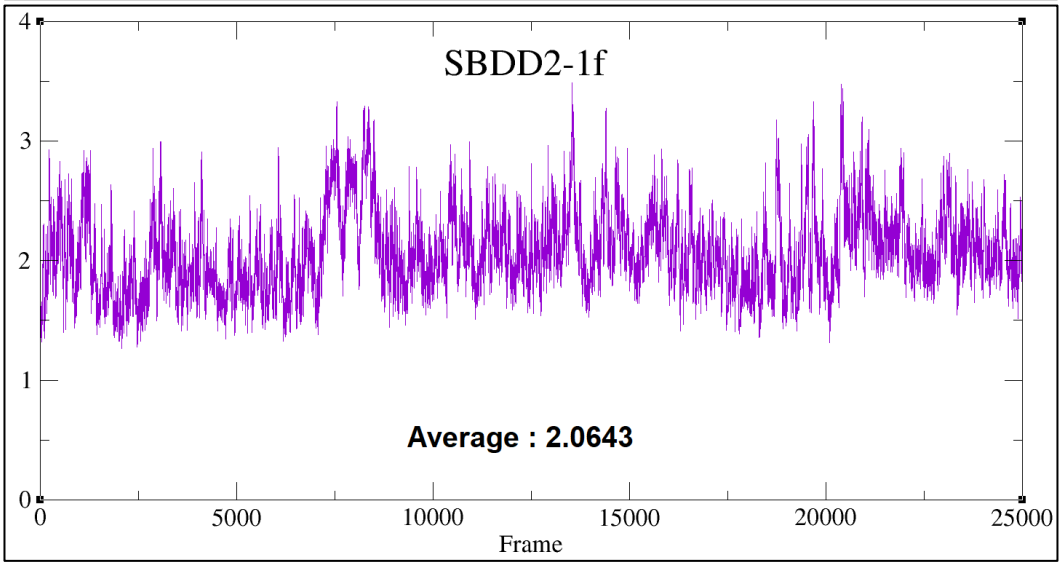

SBDD2-1f
Average : 2.0643
Frame

Figure 4.22. (continued)

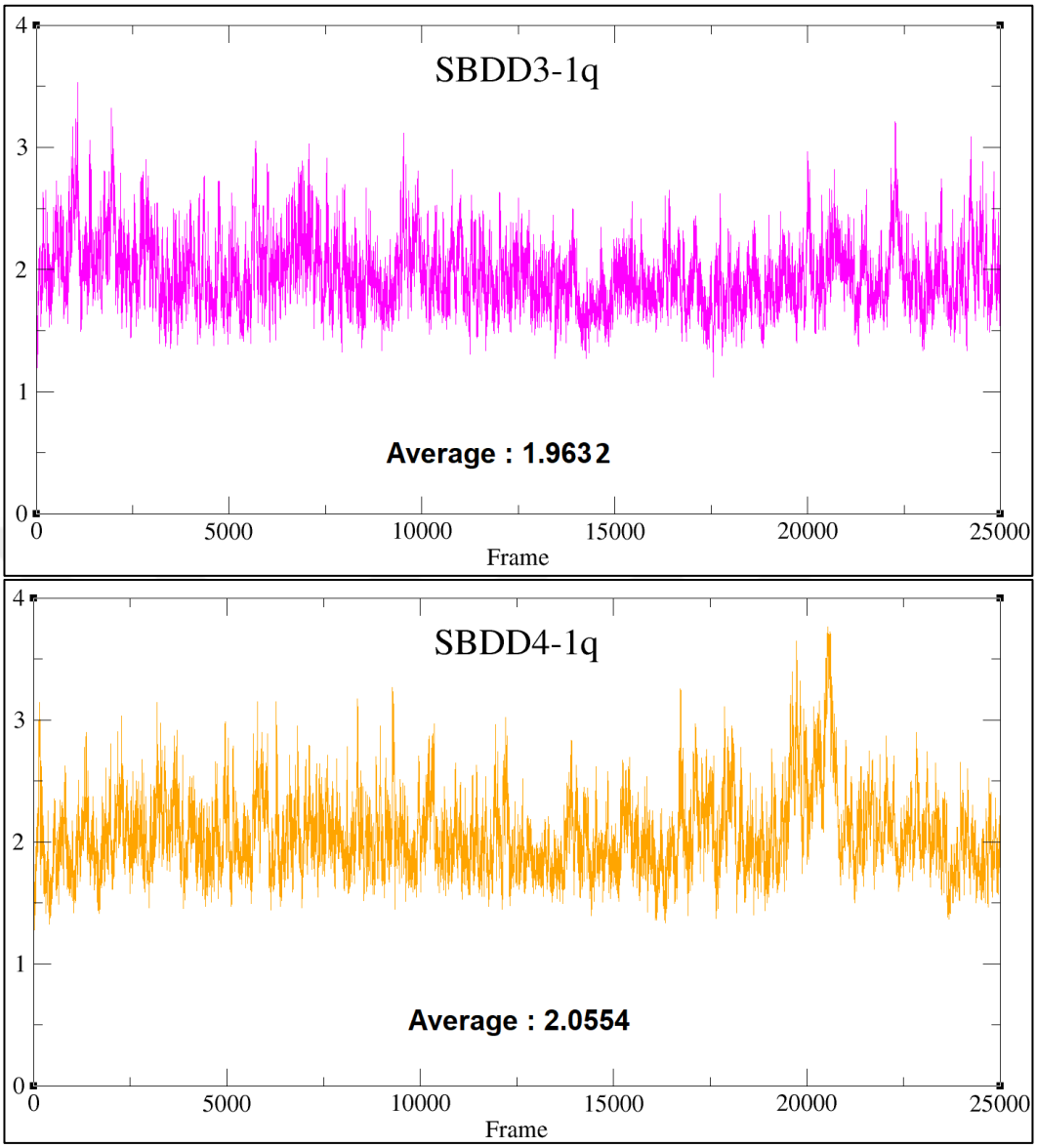


Figure 4.22. The root mean square deviation (RMSD) for all simulations.

Hydrogen bonds with a duration of 100 ns and a total of 25,000 frames were examined in table 4.13. The hydrogen bonds in the binding between DNA and compounds containing only platinum and 2-pyrimidinethiol are listed in the table. Hydrogen bonds in four groups of bonding were divided into two categories as polar and total. Polar hydrogen bonds were classified as bonds formed by the most electronegative atoms (S, O, N), and all hydrogen bonds were classified as all polar and nonpolar hydrogen bonds between ligands and DNA. Also, bond types have been added to the table to include donor and acceptor groups. LIG1 stands for 2-pyrimidinethiol in the complex.

Table 4.13. Occurrence of hydrogen bonds in molecular dynamics simulations and properties of favorite hydrogen bonds.

Complex	Average Number of Polar H-Bonds	Average Number of All H-Bonds	Number of Polar H-Bond Types	Favorite Polar H-Bond Interaction	Occupancy (%)
SBDD1-1q	0.8957	11.8712	40	Donor_LIG1:N2H-----OP1:DC9_Acceptor	33.22%
				Donor_LIG1:SH-----O5':DG10_Acceptor	20.13%
SBDD2-1f	6.2232	19.7257	38	Donor_LIG1:N1H-----O2:DT20_Acceptor	94.06%
				Donor_LIG1:N1H-----O2:DT6_Acceptor	87.01%
				Donor_LIG1:N2H-----O2:DT20_Acceptor	83.85%
				Donor_LIG1:N2H-----O2:DT6_Acceptor	80.92%
				Donor_LIG1:SH-----O3':DT8_Acceptor	20.98%
				Donor_LIG1:SH-----OP1:DC9_Acceptor	10.36%
SBDD3-1q	9.0931	23.4838	28	Donor_LIG1:N1H-----O2:DT18_Acceptor	98.97%
				Donor_LIG1:SH-----N3:DA17_Acceptor	93.48%
				Donor_LIG1:N1H-----O2:DT8_Acceptor	89.68%
				Donor_LIG1:N1H-----O4':DC9_Acceptor	74.47%
				Donor_LIG1:SH-----O2:DC9_Acceptor	73.05%
				Donor_DG16:N2H-----S:LIG1_Acceptor	53.74%
				Donor_LIG1:SH-----O4':DT18_Acceptor	53.27%
				Donor_LIG1:SH-----N2:DG16_Acceptor	26.62%
SBDD4-1q	0.8445	16.6940	13	Donor_LIG1:N1H-----OP1:DC9_Acceptor	47.63%
				Donor_LIG1:N1H-----O3:DT8_Acceptor	12.06%

In Table 4.13, the polar hydrogen bond mean in SBDD4-1q is only 0.8445 during the entire simulation period. This mean, together with the SBDD1-1q group, is a very low value. There are 13 types of polar hydrogen bonds in SBDD4-1q. This value is very low as a percentage of incidence. The favorite hydrogen bond is with the hydrogen of the platinum-bonded nitrogen atom in LIG1. The most important conclusion here is that the polar SH group in SBDD4-1q is far from the DNA phosphate group, reducing the percentage of hydrogen bonding between this group and DNA. At the same time, the nonpolar hydrogen bond average is higher than SBDD1-1q, but lower than other groups. Although the percentage of hydrogen bonding between the platinum-bound donor nitrogen group in LIG1 and the oxygen in the DNA phosphate backbone in SBDD1-1q was lower than SBDD4-1q, the expected hydrogen bond between donor SH and acceptor oxygen was formed here as 12%. Although 40 types of polar hydrogen bonds were observed in SBDD-1q, the percentage of these bonds remained low throughout the simulation. The hydrogen bond population ratios between the SBDD3-1q complex and DNA are interesting. The average of both polar and non-polar hydrogen bonds is highest in this group. The hydrogen bond formation of the nitrogen group on the guanine base of DNA as the donor with the sulfur in the 2-pyrimidinethiol is as high as 53.74%.

This is interesting in terms of binding quality. In addition, the SH group has a high incidence of polar hydrogen bonding with DNA as a donor. The interaction of the SH group with the nitrogen atoms in the DNA bases shows that SBDD3-1q is sufficiently channeled into the DNA groove. The same situation can be said for the SBDD2-1f group. In SBDD2-1f, hydrogen bonds in which polar atoms in LIG1 are donors are high as in other groups. It can be said that SBDD2-1f and SBDD3-1q groups, which have more close rings combined and more pi-anion interactions, hold a better place in the DNA groove. It emerges from the analyzes that lignans directed into DNA by hydrophobic and pi-pi interactions converge with phosphate groups and nucleobases in DNA and are strongly held by polar hydrogen bonds.

Remarkably more efficient, endpoint free energy methods ignore the details of the coupling pathway and estimate free energy only on a collection of snapshots representing the unbound and bound states. These snapshots can be created with an MD simulation. Molecular mechanics generalized Born surface area (MM-GBSA) is among the most popular free energy calculation methods. It is possible to decompose the total free energy into subcomponents with MM-GBSA and measure their contributions separately. The binding energies of SBDD1-1q, SBDD2-1f, SBDD3-1q and SBDD4-12 at various times were calculated with the MM-GBSA.py module in AMBER 20 and the list of their energy components was transferred to table 4.14. In this table, van der Waals (VDW), electrostatic energy (EEL), the electrostatic contribution to the solvation free energy (E_{GB}), surface energy (E_{SURF}), gas phase free energy (ΔG_{Gas}), solvation free energy (ΔG_{Solv}), binding energy ($\Delta G_{Bind(GBSA)}$), quasi-harmonic entropy approximation binding energy ($\Delta G_{Bind(Entropy)}$), total vibrational entropy (ΔS_{Vib}), and total entropy (ΔS_{Tot}). The entropy contribution is simply added to the calculations. In Figure 4.23, there is a thermodynamic cycle for calculating the binding energy classically. Total free energies are calculated separately for the ligand and the receptor in the gas phase and in the solvent. The final binding energy is obtained by summing the total solvation and gas phase free energies with each other.

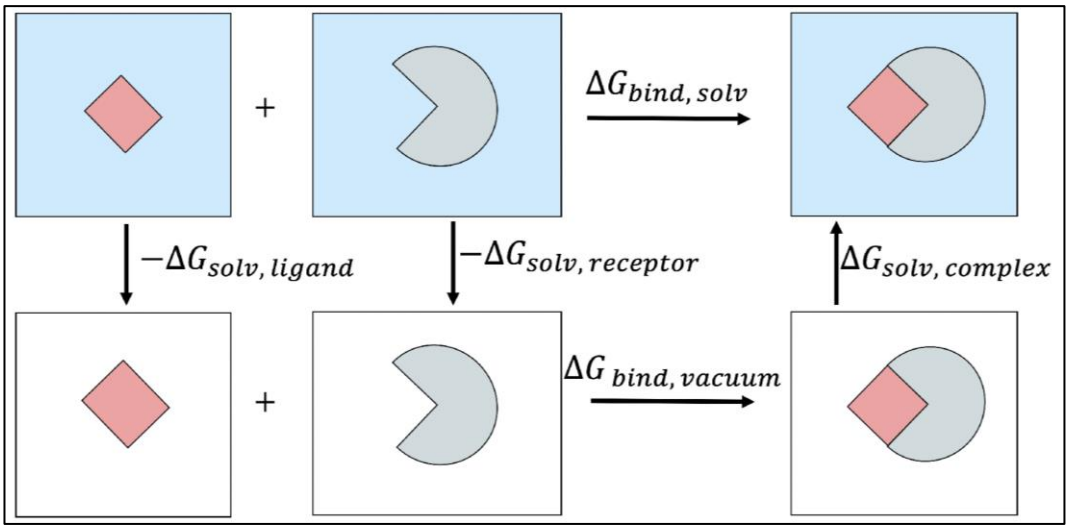


Figure 4.23. The system of thermodynamic used to calculate the binding free energy of a platinum complex-DNA in the solvent.

The first terms are vdW and EEL in table 4.14. These refer to the intermolecular interactions between the ligand and the target DNA. Average intermolecular interactions of SBDD3-1q at 100ns are stronger than other groups. Stronger interactions were detected between this complex and DNA, especially in the last 50 ns of the simulation. Looking at the docking results and hydrogen bonds analysis, it is seen that there are many polar and nonpolar hydrogen bonds with pi-anion interactions explaining these strong interactions. The group with the lowest van der Waals interactions is SBDD2-1f. This is evident when looking at the docking results. Because strong "pi" and hydrophobic interactions are less in this group. The presence of frequent rings in SBDD3-1q, the electronegative oxygen atom in the ring, and the "1q" functional group strengthen the intermolecular interactions. In direct proportion to the hydrogen bond analysis, the electrostatic interaction energies of SBDD4-1q and SBDD1-1q are lower than the other groups. Average intermolecular interactions of SBDD3-1q at 100ns are stronger than other groups. Stronger interactions were detected between this complex and DNA, especially in the last 50 ns of the simulation. Looking at the docking results and hydrogen bonds analysis, it is seen that there are many polar and nonpolar hydrogen bonds with pi-anion interactions explaining these strong interactions. The group with the lowest van der Waals interactions is SBDD2-1f. This is evident when looking at the docking results. Because strong "pi" and hydrophobic interactions are less in this group. The presence of frequent rings in SBDD3-1q, the electronegative oxygen atom in the ring, and the "1q" functional group strengthen

the intermolecular interactions. In addition, van der Waals bonds were formed by the presence of hydrophobic interaction. In direct proportion to the hydrogen bond analysis, the electrostatic interaction energies of SBDD4-1q and SBDD1-1q are lower than the other groups.

Table 4.14. Division of binding energy parameters calculated using MMGBSA by complex groups. F: First, L: Last.

NO	Parameters	F_1ns	F_ 5ns	F_10ns	F_25ns	F_50ns	L_50ns	L_25ns	L_10ns	100ns
SBDD1-1q	VDW	-45.75	-43.08	-46.44	-48.40	-49.08	-48.38	-48.38	-50.49	-48.67
	EEL	-1034.38	-1026.03	-1037.52	-1052.22	-1056.59	-1059.33	-1060.34	-1063.33	-1057.17
	E_{GB}	1054.29	1044.18	1056.93	1071.86	1076.50	1078.72	1080.00	1084.57	1076.84
	E_{SURF}	-3.66	-3.40	-3.67	-3.80	-3.84	-3.83	-3.86	-3.92	-3.82
	ΔG_{Gas}	-1080.13	-1069.11	-1083.96	-1100.63	-1105.67	-1107.71	-1108.72	-1113.83	-1105.84
	ΔG_{Solv}	1050.63	1040.79	1053.26	1068.07	1072.66	1074.88	1076.14	1080.65	1073.02
	$\Delta G_{Bind(GBSA)}$	-29.50	-28.32	-30.70	-32.56	-33.02	-32.83	-32.58	-33.18	-32.81
	$\Delta G_{Bind(Entropy}$	6.86	19.18	26.56	28.55	24.95	25.52	28.77	25.48	14.07
	ΔS_{Vib}	-11.40	-22.54	-32.32	-36.18	-33.04	-33.41	-36.42	-33.73	-21.95
	ΔS_{Tot}	-36.36	-47.50	-57.26	-61.11	-57.97	-58.34	-61.35	-58.66	-46.88
SBDD2-1f	VDW	-40.24	-39.89	-40.92	-42.33	-42.46	-42.56	-42.20	-41.51	-42.36
	EEL	-1138.85	-1133.96	-1136.45	-1147.09	-1142.66	-1142.16	-1135.60	-1134.07	-1142.42
	E_{GB}	1146.83	1141.25	1144.07	1153.65	1149.55	1148.75	1142.62	1140.52	1148.96
	E_{SURF}	-3.18	-3.16	-3.12	-3.13	-3.14	-3.17	-3.15	-3.19	-3.15
	ΔG_{Gas}	-1179.09	-1173.85	-1177.37	-1189.42	-1185.12	-1184.72	-1177.80	-1175.58	-1184.78
	ΔG_{Solv}	1143.65	1138.08	1140.95	1150.52	1146.40	1145.59	1139.47	1137.33	1145.81
	$\Delta G_{Bind(GBSA)}$	-35.43	-35.77	-36.42	-38.90	-38.71	-39.14	-38.32	-38.25	-38.97
	$\Delta G_{Bind(Entropy}$	-2.29	4.10	8.32	7.71	5.66	5.13	8.14	6.91	-1.31
	ΔS_{Vib}	-9.37	-16.09	-20.96	-22.85	-20.61	-20.50	-22.69	-21.39	-13.89
	ΔS_{Tot}	-33.14	-39.86	-44.74	-46.62	-44.37	-44.26	-46.46	-45.16	-37.66
SBDD3-1q	VDW	-53.23	-48.89	-48.91	-49.25	-50.93	-53.88	-51.53	-48.25	-52.21
	EEL	-1136.91	-1136.15	-1137.81	-1135.54	-1135.83	-1164.19	-1156.18	-1145.90	-1149.33
	E_{GB}	1150.25	1146.05	1147.17	1145.45	1147.29	1175.31	1166.35	1154.72	1160.46
	E_{SURF}	-3.83	-3.64	-3.63	-3.65	-3.73	-3.79	-3.74	-3.68	-3.77
	ΔG_{Gas}	-1190.14	-1185.04	-1186.71	-1184.80	-1186.76	-1218.07	-1207.71	-1194.15	-1201.54
	ΔG_{Solv}	1146.41	1142.41	1143.53	1141.80	1143.56	1171.52	1162.60	1151.04	1156.69
	$\Delta G_{Bind(GBSA)}$	-43.73	-42.63	-43.18	-43.00	-43.20	-46.55	-45.11	-43.11	-44.85
	$\Delta G_{Bind(Entropy}$	-8.53	1.66	8.57	11.81	9.46	6.97	10.17	9.15	0.09
	ΔS_{Vib}	-10.59	-19.69	-27.15	-30.20	-28.05	-28.90	-30.67	-27.65	-20.33
	ΔS_{Tot}	-35.20	-44.29	-51.76	-54.81	-52.66	-53.52	-55.28	-52.26	-44.94
SBDD4-1q	VDW	-49.40	-47.79	-47.60	-47.82	-47.81	-43.15	-42.51	-43.83	-45.56
	EEL	-1103.35	-1102.44	-1096.72	-1098.58	-1101.73	-1096.96	-1096.62	-1108.52	-1100.38
	E_{GB}	1122.99	1121.04	1114.94	1117.02	1120.55	1112.80	1112.62	1125.40	1117.83
	E_{SURF}	-3.63	-3.70	-3.70	-3.75	-3.78	-3.64	-3.63	-3.67	-3.72
	ΔG_{Gas}	-1152.75	-1150.24	-1144.32	-1146.40	-1149.53	-1140.11	-1139.13	-1152.34	-1145.95
	ΔG_{Solv}	1119.36	1117.34	1111.24	1113.27	1116.77	1109.16	1108.99	1121.73	1114.11
	$\Delta G_{Bind(GBSA)}$	-33.38	-32.90	-33.08	-33.14	-32.76	-30.94	-30.14	-30.61	-31.83
	$\Delta G_{Bind(Entropy}$	3.17	15.75	26.00	30.94	28.43	36.15	37.56	28.47	22.05
	ΔS_{Vib}	-11.94	-24.04	-34.47	-39.46	-36.58	-42.49	-43.08	-34.46	-29.28
	ΔS_{Tot}	-36.56	-48.65	-59.08	-64.08	-61.19	-67.09	-67.69	-59.08	-53.89

Despite entropy contribution, it is seen that SBDD3-1q (-44.85 kcal/mol for 100ns) has the best binding. This binding increased slightly in the last 50 ns. The lowest binding energy is at SBDD4-1q (-31.83 kcal/mol for 100ns). A parallel MM-GBSA binding energy values emerged with the mentioned hydrogen bond and van der Waals interactions. Also, the SBDD2-1f group has the best binding affinity (-38.97 kcal/mol for 100ns) after SBDD3-1q, as expected. The high total entropy value (-53.89 kcal/mol for 100ns) in SBDD4-1q decreased the entropy-doped binding affinity compared to the others (more positive binding energy-22.05 kcal/mol for 100 ns). The entropy in here can be explained by the flexibility movement of the "SH" group away from the DNA groove. SBDD3-1q and SBDD2-1f, which have relatively low entropy contribution, have stronger entropy-doped binding affinities. While the change in entropies in the first 1ns was less, the average entropy values increased during the simulation. In particular, the high entropy values in SBDD1-1q and SBDD4-1q indicate that conformational changes are greater than in the other two groups. This prevents the masking of the ligand to some extent. SBDD2-1f is the group with the least irregular movement with an average entropy value of -37.66 kcal/mol in 100ns. It also has an entropy contribution binding energy value of -1.31 kcal/mol.

In figure 4.24, the values of the MM-GBSA binding energies between the platinum-containing complexes and DNA without the entropy contribution in all frames and the frequency distribution graphs of these binding energies are shown. As can be seen from these graphs, SBDD3-1q, the group with the best binding affinity, shows stronger binding energy around -45 kcal/mol throughout the simulation. In addition, SBDD1-1q and SBDD-4q groups, which have lower binding strength, peaks around -30 kcal/mol.

DNA helical parameters were generated with the AMBER 20 Cpptraj module to analyze angular changes in nucleobases at binding sites. Figure 4.24 shows the visual forms of these helical parameters. The parameters examined for angular changes are expressed in the figure 4.25.

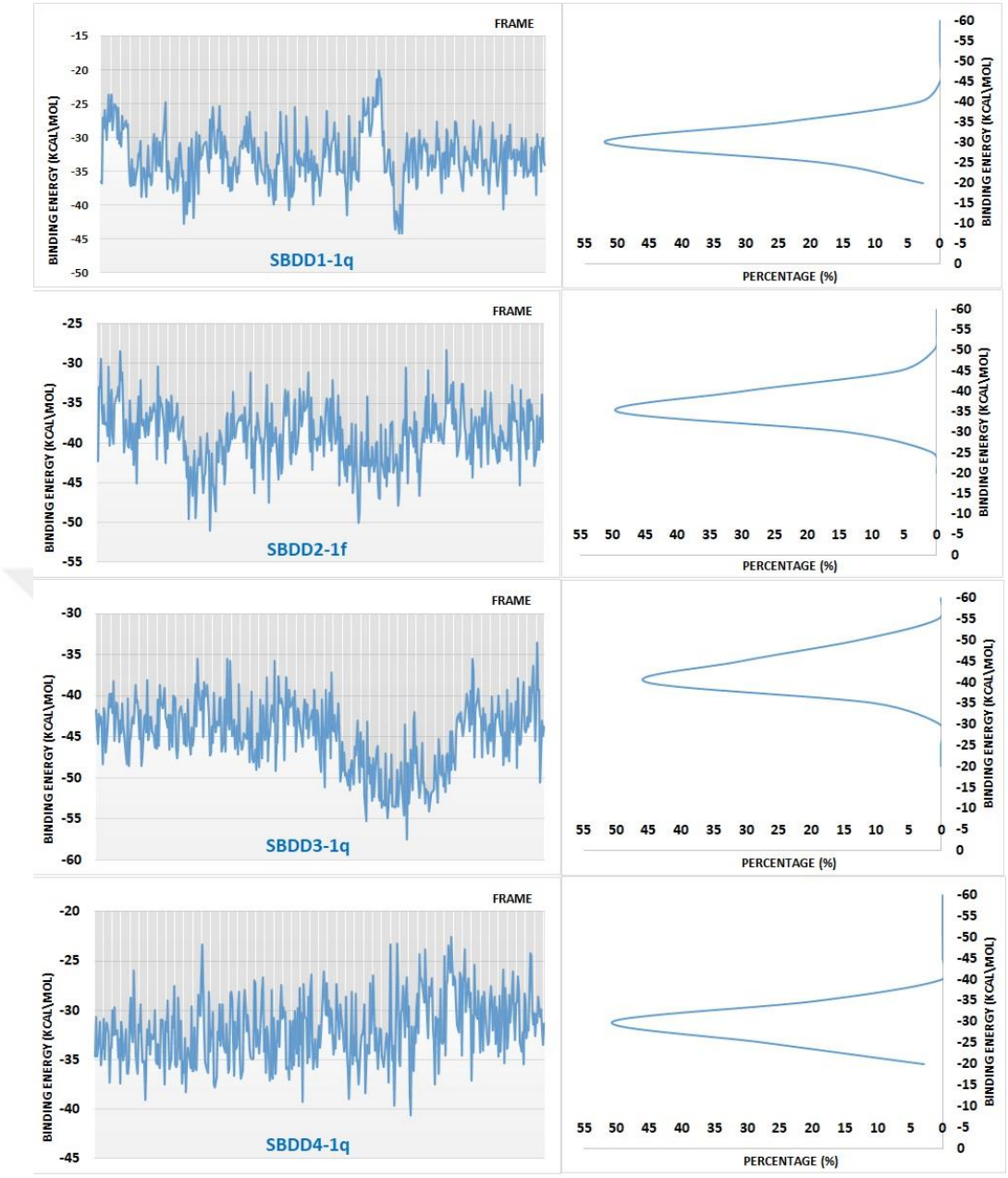


Figure 4.24. MM-GBSA binding energy (kcal/mol) and integrated distributions.

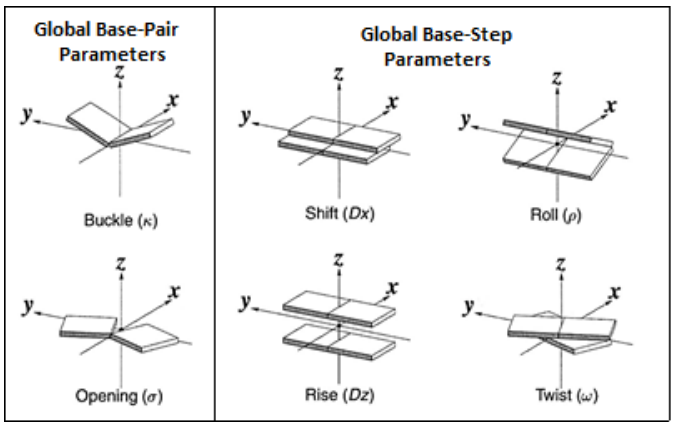


Figure 4.25. DNA helical parameters.

The frequency distributions of the DNA global base-pair parameters and global base-step parameters have been analyzed for 1DNE, SBDD1-1q, SBDD2-1f, SBDD3-1q, and SBDD4-1q and are shown in figure 4.26 and 4.27, respectively.

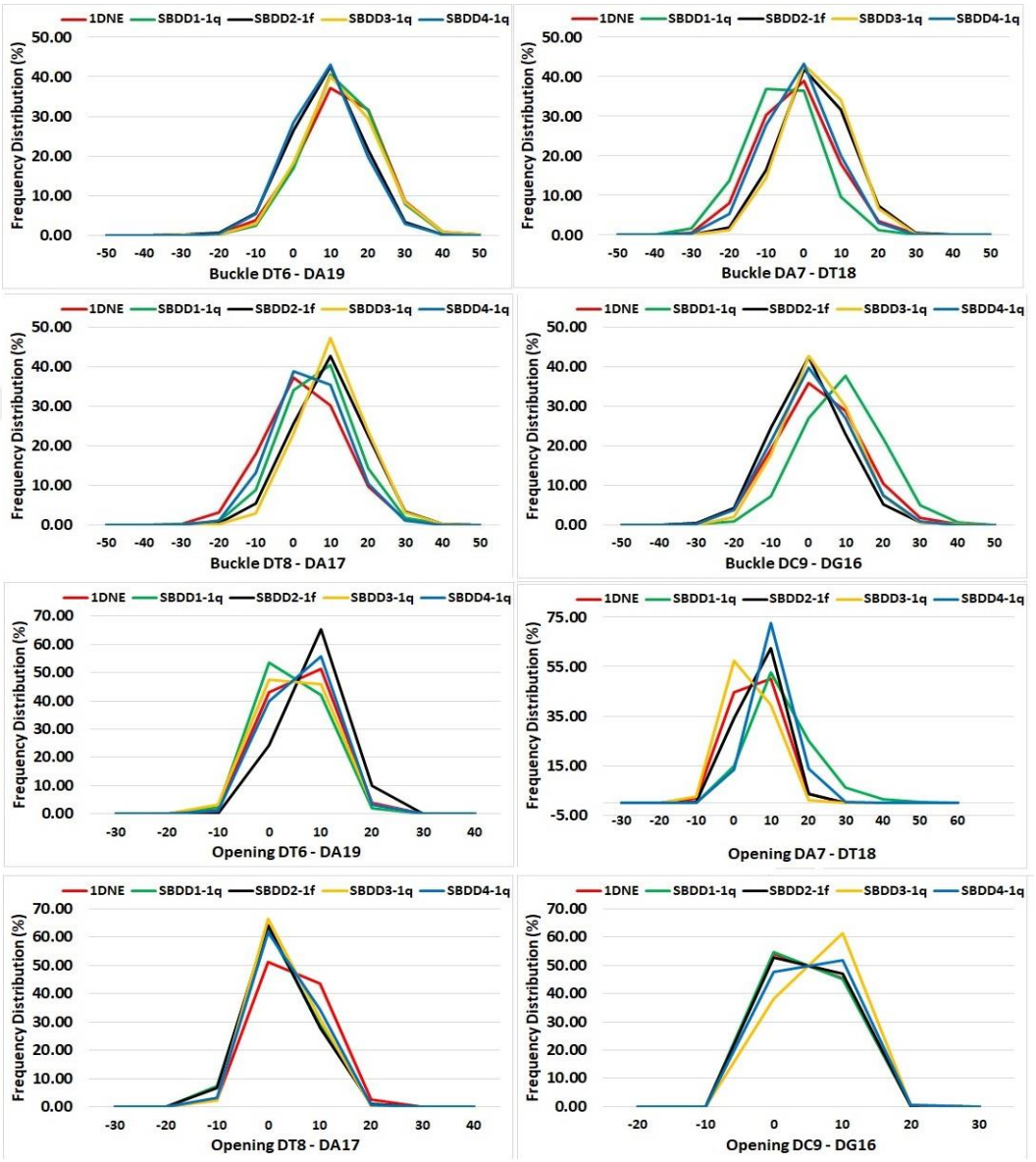


Figure 4.26. Selected frequency distributions of DNA base-pair for the central binding base pairs of all simulated complex.

For the buckle angle values in figure 4.26, the deviation of all groups from 1DNE is negligible in the DT6:DA19 base pair. However, in the DA7:DT18 base pair, only the SBDD4-1q group does not deviate from 1DNE. Other groups have a deviation of ~10°. In the DT8:DA17 base pair, 1DNE has two strong peaks, while SBDD2-1f and SBDD3-1q have distributions only at ~10°. SBDD1-1q shows a

different distribution by ~10° than other groups and 1DNE in DC9:DG16 base pair. When the opening angle base par parameters are controlled, in base pair DT6:DA19, all complexes and 1DNE have two strong peaks at ~0° and ~10°. Only the peak of SBDD2-1f at ~10° has a greater percentage of dispersion. The DA7:DT18 base pair has peaks at ~0° and ~10°. While these peaks have close distribution in 1DNE, those of the other complexes except SBDD3-1q have a high percentage of distribution at ~10°. In DC9:DG16, on the other hand, the SBDD3-1q group has a strong peak at ~10°, unlike the others.

In figure 4.27, where the base step parameters are evaluated, it is understood from the graph that the deviation of the rise value from 1DNE occurs only in the adjacent bases of DA7-DT8. This deviation is approximately 1Å in the SBDD1-1q and SBDD2-1f groups. When the roll angle values are examined, 1DNE and other groups distributed at ~10° in DT6-DA7 steps where SBDD1-1q is ~5°. In the DA7-DT8 base step, SBDD1-1q peaks at ~10°, while other groups and 1DNE show a distribution at 0°. In the DT8-DC9 base partnership, there are two peaks in the SBDD3-1q complex, ~10° and ~5°. Other groups and 1DNE indicate only distribution at ~10°. The deviations of the shift values are different from 1DNE in DT6-DA7. While the peak distribution in the 1DNE structure is between ~-0.5Å and ~1Å, this distribution is spread over a wide area between ~-1Å and ~2Å in SBDD3-1q. The spread occurred in SBDD4-1q is between ~2Å and ~3.5Å. SBDD2-1f is like 1DNE. In the DA7-DT8 base step, SBDD3-1q and SBDD2-1f with 1DNE similarly peaks at ~0Å, while other groups form a distribution at ~-0.5Å. Here, SBDD2-1f also has a peak at ~-0.5Å. In DT8-DC9, 1DNE has two large peaks at ~1Å and ~2Å, while SBDD1-1q is located at ~2Å from these peaks. Others peak around ~1Å. The deviations in the twist angle values; in DT6-DA7, 1DNE has two large peaks at ~40° and ~50°, while SBDD1-1q is located at ~50° from these peaks. Other groups indicate distributions around ~40°. In DA7-DT8, 1DNE consists of peaks at ~30° and ~40°, while the peak of SBDD1-1q is placed at ~30° from these peaks. Other groups have distributions around ~40°. In DT8-DC9, all groups except SBDD3-1q peak at ~40° like 1DNE. The SBDD3-1q group has a strong peak at ~50°. The SBDD2-1f complex has a stronger peak of ~50° apart from ~40°.

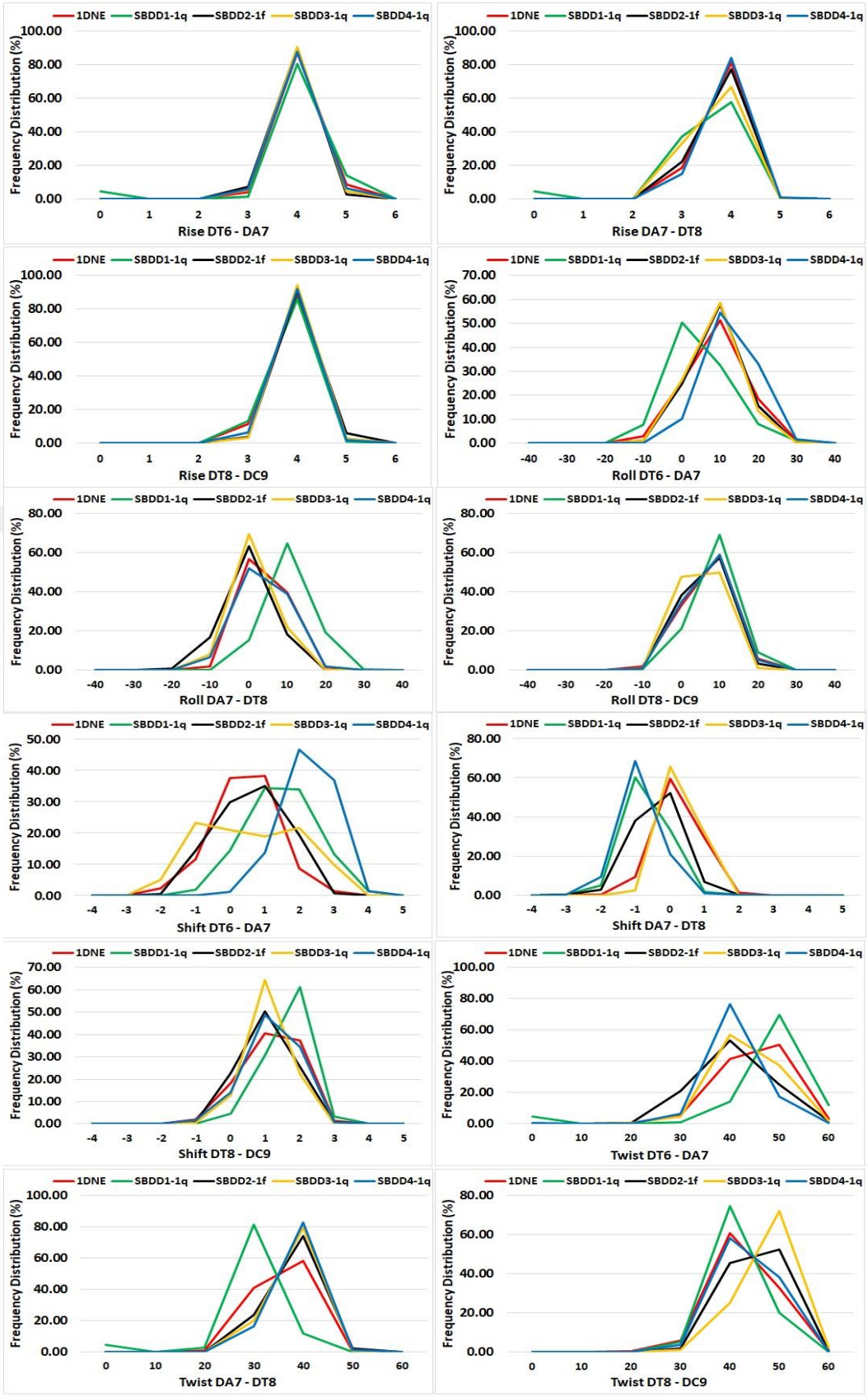


Figure 4.27. Selected frequency distributions of DNA base step for the central binding base steps of all simulated complex.

As a result of detailed examination of DNA helical base parameters, it was revealed how intermolecular interactions affect ligand-bound DNA differently from pure DNA (1DNE). The deviations in the base parameters especially reveal the effect of hydrogen bonds.

In this study, the possible interaction between target DNA and platinum-containing compounds was examined in all its lines. It has been determined from the literature that metal-containing compounds that are square planar and without leaving groups can settle in the groove. Novel square planar platinum compounds were designed with ligands and 2-pyrimidinethiol, which were determined as carriers in the literature. Structures in which platinum is bound to four nitrogen atoms were revealed. To add to these structures, 21 functional groups were determined from literature. These groups are most used and have high biological effect. The most stable states of the geometric structures of the complexes were found with OLYP DFT method and the selecting basis set combination. 84 compounds prepared with some scripts were automatically docked into the DNA groove using AutoDock 4.2.6. Virtual screening was performed for compounds containing platinum metal. The best poses were found and prepared for simulation. Molecular dynamics simulations were performed with AMBER20, one of the most used programs in the literature. Molecular dynamic simulations cannot be made automatically for metal-containing complexes. Because the bond structure of transition metal atoms is complex. Many different bonding forms occur in metal atoms with d-orbitals and f-orbitals. Existing programs cannot automatically predict the properties of bonds in metal-containing compounds. For this, it is necessary to establish the bond property parameters and force fields of these compounds one by one. In this study, the best complexes of the four groups taken from the docking were successfully transferred to the simulations with the MCPB.py plugin in AMBER 20. After the simulations were run at 100 ns, the strength of the bonds was evaluated by multiple assays. Images of simulations between ligand and DNA at 0, 50 and 100 nanoseconds are in Appendix 5. Although the MM-GBSA method used for binding does not provide accurate information about the binding affinity, it is sufficient to explain the binding differences and interactions between the complexes. In addition, studies that could take days have been concluded in a short time with the GPU-based computation used.

5. CONCLUSION

Increasing diseases and human deaths in recent years have pushed scientists to find quick and permanent solutions for treatment. The classical drug development method is insufficient to solve the current disease-related problems. Because this method is expensive, time-consuming and has the potential to fail as a result. Therefore, rational drug design method has started to be discussed in the last 30 years. This method has become popular with the development of technology and the fact that increasing experimental information can be easily retrieved from databases. Biological problems that are difficult to overcome have begun to be solved gradually with the effective use of computers with many useful programs. Bioinformatics has become a frequently applied field especially for the development of new treatments and drugs specific to cancer types. Platinum-derived compounds, which are the subject of this study, are widely used in chemotherapy as drugs against cancer types. Studies are still ongoing to develop these platinum-derived drugs.

1. The limitations and advantages of the ORCA 4.2 program in metal-containing complexes were evaluated in the main part of this study. The keywords and calculation algorithms of the ORCA 4.2 program were compared. The RIJCOSX approximation and high grid integral numbers showed good results in the calculations of heavy metal-containing complexes. The success of ZORA relativistic full electron basis set and combinations of metal ECP (only for electrons participating in bonding) basis set in complexes was tested together with 139 DFT methods for different categories. Suitable method and basis set combination was determined statistically by various graphs and deviation analyzes. The selection of the best DFT method using the ZORA-def2-TZVP relativistic full electron basis set was used in the ligand-based drug design part. In particular, the obvious success of DFT functionals based on the GGA algorithm stands out. HCTH407P functional was chosen for all complexes to be used in ligand-based drug design. Also, the mentioned comparisons were made for platinum and 2-pyrimidinethiol complexes used in structure-based drug design. Differently, metal ECP basis set combinations were

compared here as well as DFT methods. As a result of the analyzes, it was determined that while the quadruple-zeta basis set combinations had good results in bond lengths containing metal, triple-zeta basis sets were more successful in the angles where the metal was in the center. As a result of the analysis, OLYP was chosen as the best DFT method, HayWadt for ECP, LanL2TZ for metal basis set, and def2-TZVP for other atoms in structure-based drug design. Because of the use of quantum descriptors that include all electrons in the calculation for QSAR calculations, the basis set, such as ZORA, which causes longer calculations in terms of time, has been used in ligand-based drug design. To obtain more realistic results, ECP basis sets that freeze the electrons that do not participate in the bond and can make calculations in a shorter time are not selected in ligand-based drug design. It has been a detailed study that can be applied to researchers using metal-containing compounds. These comparisons are invaluable to researchers who use metal-containing compounds.

2. The QSAR studies for metal-containing compounds are scarce in the literature. Using OECD principles, which are less referenced in other studies, QSAR studies were performed and estimated cytotoxic activity values were discovered for newly designed drug candidate compounds. The insufficient database of experimental cytotoxic test data for platinum-containing compounds, and the inability to filter the databases for disease-based studies appear to be a problem. If cytotoxic data are to be obtained from different studies in the literature review, it should be noted whether the experimental procedure used is similar. Considering these details, a total of 154 complexes in 7 groups were taken from the literature. Geometric optimizations were carried out successfully with the specified DFT method. Around 1400 descriptors have been calculated. There are deficiencies in the programs that calculate descriptors in QSAR studies. It has been determined that these programs need to calculate the descriptors based on the information in the output files of the structure with this study. Standard descriptors do not fully detail the structure. Therefore, it would be interesting to produce a machine learning supported descriptive computing software that can automatically extract the desired data from the output files of many quantum chemical computing programs. Many descriptors detected

using the optimized 3D geometric structure have been successful in generating QSAR models. QSAR models were created for 7 groups and the best models were statistically selected using the OECD's 4 principles. In the QSAR studies of metal-containing compounds, we identified one of the shortcomings was that selections were not made by applying the OECD's four rules. One of the most important parts in model selection in OECD is external validation. This is a vital verification method to accurately predict the activity of compounds of unknown activity. We acted transparently to validate the models we created for each group of complexes in our study. These models were applied to 30 newly formed platinum and 2-pyrimidinethiol-containing complexes in each group and estimated low cytotoxic values were obtained for 11 complexes. These successful complexes are of interest for further in-vitro studies. In addition, it has been tried to express with all points how QSAR studies of metal-containing compounds can be done more accurately with this study.

3. Studies in which platinum-containing compounds are covalently bound with DNA are generally predominant in the literature. Changes in structure and the nature of this binding were tried to be understood in these studies of molecular dynamics. Groove binding modeling between DNA and platinum-containing compounds is scarce in the literature. DNA intercalation of platinum-containing compounds is not yet fully understood. It is thought that this bonding has a weaker and less toxic effect than covalent bonding. Molecular docking calculations are commonly used for many metal or metal-free structures. Organic compounds can be docked into the target macromolecule and then simulated with virtual screening at the same time, usually by means of ready-made programs and machine learning, which is one of the new techniques, and many studies on this subject are available in the literature. However, no study could be identified in which metal-containing compounds were subjected to docking by virtual screening. This stands out as a major shortcoming. The interactions of these compounds with macromolecules cannot be easily done by virtual screening method. Due to the nature of the bond structure of the transition metal, it is not possible to automatically participate in docking and simulations of metal-containing compounds in existing programs. As a continuation of

docking, there is no program yet to quickly implement this method in molecular dynamics simulations for metal containing complexes. Today, these calculations are made separately and transferred to the simulation. For the analysis of molecular dynamics simulations, apart from the binding site, the effect of intermolecular interactions on binding should also be examined. For these reasons, groove binding of platinum-containing complexes with DNA was modeled in the last part of the study. 21 different biologically active functional groups were used for 84 complexes designed in DNA intercalation. After geometric optimization of these complexes, molecular docking calculations that find DNA binding sites were performed for these complexes. We succeeded to implement the virtual screening method for molecular docking. The binding site and intermolecular interactions between four metal-containing compounds and DNA were investigated in this study. One of the functional groups, "1q" is rich in methyl groups and contains electronegative oxygen atoms, and it was determined that the complexes containing these functional groups had better binding energy. In addition, because of molecular dynamics simulations, it was found that the SBDD3-1q complex is better located and tightly bound in the DNA groove. It was revealed that this complex has more and stronger electrostatic interactions with DNA. As a result, it is obvious that especially machine learning-based programs should be developed to make the interactions of metal-containing compounds with macromolecules easily and quickly. In the current situation, it is necessary to shorten and automate the preparation time for simulations. All aspects of the results should be evaluated, and the scoring of the binding should be done with correct statistical methods in addition to the preparation time for properties of metal containing complexes.

4. We have shown the effect of metal-containing compounds in new drug discovery with the theoretical studies and detailed analyzes in all these parts. We believe that all the arguments used in this study have shed light on future research.

6. REFERENCES

"Vancouver citation system was used in this thesis."

1. Boyle P, Levin B. International agency for research on cancer. World cancer report, 2008.

2. Hanahan D, Weinberg RA.The Hallmarks of Cancer CellVolume 100, 2000;7:57-70.

3. Peyrone M. Über die einwirkung des ammoniaks auf platinchlorid, Annalen derChemie und Pharmacie und Pharmacie 1844; LI, 1-29.

4. Jamieson ER, Lippard SJ. Structure, recognition, and processing of cisplatin-DNA adducts. Chem Rev, 1999;99:2467-2498.

5. Rosenberg B, Van-Camp L, Krigas T. İnhibition of cell division in Esherichia Coli by electrolysis product from a platinum electrode. Nature, 1965; 205: 698-699.

6. Wang D, Lippard SJ. Celluler Processing of Platinum Anticanser Drugs. Nat. Rev. Drug. Dis-cov. 2005 Apr;4(4):307-20.

7. Amable L. Cisplatin resistance and opportinunities for precision medicine. Pharm Res, 2016;106:27-36.

8. Mehmood RK. Review of Cisplatin and oxaliplatin in current immunogenic and monoclonal antibody treatments. Oncol Rev. 2014 Sep 23;8(2):256.

9. Kintzel PE. Anticanser drug-induced kidnet disorders. Drug Safety, 2001; 24:19-38.

10. Sandman KE, Lippard SJ. Cisplatin, VHCA/Wiley–VCH, Zurich/Weinheim, 1999, p. 523.

11. Barnard CFJ, Raynaud FI, Kelland LR. Metallopharmaceuticals I, DNA Interactions, Springer, Berlin, 1999, p. 45.

12. van der Vijgh VJF. Clin. Pharmacokinet. 21 (1991) 242.

13. Greene MH. Is Cisplatin a Human Carcinogen. J. Nat. Cancer Inst 1992;84:306–312.

14. Travis LB, Curtis RE, Storm H, Hall P, Holowaty E, Van Leeuwen FE, Kohler BA, Pukkala E, Lynch CF, Andersson M, Bergfeldt K, Clarke EA, Wiklund T, Stoter G, Gospodarowicz M, Sturgeon J, Fraumeni JF Jr. Boice JD Jr. Risk of second malignant neoplasms among long-term survivors of testicular cancer. J. Natl. Cancer Inst 1997;89:1429–1439.

15. Silva MJ, Costa P, Dias A, Valente M, Louro H, Boavida MG. Comparative analysis of the mutagenic activity of oxaliplatin and cisplatin in the Hprt gene of CHO cells. Environ. Mol. Mutagen 2005;46:104–115.

16. Bassett E, King NM, Bryant MF, Hector S, Pendyala L, Chaney SG, Cordeiro-Stone M. The role of DNA polymerase η in translesion synthesis past platinum-DNA adducts in human fibroblasts. Cancer Res 2004;64:6469–6475.

17. Reedijk J. Platinum anticancer coordination compounds: study of DNA binding inspires new drug design. European Journal of Inorganic Chemistry, 1303-1312, 2009.

18. Gutierrez-Valero MD, Romero-Molina M, Lopez-GarzSn R, Salas-Peregri'n JM. Study of the complex formation between 4, 6-diamine-1, 2-dihydro-2-thiopyrimidine and cobalt(II), cobalt(III), nickel(II) and copper(II) ions. Trans. Met. Chem., 1988,13, 451.

19. Holy A, Votruba I, Jost K. Metal complexes of substituted pyrimidines. FEBS Lett., 1972, 22, 287.

20. Nowak M, Rostkowska H, Lapinski L, Leszczynski J, Kwiatkowskil J, Spec. Acta, 1991, 47A,339.

21. Kapetanovic IM. Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. Chem-Biol. Interact., 2008, 171(2), 165-176.

22. Shaikh SA, Jain T, Sandhu G, Latha N, Jayaram B. From drug target to leads-sketching a physicochemical pathway for lead molecule design in silico. Curr. Pharm. Design, 2007, 13(34), 3454-3470.

23. Adamo C. Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. J Chem Phys 1998, 108, 664.

24. (a) Burke K, Perdew JP, Wang Y. In Electronic Density Functional Theory: Recent Progress and New Directions; Dobson JF, Vignale G, Das MP, Eds.; Plenum: New York, 1998; (b) Perdew JP. Burke K, Wang Y. Phys Rev B 1996, 54, 16533; (c) Perdew JP, Wang Y. Phys Rev B 1992, 45, 13244. Gill, P. M.W. Mol Phys 1996, 89, 433.

25. Lee C, Yang W, Parr RG. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. Phys Rev B 1988, 37, 785.

26. (a) Becke ADJ. Chem Phys 1993, 98, 5648; (b) Becke ADJ. Chem Phys 1996, 104, 1040.

27. Parr RG, Yang W. Density Functional Theory of Atoms and Molecules. Oxford Univ. Press: Oxford, 1989.

28. Niu S, Hall MB. DFT Study of H−H Activation by Cp2 LnH d0 Complexes. Chem Rev 2000, 100, 353.

29. (a) Siegbahn PEM, Crabtree RHJ, Am Chem Soc 1999, 121, 117; (b) Siegbahn PEM. Crabtree RH. J Am Chem Soc 1997, 119, 3103; (c) Siegbahn PEM, Crabtree RHJ. Am Chem Soc 1996, 118, 4442; (d) Siegbahn PEM. Inorg Chem 1999, 38, 2880; (e) Siegbahn PEM, Blomberg MRA. Chem Rev 2000, 100, 421.

30. Salahub DR, Castro M, Fournier R, Calaminici P, Godbout N, Goursot A, Jamorski C, Kobayashi H, Martinez A, Papai I, Proynov E, Russo N, Sirois S, Ushio J, Vela A. In Theoretical and Computational Approaches to Interface Phenomena. Sellers, H.; Olab, J., Eds.; Plenum: New York, 1995, p. 187.

31. Musaev DG, Morokuma K. In Advances in Chemical Physics. Rice, S. A.; Prigogine, I., Eds.; John Wiley & Sons: New York, 1996, p. 61, vol. XCV.

32. Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S. QSARINS: a new software for the development, analysis and validation of QSAR MLR models. J. Comput. Chem., 34 (2013), pp. 2121-2132.

33. Julia Lorenzo, Ángel M. Montana. The molecular shape and the field similarities as criteria to interpret SAR studies for fragment-based design of platinum(IV) anticancer agents. Correlation of physicochemical properties with cytotoxicity. Journal of Molecular Graphics and Modelling, Volume 69, September 2016, Pages 39-60.

34. Igor V. Tetko, Hristo P. Varbanov, Markus Galanski, Mona Talmaciu, James A. Platts, Mauro Ravera, Elisabetta Gabano. Prediction of logP for Pt(II) and Pt(IV) complexes: Comparison of statistical and quantum-chemistry based approaches. Journal of Inorganic Biochemistry 156 (2016) 1–13.

35. Pubalee Sarmah, Ramesh C. Deka. DFT-based QSAR and QSPR models of several cis-platinum complexes: solvent effect. J Comput Aided Mol Des (2009) 23:343–354.

36. Athanassios C. Tsipis and Ioannis N. Karapetsas. ^{195}Pt NMR parameters as strong descriptors in one-parameter QSAR models for platinum based antitumor compounds. Magn. Reson. Chem. (2017).

37. Giuseppe Ermondi, Giulia Caron, Mauro Ravera, Elisabetta Gabano, Sabrina Bianco, James A. Platts and Domenico Osella. Molecular interaction fields vs. quantum-mechanical based descriptors in the modelling of lipophilicity of platinum(IV) complexes. Dalton Trans., 2012, 00, 1–8.

38. Hristo P. Varbanov, Michael A. Jakupec, Alexander Roller, Frank Jensen, Markus Galanski, and Bernhard K. Keppler. Theoretical Investigations and Density Functional Theory Based Quantitative Structure−Activity Relationships Model for Novel Cytotoxic Platinum(IV) Complexes. J. Med. Chem. 2013, 56, 330−344.

39. Ferreira LG, dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. Molecules. 2015;20:13384–13421.

40. Anderson AC. The process of structure-based drug design. Chem. Biol. 2003;10:787–797.

41. Kalyaanamoorthy S, Chen YP. Structure-based drug design to augment hit discovery. Drug Discov. Today. 2011;16:831–839.

42. Yocheved Gilad and Hanoch Senderowitz. Docking Studies on DNA Intercalators. J. Chem. Inf. Model. 2014, 54, 96−107.

43. Ceyda Icsel, Veysel T. Yilmaz, Yunus Kaya, Hale Samli, William TA. Harrison and Orhan Buyukgungor. New palladium(II) and platinum(II) 5,5-diethylbarbiturate complexes with 2-phenylpyridine, 2,2'-bipyridine and 2,2'-dipyridylamine: synthesis, structures, DNA binding, molecular docking, cellular uptake, antioxidant activity and cytotoxicity. Dalton Trans. 2015 Apr 21;44(15):6880-95.

44. Ishani Mitra, Subhajit Mukherjee, Venkata P. Reddy, Subrata Dasgupta, Jagadeesh C. Bose K, Sandip Mukherjee, Wolfgang Linert and Sankar Ch. Moi. Benzimidazole based Pt(II) complexes with better normal cell viability than cisplatin: synthesis, substitution behavior, cytotoxicity, DNA binding and DFT study. RSC Adv., 2016, 6, 76600–76613.

45. Jianhui Zhu,Yongmei Zhao, Yanyan Zhu, Ziyi Wu, Miaoxin Lin, Weijiang He, Yan Wang, Guangju Chen, Lei Dong, Junfeng Zhang, Yi Lu, and Zijian Guo. DNA Cross-Linking Patterns Induced by an Antitumor-Active Trinuclear Platinum Complex and Comparison with Its Dinuclear Analogue. Chem. Eur. J. 2009, 15, 5245 – 5253.

46. Hongwei Yue, Bo Yang, Yan Wang and Guangju Chen; Investigations of the Binding of $[Pt_2(DTBPA)Cl_2](II)$ and $[Pt_2(TPXA)Cl_2](II)$ to DNA via Various Cross-Linking Modes. Int. J. Mol. Sci. 2013, 14, 19556-19586.

47. Shanshan Cui, Yan Wang, Guangju Chen. Disturbance of DNA conformation by the binding of testosterone-based platinum drugs via groove-face and intercalative interactions: a molecular dynamics simulation study. BMC Struct Biol. 2013 Mar 22;13:4.

48. Nathália MP Rosa, Júlio AF Arvellos, Luiz Antônio S Costa. Molecular dynamics simulation of non-covalent interactions between polynuclear platinum(II) complexes and DNA. JBIC Journal of Biological Inorganic Chemistry 25, pages 963–978 (2020).

49. Daniele Veclani, Marilena Tolazzi, José P. Cerón-Carrasco, Andrea Melchior. Intercalation Ability of Novel Monofunctional Platinum Anticancer Drugs: A Key Step in Their Biological Action. J. Chem. Inf. Model. 2021, 61, 9, 4391–4399.

50. Anjali Soni, Pooja Khurana, Tanya Singh, B Jayaram. A DNA intercalation methodology for an efficient prediction of ligand binding pose and energetics. Bioinformatics, Volume 33, Issue 10, 15 May 2017, Pages 1488–1496.

51. Intini FP, Boccarelli A, Francia VC et al. Platinum complexes with imino ethers or cyclic ligands mimicking imino ethers: synthesis, in vitro antitumour activity, and DNA interaction properties. J Biol Inorg Chem 9, 768–780 (2004).

52. Angelina Boccarelli, Francesco P Intini, Rossella Sasanelli, Maria F Sivo, Mauro Coluccia, Giovanni Natile. Synthesis and in Vitro Antitumor Activity of Platinum Acetonimine Complexes. J. Med. Chem. 2006, 49, 2, 829–837.

53. Jordi de Mier-Vinué, Marina Gay, Ángel M Montaña, Rosa-Isabel Sáez, Virtudes Moreno, Jana Kasparkova, Oldrich Vrana, Pavla Heringova, Viktor Brabec, Angela Boccarelli, Mauro Coluccia, Giovanni Natile. Synthesis, Biophysical Studies, and Antiproliferative Activity of Platinum (II) Complexes Having 1,2-Bis(aminomethyl)carbobicyclic Ligands. Journal of Medicinal Chemistry 2008 51 (3), 424-431.

54. Jana Hildebrandt, Norman Häfner, Helmar Görls, Daniel Kritsch, Giarita Ferraro, Matthias Dürst, Ingo B Runnebaum, Antonello Merlino, Wolfgang Weigand. Platinum(ii) O,S complexes as potential metallodrugs against Cisplatin resistance. Dalton Trans., 2016,45, 18876-18891.

55. Carla Francisco, Sofia Gama, Filipa Mendes, Fernanda Marques, Isabel Cordeiro dos Santos, António Paulo, Isabel Santos, Joana Coimbra, Elisabetta Gabano, Mauro Ravera. Pt(II) complexes with bidentate and tridentate pyrazolyl containing chelators: synthesis, structural characterization and biological studies. Dalton Trans., 2011,40, 5781-5792.

56. Katia G Samper, Sierra C Marker, Pau Bayón, Samantha N MacMillan, Ivan Keresztes, Òscar Palacios, Justin J Wilson. Anticancer activity of hydroxy- and sulfonamide-azobenzene platinum(II) complexes in cisplatin-resistant ovarian cancer cells. Journal of Inorganic Biochemistry, Volume 174, September 2017, Pages 102-110.

57. Mauro Ravera, Elisabetta Gabano, Manuele Sardi, Giuseppe Ermondi, Giulia Caron, Michael J McGlinchey, Helge Müller-Bunz, Elena Monti, Marzia B. Gariboldi, Domenico Osella. Synthesis, characterization, structure, molecular modeling studies and biological activity of sterically crowded Pt(II) complexes containing bis(imidazole) ligands. Journal of Inorganic Biochemistry, Volume 105, Issue 3, 2011, Pages 400-409.

58. Dimitra Kovala-Demertzi, Paras Nath Yadav, Mavroudis A Demertzis, Mauro Coluccia. Synthesis, crystal structure, spectral properties and cytotoxic activity of platinum(II) complexes of 2-acetyl pyridine and pyridine-2-carbaldehyde N(4)-ethyl-thiosemicarbazones. Journal of Inorganic Biochemistry, Volume 78, Issue 4, 2000, Pages 347-354.

59. Gordon Lowe, Anne Sophie Droz, Tirayut Vilaivan, George W Weaver, Jenny J Park, Jonathan M Pratt, Lindsay Tweedale, Lloyd R Kelland. Cytotoxicity of 2,2‘:6‘,2‘ ‘-Terpyridineplatinum(II) Complexes against Human Ovarian Carcinoma. Journal of Medicinal Chemistry 1999 42 (16), 3167-3174.

60. Ana Zamora, Sergio A. Pérez, Venancio Rodríguez, Christoph Janiak, Gorakh S. Yellol, and José Ruiz. Dual Antitumor and Antiangiogenic Activity of Organoplatinum(II) Complexes. Journal of Medicinal Chemistry 2015 58 (3), 1320-1336.

61. Elena Monti, Marzia Gariboldi, Alessandro Maiocchi, Emilio Marengo, Claudio Cassino, Elisabetta Gabano, Domenico Osella. Cytotoxicity of cis-Platinum(II) Conjugate Models. The Effect of Chelating Arms and Leaving Groups on Cytotoxicity: A Quantitative Structure-Activity Relationship Approach. Journal of Medicinal Chemistry, 2005, 48 (3), 857-866.

62. Geng Xu, Zheng Yan, Nan Wang, Zhanzhu Liu. Synthesis and cytotoxicity of cis-dichloroplatinum (II) complexes of (1S,3S)-1,2,3,4-tetrahydroisoquinolines. European Journal of Medicinal Chemistry, Volume 46, Issue 1, 2011, Pages 356-363.

63. Sangeeta Roy Chaudhuri, Goran N Kaluđerović, Martin Bette, Jürgen Schmidt, Harry Schmidt, Reinhard Paschke, Dirk Steinborn. Synthesis, characterization and cytotoxicity studies of platinum(II) complexes with amino acid ligands in various coordination modes. Inorganica Chimica Acta, Volume 394, 2013, Pages 472-480.

64. Till Walther, Renate Herzog, Milena R. Kaluđerović, Christoph Wagner, Harry Schmidt, Goran N. Kaluđerović. Traceable platinum(II) complexes with alkylene diamine-derived ligands: synthesis, characterization and in vitro studies. Journal of Coordination Chemistry, 71:2, 2018, 243-257.

65. Matesanz AI, Albacete P, Souza P. Synthesis, and characterization of a new bioactive mono(thiosemicarbazone) ligand,based on 3,5-diacetyl-1,2,4-triazol diketone and its palladium and platinum complexes. Polyhedron, Volume 109, 2016, Pages 161-165.

66. Ana I Matesanz, Inés Leitao, Pilar Souza. Palladium(II) and platinum(II) bis(thiosemicarbazone) complexes of the 2,6-diacetylpyridine series with high cytotoxic activity in cisplatin resistant A2780cisR tumor cells and reduced toxicity. Journal of Inorganic Biochemistry, Volume 125, 2013, Pages 26-31.

67. Ana I Matesanz, Pilar Souza. Novel cyclopalladated and coordination palladium and platinum complexes derived from α-diphenyl ethanedione bis(thiosemicarbazones): Structural studies and cytotoxic activity against human A2780 and A2780cisR carcinoma cell lines. Journal of Inorganic Biochemistry, Volume 101, Issue 10, 2007, Pages 1354-1361.

68. Ana I Matesanz, Sandra Tapia, Pilar Souza. First 3,5-diacetyl-1,2,4-triazol derived mono(thiosemicarbazone) and its palladium and platinum complexes: Synthesis, structure, and biological properties. Inorganica Chimica Acta, Volume 445, 2016, Pages 62-69.

69. Ana I Matesanz, Carolina Hernández, Pilar Souza. New bioactive 2,6-diacetylpyridine bis(p-chlorophenylthiosemicarbazone) ligand and its Pd(II) and Pt(II) complexes: Synthesis, characterization, cytotoxic activity and DNA binding ability. Journal of Inorganic Biochemistry, Volume 138, 2014, Pages 16-23.

70. Haiyan Yu, Shaohua Gou, Zhimei Wang, Feihong Chen, Lei Fang. Toward overcoming cisplatin resistance via sterically hindered platinum(II) complexes. European Journal of Medicinal Chemistry, Volume 114, 2016, Pages 141-152.

71. Shamsuddin S, Ikuo Takahashi, Zahid H Siddik, Abdul R Khokhar. Synthesis, characterization, and antitumor activity of a series of novel cisplatin analogs with cis-1,4-diaminocyclohexane as nonleaving ammine group. Journal of Inorganic Biochemistry, Volume 61, Issue 4, 1996, Pages 291-301.

72. Irene Würtenberger, Bernhard Angermaier, Brigitte Kircher, Ronald Gust. Synthesis and In Vitro Pharmacological Behavior of Platinum(II) Complexes Containing 1,2-Diamino-1-(4-fluorophenyl)-2-alkanol Ligands. Journal of Medicinal Chemistry, 2013, 56 (20), 7951-7964.

73. Carolin Mügge, Claudia Rothenburger, Antje Beyer, Helmar Görls, Chiara Gabbiani, Angela Casini, Elena Michelucci, Ida Landini, Stefania Nobili, Enrico Mini, Luigi Messori, Wolfgang Weigand. Structure, solution chemistry, antiproliferative actions and protein binding properties of non-conventional platinum(II) compounds with sulfur and phosphorus donors. Dalton Trans., 2011,40, 2006-2016.

74. Marina Gay, Ángel M Montaña, Consuelo Batalla, Juan M Mesas, María-Teresa Alegre. Design, synthesis and SAR studies of novel 1,2-bis(aminomethyl)cyclohexane platinum(II) complexes with cytotoxic activity. Studies of interaction with DNA of iodinated seven-membered 1,4-diaminoplatinocycles. Journal of Inorganic Biochemistry, Volume 142, 2015, Pages 15-27.

75. Shakked Z, Guerstein-Guzikevich G, Eisenstein M, Frolow F, Rabinovich D. The conformation of the DNA double helix in the crystal is dependent on its environment. Nature, 1989, 342, 456-60.

76. Neidle S. Oxford Handbook of Nucleic Acid Structure; Oxford University Press: New York, NY, 1999; p. 8.

77. G. Cervantes, M. J. Prieto, and V. Moreno. AntitumorActivity ofa Pt(III) Derivative of2-pyrimidinethiol. Metal-Based Drugs, Vol. 4, No. 1, 1997, 9-18.

78. Nour T Abdel Ghani, Ahmed M Mansour. Novel Pd(II) and Pt(II) complexes of N,N-donor benzimidazole ligand: Synthesis, spectral, electrochemical, DFT studies and evaluation of biological activity. Inorganica Chimica Acta, 373, (2011) 249–258.

79. Malik M, Michalska D. Assessment of new DFT methods for predicting vibrational spectra and structure of cisplatin: Which density functional should we choose for studying platinum(II) complexes? Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 125, 2014, 431-439.

80. Louise M Debefve, Christopher J Pollock. Systematic assessment of DFT methods for geometry optimization of mononuclear platinum-containing complexes. Phys. Chem. Chem. Phys., 2021, Advance Article.

81. Hristo P. Varbanov, Michael A. Jakupec, Alexander Roller, Frank Jensen, Markus Galanski, and Bernhard K. Keppler. Theoretical Investigations and Density Functional Theory Based Quantitative Structure–Activity Relationships Model for Novel Cytotoxic Platinum(IV) Complexes. Journal of Medicinal Chemistry 2013 56 (1), 330-344.

82. Hemant Kumar Srivastava, Mukesh Chourasia, Devesh Kumar, G Narahari Sastry. Comparison of Computational Methods to Model DNA Minor Groove Binders. J. Chem. Inf. Model., 2011, 51, 558–571.

83. Frank Neese, Frank Wennmohs, Ute Becker, Christoph Riplinger. The ORCA quantum chemistry program package. J. Chem. Phys., 152, 224108 (2020).

84. Neese F. The ORCA program system. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2012, Vol. 2, Issue 1, Pages 73-78.

85. Neese, F. Software update: the ORCA program system, version 4.0. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2017, Vol. 8, Issue 1, p. e1327.

86. Becke AD. A multicenter numerical integration scheme for polyatomic molecules. J. Chem. Phys., 1988, 88, 2547.

87. Treutler O, Ahlrichs RJ. Efficient molecular numerical integration schemes. J. Chem. Phys., 1994, 102, 346.

88. Boerrigter PT, Te Velde G, Baerends JE. Three dimensional numerical integration for electronic structure calculations. International Journal of Quantum Chemistry, 1988, 33(2), 87-113.

89. Baerends EJ, Ellis DE, Ros P. Self-consistent molecular Hartree Fock Slater calculations I. The computational procedure. Chem. Phys., 1973, 2, 41.

90. Dunlap BI, Connolly JWD, Sabin JR. On some approximations in applications of Xα. theory. J. Chem. Phys., 1979, 71, 3396.

91. Van Alsenoy C. Ab initio calculations on large molecules: The multiplicative integral approximation. J. Comput. Chem., 1988, 9, 620.

92. Kendall RA, Früchtl HA. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. Theor. Chem. Acc., 1997, 97, 158.

93. Eichkorn K, Treutler O, Ohm H, Häser M. Ahlrichs R. Auxiliary basis sets to approximate Coulomb potentials. Chem. Phys. Lett., 1995, 240, 283.

94. Eichkorn K, Weigend F, Treutler O, Ahlrichs R. Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. Theor. Chem. Acc., 1997, 97, 119.

95. Whitten JL. Coulombic potential energy integrals and approximations. J. Chem. Phys., 1973, 58, 4496.

96. Qing-Kun Wang, Shao-Ping Pu, Yan-Wei Cong, Yong-Nian Li, Chun-Fang Luan. cis-Diammine(glycolato-κ^2O1,O2)platinum(II). Acta Cryst., 2009, E65, m1687.

97. Kovala-Demertzi D, Demertzis MA, Miller JR, Papadopoulou C, Dodorou C, Filousis G. Journal of Inorganic Biochemistry, 2001, 86, 555.

98. van Lenthe E, Ehlers AE, Baerends EJ. Geometry optimization in the Zero Order Regular Approximation for relativistic effects. Journal of Chemical Physics, 110, 8943, (1999).

99. Visscher L, van Lenthe E. On the distinction between scalar and spin-orbit relativistic effects. Chemical Physics Letters, 306, 357 (1999.

100. van Lenthe E, Baerends EJ, Snijders JG. Relativistic regular two-component Hamiltonians. Journal of Chemical Physics 99, 4597 (1993).

101. van Lenthe E, Baerends EJ, Snijders JG. Relativistic total energy using regular approximations. Journal of Chemical Physics 101, 9783 (1994).

102. van Lenthe E, Baerends EJ, Snijders JG. The zero-order regular approximation for relativistic effects: The effect of spin-orbit coupling in closed shell molecules. Journal of Chemical Physics 105, 6505 (1996).

103. van Lenthe E, van Leeuwen R, Baerends EJ, Snijders JG. Relativistic regular two-component Hamiltonians. International Journal of Quantum Chemistry 57, 281 (1996).

104. Dolg M, Stoll H, Savin A, Preuss H. Energy-adjusted pseudopotentials for the rare earth elements. Theor. Chim. Acta, 75, 173 (1989).

105. Dolg M, Stoll H, Preuss H. A combination of quasirelativistic pseudopotential and ligand field calculations for lanthanoid compounds. Theor. Chim. Acta 85, 441 (1993).

106. Tian Lu and Feiwu Chen. Multiwfn: A multifunctional wavefunction analyzer. J. Comput. Chem., 33, 580-592 (2012).

107. Chun Wei Yap. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. Volume 32, Issue7, 2011, 1466-1474.

108. Gramatica P, Chirico N, Papa E, Kovarich S, Cassani S. QSARINS: A New Software for the Development, Analysis, and Validation of QSAR MLR Models. Journal of Computational Chemistry, Software news and updates, 2013, 34, 2121-2132.

109. Gramatica P, Cassani S, Chirico N. QSARINS-Chem: Insubria Datasets and New QSAR/QSPR Models for Environmental Pollutants in QSARINS. Journal of Computational Chemistry, Software news and updates, 2014, 35, 1036–1044.

110. Andrae D, Häußermann U, Dolg M, Stoll H, Preuß H. Energy-adjusted ab initio pseudopotentials for the second and third row transition elements. Theor. Chim. Acta 77, 123-141 (1990).

111. Weigend F, Ahlrichs R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. Phys. Chem. Chem. Phys. 7, 3297 (2005).

112. Martin JM, Sundermann A. Correlation consistent valence basis sets for use with the Stuttgart-Dresden-Bonn relativistic effective core potentials: The atoms Ga-Kr and In-Xe. J. Chem. Phys. 114, 3408-3420 (2001).

113. Hay P, Wadt WR. Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. J. Chem. Phys. 82, 299-310 (1985).

114. Stevens WJ, Krauss M, Basch H, Jasien, PG. Relativistic compact effective potentials and efficient, shared- exponent basis sets for the third-, fourth-, and fifth-row atoms. Can. J. Chem. 70, 612-630 (1992).

115. Sayantan B, Nihar R, Dev PA. An overview of recent advances in duplex DNA recognition by small molecules. Beilstein J. Org. Chem. 2018, 14, 1051–1086.

116. Peter Ertl. An algorithm to identify functional groups in organic molecules. Journal of Cheminformatics volume 9, Article number: 36 (2017).

117. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry, 30(16), 2009, 2785–2791.

118. Case DA, et al. Amber 2021. University of California Press, 2021

119. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package.WIREs Comput. Mol. Sci., 2013, 3, 198-210.

120. Case DA, et al. The Amber biomolecular simulation programs. J. Computat. Chem., 2005, 26, 1668-1688.

121. Cheatham TE, and Case DA.Twenty-five years of nucleic acid simulations. Biopolymers, 2013, 99, 969-977.

122. Pengfei Li, Kenneth M Merz Jr. MCPB.py: A Python Based Metal Center Parameter Builder. J. Chem. Inf. Model. 2016, 56, 4, 599–604.

123. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee TA. Point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem. 2003, 24, 1999–2012.

124. Lee, MC, Duan Y. Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. Proteins Struct. Funct. Bioinf. 2004, 55, 620–634.

125. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M. Refinement of the AMBER force field for nucleic acids: Improving the description of [alpha]/[gamma] conformers. Biophys. J. 2007, 92, 3817–3829.

126. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J. Comput. Chem. 2004, 25, 1157–1174.

127. Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem. 1992, 13, 952–962.

128. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. Journal of chemical information and modeling, 2011, 51(1), 69-82.

129. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods: II. The accuracy of ranking poses generated from docking. Journal of computational chemistry, 2011, 32(5), 866.

130. Xu L, Sun H, Li Y, Wang J, Hou T. Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models. The Journal of Physical Chemistry B, 2013, 117(28), 8408-8421.

131. Sun H, Li Y, Tian S, Xu L, Hou T. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. Physical Chemistry Chemical Physics, 2014, 16(31), 16719-16729.

132. Athanassios C Tsipis, Ioannis N Karapetsas. ^{195} Pt NMR parameters as strong descriptors in one-parameter QSAR models for platinum-based antitumor compounds. Magn Reson Chem. 2017 Jul;55(7):662-669.

133. Golbraikh A, Tropsha A. Beware of q^2!. J. Mol. Graph. Model. 2002, 20, 269.

134. Chirico N, Gramatica P. Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient. J. Chem. Inf. Model. 2011, 51, 2320.

135. Chirico N, Gramatica P. Real External Predictivity of QSAR Models. Part 2. New Intercomparable Thresholds for Different Validation Criteria and the Need for Scatter Plot Inspection. J. Chem. Inf. Model. 2012, 52, 2044.

136. Gramatica P. External Evaluation of QSAR Models, in Addition to Cross-Validation Verification of Predictive Capability on Totally New Chemicals. Mol. Inform. 2014, 33, 311–314.

137. Gramatica P. On the development and validation of QSAR models. Methods Mol. Biol. 2013, 930, 499–526.

138. Atkinson, AC. Plots, transformations, and regression; an introduction to graphical methods of diagnostic regression analysis. No. 04; QA278. 2, A8. 1985.

7. APPENDICES

APPENDIX 1 Using Input for “Compound” in ORCA 4.2 During Calculations

```
* xyz 0 1
Pt      -0.168684      0.623983      1.844388
-

*
%Compound
New_Step
# A1-03
! TightOpt RKS RIJCOSX TIGHTSCF GRID7 FINALGRID7 SARC/J PRINTBASIS ZORA-def2-TZVP ZORA GridX9 AIM
ALLPOP
%pal nprocs 128 end
%base "A1-03"
%scf
  MaxIntMem 5000
end
%geom
  MaxIter 125
end
%method
  method dft
  functional gga_xc_hcth_407p
  WeightScheme Weight_Becke
  SpecialGridAtoms 78
  SpecialGridIntAcc 10
end
%basis
  newgto Pt "SARC-ZORA-TZVP" end
  DelECP Pt
  DelECP Cl
  DelECP O
  DelECP N
  DelECP C
  DelECP H
end
* xyz 0 1
Pt      -0.168684      0.623983      1.844388
-
*
Step_end
New_Step
# A1-04
! TightOpt RKS RIJCOSX TIGHTSCF GRID7 FINALGRID7 SARC/J PRINTBASIS ZORA-def2-TZVP ZORA GridX9 AIM
ALLPOP
%pal nprocs 128 end
%base "A1-04"
%scf
  MaxIntMem 5000
end
%geom
  MaxIter 125
end
%method
  method dft
  functional gga_xc_hcth_407p
  WeightScheme Weight_Becke
  SpecialGridAtoms 78
  SpecialGridIntAcc 10
end
%basis
  newgto Pt "SARC-ZORA-TZVP" end
  DelECP Pt
  DelECP Cl
  DelECP O
  DelECP N
  DelECP C
  DelECP H
end
* xyz 0 1
Pt      -0.961065450      1.551545850      0.871934820
-
*
Step_end
END
```

APPENDIX 2 An Example of a Script Run in Tübitak TRUBA.

It is stated that 16 computers with 128 cores are used in total on “mid1” high computing server and the linux version of the ORCA program is defined in the hyazici user directory. The openmpi directory in TRUBA for parallel computations is also shown.

```
#!/bin/bash
#SBATCH -p mid1
#SBATCH -A hyazici
#SBATCH -n 128
#SBATCH -N 16
#SBATCH --time=3-24:00:00
#SBATCH -o slurm.%j.out
#SBATCH -e slurm.%j.err

export PATH=/truba/sw/centos7.3/lib/openmpi/3.0.0-gcc-7.0.1/bin:$PATH
export LD_LIBRARY_PATH=/truba/sw/centos7.3/lib/openmpi/3.0.0-gcc-7.0.1/lib:
$LD_LIBRARY_PATH
export orcadir=/truba/home/hyazici/orca_4_2_0_linux_x86-64_openmpi314
export RSH_COMMAND="/usr/bin/ssh -x"
export PATH=$orcadir:$PATH
export LD_LIBRARY_PATH=$orcadir:$LD_LIBRARY_PATH

echo "SLURM_NTASKS= $SLURM_NTASKS"
echo "SLURM_TASKS_PER_NODE= $SLURM_TASKS_PER_NODE"
echo "SLURM_NNODES= $SLURM_NNODES"
echo "SLURM_NODELIST= $SLURM_NODELIST"
echo "SLURM_CPU_BIND= $SLURM_CPU_BIND"
echo "SLURM_JOB_CPUS_PER_NODE= $SLURM_JOB_CPUS_PER_NODE"

$orcadir/orca GroupA-1.inp > GroupA-1.out
```

APPENDIX 3 A Multiwfn Script Example for Quantum Mechanics Calculation.

Each of the numbers to the left of the script corresponds to a command and calculation in Multiwfn 3.7. For this script "txt" file, which is run with the "bat" file, the "molden" wavefunction input from ORCA and belonging to the complex is required.

```
100   # Other functions (Part 1)
4     # 1-Integrate a function - Electron density
1     # Electron density
4     # 2-Integrate a function - Gradient norm of electron density
2     # Gradient norm of electron density
4     # 3-Integrate a function - Laplacian of electron density
3     # Laplacian of electron density
4     # 4-Integrate a function - Hamiltonian kinetic energy density K(r)
6     # Hamiltonian kinetic energy density K(r)
4     # 5-Integrate a function - Lagrangian kinetic energy density G(r)
7     # Lagrangian kinetic energy density G(r)
4     # 6-Integrate a function - Electrostatic potential from nuclear charges
8     # Electrostatic potential from nuclear charges
4     # 7-Integrate a function - Electron Localization Function (ELF)
9     # Electron Localization Function (ELF)
4     # 8-Integrate a function - Localized orbital locator (LOL)
10    # Localized orbital locator (LOL)
4     # 9-Integrate a function - Local information entropy
11    # Local information entropy
4     # 10-Integrate a function - Reduced density gradient (RDG)
13    # Reduced density gradient (RDG)
4     # 11-Integrate a function - RDG with promolecular approximation
14    # RDG with promolecular approximation
4     # 12-Integrate a function - Sign(lambda2)*rho
15    # Sign(lambda2)*rho
4     # 13-Integrate a function - Sign(lambda2)*rho with promolecular approximation
16    # Sign(lambda2)*rho with promolecular approximation
4     # 14-Integrate a function - Correlation hole for alpha, ref. point:0.0 0.0 0.0
17    # Correlation hole for alpha, ref. point:0.0 0.0 0.0
4     # 15-Integrate a function - Average local ionization energy (ALIE)
18    # Average local ionization energy (ALIE)
4     # 16-Integrate a function - Source function, mode: 1, ref. point:0.0 0.0 0.0
19    # Source function, mode: 1, ref. point:0.0 0.0 0.0
4     # 17-Integrate a function - Delta_g (promol. approx.) 0.0 0.0
22    # Delta_g (promol. approx.) 0.0 0.0
4     # 18-Integrate a function - Delta_g (Hirshfeld partition)
23    # Delta_g (Hirshfeld partition)
0     # Return to main menu
q     # Quit
```

APPENDIX 4 Molecular Dynamic Simulation Inputs.

min_init.in

```
&cntrl
  imin = 1,
  maxcyc = 2500, ncyc = 1000,
  ntb = 1, ntr = 1, cut = 9
&end
Hold the DNA & DRG fixed
 500.0
 RES 1 26
&end
```

min.in

```
&cntrl
  imin = 1, ntb = 1, ntr = 0, cut = 9
  maxcyc = 5000, ncyc = 1500,
&end
&ewald
  ew_type = 0, skinnb = 1.0,
&end
```

heat.in

```
&cntrl
  imin=0, ntx=1,
  ntwr=500, ntwx=500, ntwe=500, nscm=5000, ntpr=500,
  ntf=2, ntc=2, ntb=1, ntp=0, ntt=1, ntr=1, nmropt=1,
  nstlim=150000, t=0.0, dt=0.002,
  cut=9.0, tempi=0.0, restraintmask=':1-26', restraint_wt=25.0,
&end
  &wt type='TEMP0', istep1=0, istep2=100000, value1=0.0, value2=300.0,
&end
  &wt type='TEMP0', istep1=100001, istep2=150000, value1=300.0, value2=300.0,
&end
  &wt type='END',
&end
```

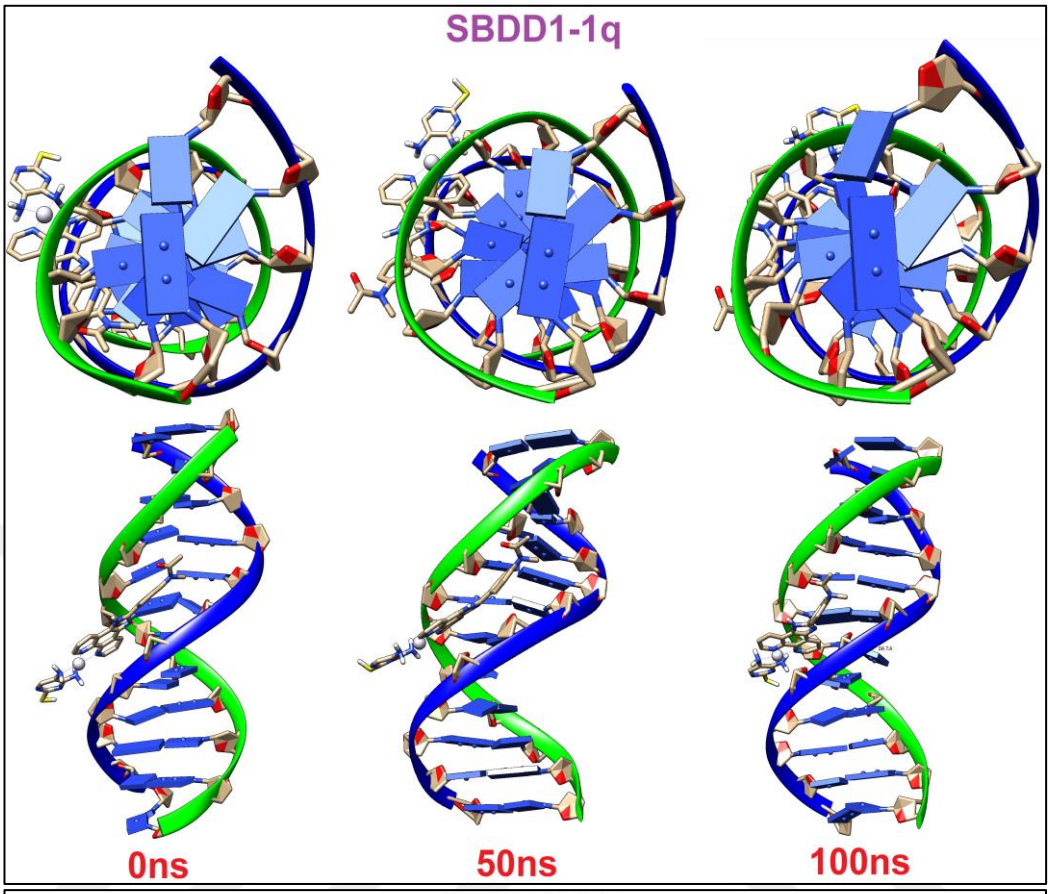
eq.in

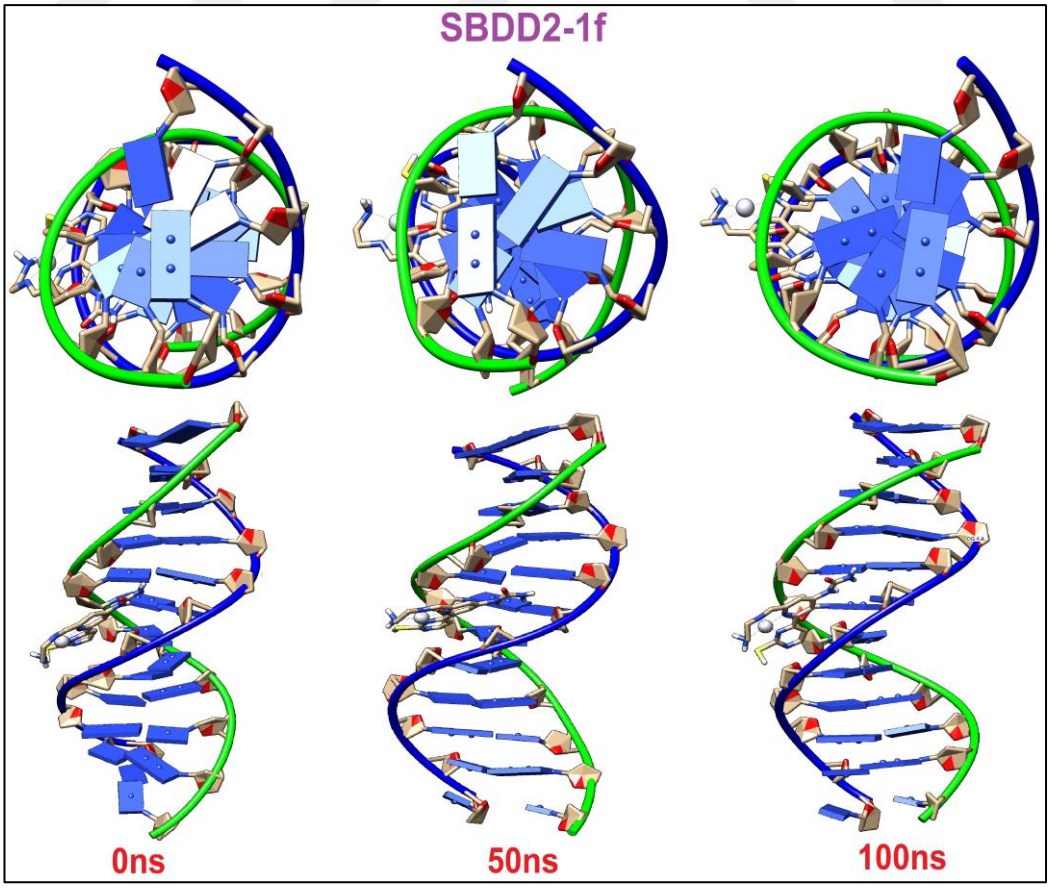
```
&cntrl
  imin=0, ntx=5,
  ntpr=500, ntwr=500, ntwx=500, ntwe=500, nscm=5000,
  ntf=2, ntc=2, ntb=2, ntp=1, ntt=1, ntr=1, irest=1,
  tautp=0.2, taup=0.2, nstlim=500000, t=0.0, dt=0.002,
  cut=9.0, tempi=300.0, temp0=300.0, restraintmask=':1-26', restraint_wt=0.5,
&end
&ewald
  ew_type = 0, skinnb = 1.0,
&end
```

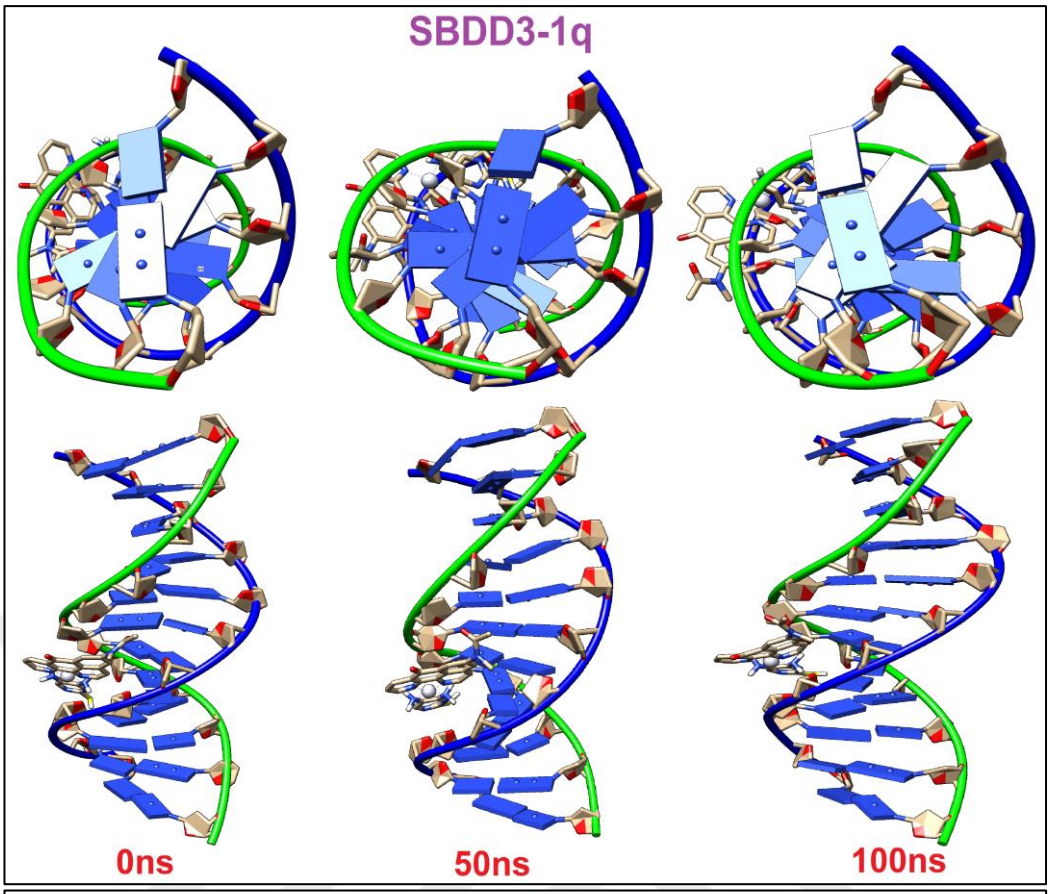
md.in

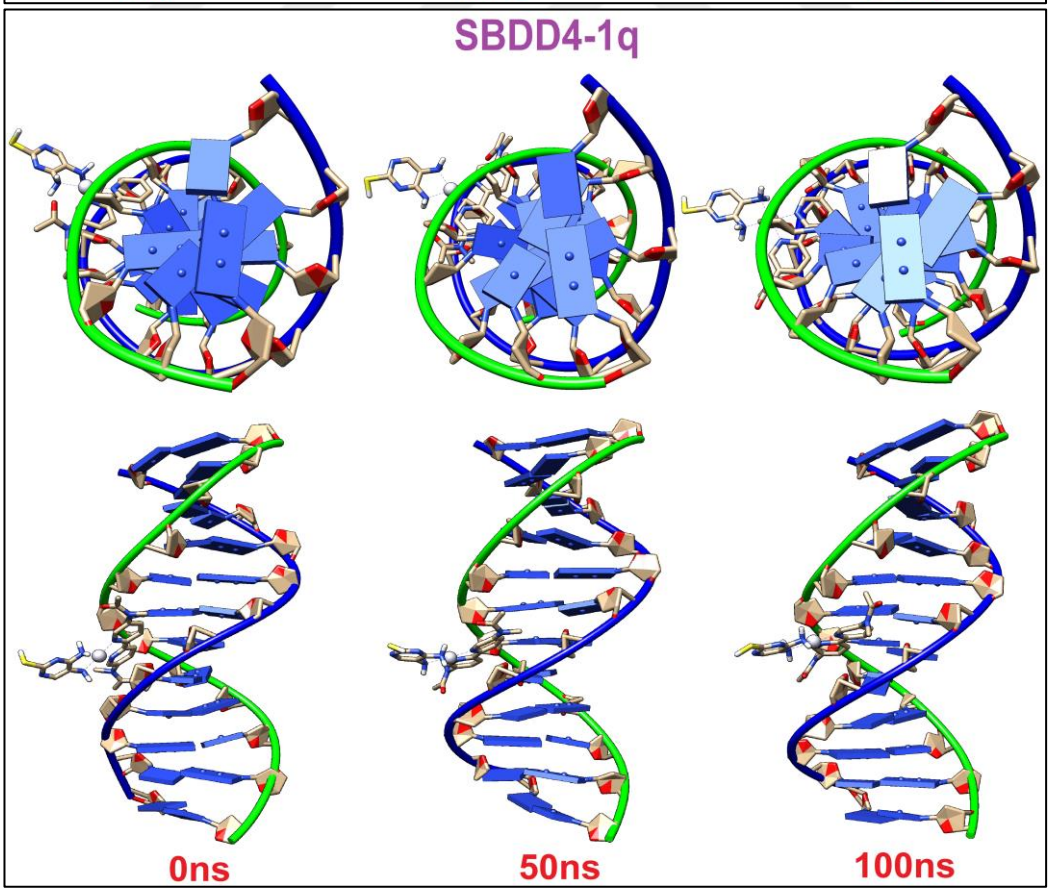
```
&cntrl
  imin=0, ntx=5,
  ntpr=2000, ntwr=2000, ntwx=2000, ntwe=2000, nscm=20000,
  ntf=2, ntc=2, ntb=2, ntp=1, ntt=1, irest=1, ntrx=1, iwrap=1,
  nstlim=50000000, t=0.0, dt=0.002, tautp=1.0, taup=2.0,
  cut=9.0, pres0=1.0, ioutfm=1,
&end
&ewald
  ew_type = 0, skinnb = 1.0,
&end
```

APPENDIX 5 Images of SBDD1-1q, SBDD2-1f, SBDD3-1q and SBDD4-1q Complexes at 0, 50 and 100 Nanoseconds.






SBDD3-1q
0ns
50ns
100ns


SBDD4-1q
0ns
50ns
100ns