REPUBLIC OF TURKEY

ACIBADEM MEHMET ALİ AYDINLAR UNIVERSITY

INSTITUTE OF HEALTH SCIENCES

IN SILICO IDENTIFICATION OF POTENTIAL INHIBITORS TARGETING N TERMINAL OF HUMAN REPLICATION PROTEIN A FOR CANCER THERAPY

SÜLEYMAN SELİM ÇINAROĞLU

MASTER THESIS

DEPARTMENT of MEDICAL BIOTECHNOLOGY

SUPERVISOR

Assoc. Prof. Dr. Emel TİMUÇİN

İSTANBUL-2018

Program:	Medical Biotechnology
Thesis title:	In Silico Identification of Potential Inhibitors Targeting N Terminal of Human Replication Protein A for Cancer Therapy
Students' name and surname:	Süleyman Selim Çınaroğlu
Date of defense:	05/07/2018

This is to certify that I have examined this copy of master thesis. I have found that she/he prepared after fulfilling requirements specified in the associated legislations before the final examining committee whose signatures are below.

Jury president	Assoc. Prof. Dr. Emel Timuçin
Supervisor of the thesis	Assoc. Prof. Dr. Emel Timuçin
Jury member	Prof. Dr. Osman Uğur Sezerman
Jury member	Assist. Prof. Dr. Günseli Bayram Akçapınar

This thesis has been approved by the above jury and it has been accepted by decision of Health Sciences Board of Directors.

Prof. Dr Uğur Özbek

Director of the Institute

Acıbadem Mehmet Ali Aydınlar University

REPUBLIC OF TURKEY

ACIBADEM MEHMET ALİ AYDINLAR UNIVERSITY

INSTITUTE OF HEALTH SCIENCES

IN SILICO IDENTIFICATION OF POTENTIAL INHIBITORS TARGETING N TERMINAL OF HUMAN REPLICATION PROTEIN A FOR CANCER THERAPY

SÜLEYMAN SELİM ÇINAROĞLU

MASTER THESIS

DEPARTMENT of MEDICAL BIOTECHNOLOGY

SUPERVISOR

Assoc. Prof. Dr. Emel TİMUÇİN

İSTANBUL-2018

DEDICATION

I dedicate to my lovely family and my dear friend, Muazzez Ceren.

CONTENTS

DECLARATION

I hereby declare that, this thesis has been written by me based on the data obtained in line with the scientific rules and ethical principles of responsible conduct of research. All information, data, comments, analyses have been collected and processed through scientific, academic writing style, and literature used have been duly shown by giving reference to the original sources in accordance with the publication ethics. I also announce and emphasize that I have not violated any rules secured by patent and copyrights whilst the conduct and writing of this research.

05/07/2018

Süleyman Selim ÇINAROĞLU

FOREWORD AND ACKNOWLEDGMENTS

Today, cancer research is the most important part of scientific research with increasing the number of different cancer types. The side effects of chemotherapeutic drugs are one of the hardest parts of the treatment of cancer patients. The necessity of new and less side-effect treatment strategies is gaining importance at this stage. In this work, an extensive virtual screening study of two databases that composed of more than 750 million molecules was performed to identify new potential cancer drugs. Multiple tools were tested using the reference set which is already known inhibitors of RPA70N, enabling us to choose the best approach for screening large databases. Overall, this comprehensive methodology will present an alternative approach, highlighting the potential of screening large databases for identification of potential inhibitors at low costs.

First and foremost, I would like to thank my family for their valuable support over all the years during this scientific journey has taken.

Also, I would like to thank Assoc. Prof. Dr. Emel Timuçin wholeheartedly, not only for his tremendous academic support but also for giving me so many wonderful opportunities. I appreciate all her contribution of time, supporting me through her ideas, funding, encouragements and her patience which allowed me to progress me as a researcher.

Süleyman Selim Çınaroğlu

ABBREVIATION

ADMET	Absorption, Distribution, Metabolism, Excretion, and Toxicity
ATR	Ataxia Telangiectasia Mutated and Rad3
CADD	Computer-aided Drug Design
DBD	DNA Binding Domain
DDR	DNA Damage Response
GAFF	General AMBER Force Field
HTS	High Throughput Screening
LiSiCa	Ligand Similarity using Clique Algorithm
MD	Molecular Dynamics
MM-PBSA	Molecular Mechanics with Poisson-Boltzmann and Surface Area
OB	Oligosaccharide/oligonucleotide Binding
PDB	Protein Data Bank
PME	Particle Mesh Ewald
PPI	Protein-Protein Interaction
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RPA	Replication Protein A
RPA70N	Replication Protein A N Terminal
SBDD	structure-based drug discovery
SHAFTS	SHApe-FeaTure Similarity
ssDNA	Single-strand DNA

TABLE LIST

FIGURE LIST

SUMMARY

During replication stress, Replication Protein A (RPA) initiates the DDR activation by binding to single-stranded DNA (ssDNA), while it also mediates the DDR through recruiting protein partners. Given these pivotal roles, RPA has become an attractive target for cancer drug discovery. Hitherto, many efforts have been devoted to identify inhibitors of RPA, employing different methods ranging from high throughput screening to fragment-based approaches. Although these studies led to the identification of RPA inhibitors, large molecule databases are awaiting to be screened, marking a possibility to identify more potent inhibitor(s). To this end, here we report the virtual screening of the ZINC15 database which is composed of more than 700 million small molecules. Particularly the RPA70N domain is used as the target, enabling inhibition of the protein-protein interactions (PPIs) formed by RPA. Prior to the screening, we assessed the performance of multiple docking tools by using benchmark ligand sets which were experimentally characterized. The best-performed tool, LeDock ($r=0.745\pm0.08$) was used for the large screen, and the ligands were filtered according to their docking scores and also the presence of a negatively charged group which was considered to be critical in binding to the positively charged amino acids located in the RPA70N cleft. 20 selected ligands were analyzed in molecular dynamics simulations followed by MM-PBSA prediction of binding free energy. The performance of the MM-PBSA method was also tested by using the benchmark set and the results showed a well-agreement ($r=0.92$) between the binding free energy predictions and experimental values. These validated tools led us to identify one promising ligand targeting RPA70N with a higher binding affinity and better drug-likeness features than any of the known inhibitors. Overall we surmise this ligand as being a inhibitor to target RPA70N more efficiently, reflecting its potential in reducing the side-effects associated with other RPA inhibitors.

Keywords: Replication Protein A, Virtual Screening, Molecular Docking, Molecular Dynamics, MM-PBSA

ÖZET

Kanser Tedavisi İçin İnsan Replikasyon Protein A N-Terminal Bölgesini Hedefleyen Potansiyel İnhibitörlerin *in silico* Tanımlanması

Çoğalma stresi sırasında, Replikasyon Proteini A (RPA), tek-iplikli DNA'ya (ssDNA) bağlanarak DNA hasar cevabı (DDR) aktivasyonunu başlatırken, aynı zamanda diğer protein partnerlerinin de DDR'ye dahil olmasına aracılık eder. Bu önemli rolü göz önünde bulundurulduğunda, RPA, kanser ilaçlarının keşfi ve geliştirilmesi için alternatif bir hedef olarak karşımıza çıkmaktadır. Şimdiye kadar, yüksek verimli taramadan fragment tabanlı yaklaşımlara kadar farklı yöntemler kullanarak, RPA inhibitörlerini tanımlamaya yönelik birçok çalışma yapılmıştır. Bu çalışmalar RPA inhibitörlerinin tanımlanmasına yol açmış olsa da daha etkin inhibitörlerin belirlenmesi için büyük veri tabanlarının taranması beklenmektedir. Bu amaçla bu çalışmada 750 milyondan fazla küçük molekülden oluşan ZINC15 veri tabanının sanal olarak taranması yapılmıştır. Özellikle RPA tarafından oluşturulan protein-protein etkileşimlerinin (PPI) inhibisyonu için RPA70N bölgesi hedef olarak kullanılmıştır. Taramadan önce, deneysel olarak tespit edilen referans ligand setlerini kullanılarak çoklu docking programlarının doğruluğu değerlendirilmiş ve en iyi performans gösteren program LeDock ($r = 0.745 \pm 0.08$) geniş tarama için kullanılmıştır. RPA70N bağlanma bölgesinde bulunan pozitif yüklü amino asitlere bağlanmada kritik olduğu düşünülen negatif yüklü bir grubun varlığı, bağlanmada kritik olduğu düşünülmüştür. Seçilen 20 ligand, moleküler dinamik simülasyonlarında MM-PBSA'nın tekniği ile serbest bağlanma enerjisinin tahmin edilebilmesi için analiz edildi. MM-PBSA yönteminin performansı da referans inhibitörler kullanılarak test edildi ve sonuçlar bağlanma serbest enerji tahminleri ve deneysel değerler arasında iyi bir korelasyon ($r = 0.92$) görüldü. Doğruluğu onaylanmış methotlar ile RPA70N'yi hedefleyen, bilinen inhibitörlerden daha yüksek bir bağlanma afinitesi ve daha iyi ilaç özellikleri bulunan bir tane potansiyel inhibitörün tanımlanmasını sağlanıştır. Genel olarak, bu potensiyel inhibitörün, RPA70N'yi daha etkin bir şekilde hedefleyen bir inhibitör olarak görüyoruz

ve diğer RPA inhibitörleriyle ilişkili olan yan etkileri daha az gösterme potansiyeli barındırmaktadır.

Anahtar sözcükler: Replikasyon Protein A, Sanal Tarama, Moleküler Docking, Moleküler Dinamik, MM-PBSA

1. INTRODUCTION

DNA is constantly exposed to DNA damaging agents externally and internally. Cells evolved to recruit a number of strategies to deal with the high rate of DNA damage that can occur in every cell every day. DNA damage response (DDR) is the collective term referring to all of the signalling and enzymatic activities found in cell-cycle arrest, regulation of DNA replication, and the repair of DNA damage (1–5). Defects in DDR pathways are associated with many cancer types (3,5,6). Indeed, this phenomenon regarding the greater propensity of cancer cells to accumulate DNA damage than normal cells ground the DNA damaging cancer therapies such as chemotherapy and radiotherapy (7). Because such therapy strategies are often accompanied with grievous side effects, targeted inhibition of DDR has become an intriguing area of research for cancer drug discovery, carrying the potential to minimize the side-effects.

Recently more than 450 different proteins have been reported to be involved in DDR pathways (8), underscoring numerous potential targets for drug discovery. Replication stress which is elevated in cancer cells compared with the normal cells (9), is the uncoupling of the DNA polymerase from the replisome helicase activity (10,11). During replication stress, extended single-strand DNA (ssDNA) is generated at the replication fork, leading to stalled replication forks. DDR is initially activated against replication stress by the binding of replication protein A (RPA) to these extended ssDNA regions (10). Subsequent to its ssDNA binding, RPA recruits several replication stress response mediators involved in DDR such as the kinase Ataxia Telangiectasia mutated and Rad3-related (ATR) (12), RAD9-HUS1-RAD1 (9-1-1) complex, ATRIP of ATR (13,14) and the replication fork remodeling protein SMARCAL1 (15). Owing to its binding to ssDNA, RPA acts as the initiator of DDR; while through protein-protein interactions (PPIs) formed by RPA, it can mediate the replication stress response involved in DDR. Given the dual roles of RPA as an initiator and regulator of DDR during replication stress, the structure and function of RPA have been extensively studied. This protein has been identified to be a hetero-trimeric protein consisting of three subunits, namely RPA70, RPA32, and RPA14

(16). All of these subunits are essential for the RPA function in DDR, such that loss of any subunits has been shown to be lethal (17). Taken together with these observations, the finding showing even conservative mutations in the RPA gene caused DDR defects (18–20), proposes RPA as an interesting target for drug discovery for cancer therapy.

Given the critical intermolecular interactions of RPA in DDR, this protein stood out as a potential target for cancer drug discovery. All of the subunits can bind to ssDNA with different affinities, thereof involved in initiating DDR (21–24). Despite the binding of different RPA subunits to ssDNA, N-terminal domain of RPA70 subunit (RPA70N) is primarily involved in forming PPIs to mediate DDR during replication stress (25). Although targeting the initiator function of RPA is a promising way of sensitization of cancer cells against replication stress, it may be a challenging route because of the ubiquitous binding of RPA to ssDNA from multiple subunits. On the other hand, inhibiting PPIs formed by RPA is more explicit since only a single subunit is involved in the recruitment of DDR mediators. Thus, PPIs formed and mediated by RPA70N have been widely studied to identify potential PPI inhibitors for cancer therapy.

A vast number of efforts primarily leads to the identification of small molecule inhibitors of RPA70N by different experimental methods including high throughput screening (HTS), fragment-, and structure-based methods. While these experimental methods are robust, they are also costly and labour intensive particularly for screening large compound libraries. Virtual screening studies may present an alternative way to screen large databases with reduced cost. Although great efforts have been dedicated to the discovery of RPA70N inhibitors by experimental methods, the currently available large databases still await to be screened, underscoring the possibility of the presence of more potent RPA PPI inhibitors.

To this end, here we present an extensive virtual screening study of two databases, ZINC12 and ZINC15 that composed of more than 700 million molecules (Figure 1). Multiple tools were tested by using a benchmark ligand set which is already known inhibitors of RPA70N, enabling us to choose the best approach for screening large

libraries. Dynamical analysis of the RPA70N and inhibitor interactions revealed the further evidence on the underlying molecular mechanism behind RPA70N inhibition, while the insights from dynamical analysis further guided us to pursue a rational design of the selected inhibitors. Through rationalizing the functional group in the selected inhibitors, we report one promising inhibitor which carries higher binding potential and better drug-likeness features than any of the known inhibitors of RPA. Overall, our comprehensive methodology will present an alternative approach, highlighting the potential of screening large databases for identification of potential inhibitors at low costs.

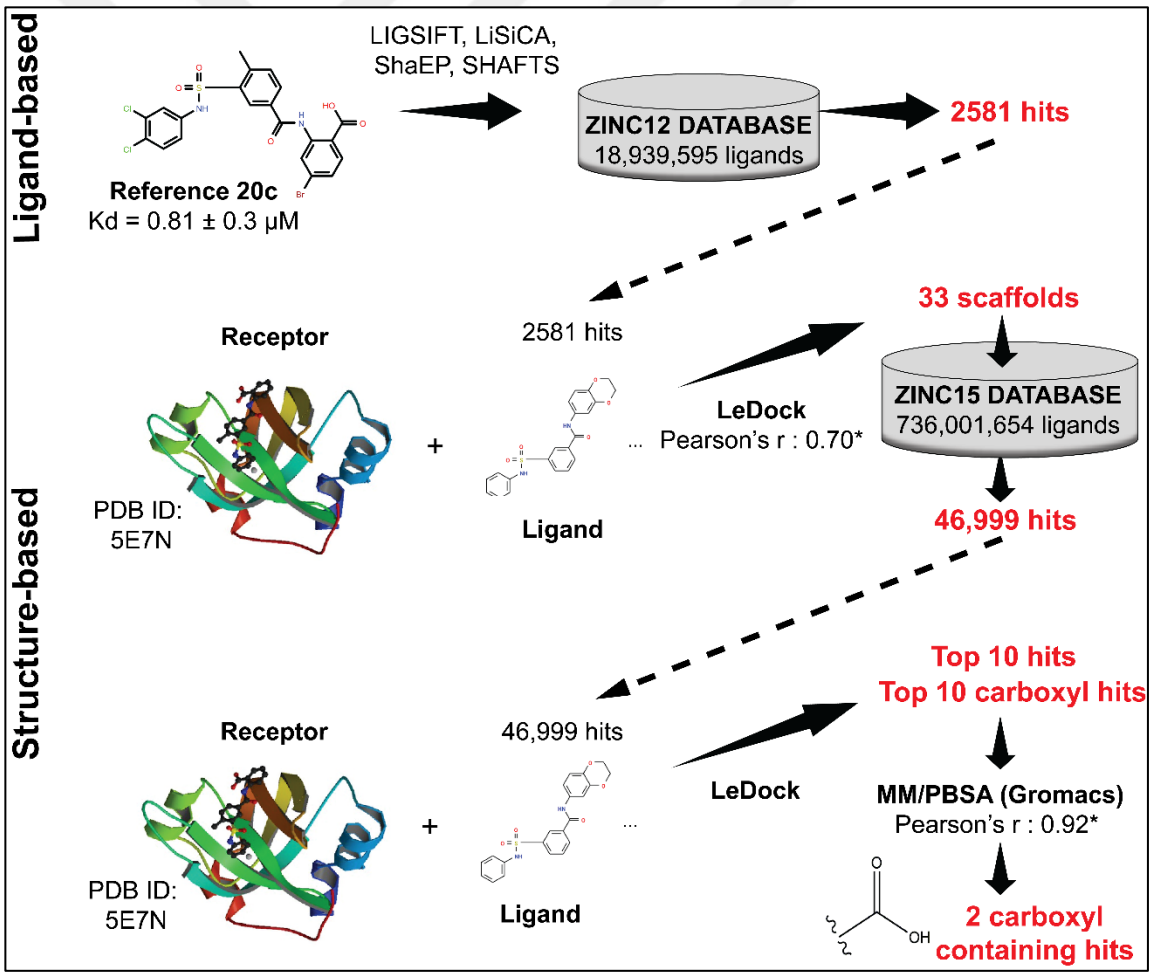

Figure 1: Multi-stage virtual screening strategy for the identification of the potential inhibitors targeting RPA70N. * Validated tools using the reference set.

2. BACKGROUND

2.1. Replication protein A

Replication protein A (RPA) is the primary single strand DNA (ssDNA) binding protein complex and is essential for processing of ssDNA intermediates found in replication, repair, and recombination (1). It is the fundamental component in DNA metabolism so that silencing of the entire RPA protein with using siRNA is cytotoxic to cells (2,3). There is a great amount of RPA protein in cells, and it binds to ssDNA with sub-nanomolar (nM) affinity (4,5). Thus, any single-stranded DNA formed in the genome is immediately coated by RPA in a non-sequence specific pattern. RPA is a heterotrimeric protein consisted of three subunits of 70-, 32-, and 14-kDa (RPA70, RPA32, and RPA14, respectively) (Figure 2)(6). The loss of any subunits in RPA is lethal for cells and non-deleterious mutations in RPA gene can cause DNA repair defects and genome instability (7–9). All RPA subunits contain one or more DNA-binding domains (DBD) composed of an oligosaccharide/oligonucleotide binding (OB) folds (10–13). The RPA14 subunit contains an OB-fold but it does not bind ssDNA and plays an important role in the stability of the RPA heterotrimerization core (15). The RPA32 subunit also contains an OB fold to bind ssDNA and several phosphorylation sites that regulate DNA metabolism and interact with other proteins (26). The RPA70 subunit is comprised of four domains (A, B, C, and N) (7). RPA70A and RPA70B domains have high ssDNA binding affinity and account for the binding of RPA to ssDNA (14). RPA70C domain also contains DBD but it binds with lower affinity (15). In addition to binding ssDNA, RPA interacts with proteins involved in DNA metabolism (16,17).

During normal DNA replication, RPA70 interacts with SENP6 (SUMO-specific protease) which keeps RPA70 in a hyposumoylated state. However, SENP6 is dissociated from RPA70 when DSBs occur in DNA. This allows RPA70 to be the sumoylated state to initiate homologous recombination. Thus, the sumoylation status of RPA70 is a critical part in DNA repair through homologous recombination (27).

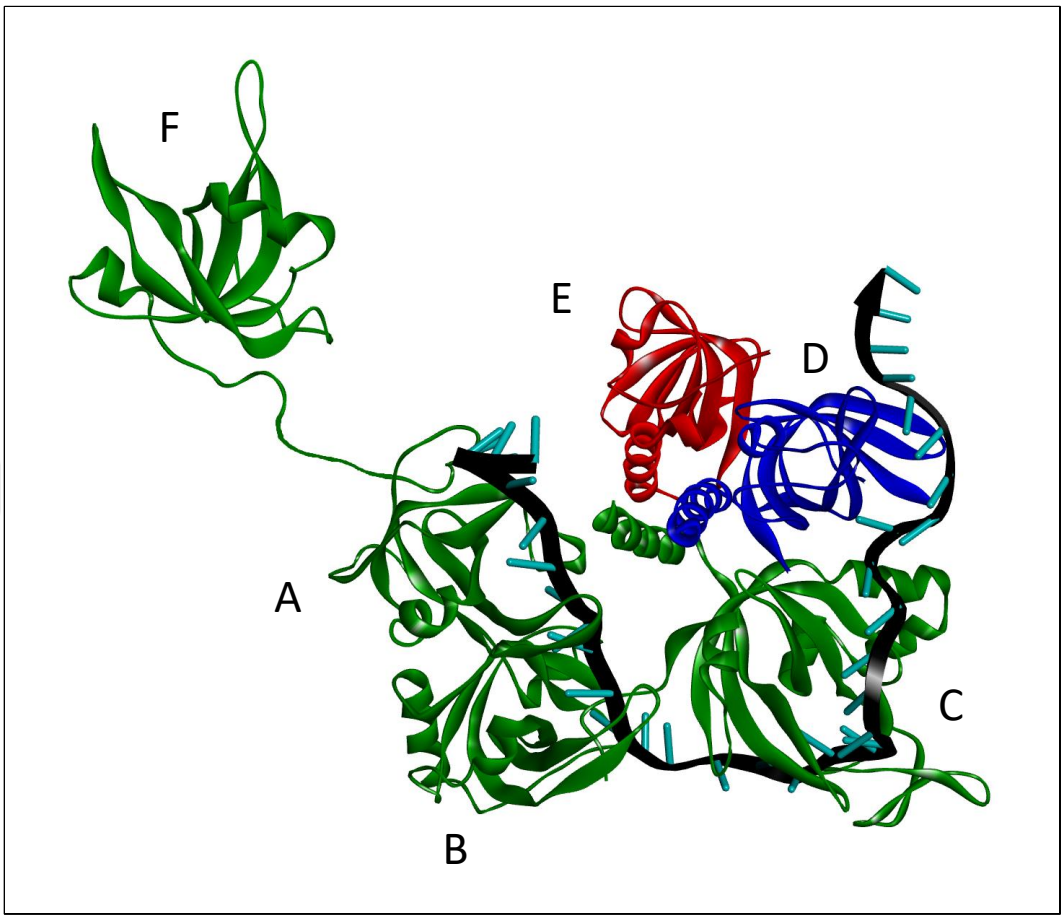

Figure 2: Replication Protein A structural model that is composed of 6 DBD. The red is RPA14 and the blue is RPA32. The green one is the RPA70 subunit is comprised of four domains (A, B, C, and F).

The N-terminal domain of the RPA70 subunit (RPA70N) is a crucial domain for protein-protein interactions and this domain plays a critical role in DNA damage response and repair proteins by way of a conserved motif of RPA70N (18). RPA70N interacts with several partner proteins that are important for mediating the DNA damage response, including ATRIP (ATR-interacting protein), MRE11, RAD9, and p53 (18,19). Disruption of the protein-protein interactions between these proteins and RPA70N by mutation of every interaction partner leads to decrease ATR signaling and increases sensitivity to DNA damage and replication stress (20–22). Selective inhibition of the RPA70N domain specifically disrupts these protein-protein interactions, leading to the inhibition of DNA repair functions and impairment of the DNA damage response (18,23). Based on this role of the RPA70N domain through protein-protein interactions in the DNA damage response,

targeted inhibition of this function is an attractive pathway for new therapeutic strategies in cancer (28). DNA damage response and repair are up-regulated in response to chemotherapy and radiotherapy. Because of the fact that cancer cells have increased DNA repair activity and higher levels of replicative stress than normal cells and these contribute resistance to chemotherapy and radiotherapy (24,25,29,30). Small molecules that specifically target the domain on RPA70N to inhibit the protein-protein interaction have potential to increase replicative stress and inhibit the DNA damage response and sensitize cancer cells to DNA-damaging agents without affecting other functions of RPA.

A small number of small molecules that inhibit RPA70N have already been reported in the literature. Firstly, Oakley and his colleagues reported Fumaropimaric Acid as the first inhibitor of the interaction between RPA70N and RAD9 (23). In this study, they performed a plate-based ELISA-like HTS assay to assess the disruption of the interactions. Similarly, using the previously described HTS assay, the Oakley laboratory identified another inhibitor, HAMNO (31). Unlike the first inhibitor, HAMNO is neutral at pH 7. Also, it is the first inhibitor of RPA70N that has been shown in vitro and in vivo experiments. Also, Fesik and his colleagues have made some efforts to identify new inhibitors of RPA70N using fragment-based techniques. Firstly, they screened the 15,000-member fragment library for binding to RPA70N using NMR. From this study, they have discovered potential inhibitors of the RPA70N–ATRIP protein-protein interaction without inhibition of RPA–ssDNA interaction (32). Also, they have described the optimized compounds with submicromolar binding affinity using the fragment-based screening and preliminary linking efforts to combine fragments (33). Lastly, they have identified anthranilic acid-based inhibitors using HTS and optimized using iterative medicinal chemistry and structure-based design techniques (Figure 3). These efforts led them to the discovery of 20c (4-Bromo-2-(3-(N-(3,4-dichlorophenyl)sulfamoyl)-4-methyl benzamide)benzoic acid), which binds to RPA70N with an affinity of 812 nM and exhibits sufficient permeability and solubility characteristics for cellular studies (34). This

compound may offer a promising starting point molecule for the discovery of potential useful RPA inhibitors.

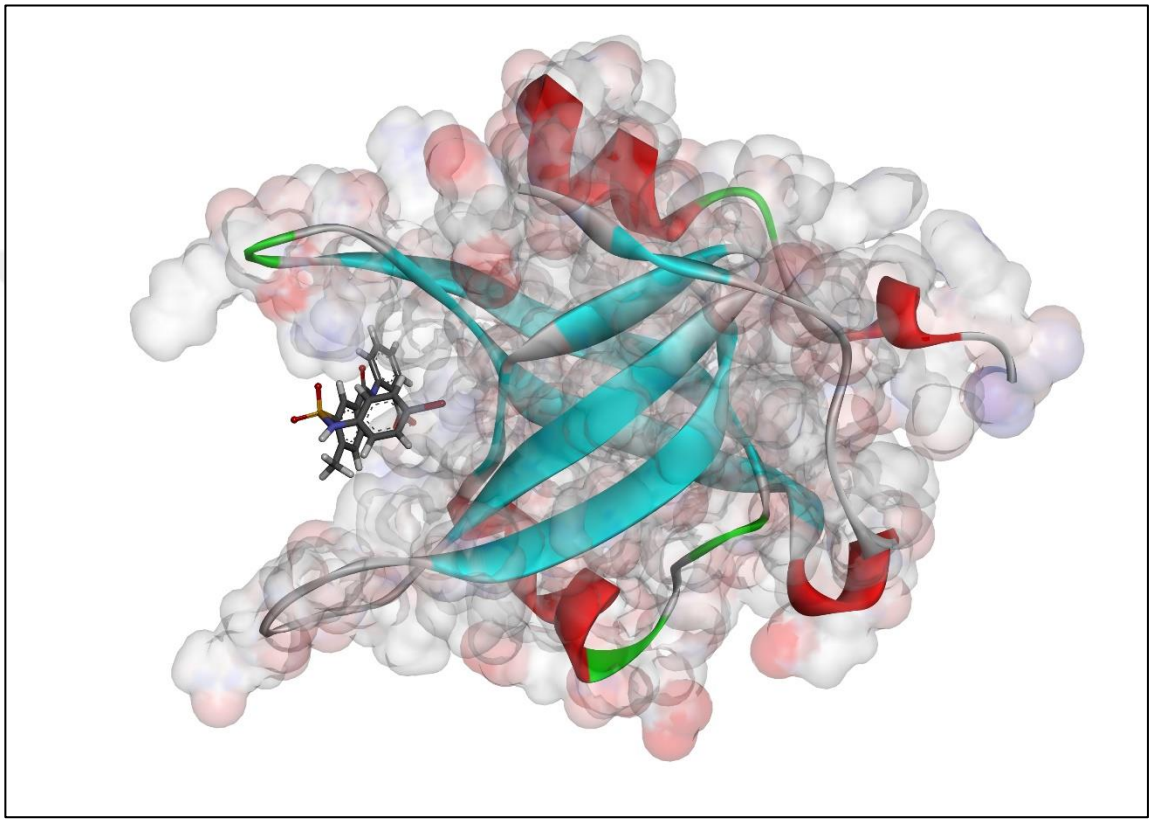

Figure 3: The binding mode of anthranilic acid-based inhibitors on the N terminal of RPA70. This crystal structure contains compound 11 identified by Fesik and his colleagues. (PDB ID: 5E7N)

2.2. Computer-Aided Drug Design and Discovery

Computer-aided drug design (CADD) is a very effective tool in the identification of potential drug candidates (35). CADD can increase the hit number of potential drug candidates because it uses a more targeted searching than traditional high throughput screening (HTS) and combinatorial chemistry. In a drug discovery effort, CADD is commonly used for three main purposes: to filter large ligand databases into smaller libraries of predicted more active ligands that may be experimentally tested; to lead to the optimization of hit compounds, whether to increase its binding affinity or optimization of pharmacokinetics properties including absorption, distribution, metabolism, excretion, and toxicity properties (ADMET) and drug metabolism; to design potential drug candidates (36,37). CADD consist of two common categories: ligand-based and structure-based. Structure-based CADD depends on the presence of the target structure for calculation of interaction energies between receptor and ligands, whereas ligand-based CADD uses active and inactive compounds through chemical similarity searching. Structure-based CADD generally prefers high-resolution structures of the target molecules. Ligand-based CADD is commonly used when insufficient structural information of target molecules. The main goal of structure-based CADD is to design and discover compounds that bind to the target with better binding affinity, valuable reduction of binding free energy, improvements of ADMET properties, and reduction off-target side effects. A successful effort of these applications will result in a ligand that has been corrected *in vitro* and *in vivo* and its binding site has been verified, ideally determining a crystal structure of the complex. The most common use of CADD is the screening of ligand libraries, which is known as virtual high-throughput screening (vHTS). Eliminating a huge number of inactive ligands saves money and time, since the size of the experimental HTS is significantly decreased using this technology. It is very important that vHTS technologies do not target to directly identify a new drug that is ready to test clinically, but rather to identify hits that have not been related to the target before. Development of new drugs may cause a huge amount of expenses in the range of 300 million to 3 billion

dollars anywhere. The significant low cost of CADD compared with traditional drug development makes these technologies very attractive to focus, decrease, and vary the chemical library. *De novo* drug designing is another method in CADD technologies and it includes the designing of new molecules. Like vHTS, the main role of *de novo* drug designing is not to design a new single molecule with very good binding affinity and acceptable ADMET properties but rather to design new compounds which can be sequentially improved.

vHTS needs so much computer power with high-performance to screen large ligand databases and identify ligands for experimental testing. Current computers and algorithms allow screening thousands of molecules using structure-based methods per day using parallel processing (38). Ligand libraries can be available in different sizes and forms including general databases which can be used to screen for any target macromolecule. Some databases focus a specific family of targets or ligand family. Generally, vHTS uses ligand databases containing drug-like small molecules that have been synthesized or can be easily synthesized from a starting molecule already available. For this aim, several databases are available with containing various information including available chemical ligands, drugs, enzymes, reactants, carbohydrates, and natural products etc. (39). Some commonly used compound databases are listed in Table 1.

Ligand databases are frequently built via enrichment ligands for drug-likeness or specific physiochemical properties for the target molecule. Even the presence of very fast docking algorithms, docking millions of ligands needs significant resources, and time which can be reduced through the eliminating non-drug like and undesirable ligands.

Table 1: Commonly used ligand databases

Database	Description	Reference
ZINC	A free database of commercially-available compounds for virtual screening	(40,41)
KEGG	Ligand Database Molecular building blocks of life in the chemical space	(42,43)
DrugBank	A unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information	(44)
PubChem	A database of chemical molecules and their activities against biological assays	(45)
ChemSpider	A free chemical structure database providing fast text and structure search access to over 60 million structures from hundreds of data sources	(46)
NLDB	A database of automatically corrected and predicted 3D protein-ligand interactions in the enzymatic reactions of various metabolic pathways	(47)
BindingDB	A public, web-accessible database of measured binding affinities, focusing chiefly on the interactions of protein considered to be drug-targets with small, drug-like molecules	(48)

Structure-based CADD needs the structure of the target molecule. The Protein Data Bank (PDB) is the most frequently used database for protein structure. PDB currently contains more than 139,000 protein structures by April 2018. The presence of 3D structures allows to diligently inspect the binding site of the receptor. Current structure-based drug discovery (SBDD) methods enable to design ligands having the important properties for efficient binding to the target receptor (49,50). Starting with a known target macromolecule, *in silico* methods are performed to identify novel small molecules. These *in silico* procedures are followed by the synthesis of the most potential molecules and experimental evaluations of biological features (51,52). Molecular docking is the most common technique used in SBDD because of the accurate prediction of ligand binding pose within the receptor binding pocket (53).

In recent years, we have been aware of the significance of analyzing and optimizing the absorption, distribution, metabolism, excretion, and toxicity (ADME-Toxicity) properties of drug candidates during the first stages of drug discovery. Lots of *in silico* tools have

been developed to predict some important ADME-Toxicity features. These *in silico* tools have the same aim of predicting ADME-Toxicity properties from molecular structure (54).

The aim of drug design and discovery is to identify a new pharmaceutical ligand that binds to a receptor. Traditionally, drug discovery efforts are carried out experimentally. These processes are both time consuming and costly. So, computational techniques have been improved with the aim of reducing the cost and time to design and develop new compounds. The most commonly used computational techniques in drug design and discovery are docking and scoring (55), whereby the binding pose of the drug is predicted, followed by the estimation of the binding affinity. These techniques are very efficient but not so accurate. They can be used to predict binding poses and distinguish between the best binder and non-binder compounds. The most popular method is molecular mechanics with Poisson-Boltzmann and surface area solvation method (MM-PBSA) (56,57). The method is used in a lot of applications, including protein design, conformation stability, protein-protein interaction, and prediction of ligand-binding affinities (58–62). The MM-PBSA calculations can be done using a single minimized structure or a large number of molecular dynamics snapshots. The first tool for the MM-PBSA method was firstly developed within the AMBER software (63). Also, another tool has also been published for the publicly available GROMACS, NAMD and APBS software (56,64,65). MM-PBSA is a useful method for postprocessing of docking conformations or to rationalize different conformations.

Here, an extensive virtual screening study of two databases, ZINC12 and ZINC15 that composed of more than 700 million molecules was performed to identify potential inhibitors targeting RPA70N (Figure 1). Multiple tools were correlated by using already known inhibitors of RPA70N, enabling us to select the best tools for screening large libraries. Molecular dynamical analysis of interactions between the RPA70N and inhibitors elicited the detailed evidence on the underlying molecular mechanism behind RPA70N inhibition.

3. MATERIALS AND METHODS

3.1. Data collection and preparation

3D structure of the reference compounds (Reference Compounds: 11, 15, 16, 20c, 20f, 20i, 21, 28, 32, 34) (Figure 4) was generated using JSME molecular editor (34,66). Open Babel v2.4.1 was used for conversion of file formats of chemical and structure files (67). The Dock Prep tool of the Chimera program was used to optimize the compounds for the docking studies (68). The X-ray crystal structures of RPA70N were obtained from the protein databank (PDB codes: 5E7N, 4LUZ, 4LWC, 4O0A, 4R4C, 4R4I, 4R4O, 4R4Q, 4R4T). The ZINC12 and ZINC15 ligand libraries which store commercially-available chemical compounds were used for virtual screening studies (40,41).

3.2. Ligand-based Virtual Screening

The ligand-based virtual screening was performed by using four different approaches: Ligand Similarity using Clique Algorithm (LiSiCA) (69), LIGSIFT (70), ShaEP (71), and SHApe-FeaTure Similarity (SHAFTS) (Figure 5) (72). Compound 20c with having the best binding constant to RPA70N was used as the reference molecule for the similarity search in all software. All purchasable compounds from ZINC12 database were used as a search library. All of the similarity scores obtained from 4 different algorithms were ranked separately and the top scored 1000 ligands from each screen were combined to be tested in the structure-based screening step.

3.3. Structure-based Virtual Screening

Before the structure-based virtual screening, 5 different docking tools including AutoDock4.2 (73,74), AutoDock Vina (75), UCSF DOCK6 (76,77), PLANTS (78–80), and LeDock (81,82) were analyzed for the determination which docking program is better for the screening in RPA70N. For this analyzing, 5E7N structure as the receptor and the reference compounds as the ligands were used and the center of 5KR was adjusted as the binding site center. The binding site covered 20 Å of dimension or diameter in all docking tools.

AutoDock and Autodock Vina: The receptor and compound files were prepared using AutoDock Tools (73) for Autodock and AutoDock Vina. The remaining parameters were maintained at their default values in both tools.

UCSF DOCK6: The Dock Prep tool of the Chimera program (68) was used to prepare the receptor structure. The molecular surface of the receptor was generated with a probe radius of 1 Å. The maximum sphere radius was 4.0 Å, and the minimum sphere radius was 1 Å. Subset spheres representing the binding site for RPA70N were identified within 8.0 Å of 5KR crystal pose in 5E7N. The grid-based scoring function based on the non-bonded terms of the molecular mechanic force field was used as the primary scoring function.

PLANTS: The SPORES (83,84) tool was used for protonation of the protein and all ligands. The piecewise linear potential (PLP) scoring function was used and other parameters were remained by default (79).

LeDock: LePro was used for the preparation of the receptor structures to be docked in LeDock. All of the heterogenous atoms were removed from the structures and the hydrogens were added by LePro (81,82).

All the ligands were ranked according to the best binding score obtained from the docking tools. Each docking was repeated three times. The Pearson correlation coefficient was calculated for the docking scores and experimental binding affinities. Root mean square deviation (RMSD) of heavy atom distance was calculated by comparing the crystal coordinates with the best pose of docking program. The docking program that produced

the strongest correlation with the experimental values was selected to be recruited for ranking of the ligands, while the program that produced the lowest RMSD from the crystal structure was used during selection of the poses.

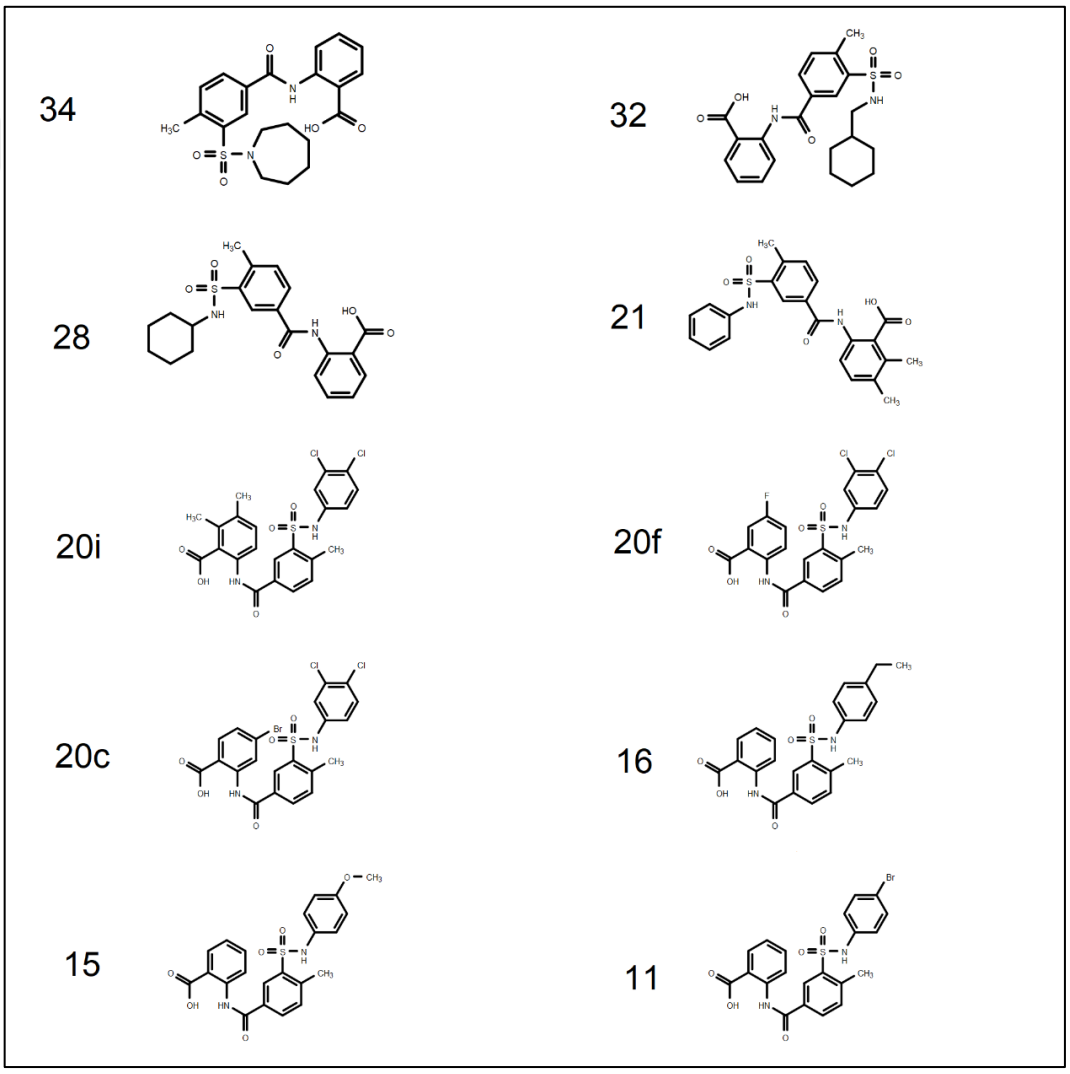

Figure 4: Reference molecules were used to evaluate the scoring function of docking tools in this study.

3.4. ADME-Toxicity Analysis

ADME properties were predicted by using SwissADME tool (85). This server enables us to calculate important physicochemical, pharmacokinetic, drug-like parameters for small molecules. Toxicity prediction was performed using Mcule Toxicity Checker web server (86). This tool applies a searching for substructures commonly found in toxic molecules and promiscuous ligands.

3.5. Molecular Dynamics Simulations

Molecular dynamics simulations performed for all 20 protein-ligand complexes along with the first benchmark set using Gromacs 4.6.5 adopting the force field the AMBER ff99SB to generate the topologies (87–89). The topology file for the ligands was generated using the general AMBER force field (GAFF) (90,91). The protein-ligand complexes were then solvated using the SPC water model and placed in the center of a cubic box. Minimum 1.0 Å distance was kept on the edge of the box and protein that can fully immerse in water and rotate freely. Particle Mesh Ewald (PME) method was used for the electrostatic energy calculation (92). Before minimization, the system was neutralized by adding Na+ or Cl- ions. The steepest descent approach was used for each protein-ligand complex for energy minimization without any constraints (93,94). The time step was maintained at 2 fs for the molecular dynamics. 10 Å cut-off distance was used for predicting the short-range non-bonded interaction. Then the system was slowly heated to a constant temperature of 310 K during 100 ps simulation. After stabilizing the system temperature at 310 K, 100 ps of simulation was performed by using an NpT ensemble with the system pressure at 1 bar. The final MD run was set to 20 ns at constant pressure and temperature for each protein-ligand complex, and trajectories were collected at every 100 ps for further analysis. All of the trajectories were analyzed by means of backbone RMSDs, Cα fluctuations, intermolecular hydrogen bonds, number of contacts and distance. For visual inspections of the structures and trajectories, Discovery Studio, VMD and/or Chimera were used (68,95,96).

3.6. Binding Free Energy Calculations

The binding free energy of protein-ligand complexes was calculated using the Molecular Mechanics Poisson–Boltzmann surface area (MM-PBSA) method (65,97,98). The binding free energy of protein-ligand complexes was calculated with using the g_mmpbsa tool from all snapshots extracted every 0.01 ns from the last 1 ns. The temperature used for Poisson–Boltzmann calculation was set to 310 K. The binding free energy (ΔG_{bind}) is calculated using by the four individual energy components from the equation 2,

Eq. 1 $$\Delta G_{bind} = E_{MM} + G_{solv} - T\Delta S = E_{vdW} + E_{elec} + G_{polar} + G_{nonpolar} - T\Delta S$$

E_{MM} is the molecular mechanics free energy in the gas phase, including electrostatic (E_{elec}) and van der Waals (E_{vdW}) contributions. G_{solv} is the solvation free energy, including polar (G_{polar}) and nonpolar ($G_{nonpolar}$) contributions. $T\Delta S$ represents the conformational entropy contribution at temperature T.

4. RESULTS

4.1. Virtual Screening

Computational approaches for the discovery of potential RPA70N inhibitors started with a ligand-based virtual similarity screening followed by molecular docking, which is a structure-based technique. Compound 20c was used as the reference compound for similarity searching in the database that comprises a total of 18,939,595 small molecules obtained from ZINC12 database (40). The ligand-based virtual screening was performed using four different virtual screening algorithms (Figure 5). All similarity scores obtained from all software were ranked separately, and 1000 top-scored ligands from each ranking were preserved. Finally, top-ranked ligands formed a ligand library consist of 2581 small molecules.

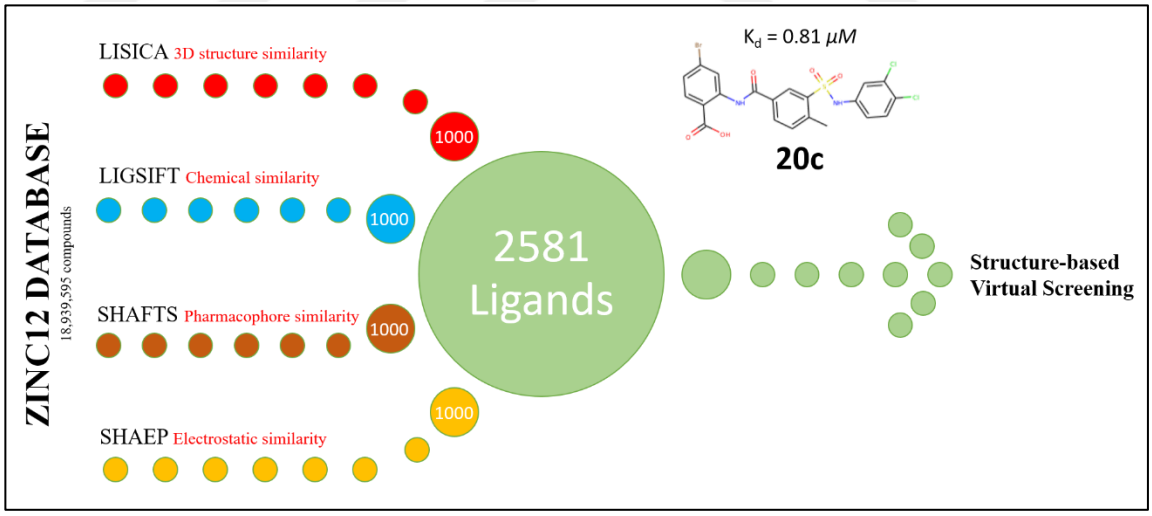

Figure 5: Ligand-based virtual screening workflow

After the ligand-based virtual screening, molecular docking studies were performed using these top-ranked ligands. It is crucial to use the most accurate software for molecular docking. Therefore, the Pearson correlation coefficient between the docking scores and experimental binding affinities was calculated to validate the prediction accuracy and

reliability of the scoring function in molecular docking programs. For this calculation, 10 reference inhibitors (Figure 4) with the significant difference between the binding affinities were randomly selected from the previous work (34). Then each inhibitor was docked three times with five different freely available molecular docking programs. Average scores for all inhibitors are listed in Table 1. According to the results, LeDock had the best scoring power with 0.745 ± 0.08 correlation compared to other programs in this case. This result shows us this software is very suitable for structure-based virtual screening to rank the ligands against to RPA70N.

Table 2. Assessment of scoring of different docking program in the reference compounds set of RPA70N inhibitors. Bivariate correlation is calculated between K_d and docking scores. Docking is repeated three times for each inhibitor and "mean±SD" correlation coefficient is given.

ID	K_d	LeDock*	DOCK6**	Vina*	PLANTS***	AutoDock*
11	30	-5.37 ± 0.04	-27.96 ± 1.40	-6.23 ± 0.06	-66.74 ± 4.17	-6.31 ± 0.11
15	234	-5.19 ± 0.09	-30.99 ± 1.57	-6.50 ± 0.26	-68.85 ± 2.45	-6.11 ± 0.14
16	95	-5.12 ± 0.16	-32.05 ± 1.03	-6.43 ± 0.06	-72.36 ± 1.39	-6.23 ± 0.19
20c	0.81	-6.37 ± 0.25	-30.46 ± 0.92	-6.80 ± 0.00	-68.31 ± 3.87	-6.74 ± 0.25
20f	15	-6.32 ± 0.07	-28.20 ± 6.03	-6.53 ± 0.23	-70.46 ± 0.02	-6.29 ± 0.09
20i	62	-6.29 ± 0.17	-26.20 ± 1.94	-7.40 ± 0.00	-67.74 ± 2.13	-6.30 ± 0.28
21	150	-5.34 ± 0.31	-23.99 ± 2.24	-6.70 ± 0.17	-69.19 ± 0.03	-6.50 ± 0.22
28	193	-4.84 ± 0.20	-28.90 ± 6.42	-6.67 ± 0.12	-69.04 ± 0.07	-6.26 ± 0.06
32	126	-5.17 ± 0.26	-32.87 ± 3.28	-6.03 ± 0.32	-72.08 ± 0.23	-6.37 ± 0.09
34	208	-4.91 ± 0.14	-23.75 ± 2.01	-6.10 ± 0.00	-70.05 ± 0.00	-6.73 ± 0.16
Pearson's *r*		0.745 ± 0.08	0.147 ± 0.10	0.247 ± 0.06	-0.171 ± 0.37	0.331 ± 0.26

Docking scores are the raw values directly obtained as the output of docking

* Energy (kcal/mol) ** Grid Score (a.u.) *** ChemPLP Score

After this validation, top-ranked ligands (2581) obtained from ligand-based virtual screening were docked to RPA70N using LeDock. After this docking study, all molecules were ranked according to their docking scores, then top-ranked 100 ligands were analyzed to obtain information which chemical scaffolds exist in this ranking (Table 2 and Figure 6). All small molecules having these chemical scaffolds were extracted from ZINC15 (41), which is a superior version of ZINC12 (40) in most ways. All molecules (46999) obtained from ZINC15 were docked to RPA70N again using LeDock.

Table 3: Chemical scaffolds obtained from the first structure-based virtual screening. In this table, the first column represents a number of small molecules having scaffold and the second contain SMILES for each scaffold.

# of Ligands	SMILES
539	O=C(Nc1cccc(S(=O)(=O)Nc2ccccc2)c1)c1ccncc1
22575	O=C(Nc1ccccc1)c1cccc(S(=O)(=O)Nc2ccccc2)c1
153	O=C(Nc1ccc2ocnc2c1)c1cccc(S(=O)(=O)Nc2ccccc2)c1
42	O=C(c1nc(-c2cccc(S(=O)(=O)Nc3ccccc3)c2)no1)N1CCCCCC1
7267	O=C(Nc1ccccc1)c1cccc(NS(=O)(=O)c2ccccc2)c1
250	O=C(Nc1cccc(S(=O)(=O)Nc2ccccc2)c1)c1cnccn1
1663	O=C(Nc1ccccn1)c1cccc(S(=O)(=O)Nc2ccccc2)c1
57	O=C(c1nc(-c2cccc(S(=O)(=O)Nc3ccccc3)c2)no1)N1CCCC1
74	O=C(Nc1cccc(S(=O)(=O)Nc2ccccc2)c1)c1ccc2[nH]nnc2c1
1765	O=C(Nc1cccc(S(=O)(=O)Nc2ccccc2)c1)c1ccccn1
59	O=C(Nc1cccc(S(=O)(=O)Nc2ccccc2)c1)c1ccc(-n2cnnn2)cc1
1	O=C(c1nc(-c2cccc(S(=O)(=O)Nc3ccc4c(c3)OCO4)c2)no1)N1CCCC1
1180	O=C(N=C(NCc1cn[nH]c1)Nc1ccccc1)c1ccccc1
5	O=C1NC(=O)C(=NNc2cccc(S(=O)(=O)Nc3ccccc3)c2)C(=O)N1
54	O=C(c1nc(-c2cccc(S(=O)(=O)Nc3ccccc3)c2)no1)N1CCCCC1
2263	O=C(Nc1cccnc1)c1cccc(S(=O)(=O)Nc2ccccc2)c1
74	O=C(Nc1cccc(S(=O)(=O)Nc2ccccc2)c1)c1cc(C2CC2)[nH]n1
123	O=C(Nc1ccc2c(c1)OCCO2)c1cccc(S(=O)(=O)Nc2ccccc2)c1
548	O=C(N=c1cc[nH]cc1)c1cccc(S(=O)(=O)Nc2ccccc2)c1
263	O=C(Nc1nc2ccccc2s1)c1cccc(S(=O)(=O)Nc2ccccc2)c1
440	O=C(Nc1cccc(S(=O)(=O)Nc2ccccc2)c1)c1cncs1
280	O=C(Nc1cccs1)c1cccc(S(=O)(=O)Nc2ccccc2)c1
7065	O=C(Nc1cccc(S(=O)(=O)Nc2ccccc2)c1)c1ccccc1
90	O=C(Nc1cccc(S(=O)(=O)Nc2ccccc2)c1)c1ccc(=O)[nH]n1
23	O=C(Nc1ccccc1C(=O)NCc1ccco1)c1cccc(S(=O)(=O)Nc2ccccc2)c1
22	O=C(Nc1cccc(S(=O)(=O)Nc2ccccc2)c1)c1ccc(Cn2cncn2)cc1
2	O=C(Nc1cccc(S(=O)(=O)Nc2ccccc2)c1)c1cccc(-c2ccccc2)c1
17	O=C(Nc1cccc(S(=O)(=O)Nc2ccccc2)c1)c1ccc(N2CCCC2)cc1
39	O=C1CCC(C(=O)Nc2cccc(S(=O)(=O)Nc3ccccc3)c2)=NN1
43	O=C(c1nc(-c2cccc(S(=O)(=O)Nc3ccccc3)c2)no1)N1CCOCC1
23	O=C(Nc1ccc(S(=O)(=O)N2CCCC2)cc1)c1cccc(S(=O)(=O)Nc2ccccc2)c1

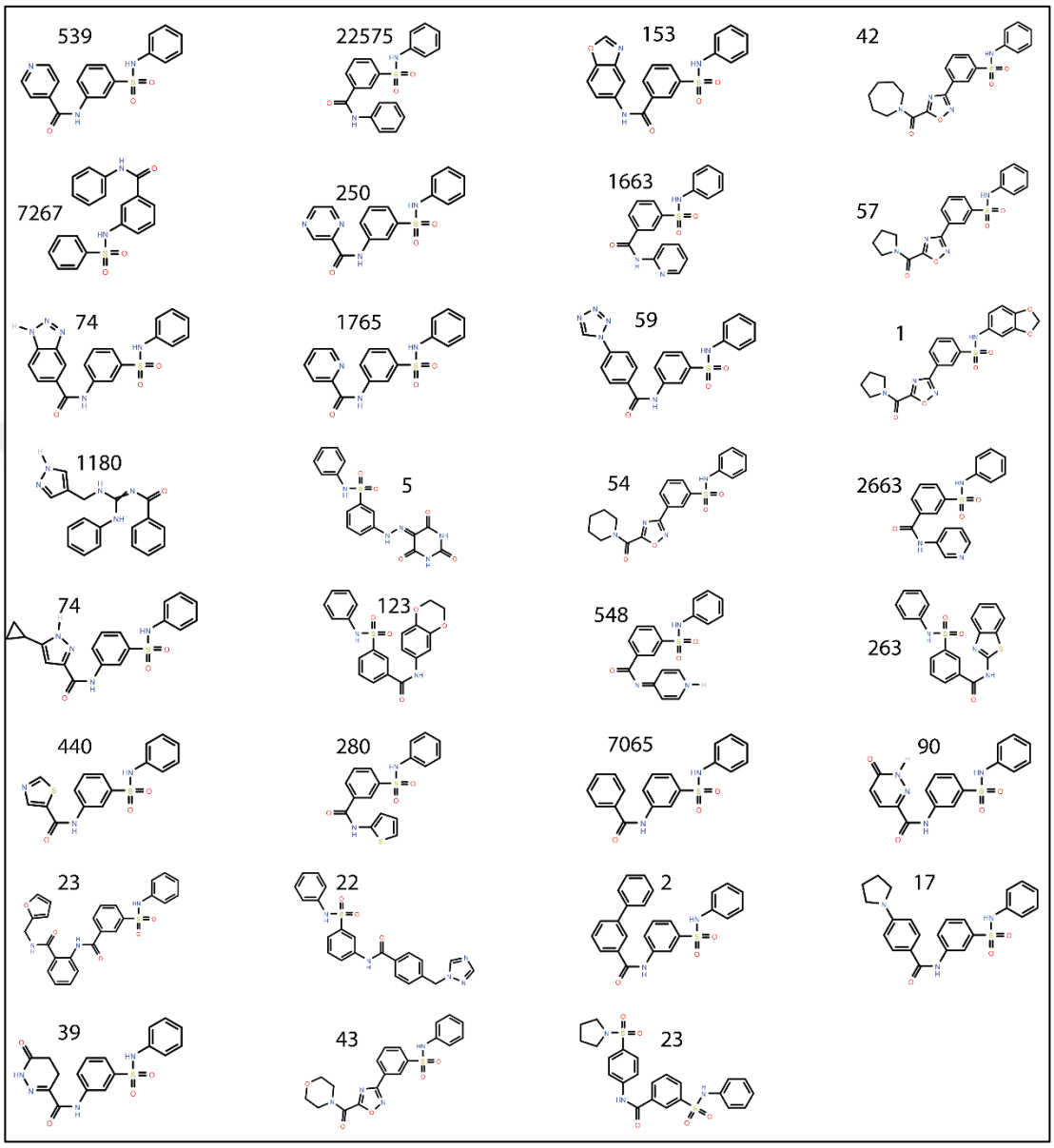

Figure 6: 31 different scaffolds representing a total of 2581 ligands that were identified to have similarity with the 20c in the ZINC12 database. The numbers display the total number of molecules that possess the represented scaffold.

Besides to scoring power of molecular docking software, the prediction accuracy of the best binding poses is another essential part of lead identification. If LeDock is not capable of predicting the best binding poses, we need to another docking tool to predict the best binding poses of top-ranked ligands. In order to validate pose prediction accuracy, we extracted RPA70N structures containing one ligand from PDB database and tried to predict binding poses of each ligand in crystal structures. These predictions were independently tested three times, and heavy-atom root-mean-squared-deviation (RMSD) between the docked poses and crystal structure poses of each ligand was calculated and evaluated. If the RMSD of the docked pose is less than 2.0 Å from the experimentally observed conformation, the prediction is regarded to be successful (82). Thus, PLANTS is the most successful software reproducing the crystallographic binding mode of each ligand. The average value of RMSD for all ligands is 0.92 ± 0.007, and this result concluded us PLANTS which is a thriving docking tool to predict the best binding poses of RPA70N ligands. Also, LeDock and AutoDock Vina, which are nearly considered successful based on their average RMSD, can be used to predict the binding mode of small molecules. Furthermore, LeDock gave us 1.23 ± 0.15 RMSD value with default parameters.

Table 4: Assessment of pose prediction accuracy for docking tools. The last row contains an overall average value for all complexes. Docking is repeated three times for each inhibitor, and the average value of RMSD is given as "mean±SD".

PDB ID	Ligand ID	LeDock	DOCK6	Vina	PLANTS	Autodock
4LUZ	1XT	1.81 ± 0.26	8.38 ± 2.53	1.46 ± 0.52	1.08 ± 0.16	2.36 ± 1.80
4LWC	1XU	1.51 ± 0.48	8.97 ± 1.19	1.74 ± 0.10	0.69 ± 0.00	7.09 ± 1.57
4O0A	2P9	1.05 ± 0.07	5.04 ± 0.15	1.29 ± 0.17	1.10 ± 0.08	5.54 ± 3.80
4R4C	3HS	1.06 ± 0.35	8.25 ± 3.64	0.78 ± 0.18	0.98 ± 0.00	3.52 ± 2.28
4R4I	3HV	1.33 ± 0.64	5.48 ± 1.03	1.39 ± 0.02	0.58 ± 0.25	4.60 ± 0.04
4R4O	3HW	1.43 ± 0.15	5.85 ± 1.31	2.38 ± 1.99	1.19 ± 0.10	5.74 ± 5.13
4R4Q	3HZ	3.73 ± 0.95	6.12 ± 0.28	2.80 ± 1.05	0.88 ± 0.18	6.04 ± 2.22
4R4T	3J0	1.11 ± 0.09	4.46 ± 0.71	1.04 ± 0.19	1.02 ± 0.13	7.96 ± 3.32
5E7N	5KR	6.65 ± 3.40	7.16 ± 1.59	9.31 ± 0.30	0.80 ± 0.40	4.29 ± 4.04
Average RMSD		2.19 ± 0.4	6.64 ± 0.5	2.46 ± 0.3	0.92 ± 0.1	5.24 ± 1.6

Table 5: Assessment of scoring function of docking tools using second reference benchmark used in pose prediction accuracy analysis. Docking is repeated three times for each inhibitor and "mean±SD" correlation coefficient is given.

PDB	Ligand	Kd	LeDock	DOCK6	Vina	PLANTS	AutoDock
4LUZ	1XT	20	-7.09 ± 0.1	-35.88 ± 3.8	-8.27 ± 0.2	-86.84 ± 0.0	-9.03 ± 0.5
4LWC	1XU	1.7	-7.43 ± 0.2	-29.53 ± 1.9	-6.90 ± 0.0	-85.18 ± 0.0	-7.05 ± 0.8
4O0A	2P9	0.19	-8.60 ± 0.2	-37.52 ± 0.4	-8.27 ± 0.2	-99.72 ± 0.0	-9.19 ± 0.1
4R4C	3HS	2.9	-8.97 ± 0.4	-29.79 ± 7.3	-9.10 ± 0.0	-100.23 ± 0.0	-10.73 ± 0.1
4R4I	3HV	0.54	-8.18 ± 0.1	-37.16 ± 2.8	-8.40 ± 0.0	-83.56 ± 0.1	-10.42 ± 0.1
4R4O	3HW	1.5	-7.61 ± 0.2	-35.52 ± 1.8	-7.53 ± 0.0	-81.70 ± 0.6	-8.44 ± 0.2
4R4Q	3HZ	8.2	-7.12 ± 0.4	-36.59 ± 0.3	-7.57 ± 0.1	-94.62 ± 0.6	-7.63 ± 0.3
4R4T	3J0	0.55	-8.33 ± 0.1	-30.58 ± 0.5	-8.13 ± 0.1	-86.25 ± 0.1	-8.65 ± 0.6
5E7N	5KR	30	-5.49 ± 0.1	-33.96 ± 2.5	-6.13 ± 0.1	-71.85 ± 0.1	-5.95 ± 0.1
Pearson's *r*			0.85 ± 0.1	-0.14 ± 0.2	0.54 ± 0.1	0.54 ± 0.0	0.55 ± 0.1

In addition to binding pose prediction, the scoring power of LeDock was re-evaluated using this dataset. For this evaluation, the Pearson correlation coefficient between the docking scores and experimental binding affinities in this dataset was calculated as 0,85 ± 0,06 r. This calculation also validated the ranking power of LeDock to score binders of RPA70N.

After molecular docking studies, top-scored 10 ligands selected for molecular dynamics studies. Also, we generated another ligand set for molecular dynamics because our binding pocket is arginine-rich and the top-selected set did not contain any ligand having carboxyl group (Table 6). Carboxyl group is very important as a functional group to bind arginine residues, so we generated top-scored carboxyl-containing ligand set. For molecular dynamics simulations, Protein-ligand complexes were generated using PLANTS that is the most successful docking tool in the assessment of pose prediction.

Table 6: Selected ligands after structure-based screening for molecular dynamics studies

The Top-selected Ligands		**The Carboxyl-containing ligands**	
ZINC ID	**Energy (kcal/mol)**	**ZINC ID**	**Energy (kcal/mol)**
ZINC000003305744	-6.98	ZINC000001196972	-5.87
ZINC000103020176	-6.76	ZINC000002860268	-6.09
ZINC000150446689	-6.85	ZINC000028571534	-5.98
ZINC000150451197	-7.28	ZINC000029749235	-6.38
ZINC000477616741	-7.36	ZINC000299859287	-5.84
ZINC000485075600	-6.89	ZINC000485282795	-6.33
ZINC000485439503	-7.39	ZINC000485657388	-6.18
ZINC000595369213	-6.74	ZINC000500370920	-5.77
ZINC000611932230	-6.97	ZINC000579026174	-5.58
ZINC000637513594	-6.75	ZINC000753854163	-5.85

4.2. Molecular Dynamics Simulations

After the all virtual screening studies, reference inhibitors and selected ligands with RPA70N were submitted to 20 ns of the molecular dynamics simulations. The aims of these simulations were to investigate the dynamic behaviour of these ligand-RPA70N complexes obtained from docking studies and obtain additional information to support the proposal of new potential inhibitors for RPA70N. The stability of all complexes was evaluated by monitoring RMSD of Cα carbon atoms of the protein with respect to the starting structure through 20 ns of molecular dynamics simulations. Root mean square fluctuation (RMSF) of side-chain atoms for all complexes was analyzed to investigate the flexibility of each residue of RPA70N. Also, number of contacts, number of hydrogen bond and distance were calculated and analyzed for all complexes. Molecular dynamics simulations were done for three benchmark sets: reference set, top-selected set and carboxyl-containing set (Table 6 and Figure 4).

4.2.1 Reference Set

Firstly, reference complexes were submitted to 20 ns of the molecular dynamics simulations. The equilibration was achieved before the beginning of 2 ns. This is a typical pattern to the molecular dynamics simulations, with fluctuations never seen above the 0.3 nm for reference complexes (Figure 7). After the equilibration, the stability of reference complexes followed with the similar trend until the end of the simulation. The average RMSD values changed between 1.36 (Compound 21) and 1.74 (Compound 15) nm. Smaller deviations represent more stable protein structure. The average RMSD values for the complexes never exceeded 0.3 nm during the simulations. This result suggested that the ligands not affect the stability of the receptor inside the binding pocket during the simulation.

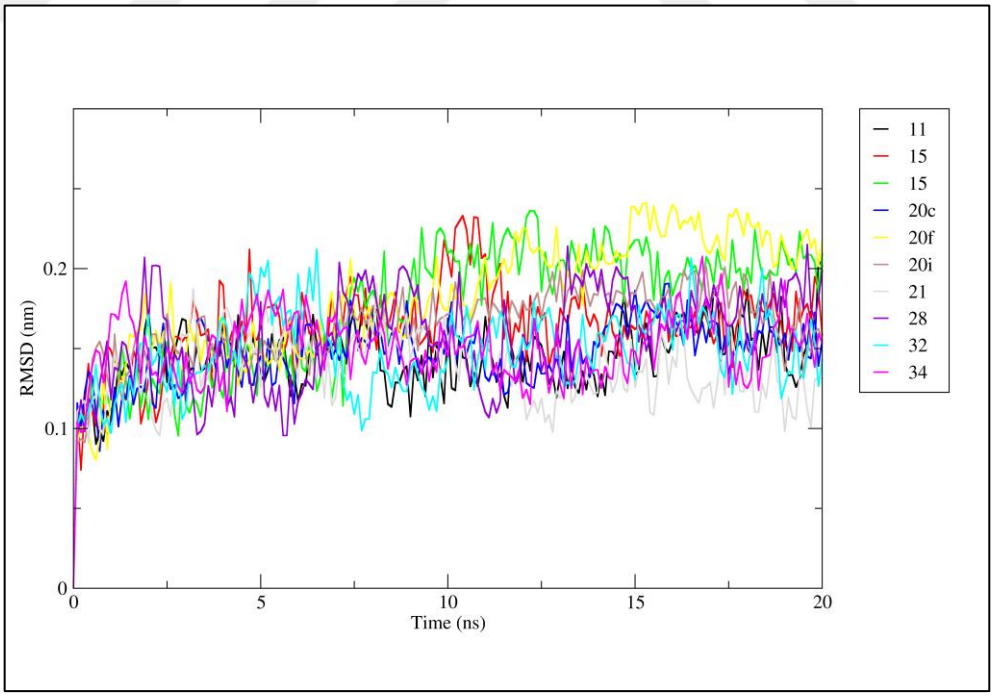

Figure 7: The RMSD of Cα carbon atoms of the protein with respect to the starting structure through 20 ns of molecular dynamics simulations for reference inhibitors.

Before the root mean square fluctuation (RMSF) analysis of side-chain atoms, the distance between the receptor and ligands was calculated to get information of the binding pocket in the receptor. According to this calculation, the binding pocket consists of 30-44, 54-61 and 80-98 amino acid residues (Figure 8). Among these residues, three amino acid residues (Arg31, Arg41 and, Arg43) are more critical for the binding and protein-protein interactions with RPA70 partners (28). All reference binders were situated in the same binding pocket during the simulation (Figure 8).

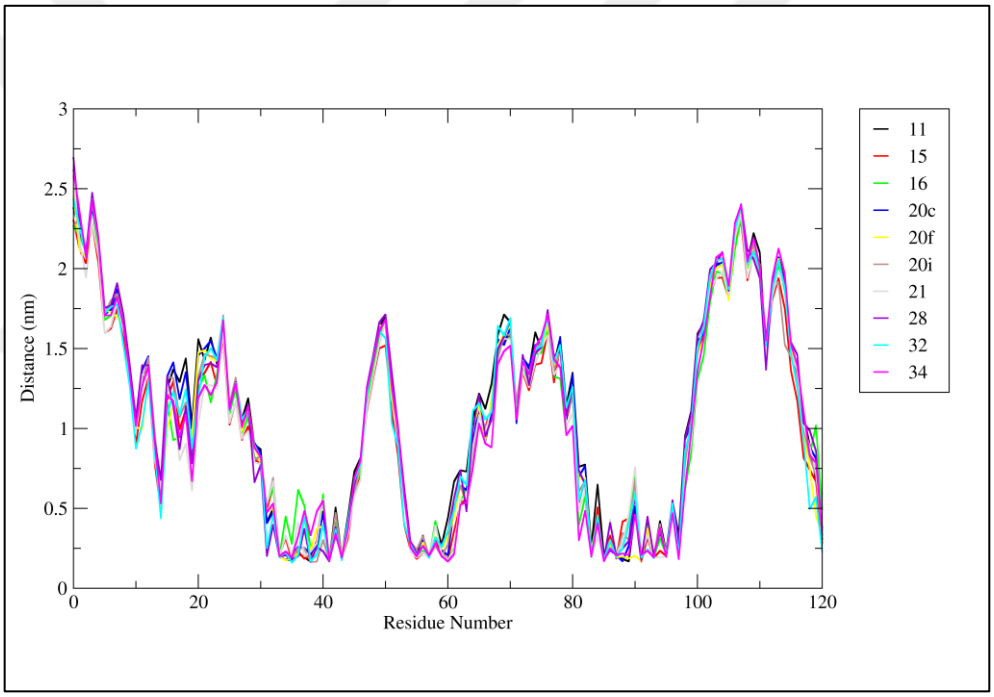

Figure 8: The average distance between the receptor and reference ligands for all residues during the molecular dynamics simulations.

Analysis of Root mean square fluctuation (RMSF) for the complexes give us information on the flexible regions of the protein. In proteins 3D structure, the loop, coils and turns showed higher fluctuation compared to the sheet and helical structures (99). The higher RMSF value pointed out the loosely organized structure, and the lower RMSF value indicated the well-organized structural regions (99). Figure 9 shows us RMSF plots for

reference inhibitors. According to Figure 9, all complexes had nearly same fluctuation pattern around the binding pocket residues.

Interestingly, compound 16 had more fluctuation in the region of 15-20 amino acid residues. Similarly, compound 20f also had higher fluctuation in a different region between 88-93 amino acid residues. The binding pocket residues, especially Arg31, Arg41 and, Arg43, rarely had RMSF value above 0.3 nm during the molecular dynamics simulations for all complexes. This indicated that the binding site of the protein was stable in the presence of the reference inhibitors and the protein-ligand complexes were acceptable for further analysis.

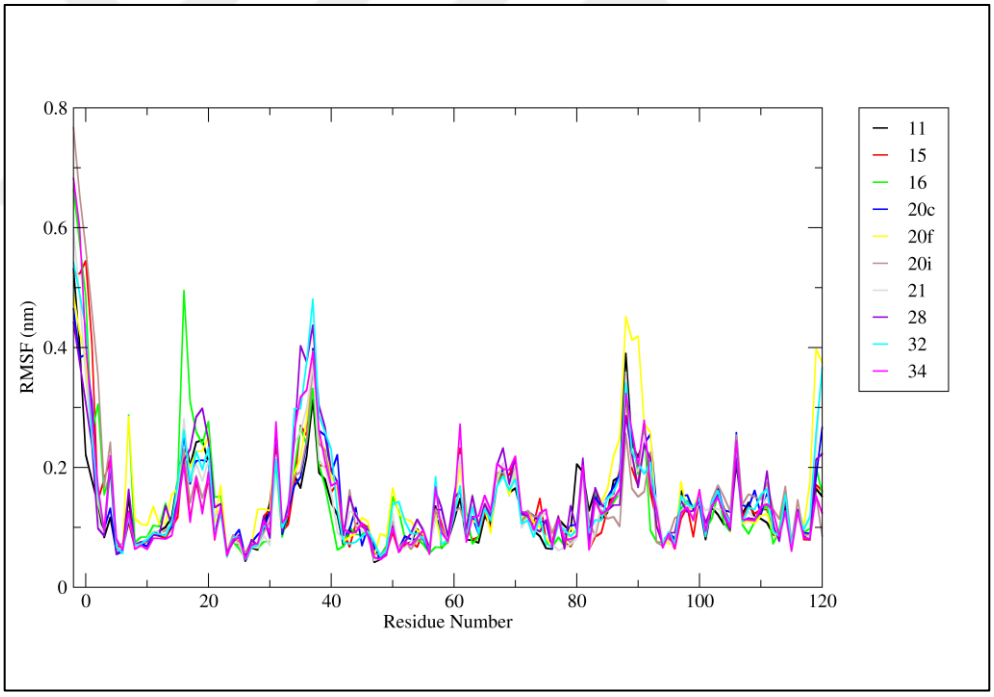

Figure 9: Root mean square fluctuation (RMSF) of side-chain atoms of the receptor for reference inhibitors.

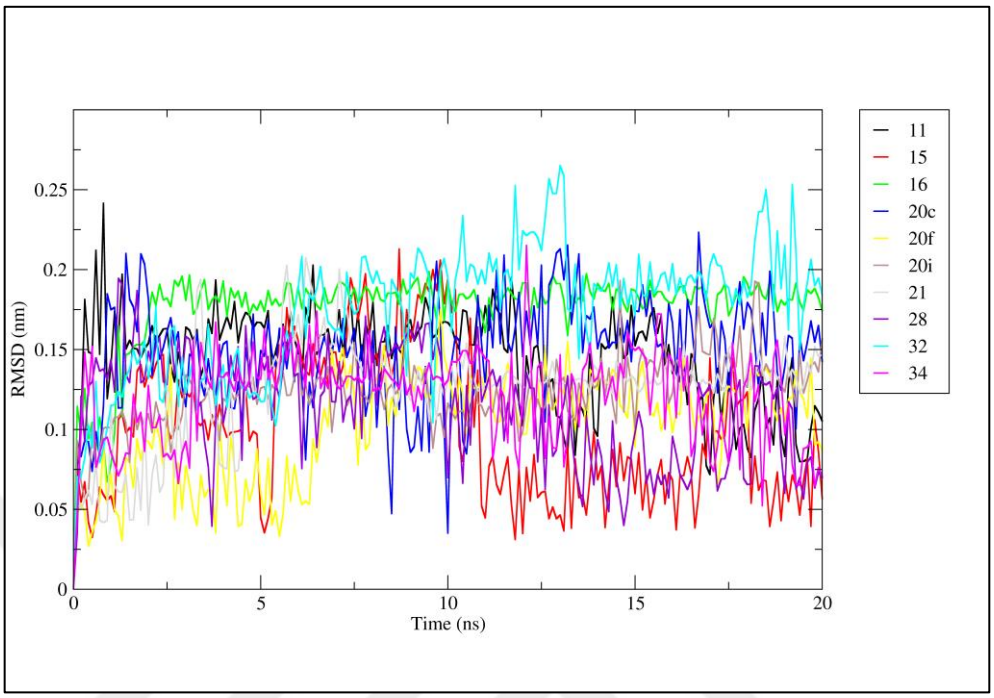

Figure 10: The RMSD of the reference ligand molecules with respect to the starting structure through 20 ns of molecular dynamics simulations.

The RMSD of reference inhibitors was presented in Figure 10, demonstrating that these ligand conformations nearly remained stable throughout the simulation for all ligands. RMSD values rarely exceeded 0.2 nm during the simulations (Figure 10). Also, average RMSD values of reference inhibitors distributed in a small range between 0.1 and 0.17 nm (Table 6). All these stability and fluctuation analysis concluded that nearly all reference structures remained stable throughout the simulation. In addition, interactions between RPA70N and inhibitors were considered, especially hydrogen bonds and contacts (< 0.6 nm). Hydrogen bonds in drug discovery are significant because of their influence on the drug-like properties. The number of hydrogen bonds was calculated and shown in Figure 11. The most number of hydrogen bonds was formed by compound 16 while the lowest number of hydrogen bonds was formed by compound 28. Compound 16 also has the highest number of contacts. Compound 20c is the best binder among the reference set, but it had the lowest number of close contacts compared to others (Figure 12 and Table 6).

Table 7: Average values of RMSD, number of hydrogen bonds and number of contacts for reference compounds throughout the simulation

Compound	11	15	16	20c	20f	20i	21	28	32	34
RMSD	0.14	0.10	0.17	0.14	0.10	0.12	0.12	0.11	0.17	0.12
H-Bonds	1.44	1.33	2.38	1.22	1	0.83	1.61	0.72	1.44	0.77
Contacts	1612	1728	1858	1389	1701	1843	1573	1438	1853	1528

Although the ligands were stable within the binding pocket throughout the molecular dynamics simulation, hydrogens bonds were not continuous and stable between protein and ligand complexes. There is no significant correlation between binding constants and hydrogen bonds numbers (Table 6). This suggested that the stability of ligands within the binding pocket would be contributed by other intra-molecular forces such as van der Waals and electrostatic forces. These will be explored in detail in the next section.

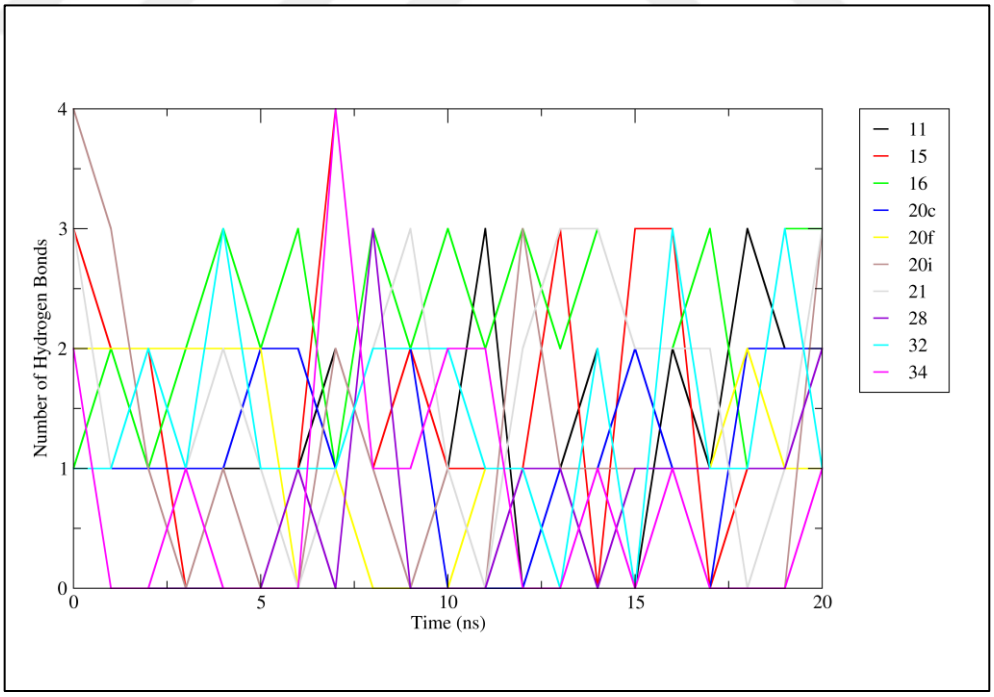

Figure 11: The number of hydrogen bonds during the simulations with 1 ns interval for the reference ligands.

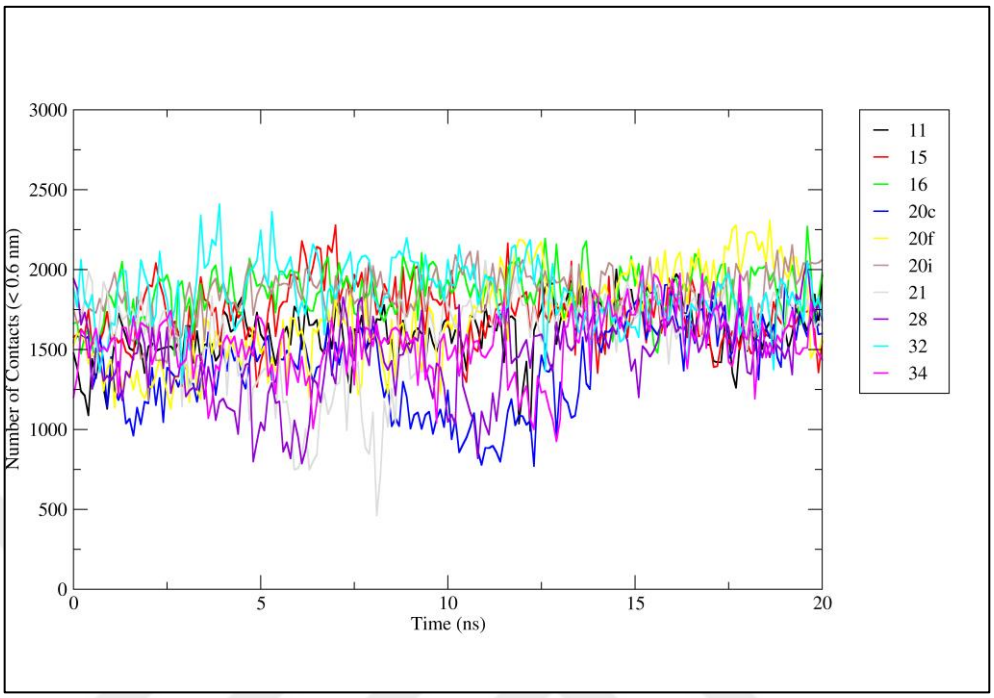

Figure 12: The number of contacts (< 0.6 nm) between molecules during the simulations for the reference ligands.

4.2.2 Top-selected Set

To examine the conformational changes of the RPA70N with the top-selected list, the root-mean-square deviation (RMSD) was calculated. Figure 13 shows the RMSD of Ca atoms of the protein as a function of time. The RMSD values that vary between 0.14 (Z.600) and 0.19 (Z.594) nm are nearly same with the reference set and rarely seen over 0.2 nm for all complexes during the simulation, indicating the structural stability of the complexes (Table 7). These selected ligands did not affect the stability of receptor inside the binding pocket during the simulation. Also, the average distance between amino acid residues of receptor and top-selected ligands was calculated to show residues involving in the binding pocket. Same as the reference ligands, these top-selected ligands were located around 30-43, 54-61 and 80-98 amino acid residues (Figure 14).

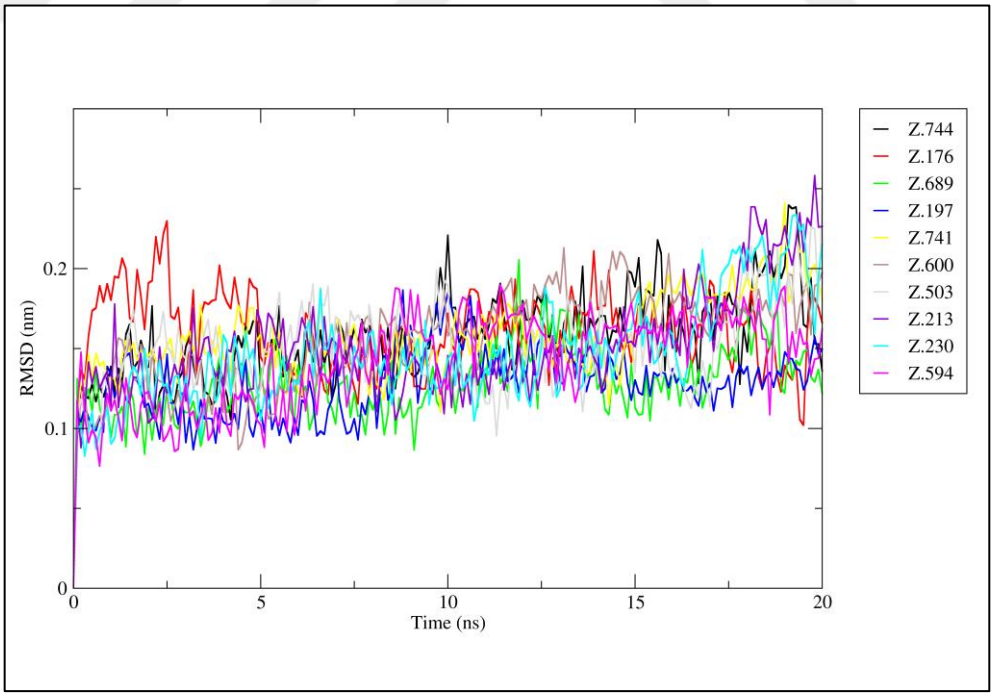

Figure 13: The RMSD of Cα carbon atoms of the protein with respect to the starting structure through 20 ns of molecular dynamics simulations for the top-selected potential inhibitors.

The RMSF values of side chain atoms against the amino acid residues for each ligand during the 20 ns trajectory data are displayed in Figure 15. Analysis of the RMSF values indicates that these top-selected ligands show nearly same fluctuation compared to the reference ligands. Only Z.741 affected side chains with more fluctuation between 67 and 73 residues. However, this fluctuation did not affect overall protein stability. These molecular simulation outputs clearly indicate that the binding orientation of top-selected ligands with RPA70N did not affect the protein stability during the simulation.

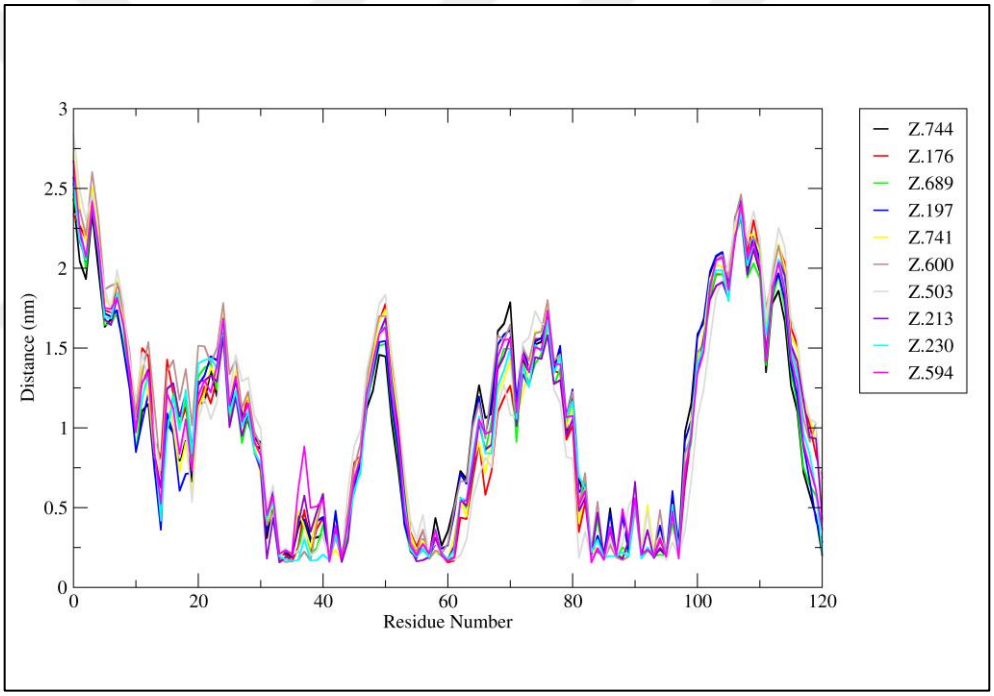

Figure 14: The average distance between the receptor and top-selected ligands for all residues during the molecular dynamics simulations.

The RMSD of top-selected potential inhibitors was presented in Figure 18, demonstrating that these ligand conformations nearly remained stable throughout the simulation for some top-selected ligands, but some of them did not remain very stable. Especially, Z.176 displayed more conformational change with 0.34 average RMSD value between 2 – 18 ns simulation period (Figure 16). However, this conformation change did not affect the

protein stability. Unlike, Z.744 and Z.197 ligands firmly remained stable throughout the simulation (Figure 18). For the others, RMSD values were around 0.2 nm during the simulations (Table 7).

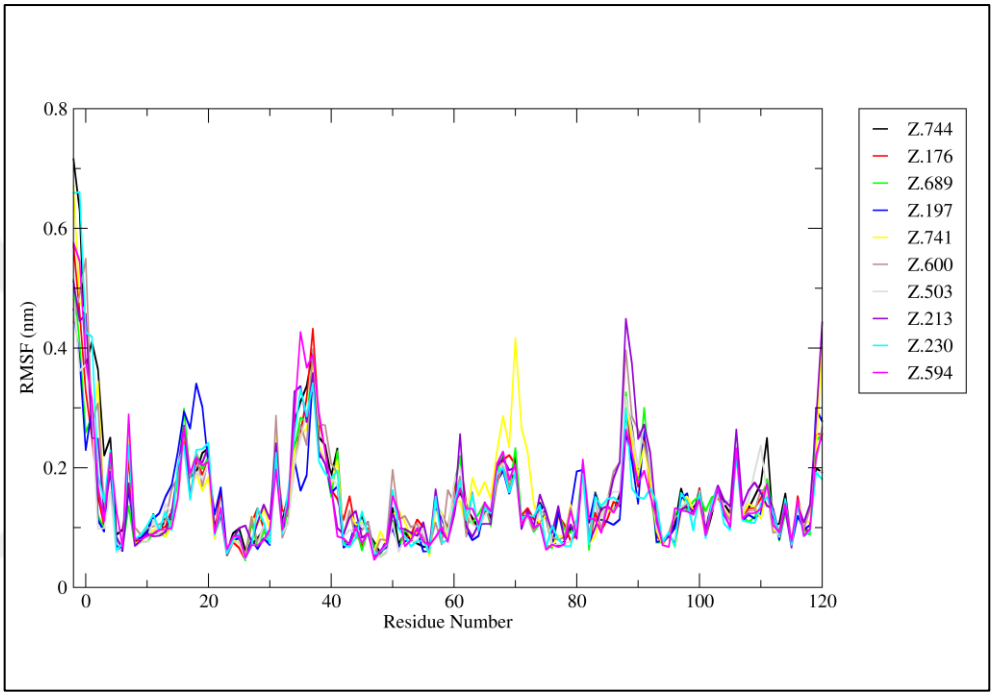

Figure 15: Root mean square fluctuation (RMSF) of side-chain atoms of the receptor for top-selected potential inhibitors.

The hydrogen bonds and close contacts between each compound of this set and RPA70N were examined (Figure 17 and Figure 18). Z.594 made more average number of hydrogen bonds during the simulation while Z.744 had the lowest average number of hydrogen bonds (Table 7). The number of hydrogen bonds also steadily continued for Z.594 during the simulation. Although Z.744 formed a lower number of hydrogen bonds, it had a nearly same number of close contacts with Z.594. The overall average number of hydrogen bond for the top-selected set was lower than the reference set. Z.689 and Z.230 interacted with RPA70N via a higher number of close contacts than the others. Z.503 had the lowest

number of close contacts, but it remained in the binding pocket using a high number of hydrogen bonds.

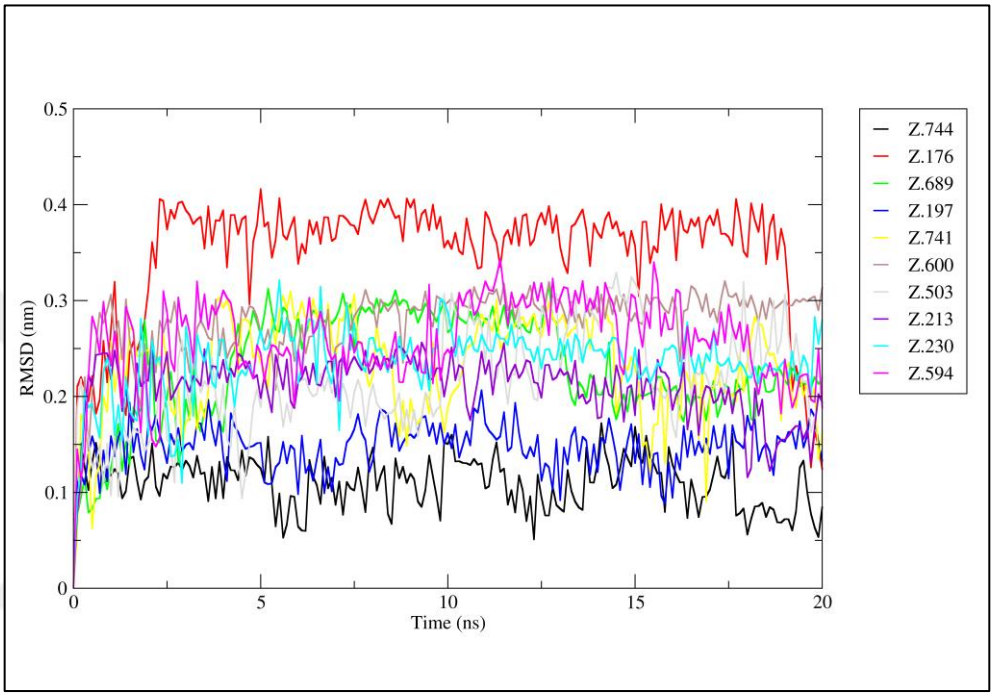

Figure 16: The RMSD of top-selected ligand molecules with respect to the starting structure through 20 ns of molecular dynamics simulations.

Table 8: Average values of RMSD, number of hydrogen bonds and number of contacts for the top-selected set throughout the simulation

Compound	**Z.744**	**Z.176**	**Z.689**	**Z.197**	**Z.741**	**Z.600**	**Z.503**	**Z.213**	**Z.230**	**Z.594**
RMSD	0.11	0.34	0.22	0.14	0.23	0.27	0.21	0.21	0.23	0.25
H-Bonds	0.33	1.14	0.76	0.47	1.04	1.04	1.71	1.95	1.61	2.09
Contacts	1827	1605	1942	1764	1222	1453	1134	1548	1972	1811

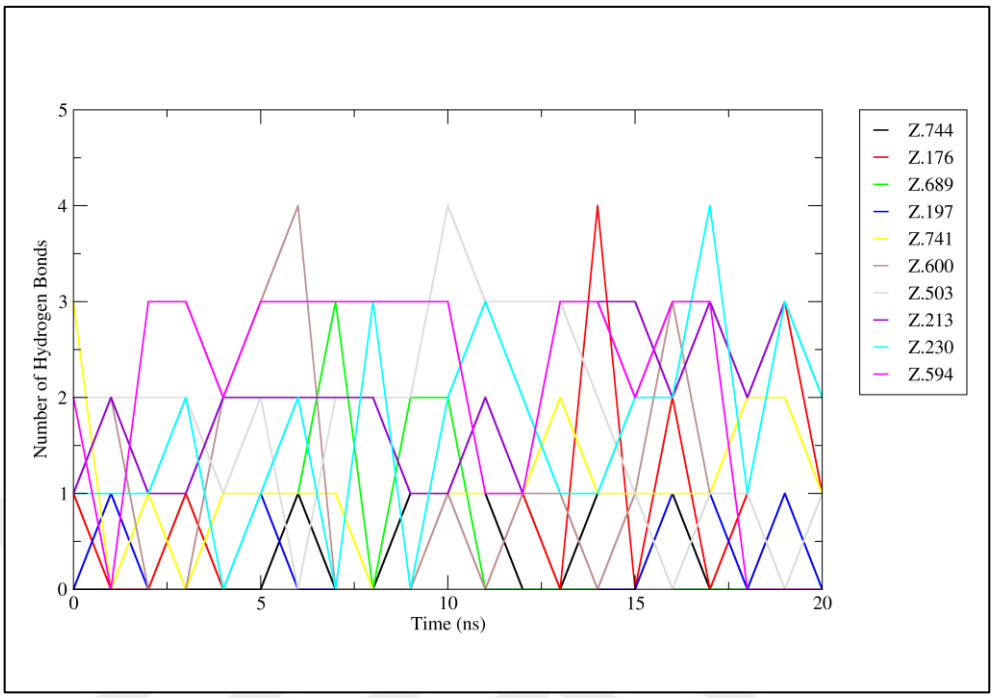

Figure 17: The number of hydrogen bonds during the simulations with 1 ns interval for top-selected ligands.

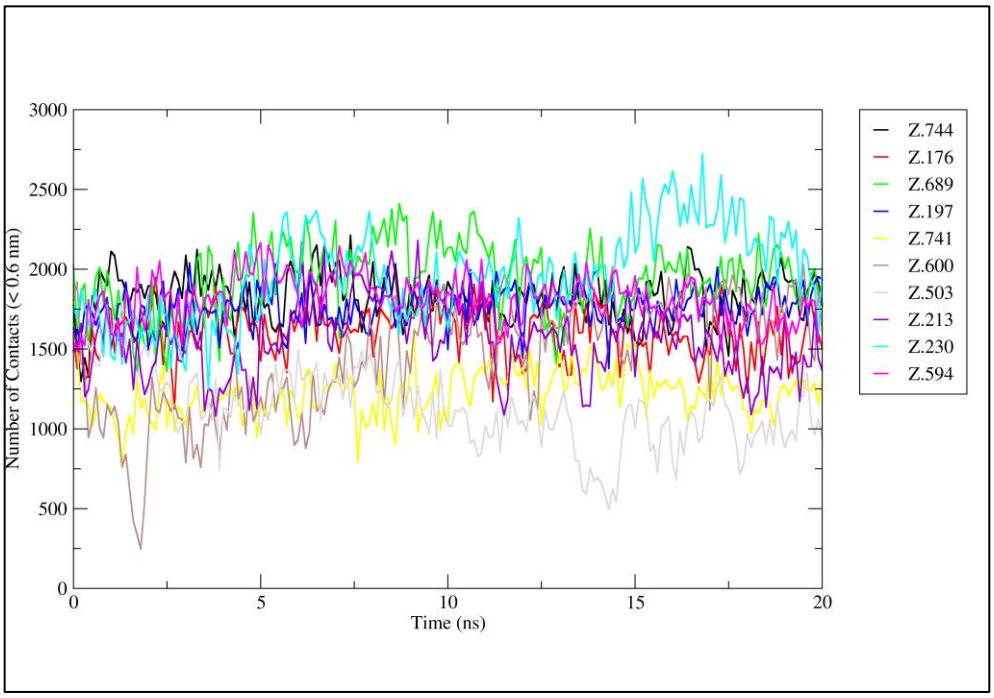

Figure 18: The number of contacts (< 0.6 nm) between molecules during the simulations for top-selected ligands.

4.2.3 Carboxyl-Containing Set

The carboxyl-containing set is the last simulation set for fishing potential inhibitors. Figure 19 presents the backbone alpha carbons RMSD of the RPA70N complexed to the carboxyl-containing potential inhibitors. Nearly all complexes remained stable throughout the simulation because of the fact that the complexes never exceeded 0.3 nm during the simulations (Figure 19). The average values of RMSD for all complexes changed between 0.14 and 0.19 nm, Z.795 and Z.163 respectively. In some complexes, RPA70N was less stable, especially Z.163 and Z.268. Ligand Z.163 affected the protein stability during 7 – 11 ns time period. But this is not very significant change on the protein stability. The average distance between protein and ligand was calculated and this calculation showed that the carboxyl-containing set shared same binding pocket with the reference set. This binding region contained 30-43, 54-61 and 80-98 amino acid residues (Figure 20).

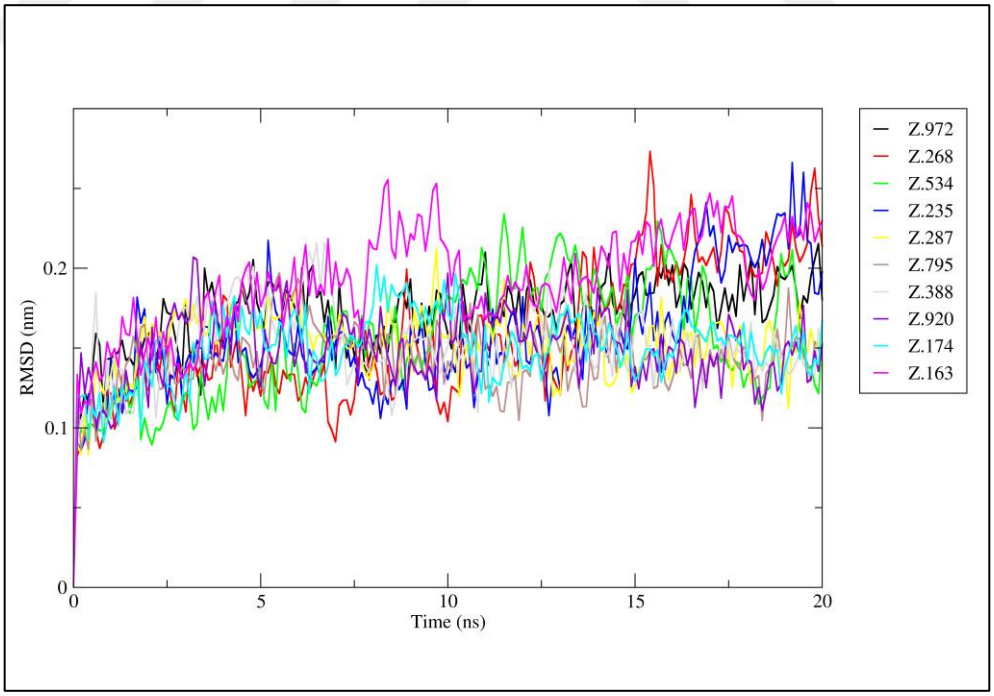

Figure 19: The RMSD of Cα carbon atoms of the protein with respect to the starting structure through 20 ns of molecular dynamics simulations for the carboxyl-containing potential inhibitors.

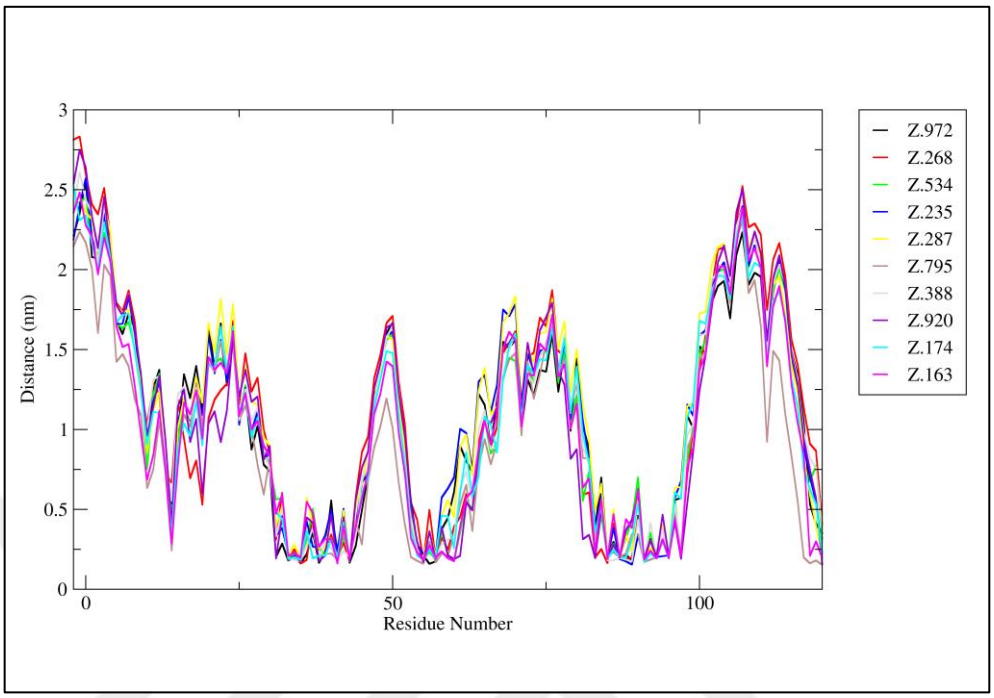

Figure 20: The average distance between the receptor and the carboxyl-containing ligands for all residues during the molecular dynamics simulations.

The RMSF values of side chain atoms against the amino acid residues of RPA70N during the 20 ns trajectory data are displayed in Figure 21. In the figure, the first and second binding regions had more RMSF compared to the reference set. This side chain fluctuation affected the protein stability more than the reference set. The RMSD of the carboxyl-containing potential inhibitors was presented Figure 24. Z.534 and Z.287 remained remarkably stable with the average RMSD of 0.06 and 0.09 nm throughout the simulation compared to the others (Table 8). Contrarily, Z.295 and Z.174 displayed higher fluctuation with the average 0.23 and 0.30 nm RMSD, respectively (Figure 18). Despite the high RMSD value for Z.163, it had very low the RMSF values of side chain atoms (Table 8 and Figure 21).

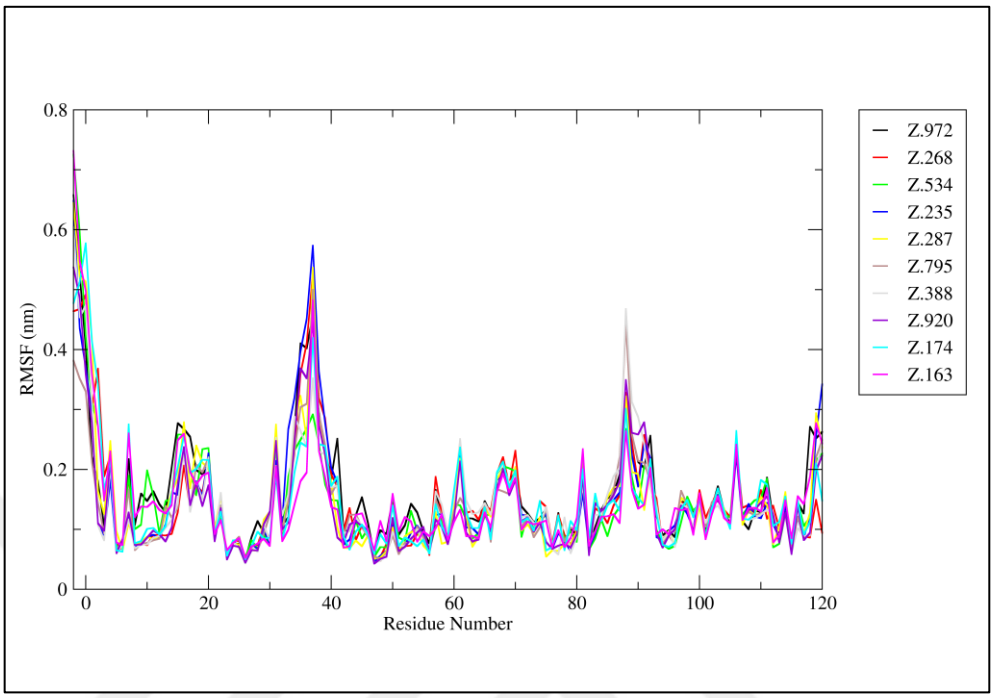

Figure 21: Root mean square fluctuation (RMSF) of side-chain atoms of the receptor for the carboxyl-containing potential inhibitors.

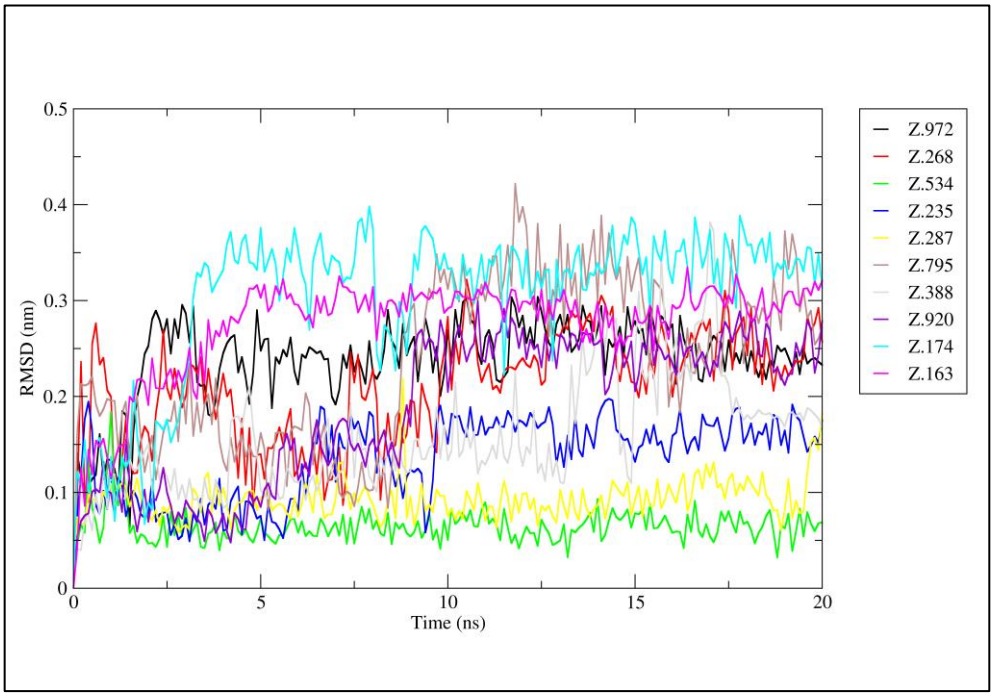

Figure 22: The RMSD of the carboxyl-containing ligand molecules with respect to the starting structure through 20 ns of molecular dynamics simulations.

Table 9: Average values of RMSD, hydrogen bonds and number of contacts for the carboxyl-containing set throughout the simulation

Compound	Z.972	Z.268	Z.534	Z.235	Z.287	Z.795	Z.388	Z.920	Z.174	Z.163
RMSD	0.23	0.21	0.06	0.13	0.09	0.23	0.15	0.18	0.30	0.27
H-Bonds	1.57	1.42	0.57	1.19	3.90	1.04	0.61	1.90	0.33	2.09
Contacts	1364	1361	1906	1445	1629	1803	1145	1433	1996	2395

The RMSD of the carboxyl-containing set was presented in Figure 18, demonstrating that these ligand conformations nearly remained stable throughout the simulation, but some of them did not remain very stable against to docked pose. Especially, Z.174 displayed more conformational change with 0.3 average RMSD value (Figure 22). However, like these conformational changes did not affect the protein stability. Beside all these dynamical analysis, intermolecular interactions were additionally analyzed. Z.287 had the most number of hydrogen bonds among these set while Z.163 possessed the most number of the close contacts (Table 8). Z.174 made the lowest average number of hydrogen bonds during the simulation while Z.174 had the higher average number of close contacts after Z.163 (Table 8).

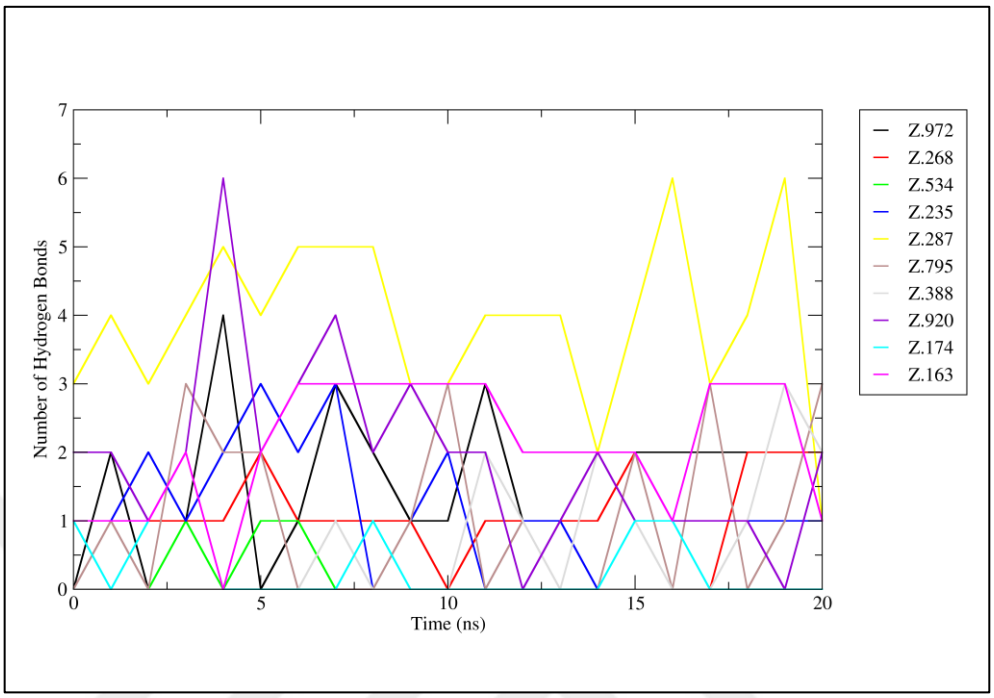

Figure 23: The number of hydrogen bonds during the simulations with 1 ns interval for the carboxyl-containing.

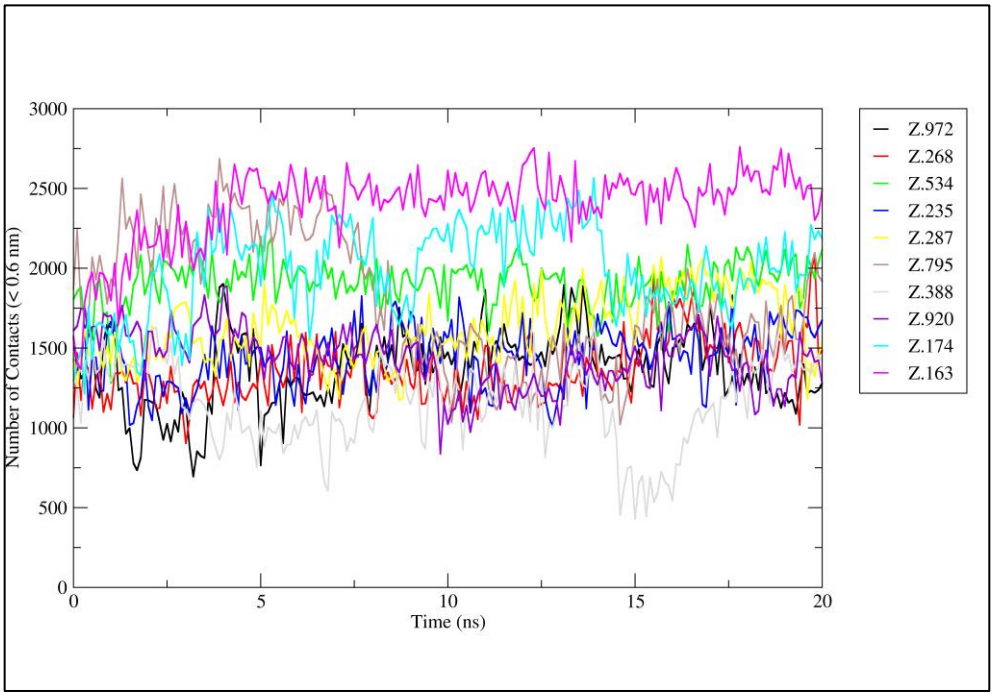

Figure 24: The number of contacts (< 0.6 nm) between molecules during the simulations for the carboxyl-containing ligands.

4.3 Binding Free Energy Calculations

The binding free energies were calculated using the MM/PBSA method to obtain more detailed information on the interactions between all ligands and RPA70N. Firstly, all of the 10 complexes of reference inhibitors were energy-minimized, equilibrated and then simulated for 20 ns using molecular dynamics simulations. The structures that were used in binding free energy calculations were collected at every 0.1 ns from the last 1 ns of 20 ns simulation. The results of the predicted binding free energies and energy components of the reference inhibitors are listed in Table 9.

Table 10: Binding energy values and individual component energy for the reference set calculated with the MM-PBSA method.

ID	K_d	ΔE_{vdW}	ΔE_{elec}	ΔG_{polar}	$\Delta G_{nonpolar}$	ΔG_{bind}
11	30	-160.5 +/11.5	-96.6 +/-24.0	144.1 +/-13.4	-16.7 +/-0.9	-129.7 +/-11.7
15	234	-138.9 +/-12.4	-42.8 +/-13.2	108.6 +/-10.8	-16.9 +/-0.8	-90.0 +/-10.7
16	95	-153.8 +/-15.7	-98.7 +/-16.3	124.0 +/- 9.8	-16.3 +/-1.3	-144.7 +/-14.1
20c	0.81	-174.0 +/-13.4	-111.5 +/-21.9	136.8 +/-11.9	-18.2 +/-0.6	-166.8 +/-12.8
20f	15	-179.5 +/-23.1	-145.4 +/-28.3	166.9 +/-32.9	-18.5 +/-1.5	-176.6 +/-10.0
20i	62	-197.3 +/-12.7	-132.6 +/-29.2	201.4 +/-14.8	-19.7 +/-0.9	-148.2 +/-23.6
21	150	-139.2 +/-10.8	-81.8 +/-23.8	113.1 +/-12.4	-15.6 +/-0.5	-123.4 +/-19.0
28	193	-139.1 +/-14.5	-68.9 +/-24.1	129.6 +/-13.3	-18.0 +/-1.9	-96.5 +/-15.0
32	126	-140.0 +/-13.4	-102.2 +/-24.7	138.3 +/-14.0	-16.4 +/-0.6	-120.3 +/-18.1
34	208	-122.4 +/-9.1	-42.0 +/-21.5	88.4 +/-13.5	-13.7 +/-0.9	-89.7 +/-18.6

MM-PBSA method was shown to produce well-correlated binding free energy values with the experimental binding constant (r=0.92). Thus, these results showed the validity of the dynamical analysis for calculation of binding free energy by MM-PBSA for the identification potential of RPA70N inhibitors. After this validation, the binding free energy for each ligand of the top-selected and the carboxyl-containing sets was calculated and listed in Table 10.

Table 11: Binding energy values and individual component energy for the top-selected and carboxyl-containing sets calculated with MM-PBSA method.

ZINC ID	ΔE_{vdW}	ΔE_{elec}	ΔG_{polar}	$\Delta G_{nonpolar}$	ΔG_{bind}
Z.744	-195.5 +/-6.4	-46.8 +/-19.5	115.5 +/-20.8	-21.0 +/-1.6	-147.9 +/-13.4
Z.176	-149.2 +/-14.3	-37.6 +/-39.9	102.5 +/-18.1	-17.9 +/-1.1	-102.3 +/-23.7
Z.689	-196.3 +/-10.8	-25.0 +/-15.3	111.5 +/-10.1	-21.2 +/-1.5	-131.0 +/-13.7
Z.197	-188.0 +/-14.7	-113.5 +/-15.2	163.7 +/-15.2	-20.9 +/-0.8	-158.9 +/-10.2
Z.741	-120.0 +/-9.1	-32.6 +/-15.2	90.1 +/-14.4	-14.5 +/-0.6	-77.0 +/-7.2
Z.600	-147.1 +/-14.5	-83.4 +/-30.4	121.6 +/-25.6	-15.8 +/-1.1	-124.8 +/-19.9
Z.503	-102.8 +/-12.5	-30.5 +/-17.4	106.9 +/-20.0	-12.8 +/-1.2	-39.3 +/-10.5
Z.213	-126.2 +/-12.8	-177.3 +/-45.2	179.5 +/-24.7	-15.5 +/-1.1	-139.6 +/-23.5
Z.230	-158.8 +/-14.5	-155.9 +/-30.2	184.8 +/-18.3	-18.5 +/-1.3	-148.5 +/-28.2
Z.594	-158.7 +/-9.9	-57.8 +/-17.4	115.2 +/-20.5	-17.6 +/-0.7	-119.0 +/-14.2
Z.972	-111.5 +/-10.4	-144.2 +/-12.7	134.0 +/-12.0	-12.2 +/-0.6	-133.9 +/-10.3
Z.268	-142.3 +/-22.4	-67.1 +/-29.6	118.1 +/-22.9	-16.9 +/-1.7	-108.3 +/-32.0
Z.534	-173.7 +/-11.2	-56.4 +/-13.4	134.6 +/-9.4	-19.4 +/-0.7	-115.0 +/-20.5
Z.235	-151.0 +/-16.5	-75.7 +/-20.6	146.5 +/-9.0	-16.3 +/-0.6	-96.6 +/-26.0
Z.287	-148.3 +/-15.4	-302.2 +/-37.1	283.1 +/-31.9	-17.9 +/-1.8	-185.2 +/-24.3
Z.795	-166.3 +/-16.7	-105.0 +/-68.0	174.6 +/-56.3	-18.1 +/-1.4	-114.7 +/-26.9
Z.388	-129.0 +/-5.6	-56.5 +/-15.8	116.5 +/-40.7	-15.4 +/-0.6	-84.4 +/-48.7
Z.920	-125.7 +/-8.9	-66.3 +/-38.8	133.3 +/-26.9	-15.3 +/-1.6	-74.1 +/-22.3
Z.174	-175.6 +/-21.4	-68.6 +/-23.2	129.3 +/-16.3	-19.4 +/-0.6	-134.4 +/-22.2
Z.163	-214.6 +/-12.6	-135.1 +/-15.2	178.5 +/-7.3	-21.9 +/-0.5	-193.2 +/-13.6

The top-selected list did not contain any ligand with the higher binding free energy value than the reference list but 3 of them can be evaluated as a potential inhibitor because of the high value of binding energy. The carboxyl-containing list possessed the most potent inhibitors according to their higher binding free energy value than the reference list. Z.163 showed the best binding energy with -193.2 kJ/mol while compound 11 and 20c showed -129.7 kJ/mol and -166.8 kJ/mol, respectively. Based on the contribution of individual energy components to the free binding energy, the van der Waal energy component contributed to Z.163 more than all others whereas electrostatic energy component contributed to Z.287 more than all others for the binding affinity. Z.287 had the highest positive enrichment of polar solvation energy.

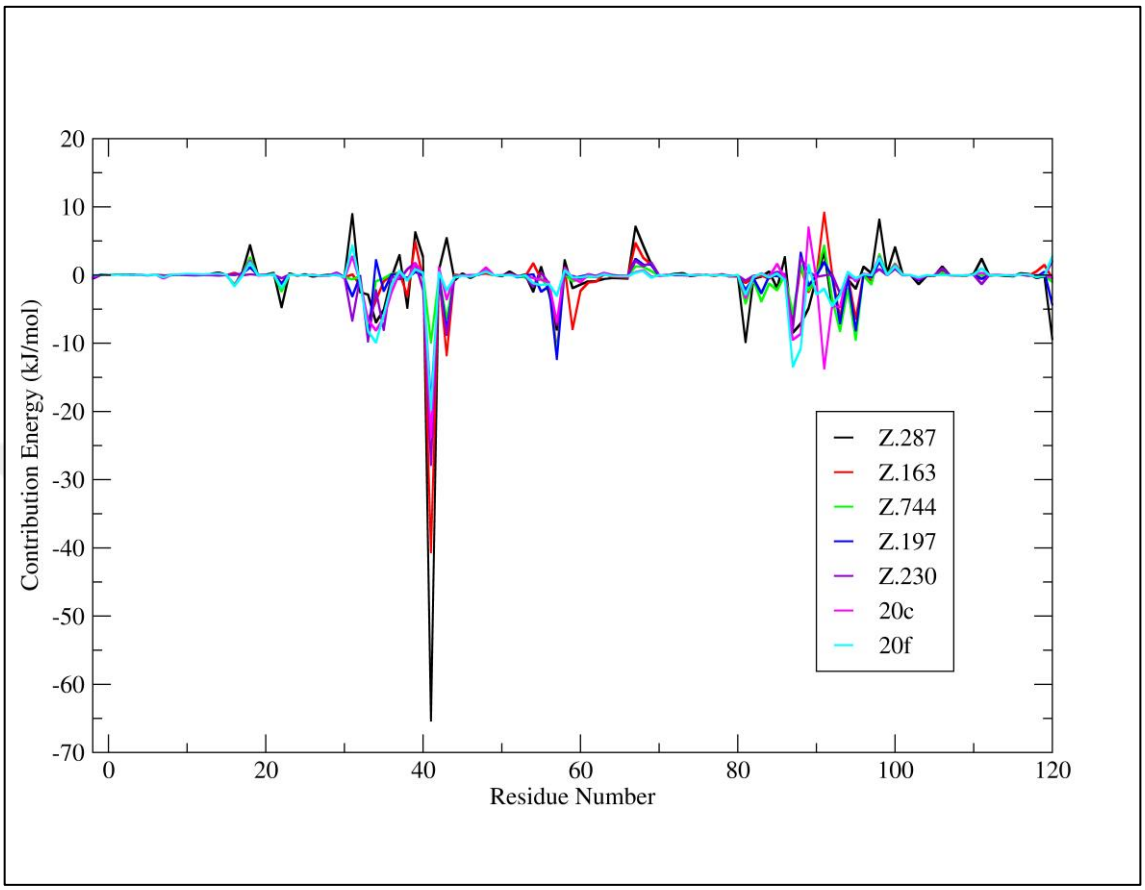

Figure 25: Contribution of each RPA70N amino acid to the binding free energy of the ligands.

To identify the key residues involved in ligand binding, the residue wise energy decomposition was plotted for the ligands displayed the highest binding free energy in Figure 25. As seen the figure, it can be observed that some residues showed a positive binding affinity while some of them showed negative binding affinity. ARG41 showed the highest binding affinity as compare to other residues. The results concluded that ARG41 play critical role in protein ligand interactions. The residues that most interacted were ILE33, THR34, GLY35, ARG41, ARG43, MET57, ARG81, LEU87, ARG91, and VAL93.

4.4. ADME-Toxicity Analysis

Table 12: Drug-likeness and toxicity analysis of ligands

ID	Lipinski	Ghose	Veber	Egan	Muegge	Total Violations*	Toxicity
20c	2	2	0	1	1	6	OK
Z.744	2	3	1	2	2	10	FAIL
Z.197	1	3	1	2	3	10	FAIL
Z.230	0	1	1	1	0	3	FAIL
Z.287	2	3	2	2	2	11	OK
Z.163	0	1	0	0	0	1	OK

*Total violations; the sum of violations from five rule-based filters.

The high absorption and well distribution of a drug in efficient time are important for its effective metabolism and action. In addition, the toxicity that often overshadows preferred ADME behaviour is another important consideration (85). Computational prediction of drug-likeness and toxicity at the drug discovery could provide opportunities for accelerating lead compounds. Five rule-based filters were used in order to ensure the improved quality of the inhibitors for future lead optimization. The ADME properties of the ligands were evaluated from its pharmacokinetics. Toxicity analysis perform searching for substructures commonly found in toxic and promiscuous ligands. According to this searching, only two potential inhibitors displayed non-toxic properties (Table 11). Although Z.287 displayed non-toxic properties, it had worse drug-likeness than the reference. Z.163 is the best candidate as an inhibitor of RPA70N because of its ADMET properties (Figure 26).

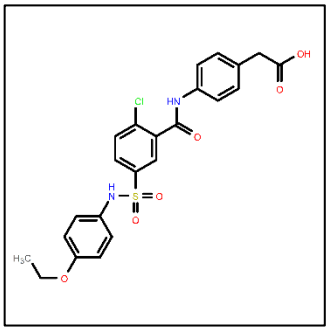

Figure 26: Chemical structure of ZINC000753854163 that is the best candidate as an inhibitor of RPA70N

5. DISCUSSION

The rationale behind conventional cancer therapy relies on the fact that cancer cells accumulate higher DNA damage than normal cells, and as a result, they are more susceptible to DNA damaging agents. Despite many efforts in the development of targeted drugs, DNA damaging therapy is still associated with considerable side effects, highlighting the necessity of identification of novel lead molecules with higher selectivity. One plausible way to induce DNA damage is to impair DNA damage response (DDR) pathways, in which Replication Protein A (RPA) plays a crucial part. Particularly, targeting the protein-protein interactions (PPIs) formed between RPA70N and other DDR mediators has become an intriguing area of research populated with a number of studies allocating different methods ranging from high throughput screening approaches to fragment-based drug design tools (28). Although these methods are robust, for screening large libraries virtual screening approaches may present valid alternatives to such experimental tools. Here we report the comprehensive screening of the large database, ZINC15 to unveil potential inhibitors of RPA70N and also document the comparison of the identified molecules with the previously identified potent inhibitors by means of binding affinity and drug-likeness. Our *in silico* results underline the presence of potential inhibitors that showed better inhibition capacity and drug-likeness features than previously known inhibitors.

A number of small molecules have been reported to inhibit RPA70N (29). Among these inhibitors, fumaropimaric acid and HAMNO, which are among the first generation inhibitors of RPA70N have been identified in HTS studies (28,30). Fragment-based screening approaches have been further applied and upon screening a 15,000-member fragment library, novel ligands have been identified, impairing particularly RPA70N-ATRIP interactions (31). Subsequently these compounds have been optimized to bind at submicromolar levels (32), albeit these group of inhibitors displays unfavorable drug-

likeness features. Other inhibitors based on anthranilic acid derivatives have been also identified by HTS and optimized by iterative medicinal chemistry and structure-based design techniques. These efforts ultimately led to the discovery of 20c which displays a high binding affinity (812 nM) and promising permeability, solubility characteristics to be used as a drug (33). This ligand, 20c, mainly because of its well established inhibitory effect at the submicromolar concentrations and superior drug-likeliness features than any other inhibitors stood out to be the most promising lead molecule for targeting RPA70N. Thus we have used 20c (34) as the reference ligand in our study. The ligands that show better binding affinity and drug-likeness were considered potential lead molecules.

Ligand-based screening is widely used for detecting similarity between small molecules. However, there is no universal strategy for quantification of the similarity.[73–75] Here we have applied multiple algorithms to collect similar ligands to 20c from the ZINC12 database. Among the similarity search methods used here, LiSiCA screened 3D structure (69), while ShaEP was used to assess the similarity of each compound in ligand library with the reference molecule by evaluating the similarity as means of spatial overlap in volume and electrostatic potentials (71). LIGSIFT evaluated the global shape of the ligands, was used to calculate the 3D shape and chemical similarities of the reference molecule with the ligand library (70). SHAFTS was applied to integrate the strength of both the pharmacophore matching and volumetric similarity approaches, as this method uses a hybrid similarity search of molecular shape and labeled chemical groups annotated as pharmacophore features for the 3D similarity calculation and ranking (72). By combining different similarity search methods, we ensured reaching all of the ligands similar to 20c (2581 ligands) in the ZINC12 database.

Structure-based screening has been widely used for drug discovery, particularly for identification of novel lead compounds when the 3D structure of the target is available (100–106). The docking program and/or free binding energy prediction method used

during structure-based screening may greatly affect the screening results. Considering the high number of the ligands to be screened by structure-based tools used in this study, application of multiple programs would be impractical and computationally costly. Nevertheless, it is critical to evaluate the performance of different programs prior to screening large databases. Therefore, we have established benchmark ligand sets to test multiple structure-based tools prior to screening.

Many docking programs have been developed and reported to be successful for a variety of protein-ligand pairs, still accurate prediction of the binding affinities and poses is a challenging task for docking (107–111). For instance, a given docking program may assign an unfavorable score to an active compound, while it predicts a favorable score for an inactive compound. Moreover, individual programs often show better ability to predict high-affinity compounds for particular receptors or ligand classes (111,112). Thus, we initially aimed to select the best-optimized docking program to apply a larger database. For assessment of docking programs, 5 widely-used programs were used to score 20 different inhibitors with experimentally known binding affinities. Only LeDock gave well-correlated (0.745 ± 0.08) binding scores with the experimental values (Table 2), while both PLANTS and LeDock were successful in the reproduction of the crystal pose (Table 4). LeDock generates these well-correlated binding scores using the evolutionary algorithm that is adopted in combination with simulated annealing search, which is used to generate the first generation of docking poses. Although other well-known and highly cited docking programs such as Autodock and DOCK6 were also employed, we note that they did not perform as well as LeDock and PLANTS did in predicting the active ligands targeting RPA70N. Overall, we stress that LeDock produced a remarkably better performance than the most widely used and cited docking programs (5). These results reinforce the necessity of assessment of multiple docking programs with the same rigor prior to screening large databases.

During dynamical analysis, we used the first benchmark set, due to the fact that the second ligand set with known 3D structures are much larger ligands than those found in the first benchmark set. Their relatively larger size makes molecular dynamics applications challenging. Moreover, the drug-likeness properties of the second set are very unfavorable compared with the first set. On the other hand, the first benchmark set including 20c are smaller in size and have favorable ADME and toxicity scores (Table 11). Therefore, we used the first benchmark set to be tested in the molecular dynamics simulations coupled with free energy predictions by MM-PBSA. This benchmark set also included the crystal structure (PDB ID: 5E7N) harboring the compound 11, which produced a similar complex to 20c at the end of 20 ns of simulation. Overall, we observed an excellent agreement between the free energy predictions and experimental values (Table 9). Highlighting the validity of binding free energy prediction tool on this benchmark ligand set, we have applied molecular dynamics simulations and the MM-PBSA method to the selected RPA70N-ligand complexes.

Structural analysis of the RPA70N structure showed that the binding cleft possessed critical positive charged amino acids. Essentially, R41, R43 and R91 stood out to play an important role in inhibition of RPA70N (28). Dynamical analysis performed here also underscored the importance of these amino acids for the reference ligand and RPA70N interactions. Moreover, we stress that N87 and K88 as being other critical amino acids in the vicinity of the reference ligands with notable contributions to the binding. Therefore, we note that along with previously identified positively charged amino acids (R41, R43, R91), N87 and K88 may in part mediate ligand binding, reflecting its significance for prospective studies concerning intermolecular interactions of RPA70N.

Inspection of the selected molecules by only LeDock scores revealed that most of the molecules do not contain any negatively charged groups that can interact with the positively charged amino acids. From this respect, we further refined our search by adding

the criterion that ensured the presence of at least one carboxyl group which might potentiate binding to the positively charged RPA70N cleft. In line with this appraisal, the ligands (Z.163 and Z.287) that were selected according to this criterion, stood out among others (Figure 25). According to the stability analysis of RPA70N-ligand complexes, Z.163 and Z.287 showed much higher binding affinities than any of the reference inhibitors used here. Further structural inspections revealed that these ligands were also the most potent inhibitors with the highest number of contacts (Table 8) and hydrogen bonds (Figure 23) than any of the ligands including the reference inhibitors. More importantly, these carboxyl-containing ligands interacted with RPA70N, particularly from the R41, R43, R91 and K88. These interactions were shown to be more potent than those formed by 20c. Thus, the ligands identified by this study (Z.163 and Z.287) carry a high potential for targeting RPA70N. Owing to the fact that the Z.163 has a better ADME and toxicity profiles than Z.287 (Table 11), we report Z.163 as being the best potential inhibitor for targeting RPA70N.

6. CONCLUSION

Identification of novel RPA70N inhibitors will lead to new routes of drug discovery which might result in more potent therapies with reduced side effects. This study employed the best-optimized methods of docking and binding free energy prediction for screening small molecules that target RPA70N. Given these validated methods, we appraise that two ligands that were identified here represent the most potent inhibitors found in the ZINC12 and ZINC15 databases including all of the inhibitors tested here. In conclusion these series of *in silico* approaches concluded that Z.163 carry high potential to be developed as a potential chemotherapeutic for targeting the PPIs formed between RPA70N and other DDR proteins.

7. BIBLIOGRAPHY

1. Wold MS. Replication protein A: A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism. Annu Rev Biochem. 1997;66:61–92.

2. Balajee AS, Geard CR. Replication protein A and gamma-H2AX foci assembly is triggered by cellular response to DNA double-strand breaks. Exp Cell Res [Internet]. 2004;300(2):320–34.

3. Wu X, Shell SM, Zou Y. Interaction and colocalization of Rad9/Rad1/Hus1 checkpoint complex with replication protein A in human cells. Oncogene [Internet]. 2005;24(29):4728–35.

4. Kumaran S, Kozlov AG, Lohman TM. Saccharomyces cerevisiae replication protein A binds to single-stranded DNA in multiple salt-dependent modes. Biochemistry. 2006;45(39):11958–73.

5. Kim C, Paulus BF, Wold MS. Interactions of Human Replication Protein A with Oligonucleotides. Biochemistry. 1994;33(47):14197–206.

6. Iftode C, Daniely Y, Borowiec JA. Replication protein A (RPA): The eukaryotic SSB. Vol. 34, Critical Reviews in Biochemistry and Molecular Biology. 1999. p. 141–80.

7. Haring SJ, Mason AC, Binz SK, Wold MS. Cellular functions of human RPA1: Multiple roles of domains in replication, repair, and checkpoints. J Biol Chem. 2008;283(27):19095–111.

8. Wang Y, Putnam CD, Kane MF, Zhang W, Edelmann L, Russell R, et al. Mutation in Rpa1 results in defective DNA double-strand break repair, chromosomal instability and cancer in mice. Nat Genet [Internet]. 2005;37(7):750–5.

9. Santocanale C, Neecke H, Longhese MP, Lucchini G, Plevani P. Mutations in the

gene encoding the 34 kDa subunit of yeast replication protein A cause defective S phase progression. J Mol Biol. 1995;254(4):595–607.

10. Murzin G. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J [Internet]. 1993;12(3):861–7.

11. Brill SJ, Bastin-Shanower S. Identification and characterization of the fourth single-stranded-DNA binding domain of replication protein A. Mol Cell Biol. 1998;18(12):7225–34.

12. Bochkarev A, Bochkareva E. From RPA to BRCA2: Lessons from single-stranded DNA binding by the OB-fold. Vol. 14, Current Opinion in Structural Biology. 2004. p. 36–42.

13. Flynn RL, Zou L. Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians. Crit Rev Biochem Mol Biol. 2010;45(4):266–75.

14. Arunkumar AI, Stauffer ME, Bochkareva E, Bochkarev A, Chazin WJ. Independent and Coordinated Functions of Replication Protein A Tandem High Affinity Single-stranded DNA Binding Domains. J Biol Chem. 2003;278(42):41077–82.

15. Bochkareva E, Korolev S, Lees-Miller SP, Bochkarev A. Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. EMBO J. 2002;21(7):1855–63.

16. Matsuda T, Saijo M, Kuraoka I, Kobayashi T, Nakatsu Y, Nagai A, et al. DNA repair protein XPA binds replication protein A (RPA). J Biol Chem. 1995;270(8):4152–7.

17. Sugiyama T, New JH, Kowalczykowski SC. DNA annealing by Rad52 Protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. Proc Natl Acad Sci [Internet]. 1998;95(11):6049–54.

18. Xu X, Vaithiyalingam S, Glick GG, Mordes DA, Chazin WJ, Cortez D. The Basic Cleft of RPA70N Binds Multiple Checkpoint Proteins, Including RAD9, To Regulate ATR Signaling. Mol Cell Biol [Internet]. 2008;28(24):7345–53.

19. Lee SE, Moore JK, Holmes A, Umezu K, Kolodner RD, Haber JE. Saccharomyces Ku70, Mre11/Rad50, and RPA proteins regulate adaptation to G2/M arrest after DNA damage. Cell. 1998;94(3):399–409.

20. Jackson SP, Bartek J. The DNA-damage response in human biology and disease. Nature [Internet]. 2010;461(7267):1071–8.

21. Wu CS, Ouyang J, Mori E, Nguyen HD, Maréchal A, Hallet A, et al. SUMOylation of ATRIP potentiates DNA damage signaling by boosting multiple protein interactions in the ATR pathway. Genes Dev. 2014;28(13):1472–84.

22. Flynn RL, Zou L. ATR: A master conductor of cellular responses to DNA replication stress. Vol. 36, Trends in Biochemical Sciences. 2011. p. 133–40.

23. Glanzer JG, Liu S, Oakley GG. Small molecule inhibitor of the RPA70 N-terminal protein interaction domain discovered using in silico and in vitro methods. Bioorganic Med Chem. 2011;19(8):2589–95.

24. Sakthivel KM, Hariharan S. Regulatory players of DNA damage repair mechanisms: Role in Cancer Chemoresistance. Biomed Pharmacother [Internet]. 2017;93:1238–45.

25. Hosoya N, Miyagawa K. Targeting DNA damage response in cancer therapy. Cancer Sci. 2014;105(4):370–88.

26. Bochkareva E, Frappier L, Edwards a M, Bochkarev a. The RPA32 subunit of human replication protein A contains a single-stranded DNA-binding domain. J Biol Chem. 1998;273(7):3932–6.

27. Dou H, Huang C, Singh M, Carpenter PB, Yeh ETH. Regulation of DNA Repair through DeSUMOylation and SUMOylation of replication protein A complex. Mol

Cell. 2010;39(3):333–45.

28. Patrone JD, Waterson AG, Fesik SW. Recent advancements in the discovery of protein–protein interaction inhibitors of replication protein A. Med Chem Commun [Internet]. 2017;8(2):259–67.

29. Pearl LH, Schierz AC, Ward SE, Al-Lazikani B, Pearl FMG. Therapeutic opportunities within the DNA damage response. Nat Rev Cancer [Internet]. 2015;15(3):166–80.

30. Santivasi WL, Xia F. Ionizing Radiation-Induced DNA Damage, Response, and Repair. Antioxid Redox Signal [Internet]. 2014;21(2):251–9.

31. Glanzer JG, Liu S, Wang L, Mosel A, Peng A, Oakley GG. RPA inhibition increases replication stress and suppresses tumor growth. Cancer Res. 2014;74(18):5165–72.

32. Patrone JD, Kennedy JP, Frank AO, Feldkamp MD, Vangamudi B, Pelz NF, et al. Discovery of protein-protein interaction inhibitors of replication protein A. ACS Med Chem Lett. 2013;4(7):601–5.

33. Frank AO, Feldkamp MD, Kennedy JP, Waterson AG, Pelz NF, Patrone JD, et al. Discovery of a potent inhibitor of replication protein A protein-protein interactions using a fragment-linking approach. J Med Chem. 2013;56(22):9242–50.

34. Patrone JD, Pelz NF, Bates BS, Souza-Fagundes EM, Vangamudi B, Camper D V., et al. Identification and Optimization of Anthranilic Acid Based Inhibitors of Replication Protein A. ChemMedChem. 2016;11(8):893–9.

35. Huang HJ, Yu HW, Chen CY, Hsu CH, Chen HY, Lee KJ, et al. Current developments of computer-aided drug design. Vol. 41, Journal of the Taiwan Institute of Chemical Engineers. 2010. p. 623–35.

36. Lagorce D, Reynes C, Camproux AC, Miteva MA, Sperandio O, Villoutreix BO. In Silico ADME/Tox Predictions. In: ADMET for Medicinal Chemists: A Practical

Guide. 2010. p. 29–124.

37. Yu H, Adedoyin A. ADME-Tox in drug discovery: Integration of experimental and computational technologies. Vol. 8, Drug Discovery Today. 2003. p. 852–61.

38. Agarwal AK, Fishwick CWG. Structure-based design of anti-infectives. Ann N Y Acad Sci. 2010;1213(1):20–45.

39. Song CM, Lim SJ, Tong JC. Recent advances in computer-aided drug design. Vol. 10, Briefings in Bioinformatics. 2009. p. 579–91.

40. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: A free tool to discover chemistry for biology. Vol. 52, Journal of Chemical Information and Modeling. 2012. p. 1757–68.

41. Sterling T, Irwin JJ. ZINC15–Ligand Discovery for Everyone. J Chem Inf Model. 2015;55(11):2324–37.

42. Tanabe M, Kanehisa M. Using the KEGG database resource. Curr Protoc Bioinforma. 2012;(SUPPL.38).

43. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Vol. 27, Nucleic Acids Research. 1999. p. 29–34.

44. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: A knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res. 2008;36(SUPPL. 1).

45. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem substance and compound databases. Nucleic Acids Res. 2016;44(D1):D1202–13.

46. Pence HE, Williams A. ChemSpider database. J Chem Educ. 2010;87:1123– 1124.

47. Murakami Y, Omori S, Kinoshita K. NLDB: a database for 3D protein–ligand interactions in enzymatic reactions. J Struct Funct Genomics. 2016;17(4):101–10.

48. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res. 2016;44(D1):D1045–53.

49. Blaney J. A very short history of structure-based design: How did we get here and where do we need to go? Vol. 26, Journal of Computer-Aided Molecular Design. 2012. p. 13–4.

50. Mandal S, Moudgil M, Mandal SK. Rational drug design. Vol. 625, European Journal of Pharmacology. 2009. p. 90–100.

51. Wilson GL, Lill M a. Integrating structure-based and ligand-based approaches for computational drug design. Future Med Chem [Internet]. 2011;3(6):735–50.

52. Fang Y. Ligand–receptor interaction platforms and their applications for drug discovery. Expert Opin Drug Discov [Internet]. 2012;7(10):969–88.

53. Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des [Internet]. 2011;7(2):146–57.

54. Tian S, Wang J, Li Y, Li D, Xu L, Hou T. The application of in silico drug-likeness predictions in pharmaceutical research. Vol. 86, Advanced Drug Delivery Reviews. 2015. p. 2–10.

55. Gohlke H, Klebe G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Vol. 41, Angewandte Chemie - International Edition. 2002. p. 2644–76.

56. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. Acc Chem Res. 2000;33(12):889–97.

57. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. J Am

Chem Soc. 1998;120(37):9401–9.

58. Sirin S, Kumar R, Martinez C, Karmilowicz MJ, Ghosh P, Abramov YA, et al. A computational approach to enzyme design: Predicting W-Aminotransferase catalytic activity using docking and MM-GBSA scoring. J Chem Inf Model. 2014;54(8):2334–46.

59. Moreira IS, Fernandes PA, Ramos MJ. Unravelling hot spots: A comprehensive computational mutagenesis study. Theor Chem Acc. 2007;117(1):99–113.

60. Réblová K, Střelcová Z, Kulhánek P, Beššeová I, Mathews DH, Van Nostrand K, et al. An RNA molecular switch: Intrinsic flexibility of 23S rRNA helices 40 and 68 5'-UAA/5'-GAN internal loops studied by molecular dynamics methods. J Chem Theory Comput. 2010;6(3):910–29.

61. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. the accuracy of ranking poses generated from docking. J Comput Chem. 2011;32(5):866–77.

62. Foloppe N, Hubbard R. Towards predictive ligand design with free-energy based computational methods? Curr Med Chem. 2006;13(29):3583–608.

63. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An efficient program for end-state free energy calculations. J Chem Theory Comput. 2012;8(9):3314–21.

64. Homeyer N, Gohlke H. FEW: A workflow tool for free energy calculations of ligand binding. J Comput Chem. 2013;34(11):965–73.

65. Kumari R, Kumar R, Lynn A. G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model. 2014;54(7):1951–62.

66. Bienfait B, Ertl P. JSME: A free molecule editor in JavaScript. J Cheminform. 2013;5(5).

67. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An Open chemical toolbox. J Cheminform. 2011;3(10).

68. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605–12.

69. Lesnik S, Stular T, Brus B, Knez D, Gobec S, Janezic D, et al. LiSiCA: A Software for Ligand-Based Virtual Screening and Its Application for the Discovery of Butyrylcholinesterase Inhibitors. J Chem Inf Model. 2015;55(8):1521–8.

70. Roy A, Skolnick J. LIGSIFT: An open-source tool for ligand structural alignment and virtual screening. Bioinformatics. 2015;31(4):539–44.

71. Vainio MJ, Puranen JS, Johnson MS. ShaEP: Molecular overlay based on shape and electrostatic potential. J Chem Inf Model. 2009;49(2):492–502.

72. Liu X, Jiang H, Li H. SHAFTS: A hybrid approach for 3D molecular similarity calculation. 1. method and assessment of virtual screening. J Chem Inf Model. 2011;51(9):2372–85.

73. Morris G, Huey R. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J …. 2009;30(16):2785–91.

74. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem. 1998;19(14):1639–62.

75. Trott O, Olson AJ. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, EfficientOptimization, and Multithreading. J Comput Chem. 2010;31(2):455–61.

76. Allen WJ, Balius TE, Mukherjee S, Brozell SR, Moustakas DT, Lang PT, et al. DOCK 6: Impact of new features and current docking performance. J Comput Chem. 2015;36(15):1132–56.

77. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, et al. DOCK 6: Combining techniques to model RNA-small molecule complexes. RNA [Internet]. 2009;15(6):1219–30.

78. Korb O, Stützle T, Exner TE. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. Lect notes Comput Sci vol 4150 Ant colony Optim swarm Intell - ANTS2006 Proc [Internet]. 2006;4150:247–58.

79. Korb O, Stützle T, Exner TE. Empirical scoring functions for advanced Protein-Ligand docking with PLANTS. J Chem Inf Model. 2009;49(1):84–96.

80. Korb O, Stützle T, Exner TE. An ant colony optimization approach to flexible protein–ligand docking. Swarm Intell [Internet]. 2007;1(2):115–34.

81. Zhao H, Caflisch A. Discovery of ZAP70 inhibitors by high-throughput docking into a conformation of its kinase domain generated by molecular dynamics. Bioorganic Med Chem Lett. 2013;23(20):5721–6.

82. Wang Z, Sun H, Yao X, Li D, Xu L, Li Y, et al. Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. Phys Chem Chem Phys [Internet]. 2016;18(18):12964–75.

83. Brink T Ten, Exner TE. Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results. J Chem Inf Model. 2009;49(6):1535–46.

84. Ten Brink T, Exner TE. PKabased protonation states and microspecies for protein-ligand docking. J Comput Aided Mol Des. 2010;24(11):935–42.

85. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep [Internet]. 2017;7:42717.

86. Kiss R, Sandor M, Szalai FA. http://Mcule.com: a public web service for drug discovery. J Cheminform [Internet]. 2012 May;4(1):P17.

87. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. Comput Phys Commun. 1995;91(1–3):43–56.

88. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. Vol. 26, Journal of Computational Chemistry. 2005. p. 1701–18.

89. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput. 2008;4(3):435–47.

90. Mukherjee G, Patra N, Barua P, Jayaram B. A fast empirical GAFF compatible partial atomic charge assignment scheme for modeling interactions of small molecules with biomolecular targets. J Comput Chem. 2011;32(5):893–907.

91. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general Amber force field. J Comput Chem. 2004;25(9):1157–74.

92. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys. 1995;103(19):8577–93.

93. Meza JC. Steepest descent. Wiley Interdiscip Rev Comput Stat. 2010;2(6):719–22.

94. Brereton RG. Steepest Ascent, Steepest Descent, and Gradient Methods. In: Comprehensive Chemometrics. 2010. p. 577–90.

95. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph. 1996;14(1):33–8.

96. BIOvIA DS. Discovery studio modeling environment. San Diego, Dassault Syst Release, 4. 2015;98–104.

97. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov [Internet]. 2015;10(5):449–61.

98. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of

nanosystems: Application to microtubules and the ribosome. Proc Natl Acad Sci [Internet]. 2001;98(18):10037–41.

99. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Vol. 9, Nature Structural Biology. 2002. p. 646–52.

100. Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, Caffrey DR, et al. Structure-based maximal affinity model predicts small-molecule druggability. Nat Biotechnol. 2007;25(1):71–5.

101. Lo Conte L, Chothia C, Janin J. The atomic structure of protein– protein recognition sites. J Mol Biol. 1999;285:2177–2198.

102. Brooijmans N, Kuntz ID. Molecular Recognition and Docking Algorithms. Annu Rev Biophys Biomol Struct [Internet]. 2003;32(1):335–73.

103. Brown SP, Muchmore SW. High-throughput calculation of protein-ligand binding affinities: Modification and adaptation of the MM-PBSA protocol to enterprise grid computing. J Chem Inf Model. 2006;46(3):999–1005.

104. Jorgensen WL. The Many Roles of Computation in Drug Discovery. Vol. 303, Science. 2004. p. 1813–8.

105. Shoichet BK. Virtual screening of chemical libraries. Vol. 432, Nature. 2004. p. 862–5.

106. Walters WP, Stahl MT, Murcko MA. Virtual screening—an overview. Drug Discov Today [Internet]. 1998;3(4):160–78.

107. Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP. Comparison of automated docking programs as virtual screening tools. J Med Chem. 2005;48(4):962–76.

108. Kellenberger E, Rodrigo J, Muller P, Rognan D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. Proteins Struct Funct Genet. 2004;57(2):225–42.

109. Mohan V, Gibbs A, Cummings M, Jaeger E, DesJarlais R. Docking: Successes and Challenges. Curr Pharm Des [Internet]. 2005 Feb 1;11(3):323–33.

110. Onodera K, Satou K, Hirota H. Evaluations of molecular docking programs for virtual screening. J Chem Inf Model. 2007;47(4):1609–18.

111. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, et al. A critical assessment of docking programs and scoring functions. J Med Chem. 2006;49(20):5912–31.

112. Park MS, Dessal AL, Smrcka A V., Stern HA. Evaluating docking methods for prediction of binding affinities of small molecules to the G protein βγ subunits. J Chem Inf Model. 2009;49(2):437–43.

RESUME

Personal Information

Name	Süleyman Selim	**Surname**	Çınaroğlu
Birth Place	Fethiye	**Birth Date**	01/03/1990
Nationality	TC	**Phone Number**	05466203575
E-mail	suleyman.selim@mail.com		

Education

	Instituton	**Graduation Year**
Undergraduate	İstanbul University	2015
High School	Tavşanlı Yavuz Selim High School	2009

Experience

Position	**Organization**	**Duration (Year - Year)**
1. Research Asistant	Eternans Biotechnology LTD.	2018 -

Languages	**Reading**	**Speaking**	**Writing**
English	Advanced	Advanced	Advanced

Foreign Language Examination Grade									
KPDS	ÜDS	IELTS	TOEFL IBT	TOEFL PBT	TOEFL CBT	FCE	CAE	CPE	YÖKDİL
									87.75

 KPDS: Kamu Personeli Yabancı Dil Sınavı; ÜDS: Üniversitelerarası Kurul Yabancı Dil Sınavı; IELTS: International English Language Testing System; TOEFL IBT: Test of English as a Foreign Language-Internet-Based Test TOEFL PBT: Test of English as a Foreign

Language-Paper-Based Test; TOEFL CBT: Test of English as a Foreign Language-Computer-Based Test; FCE: First Certificate in English;

CAE: Certificate in Advanced English; CPE: Certificate of Proficiency in English

	Quantitative	**Equally Weighted**	**Verbal**
ALES Note	89,11924	84,31513	77,73913