STRUCTURE PREDICTION OF HUMAN DAT AND ITS BINDING ANALYSIS

GİZEM TATAR
20091109009

KADIR HAS UNIVERSITY

2011

STRUCTURE PREDICTION OF HUMAN DAT
AND ITS BINDING ANALYSIS

GİZEM TATAR

B.S.Molecular Biology and Genetics, Haliç University, 2009

Submitted to the Graduate School of Kadir Has University

in partial fulfillment of the requirements for the degree of

Master of Science

Graduate in Computational Biology and Bioinformatics

KADIR HAS UNIVERSITY

Haziran 2011

KADIR HAS UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

STRUCTURE PREDICTION OF HUMAN DAT
AND ITS BINDING ANALYSIS

GİZEM TATAR

APPROVED BY:

Dr. Tuğba Arzu Özal   (Kadir Has University)   _____
(Thesis Supervisor)

Prof.Dr.Kemal Yelekçi   (Kadir Has University)   _____

Prof.Dr.Safiye Erdem   (Marmara University)   _____

APPROVAL DATE: 16/06/2011

Structure Prediction of Human DAT and its Binding Analysis

Abstract

Dopamine neurotransmitter and its receptors are crucial in cell signaling process in the brain, which is in charge of information transfer in neurons functioning in the nervous system. Therapeutics used for the treatment of related disorders such as Parkinson's and schizophrenia would be considerably improved with the availability of the three dimensional (3D) structure of the dopamine transporter (DAT) and of the binding site for dopamine and other ligands.

Therefore, in this thesis; I have studied the prospective 3D structures of the neurotransmitter molecules such as human DAT which is predicted from primary amino acid sequence using computational molecular modeling techniques.

We have determined the binding sites and relative binding affinities of several ligands with the predicted structures of DAT. These computationally obtained binding affinities and binding sites, i.e. the critical residues of DAT for binding of dopamine and the other ligand molecules, correlate well with experiments. For instance, based on the modeled structures, our calculated binding free energy (ΔG_{bind}= -7.4 kcal/mol) for dopamine with DAT is found to be the same as the experimentally observed ΔG_{bind} value of -7.4 kcal/mol.

As a conclusion, new 3D structural models of human DAT has been constructed through homology modeling. Two of these human DAT models have been used to determine the binding characteristics between DAT and the ligands by

means of computational docking. These kind of computational studies, in which new structural and mechanistic insights were obtained, are expected in future to stimulate, further biochemical and pharmacological studies with much more detailed structures and accordingly, come up with the detailed insights of the working mechanisms of DAT and other homologous transporters.

İnsan DAT proteinin Yapısal Modelenmesi ve Bağlanma Analizi

Özet

Dopamin nörotransmiter ve reseptörleri, beyindeki sinir sisteminde görevli sinir hücrelerinde bilgi transferinden sorumlu olan hücre sinyal sürecinde büyük önem taşırlar. Parkinson hastalığı ve şizofreni gibi rahatsızlıkların tedavisinde kullanılan ilaçlarda, dopamin taşıyıcısının (DAT) üç boyutlu (3D) yapılarının ve dopamin ile diğer ligandların bağlanma bölgelerinin bulunmasıyla büyük ölçüde ilerleme kaydedilebilir.

Dolayısıyla, bu tezde, hesaplamalı moleküler modelleme tekniklerinin kullanılmasıyla primer sekanstan tahmin edilen insan DAT'nın olası 3 boyutlu yapısı (3D) çalışılmıştır.

Ayrıca, bazı ligandların ve tahmin edilen 3D DAT yapılarına bağlanma bölgeleri ve bağlanma afiniteleri tanımlanmıştır. Bu hesaplamalı yöntemlerle elde edilen bağlanma afiniteleri ve bölgeleri; dopamin ve diğer ligandların bağlanması için gerekli olan önemli rezidüler, deneysel sonuçlarla örtüşmektedir. Örneğin, dopamin ile DAT'ın, modellenen yapıya göre hesapladığımız bağlanma serbest enerjisi (ΔG_{bind}= -7.4 kcal/mol), deneysel olarak gözlemlenen ΔG_{bind} -7.4 kcal/mol değerine eşittir.

Sonuç olarak, homoloji modelleme yönteminden yararlanılarak insan DAT'nın yeni bir 3D yapısı geliştirilmiştir. Bu yapının oluşturulmasının ardından, DAT ve ligandlar arasındaki bağlanma özelliklerinin hesaplamalı docking yöntemi belirlenmesi için iki adet insan DAT modeli kullanılmıştır. Yeni yapısal ve mekanik görüşlerin elde edilmiş olduğu bu tür hesaplamalı araştırmaların gelecekte biyokimyasal ve farmakolojik araştırmaları daha da ilerleteceği ve buna bağlı olarak DAT ve diğer homolog taşıyıcıların işleyişini kavramamıza olanak sağlayacak daha detaylı bilgilerle sonuçlanacağı düşünülmektedir.

Acknowledgements

Table of Contents

List of Tables

List of Figures

List of Symbols

ΔG_{bind}: Estimation of Binding Affinity

m: sequence of length

n: sequence of length

m+1: the matrix dimension

n+1: the matrix dimension

S(i,j): score

i-1: the score from the cell at position

j-1: the score from the cell at position

s[*i,j*]: the new score at position

s[*i,j*-1]: the score one cell to the left

s[*i*-1,*j*]: the score immediately above the new cell

K,λ: constants

m: length of query sequence

n: length of the entire database

S: score of the alignment

E: expect value

K_i: binding constant

V: pair-wise evaluations

ρ : probability density function

ρ : the single body distribution function for atom I and is a constant

for a given protein

ΔS_{conf}: entropy lost upon binding

L: ligand

P: protein

List of Abbreviations

STS: Ion-coupled secondary transporters superfamily

NSS: Sodium symporter

DAT: Dopamine transporter

GABA: Gama-aminobutyric acid

DA: Dopamine

MPP: 1-methly-4-phenylpyridinium

MD: Molecular Dynamics

LeuT$_{Aa:}$ Leucine transpoter

LacY: Lac permease

SERT: Serotonin transporter

NET: Noradrenalin transporter

3D: Three dimension

NMR: Nuclear magnetic resonance

hDAT: Human dopamine transporter

BLAST: Basic local alignment search

BLOSUM: Blocks of aminoacid substitution matrix

TM: Transmembrane

TMH: Transmembrane helix

DSSP: Dictionary of Protein Secondary Structure

DOPE: Discrete Optimized Protein Energy

PDF: Probability Density Function

PDB: Protein data bank

RMSD: Root-mean-square deviation

MC: Monte carlo

GA: Genetic algorithm

FF: Force field

SFs: Scoring functions

PD: Parkinson

SN: Substantia nigra

NM: Neuromelamin

GlpT: Glycerol-3-phosphate transporters

DS: Discovery studio

gi: Query sequence

E: Expect value

TMH: Transmembrane helix

rDAT: rat dopamine transporter

gpf: Grid parameter file

dpf: Docking parameter file

MC: Monte Carlo

HB: Hydrogen bonding

Chapter 1

Introduction

The focus of my thesis is the development and application of structure and function prediction methods on membrane proteins called the dopamine transporter (DAT) belonging to the ion-coupled secondary transporters superfamily (STS), taking place in a neurotransmitter sodium symporter (NSS) family **[1].**

Dopamine neurotransmitter is fundamental for cellular signalling processes that are responsible for information transfer in neurons functioning in the nervous system **[2]**. The DAT pertains to a large family of Na^+ and Cl^- dependent transporters. This family comprises other monoamine transporters, such as the norepinephrine and serotonin transporters, gama-aminobutyric acid (GABA) and glycine carriers **[3].**

There are some experimental studies performed on DAT to gain information about the working mechanism of it. And it has been proven through pharmacological studies that dopamine transporting by DAT is Na^+/Cl^- dependent **[1].** The transport process of the DAT includes the translocation of the substrate dopamine (DA) and two sodium, one chloride ions across the cell membrane **[4].** The uptake is energetically coupled to transmembrane concentration gradient of Na^+, which is maintained by Na^+ K^+ ATPase **[5].**
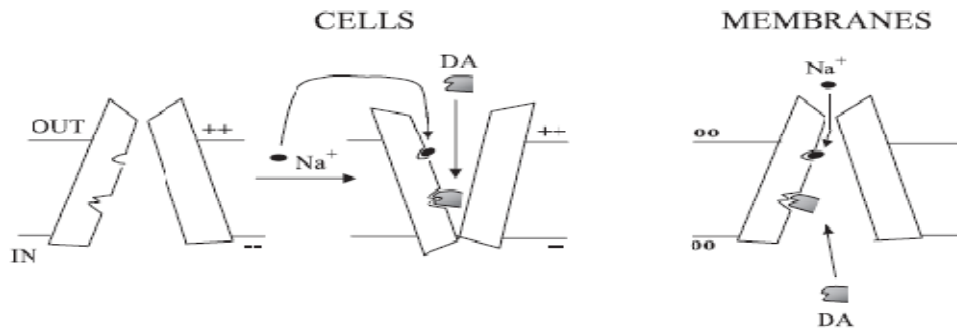
Figure 1.1: Model for Na^+ effects on the interaction of dopamine with its transporter in intact cell **[6].**

Furthermore, some other experimental studies show that the dopamine transporter ensures the return of dopamine into neurons and is a major target for a variety of pharmacologically active drugs and environmental toxins. Various substances, such as the psychostimulant amphetamine, the dopaminergic neurotoxin 1-methly-4-phenylpyridinium (MPP^+) and several sympathomimetic amines, are structurally similar to dopamine. Therefore, they make up substrates for the dopamine carrier and can be transported. In addition to this, the dopamine transporter constitutes an important molecular target for the addictive drug cocaine and to some extent, antidepressants **[2].**
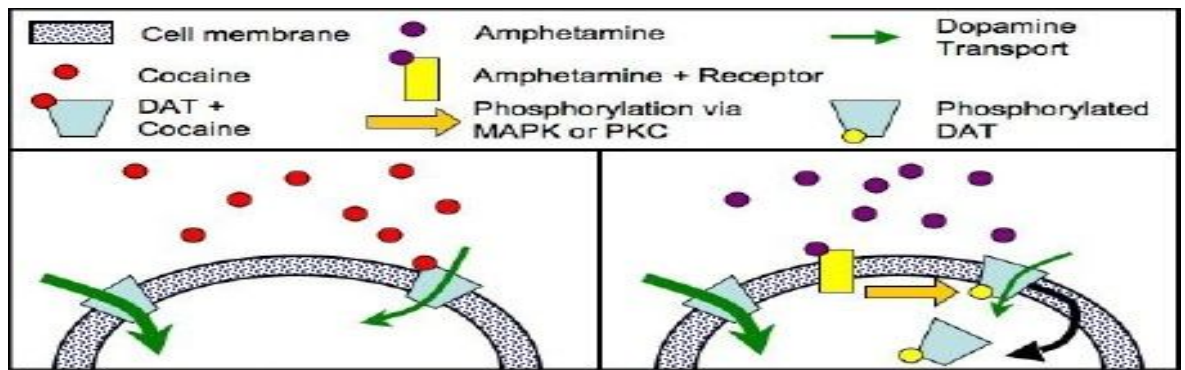
Figure 1.2: Mechanism of Cocaine and Amphetamine based DAT block: Cocaine binds DAT and slows transport and Amphetamine induces MAPK/PKC to phosphorylate DAT. Phosphorylation slows transport and triggers internalization of DAT **[7].**

Some more substantial experimental and computational studies, such as Zn^{2+} binding site engineering, site-directed mutagenesis, Molecular Dynamics (MD) simulations, have also been performed so far in order to further investigate the structure and potential mechanism of DAT **[8].**

The first structural study related to DAT is the x-ray crystal structure determination for a bacterial homolog of DAT, which is bacterial leucine transporter (LeuT$_{Aa}$) from NSS family. This bacterial homolog has recently been identified as *Aquifex aeolicus* as a complex of substrate leucine and two sodium ions (PDB entry of 2A65 at 1.65 A° resolution) **[10]**.

Consequently, initial computational studies for the structure determination of human DAT (hDAT) had been based on the above LeuT$_{Aa}$ template **[1].** And some more computational studies make use of Na^+/H^+ antiporter or Lac permease (LacY) as the template **[1].** Further molecular modelling studies had been performed on DAT, serotonin transporter (SERT), and noradrenalin transporter (NET) **[1].**

However, for the latter there is no experimental evidence supporting the previously reported models from any kind of structural or biological studies. Therefore, the structural models of DAT, SERT, and NET that have previously been reported are not good enough to be used for studying molecular interaction of psychotropic drugs with these transporters. Since, these structural models are not appropriate for the homology modelling of any member of the NSS family consequently, the predicted atomic interactions are not reliable as well.

Up to now, LeuT$_{Aa}$, which has similar structural folding and physiological features as other NSS family members, had been reported to be the most reasonable template to study the substrate binding and transporting mechanism for NSS transporters, especially for DAT and SERT **[1].**
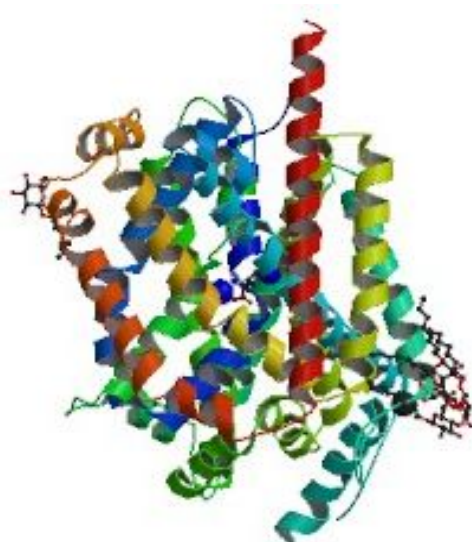
Figure 1.3: Crystal structure of LeuT$_{Aa}$, a bacterial homology of Na$^+$/Cl$^-$ dependent neurotransmitter transporters. The determination of the X-ray structure of LeuT$_{Aa}$ in complex with its substrate leucine and two Na ions (PDB entry of 2A65 at 1.65 A° resolution) has been viewed as a milestone in understanding structural and functional relationships of NSS members **[1].**

As a result, even though there are several studies reported before, the structure and working mechanism of DAT is not revealed at the atomic scale. And furthermore, since the previous studies reported to be not enough, either because of the template they have used or the resolution of the experimental studies. In this thesis, new three-dimensional (3D) structural model of hDAT has been constructed through two different methods; first template is based on an experimental structure obtained from the protein database by the means of the multiple sequence alignment, whereas the second template protein was the protein DAT of rat. Following the construction of the 3D structure, two different hDAT models have been used to dock DAT ligands with automatic computational docking.
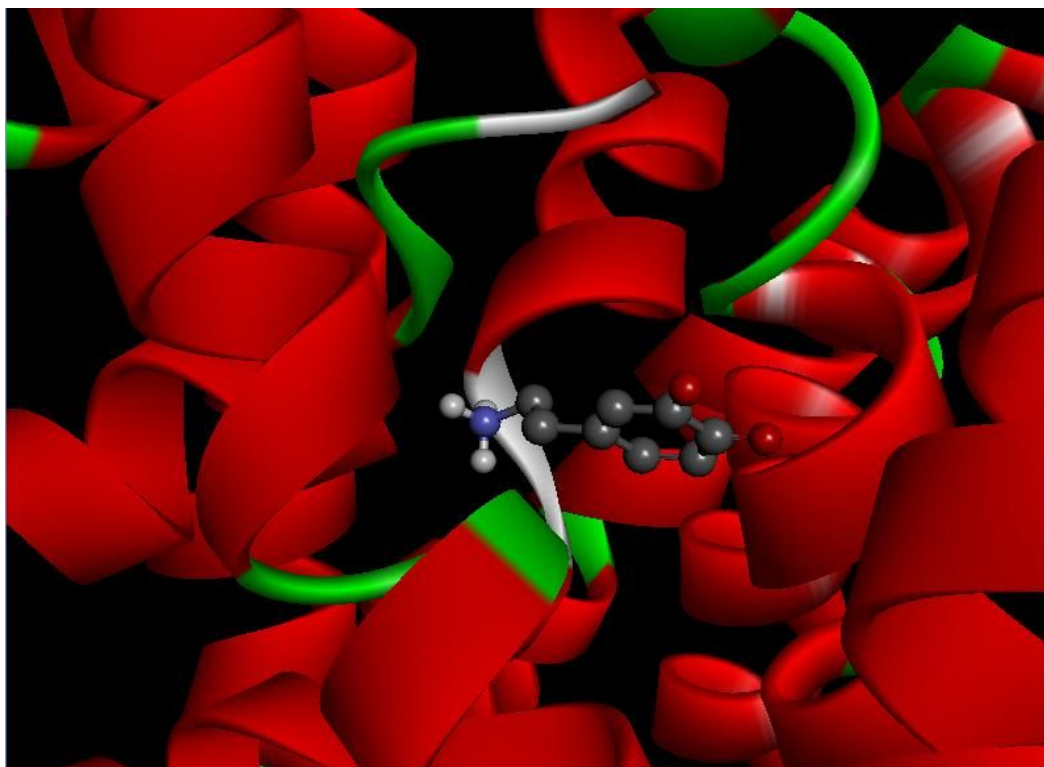
Figure 1.4: Shows model of human DAT with dopamine. Red bars indicate alpha helices and dopamine shown in ball-and-stick.

In the Chapter 2, we have given the basic theory of the homology modelling. Here we gave a brief theory of how to perform pairwise and multiple sequence alignment. Furthermore, we introduce the algorithms used for these sequence alignment; methodologies, which have two main categories as local and global alignment algorithms. Finally, in these two sections we talk about secondary structure from sequence and homology modelling of the protein structure. And in Chapter 3, we have given the theory of protein-ligand interaction which have two main categories as docking algorithm and scoring function.

In Chapter 4, we introduce the methods that we used to predict the 3D structural model of DAT. This model has been constructed through comparative homology modelling based on the most reasonable template proteins. At the end, we obtained two different hDAT.

And in Chapter 5, upon obtaining a 3D hDAT model, eight ligands that couple with the macromolecule have been docked using the AutoDock program to obtain the conformation and orientations for hDAT functional analysis**.** There are eight

ligand molecules, namely 5-hydroxydopamine, 6-hydroxydopamine, Dopamine, Amphetamine, Methamphetamine, Cocaine, Benzene-1.3-diol (benzenediol), Tyramine. These ligand molecules were tested on the predicted hDAT structure model.

Chapter 2
Theory of Homology Modelling

2.1 Introduction

Despite the newly developed techniques in molecular biology that allow rapid identification, isolation, sequencing of genes and consequently the sequences of many proteins, the task of acquiring the 3D structures of these proteins remains a time-consuming task. The 3D structure of the protein, along with the sequence, can provide us a detailed understanding of the protein's function, mechanism of action and its structure-function relation. As the main purpose of structural biology is to infer the three-dimensional structure through the sequence, different approaches are being taken in order to achieve this goal by making up for the constraints related to X-ray diffraction or Nuclear Magnetic Resonance (NMR).

In the field of computational biology there are several studies conducted to support the tertiary protein structure. The most important topic of these studies is the homology modelling method. The observation that tertiary structure is better conserved than amino acid sequence underlies this method of homology modelling [11]. This modelling enables us to detect common structural properties and especially the overall fold of proteins which have observable resemblance despite the considerable change they have undergone in the sequence. While obtaining experimental structures through methods such as X-ray crystallography and protein NMR requires a considerable amount of time and effort, homology modelling provides useful structural models for developing theories on a protein's function and directing prospective experimental work.

For building the first homology models wire and plastic models of bonds and atoms were used in the 1960s. Following this advancement, the first homology model structure, which was of the small globular protein α-lactalbumin, got published in 1968. Hen egg white lysozyme had been taken as the template protein of this model structure. Subsequently, X-ray crystallography method has shown that the model was basically precise except for the misconception that the structure of carboxy-terminal end is similar to that of lysozyme **[12].**

Basically, there are two types of methods that are used most frequently for tertiary structure prediction. The first is the homology modelling (also known as comparative or knowledge-based modelling) which functions on the basis of modelling the structure of an unknown protein with regard to the known structure of homologous protein. The other approach, which is known as threading of fold recognition, eliminates the need for a homologous protein structure because it tries to match the sequence with all known protein folds to identify the best fit **[12].**

In our thesis, we have predicted the three-dimensional structure of the human dopamine transporter (hDAT) protein by making use of the homology modelling methods. In the course of this research, we obtained two different 3D hDAT protein structures using two different homology modelling methods. Whereas the templates of the first modelling method were experimental structures, the template of the second method was a structure of the rat DAT protein.

Modelling process of a protein's three-dimensional structure consists of a few steps. Accordingly, this chapter is outlined as follows; first, an introductory information is given about sequence alignment methods, which are basically pairwise sequence alignment and multiple sequence alignment. Then, the algorithms used are explained as; local alignment and global alignment. After that, theory of the corresponding computational tools, i.e. BLAST and FASTA, are given. Definition of substitution matrices and scoring functions; which are PAM and BLOSUM, is included. Secondly, a description of the methods used for secondary structure prediction and structure assignment is given. Finally, principle of homology modelling with automatic homology modelling systems are explained.

2.1.1 Pairwise Sequence Alignment

Sequence alignments are intended to discover and illustrate the similarities, differences, or evolutionary relationships between sequences. Sequence alignment is generally divided into two category; global and local alignment. And for both global and local alignment methods there exists algorithms for performing *pairwise alignments*, that is, the alignment of just two sequences, and for performing *multiple sequence alignments,* in which more than two sequences are aligned with each other **[13].**

Pairwise sequence alignment methods are used to find the best-matching piecewise local or global alignments of two query sequences **[13]**, where as in the multiple sequence alignment multiple data within databases are considered.

In this chapter, we will present the basic sequence alignment algorithms, broadly characterized as *global alignment, semiglobal alignment,* and *local alignment,* which are also common in multiple sequence alignment.

2.1.2 Multiple Sequence Alignment

Though the alignment of all sequences together, instead of making use of a generalized representation of the sequence family, multiple sequence alignments are more effective for the comparison of similar sequences. In the cases where there are subgroups of sequences with mutual extra features that are not available in the complete sequences set, these features may disappear during the creation of a profile. This occurs when the subgroup can be used for generating a new profile which is used in the search for other sequences **[12].**

Multiple alignment programs such as ClustalW and T-Coffee execute Needleman-Wunsch global alignment for every pair of sequence and obtain from these alignments the measure used in developing the guide tree. This measure is, in fact, the evolutionary distance which is a measure of dissimilarity. The percentage of alignment sites at which various residues have been aligned is considered as the simplest evolutionary distance measure. The Needleman-Wunsch alignment of all

sequence pairs takes a very long time when there are too many sequences to be aligned. As a result, faster but more approximate methods have been suggested for obtaining a distance measure **[12].**

2.1.3 Algorithm of Sequence Alignment

There are two main types of alignment: global and local. The local alignment algorithm of Smith and Waterman (1981) is the most rigorous method by which subsets of two protein or DNA sequences can be aligned. A global alignment such as the method of Needleman and Wunsch (1970) contains the entire sequence of each protein or DNA sequence. At the sequence alignment stage of our research, we made use of the BLAST server which functions on the basis of the below mentioned algorithms **[14].**

2.1.3.1 Local Alignment Algorithm

Local alignment is extremely useful in a variety of applications such as database searching where we may wish to align domains of proteins.

For the Smith-Waterman algorithm a matrix is constructed with an extra row along the top and the extra column on the left side. Thus for sequences of lengths m and n, the matrix has dimensions $m + 1$ and $n + 1$. The score in each cell is selected as the maximum of the preceding diagonal or the score obtained from the introduction of a gap. However, the score cannot be negative: A rule introduced by Smith-Waterman algorithm is that if all other score options produce a negative value, then a zero must be inserted in the cell. Thus the score S(i,j) is given as the maximum of four possible values

1. The score from the cell at position i-1, j-1; that is, the score diagonally up to the left. To this score, add the new score at position s[i, j],which consists of either a match or a mismatch.
2. s(i, j-1)(i.e.,the score one cell to the left)minus a gap penalty
3. s(i-1,j)(i.e.,the score immediately above the new cell)minus a gap penalty.

4. The number zero. This condition assures that there are no negative values in the matrix. In contrast negative numbers commonly occur in global alignments because of gap or mismatch penalties **[14].**

BLAST (Basic Local Alignment Search)

Having primarily been designed as a local alignment search tool, BLAST identifies alignments between a query sequence and database without the introduction of gaps. The current version of BLAST allows gaps in the alignment **[14].**

By making use of BLAST, the user can select one sequence (termed as the "query") and furthermore achieve pairwise sequence alignments between the query and the whole database (termed the "target"). As a result BLAST returns just the closest matches of millions of alignments which are analysed in a BLAST search. It also enables you to compare query sequence to DNA or protein sequences in a database.

This algorithm can be described in three phases

1. BLAST compiles a preliminary list of pairwise alignment, called word pairs.
2. The algorithm scans a database for word pairs that meet some threshold score T.
3. BLAST extends the word pairs to find those that surpass a cut-off score S, at which point those hits will be reported to the user. Scores are calculated from scoring (substitution) matrices (such as BLOSUM62) along with gap penalties **[14].**

Optional BLAST Search Parameters

We will initially focus our attention on a standard protein-protein BLAST search.

Scoring Matrix: Five amino acid substitution matrices are available for blastp protein-protein searches: PAM30, PAM70, BLOSUM45, BLOSUM62 (default), and BLOSUM80.

Gap Penalties: A gap is a space introduced into an alignment to compensate for insertions and deletions in one sequence relative to another.

Expect Threshold: A threshold value T is established as a cut-off for the score of aligned words. It can be lowered to identify more initial pairwise alignments. This will increase the time required to perform the search but may increase the sensitivity. The default setting for expect value threshold is 10 for blastn, blastp, blastx and tblastn.

Filter Low-Complexity: Low-complexity sequences are defined as having commonly found stretches of amino acids (or nucleotides) with limited information content.

Word Size: For protein searches, a window size of 3 (default) or 2 may be set. When a query is used to search a database, the BLAST algorithm first divides the query into a series of smaller sequences (words) of a particular length (word size), as will be described below

Expect value: Altschul and colleagues revised the application of extreme value distribution to BLAST. The following formula defines the distribution of function scores for two random sequences *m* and *n*.

$$E = K \times m \times n \times e^{-\lambda S}$$

K,λ: constants
m: length of query sequence
n: length of the entire database
S: score of the alignment

E is the expect value of different alignment with the same or greater scores than some scores that are expected to be the results of a database search. Very small E-values of 10^{-20} or less are usually collected from closely related sequences and these scores manifest the resemblance of the database and query sequences. The fact that the E-value is determined by the sequence length, the number of sequences in the database and the alignment score should always be kept in mind. A small E-value mostly signifies a better alignment. Likewise, a higher percentage identity implies a more accurate assessment of the importance of the resemblance between the database sequence and query sequence **[14].**

2.1.3.2 Global Alignment Algorithm

This algorithm is important because it produces an optimal alignment of protein or DNA sequences, even allowing the introduction of gaps.

We can describe the Needleman-Wunsch approach for global sequence alignment in three steps **[14]:**

1. Setting up a matrix
2. Scoring the matrix
3. Identifying the optimal alignment

The algorithms for two commonly used search programs **BLAST** and **FASTA**. They make use of indexing techniques such as suffix trees and hashing to locate short stretches of database sequences highly similar or identical to parts of the query sequence **[14].**

2.1.4 Substitution Matrices and Scoring

Referring to a substitution matrix, score values for all possible pairs of residue are calculated for the aligned pairs of amino acids. Among different types of substitution matrices used over the years, there were some based on theoretical considerations, such as the number of mutations that are necessary for converting one amino acid into another, or resemblances in physicochemical characteristics.

Nevertheless, the most effective matrices, which are based on analysis of alignments of many homologs of well-studied protein from a wide range of species, use real proof of what happened during evolution **[12].**

Two methods have been used in deriving substitution matrices from multiple sequence alignments and the resultant substitution matrices PAM and BLOSUM are used for calculating the scoring matrices **[12].**

The matrices of BLOSUM (**Blo**cks of Amino Acid **SU**bstitution **M**atrix), which is a substitution matrix serving for protein sequence alignment, are based on local alignments and used for scoring alignments between evolutionarily variant protein sequences. First, BLOCKS database has been scanned for highly preserved regions of protein families (the ones without gaps in the sequence alignment) and relevant amino acid frequencies and their substitution probabilities have been calculated. Secondly, a log-odds score for 210 probable substitutions of the 20 standard amino acids has been computed. BLOSUM matrices, which are not extrapolated from related proteins as in the PAM Matrices, function on the basis of examined alignments **[15].**

Henikoff and Henikoff (1992,1996) used the BLOCKS database, which consisted of over 500 groups of local multiple alignments (blocks) of distantly related protein sequences. Thus the Henikoff's focused on conserved regions (blocks) of proteins that are distantly related to each other **[13].**

The aptitude of a range of BLOSUM and PAM matrices was put to test by Henikoff and Henikoff (1992) in order to identify proteins in BLAST searches of database. BLOSUM 62's performance at identifying a variety of proteins was found to be slightly better than BLOSUM 60 or BLOSUM70; but much better than PAM matrices. Their matrices were particularly effective for determining weakly scoring alignments. Other scoring matrices in common used in BLAST searches are BLOSUM 50 (matrix that is best for the alignment of two proteins sharing about 50% identity) and BLOSUM90 **[14].**

```
BLOSUM90          BLOSUM80  BLOSUM62              BLOSUM45
PAM30             PAM120    PAM180                PAM240

Less divergent   ◄─────────────────────►   More divergent

Mouse versus rat RBP                        Mouse versus bacterial lipocalins
```
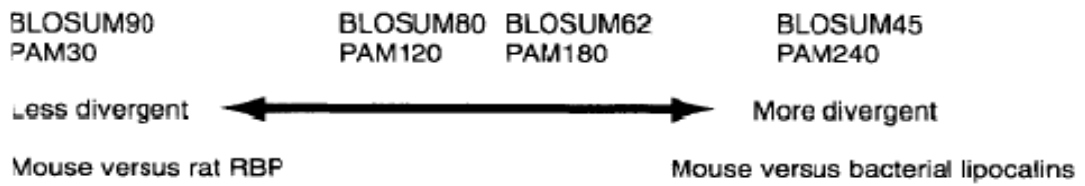
Figure 2.1: Summary of PAM and BLOSUM matrices. High-value BLOSUM matrices and low-value PAM matrices are best suited to study well-conserved proteins such as mouse and rat RBP. BLOSUM matrices with low number (e.g.; BLOSUM45) or high PAM numbers are best suited to detect distantly related proteins.

As seen in the Figure 2.1, based on data from the alignment of highly linked protein families, PAM matrices contain the presumption that substitution probabilities for closely related proteins (e.g., PAM10) can be extrapolated to probabilities for distantly related proteins (e.g., PAM250). On the other hand, BLOSUM matrices are based on empirical observations of more distantly related protein alignments.

Considering all the properties of the substitution matrices, in our research, we preferred to use BLOSUM62 matrix as the scoring matrix **[14].**

2.2 Secondary structure from sequence

In the case of an atomic-resolution of a structure, the hydrogen bonds of the biopolymer define its secondary structure. In proteins, the patterns of hydrogen bonds between backbone amide and carboxyl groups identify the secondary structure. Whereas, in nucleic acids, it is the hydrogen bonding between the nitrogenous bases that identifies the secondary structure.

An α-helix is generated when a single polypeptide chain turns around itself to form a structurally rigid cylinder. A hydrogen bond is made between every fourth amino acid, linking the C=O of one peptide bond to the N-H of another. This gives rise to a regular helix with a complete turn every 3.6 amino acids **[16].**

β-strands (sheet) are created when hydrogen bonds form between segments of polypeptide chains lying side by side. A beta strand (also β strand) is a stretch of polypeptide chain typically 3 to 10 amino acid long with backbone in an almost fully extended conformation. When the structure consists of neighbouring polypeptide chains that run in the same orientation, it is considered a *parallel β-sheet;* opposite to that of its immediate neighbours-the structure is an *antiparallel β-sheet* **[16].**
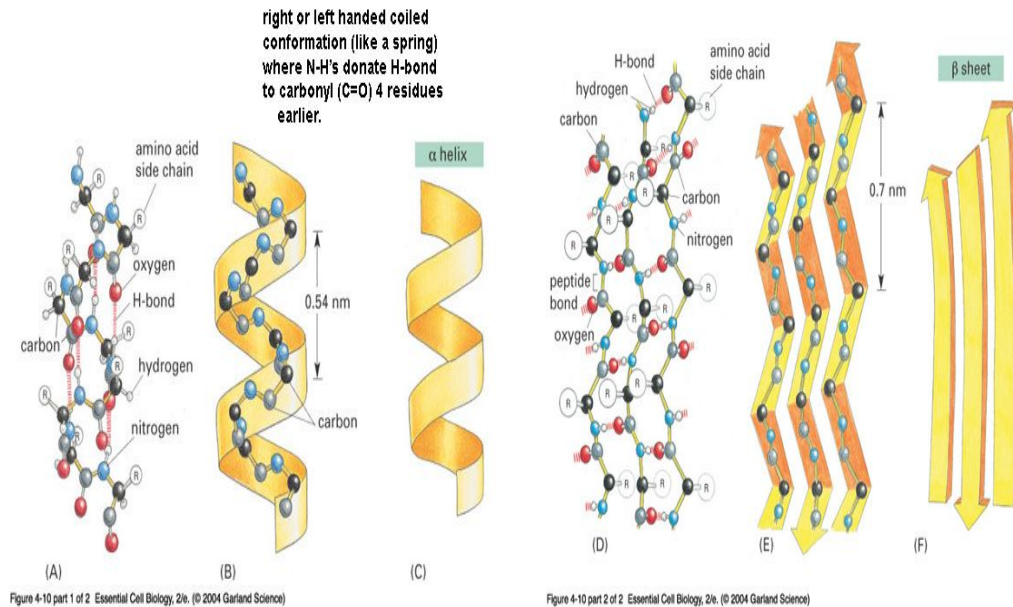
Figure 2.2: Polypeptide chains often fold into one of two orderly repeating forms known as the α helix and the β sheet.

Most transmembrane (TM) proteins span the membrane in the form of one or more α-helix, and we shall discuss these first. The properties of the lipid bilayer force a number of structural constraints on the α-helical TM segments. An average membrane thickness is 30 A^{o}, which corresponds to an α-helix of between 15 and 30 residue **[12].**
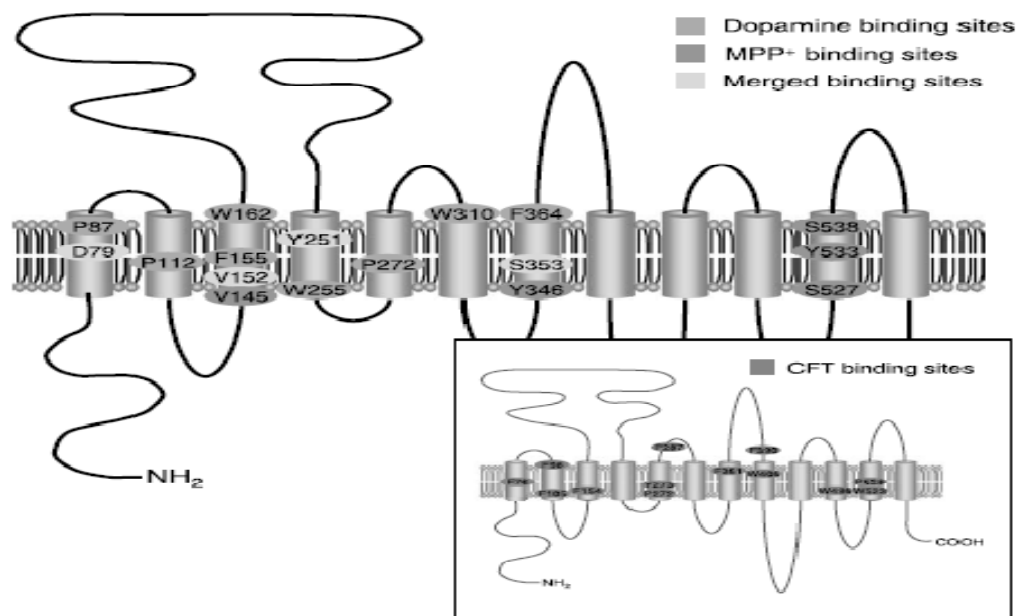
Figure 2.3: Putative topology and structural features of the DAT protein.

DAT protein structure shown in the Figure 2.3. human dopamine transporter (hDAT), which is in the content of our thesis, pertains to NSS family. Results obtained from structure analysis of previous studies indicate that DAT pertaining to NSS family are transmembrane proteins with the following characteristics; 12 TM α-helices with intracellular location of N-terminal and C-terminal, and a large internal substrate binding cavity on the midpoint of the membrane **[17].**

Methods for predicting protein secondary structure can be broadly divided into the following categories: statistical analysis, also referred to as probabilistic analysis; knowledge-based analysis; machine-learning methods; and those mainly based on neural networks. Most automated methods in use today use a mixture of these techniques, and all of them incorporate some form of statistical analysis, is employed in most of the automated methods **[12].**

Many prediction programs are designed to recognize just three different regular structural states: α-helices, β-strand, and β-turns. There are also some algorithms that even try to predict different kinds of helical or turn conformations, such a   -helices, 3_{10}-helices, and polyproline helices, and type I, II and other types of turns **[12].**

17

The most widely used of these programs is Dictionary of Protein Secondary Structure (DSSP), which assigns secondary structure according to hydrogen-bond patterns. DSSP, proposed by Wolfgang Kabsch and Chris Sander in 1983, takes the approach just described to assign helices.

Another program, STRIDE, uses both hydrogen-bond energy and backbone dihedral angles. DEFINE matches the interatomic distances in the protein with those from idealized secondary structures. However, these programs do not always produce identical results from the same data, giving slightly different secondary structure assignments. The differences are almost exclusively at the ends of structural elements. These differences in defining secondary structure elements can affect the apparent accuracy of secondary structure prediction methods. A prediction method should be trained and subsequently tested using training and testing datasets whose structural features were defined using the same assignment method **[12].**

Table 2.1: The eight structural states detailed by the DSSP method, with their descriptions.

DSSP, is commonly used to describe the protein secondary structure with single letter codes. There are eight types of secondary structure that DSSP defines:

- G = 3-turn helix (3_{10} helix). Min length 3 residues.
- H = 4-turn helix (α helix). Min length 4 residues.
- I = 5-turn helix (π helix). Min length 5 residues.
- T = hydrogen bonded turn (β-turn)
- E = extended strand in parallel and/or anti-parallel β-strand conformation. Min length 2 residues.
- B = residue in isolated β-bridge (single pair β-sheet hydrogen bond formation)
- S = bend (the only non-hydrogen-bond based assignment).

2.3 Homology modelling of protein structure

Homology modelling has several steps. The first step consists of finding structural homologues to the target protein. Most experimentally resolved structures are placed in the Protein Data Bank (PDB). Using one of the sequence-search programs (BLAST and FASTA), a database with the sequences of the PDB structures is searched with the target sequence for homologues. The protein with the highest similarity score and highest sequence identity over the largest stretch of amino acids is chosen as the template. Having identified a matching template protein, the next step should be the alignment of the target and template sequences. The most important step in the modelling process is this alignment **[12].** After alignment, the structurally conserved regions-the core-are modelled first. Modelling the core is simply achieved by transferring the *x, y,* and *z* coordinates of every matched atom within an aligned residue from the template to the target molecule. The backbone atoms are then joined together to form peptide bonds at the correct angles. It is usually possible to copy only some of the side-chain coordinates, as many side chains in the target will not be identical to those in the template. Regions with insertions and deletions are left for later. It is important to check the core for misfits. In regions that were difficult to align, one should examine whether the insertion by one or two residues would position it in a more favourable conformation **[12].**
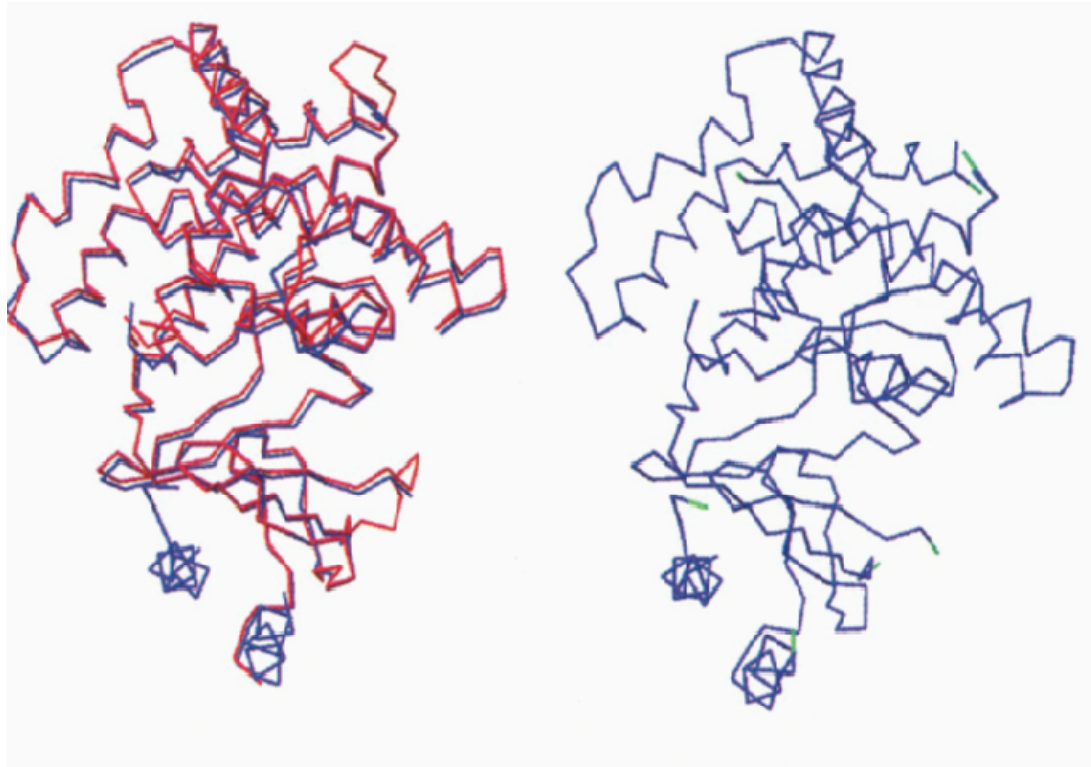
Figure 2.4: First picture, shows a target structure superposed on a template structure. Second picture, shows only the target model with the end of the core regions where an insertion has to be modelled **[12].**

Once a good alignment has been obtained, the other major procedure in model building is modelling the loops. These are the regions that usually contain insertions or deletions and the most variable in sequence. Because of their variability in both sequence and length, loops are generally the most difficult regions to model. If the target protein contains an insertion in a loop sequence relative to the template structure, there will be no template coordinates from which to model the insertion. The easiest way round this problem is if there are other structure homologs with the same insertion. It is then possible to model the missing part using their coordinates. However, an insertion is often unique to the target protein. In this case, the most widely used method for modelling loops, in both manual and automatic procedures, is to search for fragments of the same length in a database of high-resolution structures. Loops that have the lowest RMSD are selected for further evaluation **[12].**
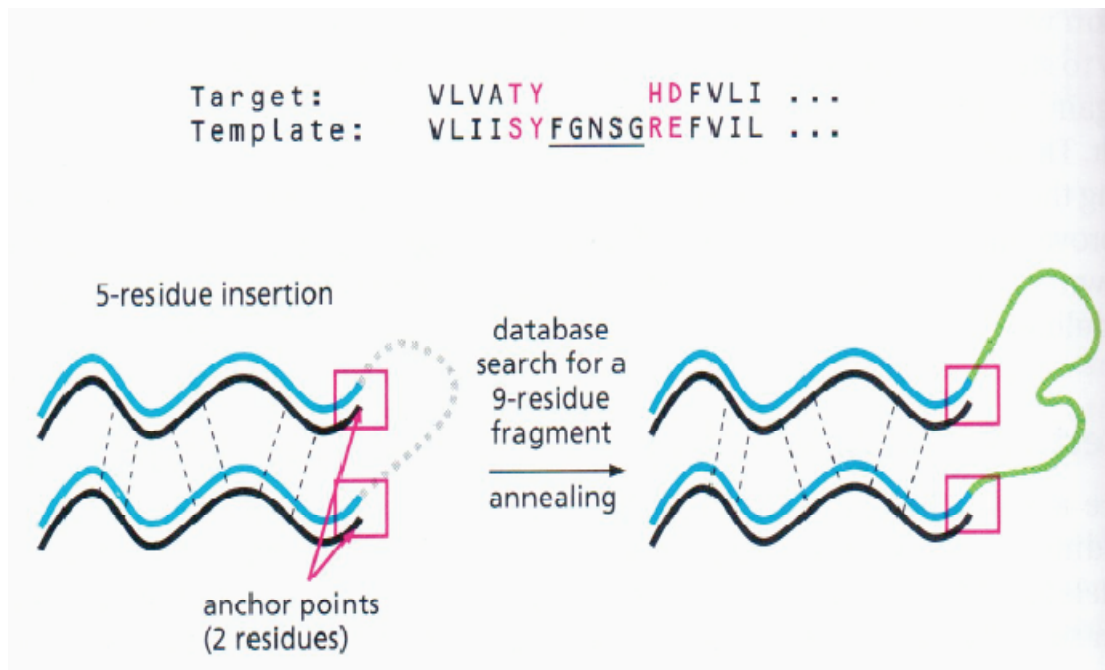
Figure 2.5: A schematic illustration viewing the database search method for building loop **[12].**

Then, to be able to build side chains in a model, however, it is necessary to have some understanding of the conformations they can adopt. A study of atomic coordinates of side-chain atoms of conserved residues in proteins with similar three-dimensional structure has found that in more than 90% of cases the side chains have the same conformations in the two proteins. To predict the side-chain conformation when the aligned residues are not identical, rotamer libraries are used **[12].**
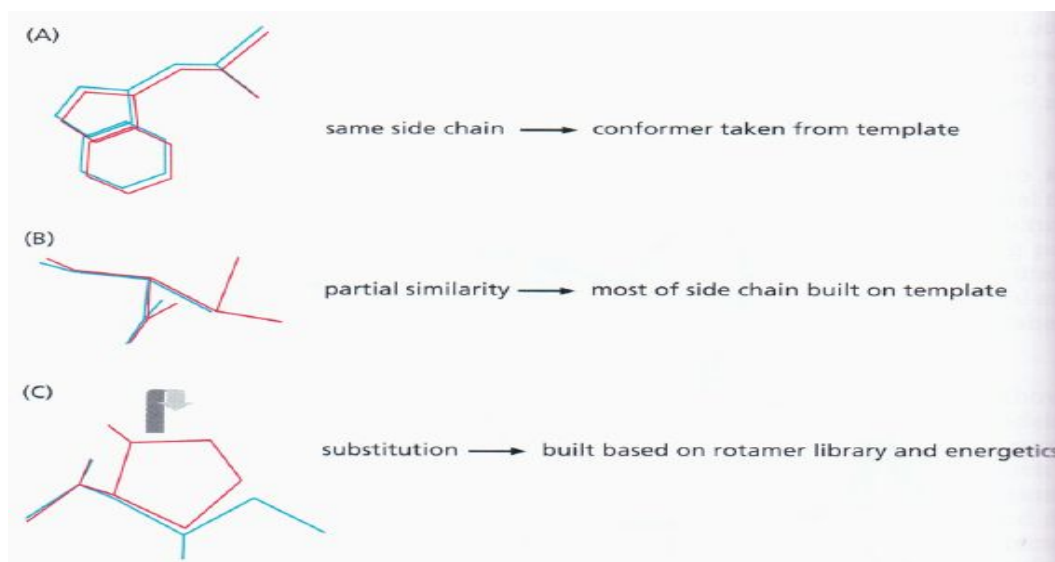
Figure 2.6: (A) In cases where the side chains at a given position are identical, the conformation of the target side chain can be extracted directly from the template. (B) In cases where the side chains are similar but not identical, most of the target side chain can be built from the template. (C) In cases where the side chains are different, the conformation of the target side chain has to be deduced from a library of rotamer structures and an assessment of the energetic **[12].**

And the final step is the model optimization. This method boils down to a sequence of rotamer prediction and energy minimization steps. At every minimization step, a few big errors (like bumps, i.e., too short atomic distances) are removed while many small errors are introduced. When the big errors are gone, the small ones start accumulating and the model moves away from the target. As a rule of thumb, today's modelling programs therefore either restrain the atom positions and/or apply only a few hundred steps of energy minimization. In short, model optimization does not work until energy functions (force fields) get more precise. Such a simulation follows the motions of the protein on a femtosecond (10^{-15}) timescale and mimics the true folding process **[18].**
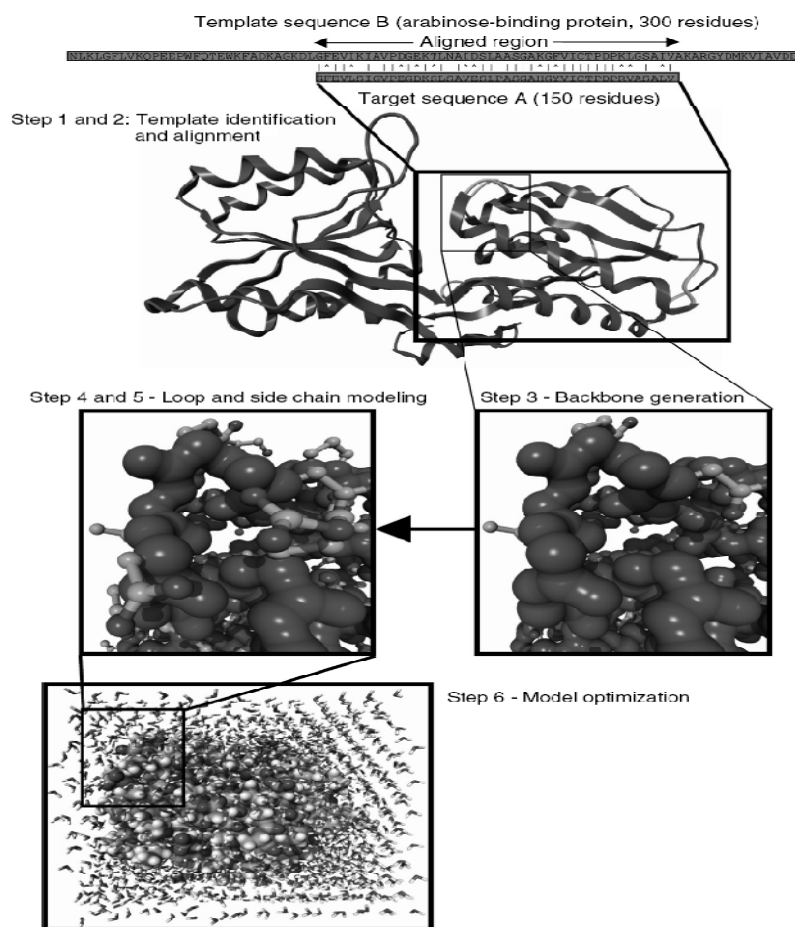
Figure 2.7: Steps in homology modelling. Example, the fragment of the template (arabinose-binding protein) corresponding to the region aligned with the target sequence forms the basis of the model (including conserved side chains). Loops and missing side chains are predicted, then the model is optimized (in this case together with surrounding water molecules) **[18].**

2.3.1 Automated Homology Modelling

Having changed modelling into a practically routine technique for the observation of the three-dimensional structure of a new protein, automated model building by homology has revolutionized the modelling assignment. Fully automatic model-building programs, such as Swiss-Model and MODELLER, brought along the advantages of objectivity, rapidity. We made use of the MODELLER program for the homology modelling in our research. Information about this program is given below **[12].**

23

2.3.1.1 MODELLER

With the aim of satisfying constraints, MODELLER develops a hypothetical target structure based on a familiar template structure. First, these spatial constraints in the form of atom-atom distances and dihedral angles are extracted from the template(s), used in the target and afterwards synthesized with general rules of protein structure like bond length and angle preferences.

The alignment enables us to identify parallel residues between the target and the template. The optimization of the target model continues until a model that best satisfies the spatial constraints is obtained. MODELLER is most commonly used for homology modelling of protein's three-dimensional structure as the program deduces structure by satisfaction of spatial constraints; yet, it can also be put into use in experimental structure determination by NMR. The output of MODELLER is a tertiary structure of a protein that satisfies a set of constraints as well as possible **[11].**

MODELLER protocol evaluates protein structure using the Probability Density Function (PDF) energy data and DOPE (Discrete Optimized Protein Energy) scoring function (in the tutorial) **[19].**

PDF (Probability Density Function)

In order to constrain a model towards the structure of the template and to preserve fine structure geometry, MODELLER creates numerous restraints. There are dissatisfied restraints at the end of a MODELLER refinement and in order to obtain an evidence of the quality of the model for that particular residue, the PDF energy (i.e., the value of the objective function) regarding each residue is summed. As parts of a model (for example residues alongside an insertion or deletion) are practically sure to have higher restraint violations, a simple interpretation of the absolute value of the energy does not exist. The energies are effective for comparisons between different models. Models with lower energy in a given region are probably better in that region. If these models have been developed from different initial alignments of the model sequence and the homologues, then lower

energies are an evidence of better alignment. Either residue or atomic based statistical potentials are used as quality indicators of protein structures and as potential energy functions in predicting protein structures.

PDF Physical Energy is more of a true energy term. It is comprised of the sum of energies for the stereo chemical pseudo-energy terms which consist of valence bonds, valence angles and torsion angles, improper torsion angles, and soft-sphere repulsion, as well as knowledge based non-bonded potentials used only for loop and mutant modelling. Higher energy indicates a larger restraint violation in the model.

PDF Total Energy is the total PDF energy, that is the sum of the scoring function value of all homology-derived pseudo-energy terms and stereo chemical pseudo-energy terms.

DOPE Scoring Function

Many statistical potentials, either residue based or atomic based are used in evaluating the quality of protein structures and as potential energy functions in predicting protein structures.

DOPE is an atomic based statistical potential in MODELLER package for model evaluation and structure prediction. The DOPE score of a protein can be viewed as a conformational energy which measures the relative stability of conformation with respect to other conformations of the same protein. It can assist in choosing the best model out of a set of predicted model structures of a protein sequence. And low scores are better.

In addition, the statistical potential is also used as part of the energy function in predicting loop models or to optimize the local structure of a mutated residue.

The pairwise statistical potential function of a protein with N atoms can be represented by the pairwise probability density function:

$$i, j : \bullet$$

where $\quad_i$ is the single body distribution function for atom i and is a constant for a given protein.

The pairwise probability density function is proportional to $\quad_i \quad_j$ and the (m,n) pair density where m and n are the atom types of atom i and atom j respectively.

The density of (m,n) pairs at distance r is $p_{mn}(r)/(4\pi r^2 \Delta r)$. For a finite system, such as a protein structure, the spherical shell can be partly outside of the protein boundary in many cases and the density should be more accurately represented by $p_{mn}(r)/n(r,a)$ where $n(r,a) = 4\pi r 2\xi(r)$

$$i, j \qquad i \quad j \quad n \quad /4\pi^2$$

The scores used in the current MODELLER are calculated as a weighted sum from a set of high resolution X-ray structures with low sequence homology:

$$i, j \qquad w \quad_{,n} \, , \quad /n \, ,$$

For all the heavy atoms in the 20 standard amino acid residue types, MODELLER determines 158 various atom types. DOPE scores are calculated for each pair of heavy atoms, excluding the N-terminal nitrogen and C-terminal oxygen atoms and tabulated for distances from 0 to 15 A^o with an interval of 0.5 A^o. Atom pairs between different residues have the pair potential; however, the calculation does not include the atom pairs in bond, angle, and torsion angle.

Then, the scores are entered into the cubic spline function. This function not only calculates the single point energy of a protein model, but also calculates the first and second derivatives of the energy function in molecular dynamics protocol. A protein's absolute score is insignificant as the DOPE energy is not standardized

depending on the proportions of the protein. Nevertheless, in model assessment, the related energies of various conformations are effective. Single chain protein structures form a potential. Making use of the score on single chain proteins is highly secure.

DOPE scores can be calculated at high and regular resolutions; for former, a bin size of 0.125 and for latter, a bin size of 0.5 is needed.

2.3.1.2 Verify protein

Verify protein protocol is mostly used in the final phase of a homology modelling for analysing a preliminary protein structure on the basis of experimental data. Compatibility of a protein's structure with its own sequence is a requirement for the functioning of this protocol.

The strategy that Verify Protein makes use of comprises building a protein's 3D profile and calculating its compatibility with the same protein's sequence. Verify Protein can also be used to calculate local 3D-1D scores in a fixed-length sliding window (typically about 5 to 20 residues long). These can be plotted against residue position to expose local regions of comparatively high or low 3D-1D compatibility. Regions with remarkably low scores are likely to be places where the backbone has been incorrectly threaded through the electron density data (if the structure is experimental) **[19].**

Chapter 3
Prediction of protein-ligand interaction in drug design

3.1 Introduction

Due to the vital role they play in critical metabolic pathways and the incomparable opportunities they offer for structure-based drug design and discovery, many of the macromolecules are considered to be potential therapeutic targets. With a growing importance in the endeavour, structure-based design has now become an integral part of medicinal chemistry **[20].**

Relenza, which is used for the treatment of flu, is the first drug to be designed in that way. It was developed through the selection of molecules that had the potential of binding the conserved regions of the enzyme neuraminidase, which is created by the flu virus to take newly formed viruses away from infected cells. The prevention of this enzyme would therefore keep new viruses from being dispersed into the body. For the prevention of the parainfluenza virus hemaglutinin-neuraminidase, a similar structure-based approach was applied **[12].** Experimental techniques and their outcomes in the database help computational drug design by providing the 3D structure of proteins. More than 35,000 crystallographic or NMR structures of proteins or nucleic acids are available in the Protein Data Bank (PDB). Furthermore, the rate of 3D macromolecular structure identification keeps increasing every year, especially with the development of new techniques such as X-ray crystallography **[20].** The interest in computational methods continues to increase for the prediction of protein-drug docking.

When trying to find out the structure of a protein-ligand complex, a problem may sometimes occur in X-ray crystallography. Through analogy with previously

identified homologous proteins, the positions of the protein atoms can be identified more directly. However, the electron density of the ligand may occasionally turn out to be too unclear to determine the atomic positions of its binding mode. Identification of the most favourable conformation of the ligand when bound to the protein molecule can be achieved through computational docking. The protein molecule may later be tested for affinity with the experimental electron density [21].

Computational docking may be used to identify the most convenient form of the ligand when bound to the protein molecule, which may then be tested for compatibility with the experimental electron density. Likewise, making use of computation, the binding of substrates, products, and transition states may be predicted. This would also allow study of enzyme mechanisms and bound states which couldn't be studied experimentally [21].

The structural and energetic recognition of a ligand and a protein helps determine the selective binding of a low-molecular-weight ligand to a particular protein. Free energy decreases when a bond between atoms is formed. ΔG is thus negative. The binding affinity can be determined from the experimentally measured binding constant K_i

$$\Delta G = -RT \ln K_i = \Delta H - T\Delta S$$

The experimentally estimated binding constant K_i ranges mostly between 10^{-2} and 10^{-12} M, equal to a Gibbs free energy of binding ΔG between -10 and -70 kj/mol in aqueous solution [22]. Beside the experimental techniques used for the determination of ΔG, there are also computational methods for that. Docking is one of these methods. Docking protocols are considered to be a combination of a search algorithm and scoring functions. The number of various search algorithms and scoring functions is comparatively large and continues to increase. The degrees of freedom of the protein-ligand system should allow for the system to be adequately tested by the search algorithm, so that it consists of true binding modes. Speed and effectiveness in covering the related conformational space are certainly the key elements of a search algorithm. In addition to these essentials, the thermodynamics of interaction of the protein-ligand system should sufficiently be covered in the

scoring function in a way that the true binding modes could be separated from all the others discovered and be classified appropriately.

3.2 Docking algorithm

In the prediction of a ligand binding pose, the effectiveness of a docking algorithm is usually evaluated in terms of the root-mean-square deviation (RMSD) between the experimentally observed heavy-atom positions of the ligands and the system is major challenge in the search for the correct pose. The number of degrees of freedom included in the conformational search is the main feature that affects the searching efficiency. Docking algorithm falls into two general categories: Flexible docking search algorithm and matching algorithm.

3.2.1 Flexible docking search algorithm

Flexible docking search algorithm three general categories of algorithms are designed to treat ligand flexibility: systematic methods; random or stochastic methods; and simulation methods **[20].**

The systematic search algorithms attempt to examine all the degrees of freedom in a molecule, and can be further divided in three main types: conformational search methods, fragmentation methods, and database methods.

Random search algorithms test the conformational space by performing random changes to a single ligand or multiple ligands. There are three main types of methods based on random algorithms: Monte Carlo methods (MC), Genetic Algorithm methods (GA), and Tabu Search methods **[20].**

Many programmes function on the basis of GAs (for example, AutoDock, DARWIN, DIVALI, GOLD, EADock, FITTED and PSI-DOCK) **[23].** Genetic algorithms (GAs) benefit from ideas obtained from genetics and the theory of biological evolution to docking. GAs firstly start from a population of diverse conformations of the ligand regarding the protein. Each conformation is characterized by a set of state variables (described as genes) that describe facets such

as translation, orientation, and conformation of the ligand related to the protein receptor. The full set of the ligands state variables is described as the genotype, whereas the automatic coordinates refer to the phenotype [20]. Different methods are employed for choosing the next generation; however, the most popular is the survival of the fittest, where the two lowest scoring conformations are passed to the next generation [23].

3.2.2 Matching algorithm

A pharmacophore representing the protein is initially designed and used to guide the docking in matching algorithms. An initial ligand conformation is created and a ligand pharmacophore is extracted from this conformation. The distance matrices (list of all the distances between the pharmacophoric points) of the ligand and protein pharmacophore are analysed for a match. When a match is found, rotational and translational vectors are computed and applied to the ligand. These vectors place the ligand in the same frame of reference as the protein. While speed is the most important strength of these approaches, their prognostic power is reduced when ligand poses are not totally refined as it makes it difficult for the correct poses to be found or for an exact score to be given. The former issue is addressed by post-processing the selected poses through an additional local search algorithm [23].
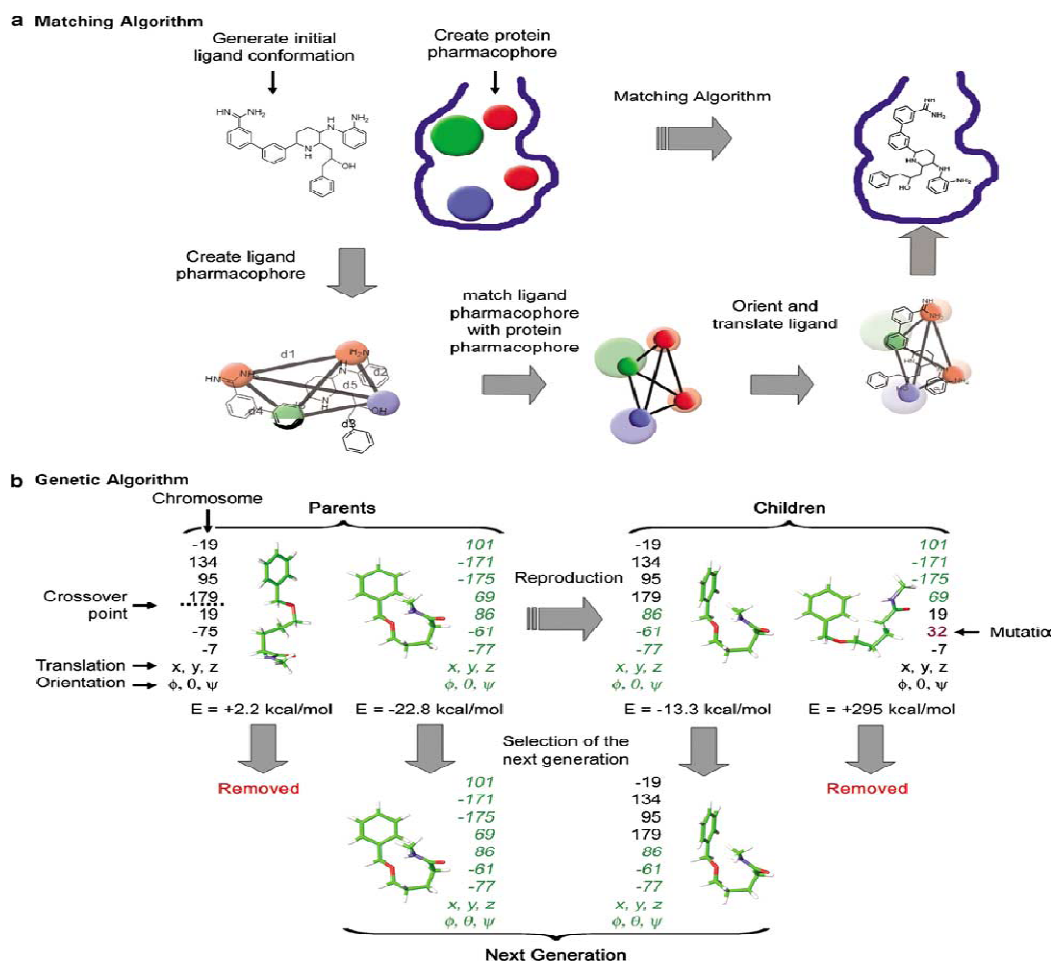
Figure 3.1: Schematic illustration of a matching algorithm (a) and genetic algorithm (b) in the context of docking of a flexible ligand **[23].**

3.3 Scoring function

For an effective docking process, the conformation of the ligand-receptor complex has to be predicted correctly, and the ranking of final structures also has to be correct. The procedure needs to have the ability to differentiate among similar conformations of the same system, as well as to predict the relative stability of diverse complexes **[24].**

The scoring functions normally used in protein-ligand docking are usually capable of predicting binding free energies within 7-10 kJ/mol,14 and can be placed into three major classes: force field-based, empirical, and knowledge-based scoring functions **[20].**

3.3.1 Force Field-Based Scoring

Standard force field (FF) normally measures the sum of two energies: the interaction energy between the receptor and the ligand, and the internal energy of the ligand. The energies are normally calculated through a combination of a Van der Waals in terms of electrostatic energy. The Van der Waals energy term is described with a Lennard-Jones potential, whereas the electrostatic term is given by a Coulombic formulation with a distance-dependent dielectric function that reduces the contribution from charge-charge interactions [20; 22].

$$\sum_{i}\sum_{j} \frac{ec_{ij}}{r_{ij}^2} - \frac{ij}{ij} \quad 332 \quad \frac{i \; j}{ij}$$

Various FF Parameters are chosen; AutoDock, DOCK and RankScore scoring functions (SFs) synthesized the van der Waals, electrostatic and hydrogen bond interaction energy [23].

A wide range of computational docking methods use empirical free energy force fields, which describe basic functional forms for ligand-protein interactions, and semi-empirical free energy force fields, which synthesize traditional molecular mechanics force fields with empirical weights and/or empirical functional forms.

These force fields ensure a fast method to rank potential inhibitor candidates or bound states based on an empirical score. This score may occasionally be calibrated to obtain a prediction of the free energy of binding [25].

The AutoDock force field makes use of a molecular mechanics approach for the evaluation of the enthalpic contributions such as dispersion/repulsion and hydrogen bonding and an empirical approach to assess the entropic contribution of changes in salvation and conformational mobility. Empirical weights are applied to each of the components based on calibration against a set of known binding constants. The final semi-empirical force field is generated to obtain an estimate of the binding constant [25].

The AutoDock 4.0 programme used new force field, which has been calibrated using a large set of diverse protein-ligand complexes, includes two major advances.

The first is the use of an upgraded thermodynamic model of the binding process, which now ensures inclusion of intra-molecular terms in the estimated free energy. Second is that the force field incorporates a full desolvation model that includes terms for all atom types, comprising the favourable energetic of desolvating carbon atoms as well as the unfavourable energetic of desolvating polar and charged atoms. The force field also includes an improved model of directionality in hydrogen bonds, now predicting the proper alignment of groups with multiple hydrogen bonds such as DNA bases [26].

This force field is composed of six pair-wise evaluations (V) and an estimate of the conformational entropy lost upon binding (ΔS_{conf}): L refers to the "ligand", P refers to the "protein" in a ligand-protein docking calculation [27].

$$\Delta G= (\quad bound \qquad unbound \qquad bound \qquad unbound \qquad bound \qquad unbound \quad \Delta \; conf$$

Each of the pair-wise energetic terms includes evaluations for dispersion/repulsion, hydrogen bonding, electrostatic, and desolvation:

$$dw \sum_{i,j} \frac{ij}{\quad_{ij}^2} - \frac{ij}{\quad_{ij}} \qquad bond \sum_{i,j} \frac{ij}{\quad_{ij}^2} - \frac{ij}{\quad_{ij}}$$

$$e \; ec \sum_{i,j} \frac{i\;j}{\quad_{ij}\quad_{ij}} \qquad o \sum_{i,j} \quad i\;j \quad j\;i\; e^{\quad_{ij}^2/2\quad^2}$$

We have chosen to use AutoDock4 program for the docking analysis in our thesis. Information related to the method and the results of this study is given in Chapter5.

34

Chapter 4

Obtaining the 3D-Structure of DAT by comparative homology modelling method

4.1 Introduction

Depending on the working mechanism of neurotransmitters, studies performed on the dopamine neurotransmitter and its receptors are fundamental for the treatment of diseases such as Parkinson's (PD) and schizophrenia. The reason of PD and Parkinson-plus syndromes can briefly be explained as the degeneration of substantia nigra (SN) pars compacta dopaminergic neurons. Due to the loss of nigrostriatal neurons, this degeneration results in the deficiency of striatal dopamine (DA) concentration. In addition to this neuronal loss, there occurs an increase in glial cells of the SN and a loss of neuromelamin (NM). As the nigrostriatal dopaminergic neurons get lost, DA content in SN and the striatum decreases similarly. Within these mechanisms, dopamine agonists were found to provide treatment for the early symptoms and signs of PD **[28].**

Therapeutics used for the treatment of disorders such as Parkinson's and schizophrenia would be considerably improved with the availability of 3D structure for the dopamine receptors and of the binding site for dopamine and other agonists and antagonists.

Some research groups have built homology molecular models for the dopamine transporter (DAT) based on the structure of LacY, (Protein Data Bank, ie., PDB entry of 1PV7 with a resolution of 3.6 A°) **[29],** glycerol-3-phosphate transporters (GlpT, PDB entry of 1PW4 with a resolution of 3.3 A°) **[30]** and

Na+/Cl- dependent neurotransmitter transporters (LeuT$_{Aa}$, PDB entry of 2A65 at 1.65 A° resolution) [10]. For example, previously, to provide structural models of DAT, Xiaqin Huang and Chang-Guo Zhan group members used computational techniques (InsightII software) which based on the X-ray crystal structure of LeuT$_{Aa}$ [1]. Unfortunately the sequence similarity of dopamine is 29% to LacY, 28% to GlpT and 23% to LeuT$_{Aa}$, so that these homology models are not accurate enough to be used to design drugs specific to the receptor subtypes. They have been useful in rationalizing the results of biochemical experiments after using the information to refine the homology model.

Due to the deficiency of adequately similar experimental 3D structures of DAT, computational methods have been applied in this thesis, in order to make predictions of the DAT three-dimensional structure, and depending on these predicted structures its binding sites and binding energies for a variety of ligands.

Martin Indarte, Jeffry D. Madura and other members, had some efforts to predict the 3D structure of rat DAT protein (SwissPort locus SC6A3_RAT; accession number P23977; NCBI accession number AAB21099) by means of comparative molecular modelling with de novo method of the Robetta server. In that study, again the inadequately similar bacterial leucine transporter protein LeuT$_{Aa}$ was employed as the template. However, with the help of the Molecular Dynamic simulations applied further they end up with a structure with important domains taking part in the right place with respect to the experimental results [31].

In this thesis, we have also built a structural model of human DAT using homology modelling based on the templates with experimentally determined 3D structures having sequence similarity less than 30% and, additionally, we had chosen rat DAT robetta [31] 3D structure which has 93.1% sequence similarity as a template. Our aim is to compare the model dependency on the templates, and furthermore examining the binding mechanism variations due to obtained models.

4.2 Method

Comparative methods are based on the construction of 3D structures starting from the amino acid sequence. Homology modelling is the most reliable computational method for predicting protein structure from its sequence. Basically, the method involves searching through the already existing protein data within the databases (PDB) and matching the similar segments of interest. The extended theory of homology modelling is given in Chapter 2 of this thesis. There are several software and online database tools. Discovery Studio (DS) is one of them, which is a comprehensive software suite for analyzing and modelling molecular structures, sequences, and other data of relevant to life science researchers. The software includes functionality for viewing and editing data along with tools for performing basic data analysis. It consists of an interactive environment for viewing and editing molecular structures, sequences, X-ray reflection data, scripts, and other data. Besides that, it also provides a rich set of viewers for displaying plots and other graphical representations of data **[19].**

This method consists of a few steps, first of which is determining templates by making use of sequence similarity search method by BLAST. The second step is aligning and superimposing the templates through structure alignment methods. The third step is the alignment of the model sequence (DAT) with the template sequences by making use of sequence alignment methods. The final step is the 3D model predicted for template protein with respect to some algorithms (given Chapter 2) implemented as MODELLER protocols.

4.2.1 Sequence Similarity Search

We start developing a homology model by determining single or multiple templates from known protein structures according to the sequence resemblance. Templates help us build 3D structural model of DAT based on x-ray crystal structures. The strategy, which is about determining the templates and aligning the model sequence with the template sequences, is based on the matches between the model sequence and the templates. Between templates and model, a high level of homology exists (mostly sequence identity over %60) then the alignment is

considered to be successful. By making use of BLAST, we are able to identify the better templates without difficulty. Correspondingly, simple multiple sequence alignment between the model sequence and templates works successfully.

In our research, we used the BLAST search sequence by similarity tool of the DS program. This tool functions on the basis of certain parameters. Similarity sequence analysis in this work has been performed using the parameters displayed on the following table.

Table 4.1: shows BLAST DS search parameters used.

Parameter Name	Parameter Value
Input Sequence	Human_gi
Input Database	PDB_nr95
Scoring Matrix	BLOSUM62
Gap Penalties	Existence:11 Extension:1
Gapped Alignment	True
E-value Cutoff	10
Filter Low Complexity	True
Maximum Hits	250
Word Size	3

Our input sequence has been set as the Human_gi (hDAT (Access No:Q01959 NCBI databank)) and PDB_nr95 database is used for sequence search of the templates. BLOSUM 62 scoring matrix, E-value (cut-off) value 10 and word size 3 are the other important parameters we have chosen. Detailed information and the theory related to these parameters given on Chapter 2, in the section 2.1.3.1 regarding the BLAST program.

When the task is complete, the sequences identified by BLAST search are automatically displayed, and the templates obtained as a result of our work are

shown in Figure 4.1 and as in the Map View, Figure 4.2 in the following Results and Discussion section.

4.2.2 Structure Alignment

The second step is to perform multiple structure alignment for a group of protein structures. There are two important matters that necessitate particular attention in order to acquire an optimal structure alignment. The first is the global optimization problem. In order to find out an optimal solution, a wide range of residue equivalence combinations in 3D space need to be searched. The second matter of concern is about identifying a target function; here, the question is "Is the alignment optimized to align the largest number of residues, or for the lowest RMSD?". Whereas there isn't a single correct solution, all the algorithms attempt to minimize the RMSD and maximize the residues to be aligned.

It is time consuming and impractical to thoroughly examine all probable alignments. For identifying the optimal solution, many heuristic methods are employed by various programs. Making use of a fast algorithm in order to choose the initial seed alignment on the basis of a hash table, the seed alignment is turned into full alignment. The similarity between two proteins is computed using the coordinates of their Cα atoms. Next, in order to perform the final multiple structure alignment, best-pairwise alignments for each pair of input structures are used **[19].**

In this work, the approaches explained above are further applied can be used by aligning a model sequence (hDAT) to templates found in the previous steps. Through the alignments of sequences with respect to a sequence profile, the system made use of the template structure and later combine the model sequence information with the template structures. A sequence profile defines the existence possibility of twenty standard amino acids at an assumed alignment position. The possibility is computed on the basis of the multiple sequence alignment. For profile calculation, a position-based weighting method (Henikoff and Henikoff, 1994) **[32]** is used to every sequence. Information on this subject is given in the Chapter 2.

Following the completion of the task, a Sequence Window (Figure 4.3) displays the alignment between the model sequence and the templates, and a Molecule Window (Figure 4.4) displays the super-imposed template structures.

4.2.3 Building a 3D model

In this section, by making use of the alignment we have performed and three template structures created with the automatic model-building program MODELLER (Sali and Blundell 1993), we have developed a 3D model belonging to our original protein sequence (hDAT). As a result, protein structures of a model sequence is created based on an alignment of the model sequence with other related known protein structures.

Table 4.2: Build Homology Models protocol parameters.

Parameter Name	Parameter Value
Input Sequence Alignment	gi:ALL
Input Model Sequence	Gi
Input Template Structure	2A65,3FZEA,1SWWB
Water	FALSE
Number of Models	2
Optimization Level	Low
Cut Overhangs	TRUE
Refine Loops	FALSE
Parallel processing	FALSE

As seen in the table above, the proteins of the previous step are used as template structure and a 3D model of the input model protein hDAT (Access No: Q01959 NCBI databank) has been created.

The sequence and the model structures alignment figures are given Figure 4.5 and Figure 4.6. PDF Total Energy and Physical Energy and DOPE scores of all the models are displayed in the Table 4.5. Detailed information on these energy and score rates is given in the section 2.3.1.1 of the Chapter 2.

4.2.4 Assessing the validity of the 3D structure

Verification of the final protein structure is performed by analyzing the validity of a hypothetical protein structure through measuring the compatibility between that structure and the protein's structure. It uses the sequence of residue information to further evaluate the model structure. The theory behind the evaluation process is given Chapter 2.

In this protocol;

Table 4.3: Verify Protein (Profiles-3D) protocol parameters.

Input Protein Molecule	gi:giM0002
Smooth Window size	10
Secondary Structure Model	Kabsch-Sander

As seen in the Table 4.3, giM0002, which is the built 3D model protein in the previous step, has been chosen as the input protein. Kabsch-Sanger model has been chosen as the secondary structure model and smooth window size has been set as 10.

4.3 Results and Discussion

In this work, we have applied the methods described above, step by step according to homology modelling protocol of the DS which is the software we used for the 3D conformation analysis of our model protein, i.e hDAT protein. This chapter, includes the results corresponding to each step and the interpretation of these results.

4.3.1 Structural model of DAT based on the protein database

Here, first of all we have performed single/multiple sequence alignment to find out the possible most appropriate templates throughout the known 3D structures of proteins existing in the databases. The results coming out of the sequence alignment analysis of the hDAT protein by means of the BLAST, which is the first step of homology modelling, is displayed in the following figures;
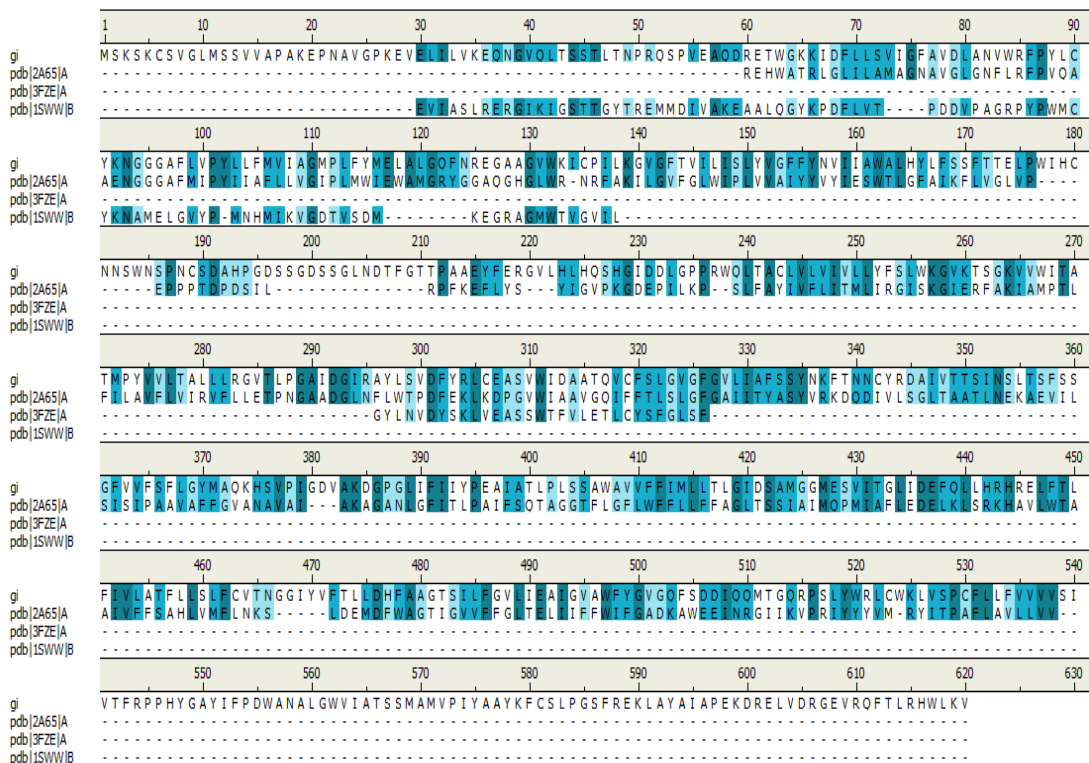
Figure 4.1: Sequence alignment results from BLAST. Blue: identical alignment similarity, light blue: weak alignment similarity, dark blue: strong alignment similarity.

Throughout the BLAST searches; hDAT sequence (query sequence:gi) matches some PBD_nr95 protein database sequences in the matching sequences are proteins with PDB ID number; 2A65, 3FZEA and 1SWWB. As seen in the figure above, the amino acid sequence of the hDAT protein is aligned with the aminoacid sequences of the proteins. The alignment covers amino acids from 30th residue of N-terminal to 326th residue of C-terminal with ISWW, from 60th residue of N-terminal to 538th residue of C-terminal with 2A65, from 296th residue of N-terminal to 326th residue of C-terminal with 3FZE. According to these results, hDAT protein is aligned mostly with 2A65 protein amino acid sequence.
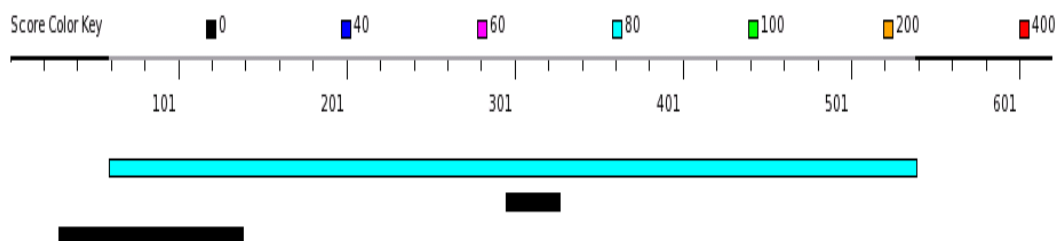
Figure 4.2: The map view displays the coverage of the hits in a map, with one line per sequence. The bars are colored according to the bit score of the hits (with above 400, red, being the best hits).

As a result, as can be seen from the figures, three successful hits occurred as a result of the protein sequence alignment analysis of hDAT protein. The BLAST result; the names of these three best fitting proteins, the organisms that they belong to, amino acid sequence lengths, alignment scores, expect values, percent identities and gap percentages are given in the following table, Table 4.4.

Table 4.4: BLAST search results for the hDAT sequence.

Protein Name	Na(+) neurotrans. symporter,(SNF family)	Protein STE5	Phosphonoacetaldehyde Hydrolase
Accesion Number	2A65	3FZE	1SWW
Organism	Aquifex aeolices VFS	Saccharomyces cerevisiae	Bacillus cereus
Length	509	185	257
Score(bits)	98.2	28.9	27.3
Expect	4e-21	3.7	9.1
Identities	119/498(23%)	13/34(38%)	28/110(25%)
Positives	212/498(42%)	19/34(55%)	50/110(45%)
Gaps	57/498(11%)	3/34(8%)	14/110(12%)

As shown on the report regarding the BLAST search, score value of the results have been calculated as bit value. Bit score is defined as;

$$E = K \times m \times n \times e^{-\lambda S}$$ (given in section 2.1.3.1)

As seen on this formula, there is an inverse proportion between the score value and the Expect Value (E). As I have mentioned in the theory, section 2.1.3.1 of the Chapter 2, the smaller the *E*-value the higher the percentage identity, and the higher the score the better the alignment is for the evaluation. These criteria are used of the similarity between the database sequence and the query sequence. According to this explanation, as the Expect Value of the Na(+) dependent neurotransmitter symporter protein (2A65) is lower than other proteins, the highest score value pertains to this protein. So that this is considered to be the best template in the database.

Na(+) dependent neurotransmitter symporter (2A65) is a part of the *Aquifex aeolicus* organism and pertains to the NSS protein family with known 3D structure

(enlightened through the x-ray crystal structure) [**1**]**.** In general, this protein is used as a template in the researches regarding the proteins pertaining to the NSS protein family (dopamine transporter and serotonin transporter). This protein (2A65) was also found as the best hit at our sequence alignment search with the hDAT that is used in our research. Therefore it was used as a template for the further steps. However, the percent sequence similarity between this protein and hDAT was calculated to be lower than other proteins. Even though similarity of the sequence between proteins are sometimes low, the 3D structures of those proteins may turn out to be homologous. For instance, X-ray crystallography makes it evident that myoglobin and beta globin have quite similar structures. Whereas human beta globin and human neuroglobin share only 22% amino acid identity, the myoglobin and alpha globin proteins share only about 26% amino acid identity [**13**]**.**

At the second step of the homology modelling, amino acid sequences between our protein and the proteins with the best-hit scores have been aligned to determine the secondary structures. The results of this method are shown in Figure 4.3 and Figure 4.4.
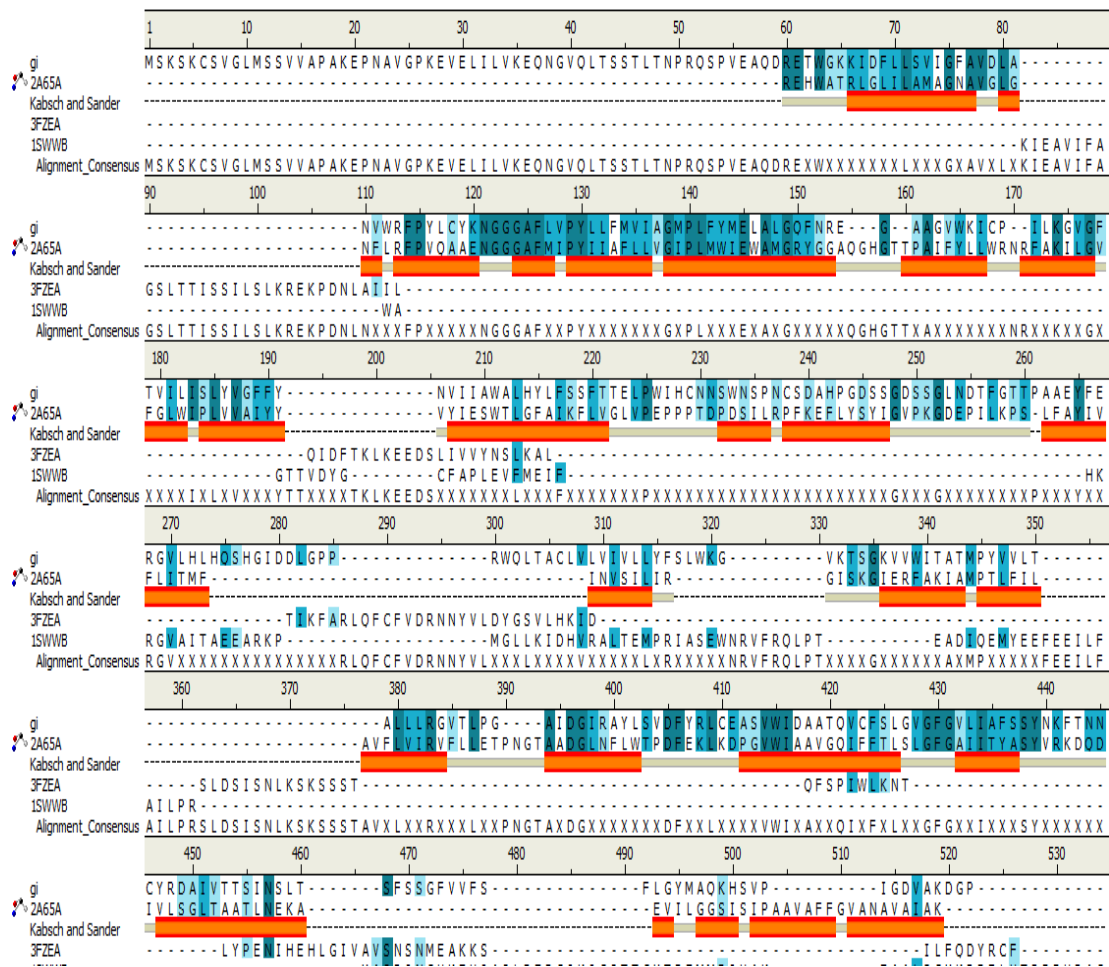
Figure 4.3: The sequence alignment analysis pertaining to the templates (2A65, 3FZEA, 1SWWB) and the model (gi: hDAT) protein. According to the secondary structure definition of Kabsch and Sanger, alpha helix regions are shown in orange. The parts of the model protein and the template proteins which haven't been aligned are defined as X.

(A)   (B)   (C)
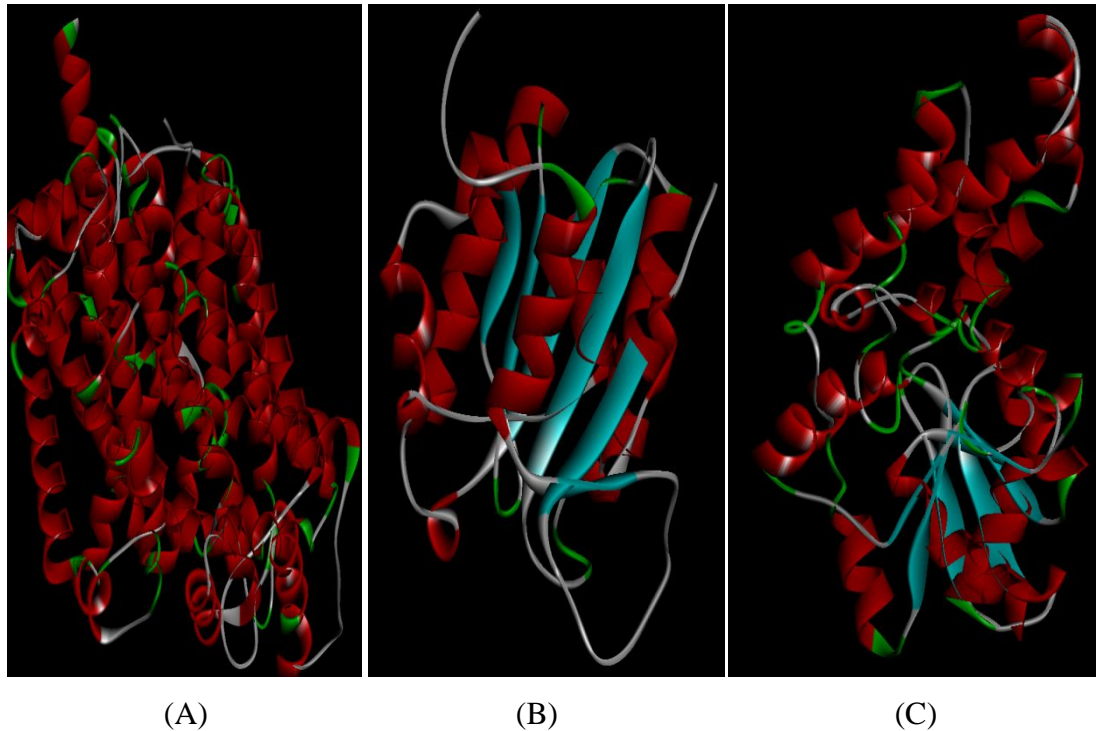
Figure 4.4: Template structures: 2A65 (A); 3FZEA (B); 1SWWB (C). Red bars indicate alpha helices, blue arrows indicate beta strands and green arrows indicate coil.

As a third step, we made use of these secondary structures of three template proteins, which have been proven, at the previous stages, to be homologous with the hDAT protein. Here, from now on we have attempts to build the 3D structure of the hDAT protein. As a result of this task, initially two different model structures have been obtained (gi.M0001 and gi.M0002). The following figures, Figure 4.5 and Figure 4.6 show the secondary structure alignments during this step and the resultant 3D model structures obtained.

Figure 4.5: The structure alignment results of the model (gi: hDAT), templates (2A65A, 3FZEA, 1SWWB) and the model proteins (gi.M0001 and gi.M0002) created through the MODELLER program. Regions colored in orange represent the alpha helix and the arrows colored in blue represent the beta-sheets.
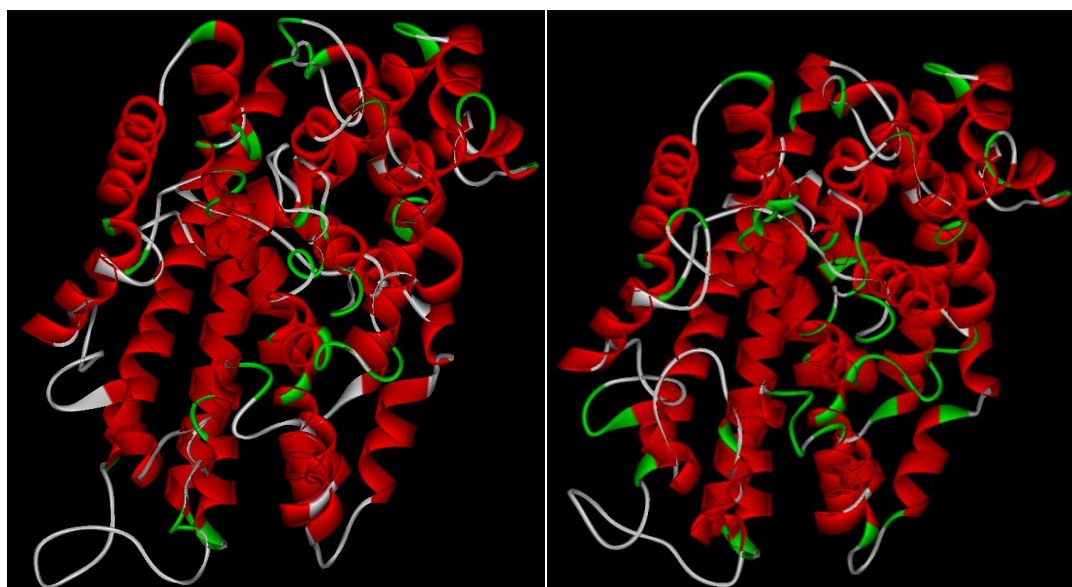
<center>(A)</center> <center>(B)</center>

Figure 4.6: The resultant 3D model structures of giM0001 (A) and giM0002 (B). Red bars indicate alpha helices and green arrows indicate coil.

As I have previously mentioned in the MODELLER section of the Chapter 2, a low energy level is the indicator of a better alignment when these models have been generated from different initial alignments and homologues. According to the results shown in Table 4.5, the empirical PDF Total Energy value of gi.M0002 is less than that of gi.M0001. Therefore, we used only gi.M0002 model for the further steps.

Table 4.5: The MODELLER PDF energy data and DOPE Scores are stored in each model structure.

Model Scores			
Name	**PDF Total Energy**	**PDF Physical Energy**	**DOPE Score**
gi.M0002	27974.937	3339.621	-62487.925
gi.M0001	31550.265	3722.596	-61833.507

After obtaining the model 3D structure, a verification process should be performed. Here, this validation was achieved by comparing the verification scores, which is a numerical way for indication of the compatibility.

Table 4.6: The Verify Protein (Profiles-3D) Verify Score, Verify Expected High Score and Verify Expected Low Score values are shown in each model structure.

Name	Verify Score	Verify Expected High Score	Verify Expected Low Score
2A65A	234.07	233.058	104.876
3FZEA	75.86	83.8575	37.7359
1SWWB	132.08	116.801	52.5603
gi.M0002	149.03	232.597	104.669

If the Verify Score result of a model protein is higher than Verify Expected Low Score value, then the model is considered to have acceptable quality. And the closer the Verify Score result is to the Verify Expected High Score value, the better the quality of the model. In our result; the model (gi.M0002) Verify score is higher than its Verify Expected low score value. Therefore; we use this as an acceptance criteria.
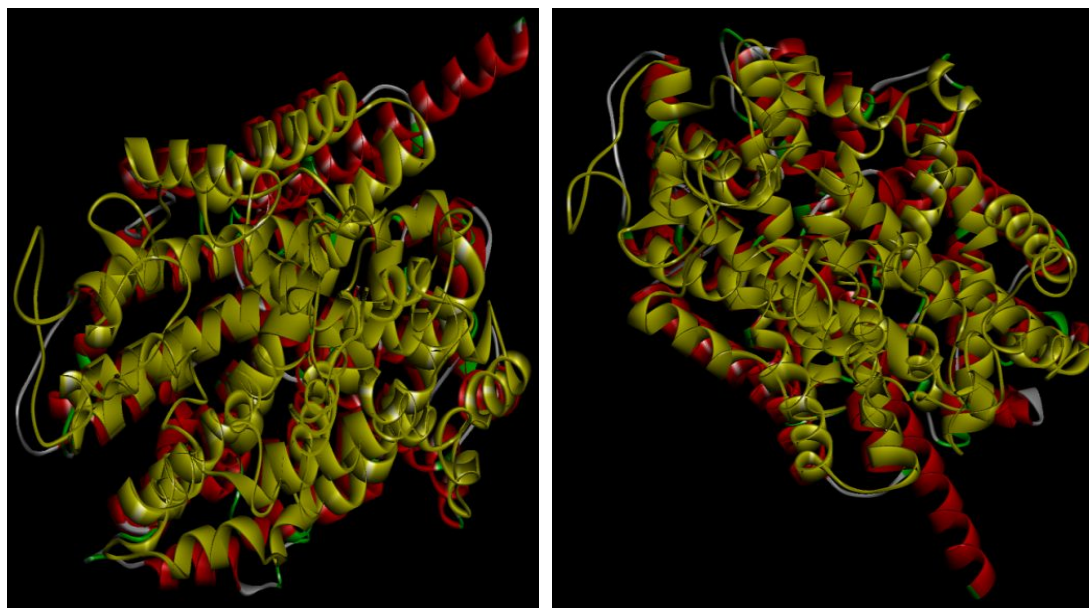
Figure 4.7: The superimposed poses of the tertiary structure prediction result for giM0002 (model protein) from MODELLER, and the 3D structure of template protein 2A65 from the two different view angles. Yellow color: giM0002 red color: 2A65.

As a conclusion, the model gi.M0002 has been obtained as the most suitable 3D structure model from the overall results of the homology modelling study we have conducted for hDAT protein so far.

A further energy minimization procedure is applied on this model by means of simulation, in order to get rid of the steric clashes and to end up with an energetically better conformation. Within this procedure, first of all, the model was ionized for neutralization purposes.

After the ionization of the model protein which was created through this task, a brief energy minimization has been performed with the NAMD program. The energy-minimized DAT structure has a RMSD of 0.89 A^0 for Cα atoms from the initial structural model. This energy minimized structure is shown in Figure 4.8.
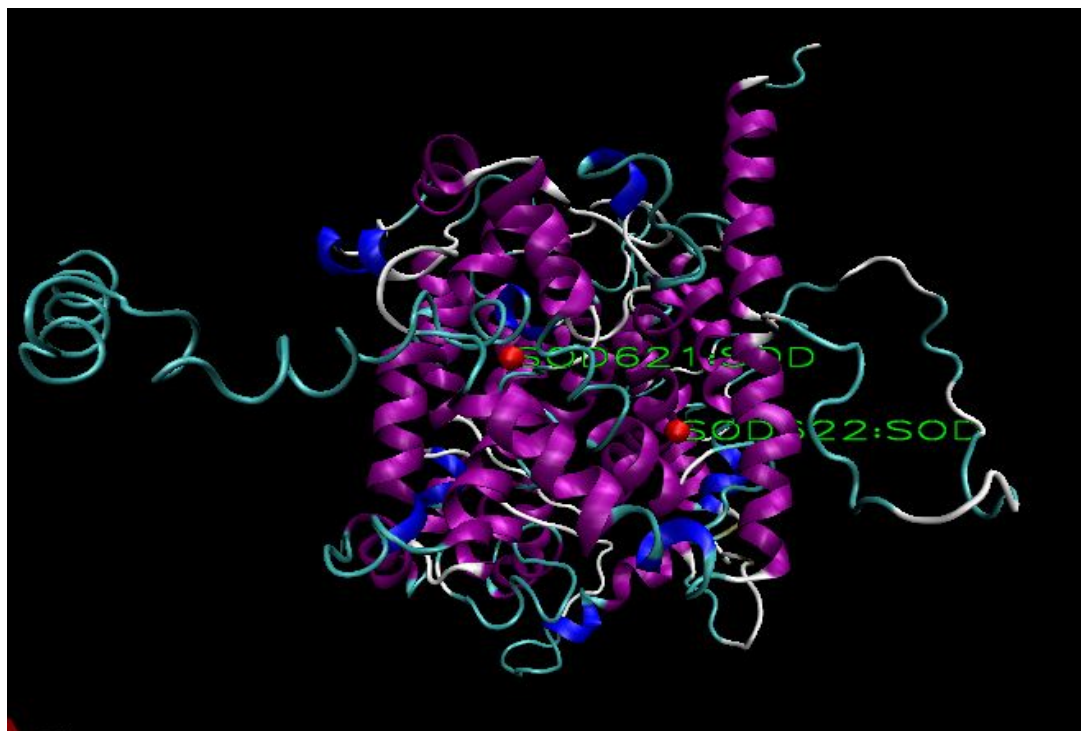
Figure 4.8: The 3D structure of the hDAT protein after the energy minimisation. Purple represents the alpha helix, blue represents the helix_3_10, green represents the turn and red represents sodium ions.

Helical structures shown in purple on the above figure are the regions that are identified as the alpha helices. The residue intervals of these regions given in Table 4.7 shows the computationally obtained intervals.

Table 4.7: The amino acid sequence positions of the Transmembrane helix (TMH) regions pertaining to the model hDAT protein by VMD (VMD used by STRIDE secondary structure codes).

TMH Number	Aminoacid seq.position	TMH Number	Aminoacid seq.position
1	68-77	13	328-335
2	86-92	14	343-356
3	99-109	15	392-400
4	113-123	16	404-429
5	142-173	17	442-455
6	192-198	18	473-483
7	214-225	19	490-498
8	246-252	20	501-509
9	263-276	21	531-544
10	279-286	22	553-568
11	290-298	23	600-615
12	308-322		

As seen in the above table, as a result of the research study we have conducted on the hDAT protein, 23 TMH (Alpha helix) regions have been identified. When we check for comparison purpose, as a consequence of the experimental studies, 12 alpha helical regions in the hDAT protein have been identified **[33].** Figure 4.9 shows the results of these experiments. And in the following table, Table 4.8 the corresponding residue intervals for alpha helices obtained experimentally are tabulated.

Figure 4.9: Nucleic acid and deduced amino acid sequences of the hDAT cDNA clone. The 12 TMH domains are boxed **[33].**

Table 4.8: Amino acid sequence positions of the TMH (alpha helix) regions detected in the hDAT protein through experimental studies.

TMH Number	Aminoacid seq.position
1	67-90
2	98-121
3	140-163
4	238-259
5	266-289
6	310-333
7	347-369
8	395-418
9	440-463
10	478-500
11	518-541
12	556-566

Even though this seems like a contradiction, actually it is not. Because when examined residue-by-residue, the regions cover more or less the same regions. To conclude, the TMH (alpha helix) regions of the model obtained from the computational 3D structure analysis that we have performed for the hDAT protein and the TMH (alpha helix) regions that are experimentally detected, have been identified as approximately similar. And in addition, TMH (alpha helix) regions have also been found on the model protein were obtained through the structural analysis tools of the computational softwares, which uses different definitions of the secondary structures. So that, this might be an other reason for the numbering.

As a result, computational and experimental studies correlates well for hDAT model obtained based on the multiple sequence alignment of proteins within the protein database.

4.3.2 Structural model of DAT based on rat DAT template

In our second homology modelling study, we used the rat DAT robetta (rDAT_robetta) protein structure as a template whose 3D structure modelling had previously been created by some other independent researchers that made use of the Robetta server and the structure template pertaining to the LeuT$_{Aa}$ **[31].**

The 3D structure modelling of the hDAT protein has been performed with the above mentioned homology modelling methods. Different from the previous study, in this process, a built 3D model structure pertaining to the rat DAT protein has been used as a template.

In the first stage, sequence alignment analysis between the amino acid sequences of the hDAT protein and the rDAT_robetta has been performed, and corresponding results are shown in Figure 4.10
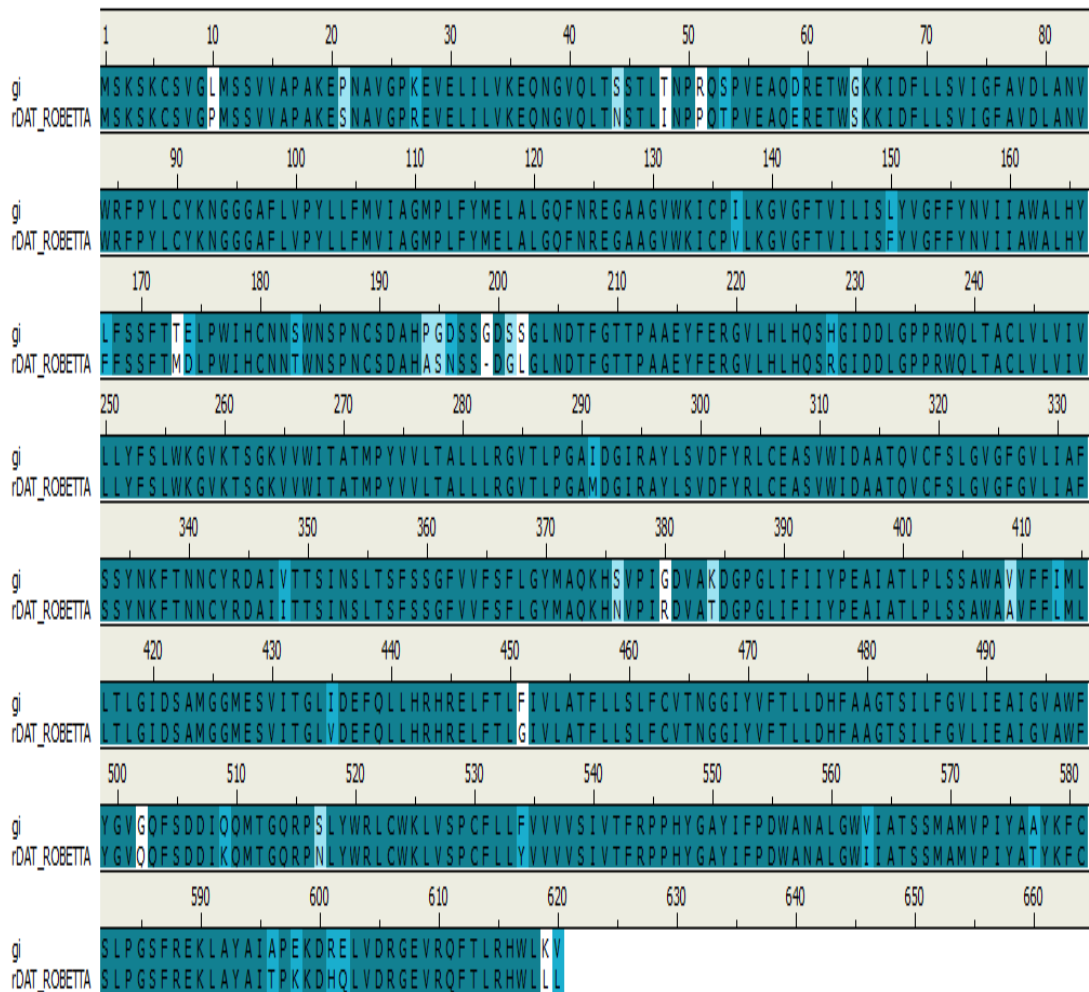
Figure 4.10: Sequence alignment results from BLAST. For BLAST searches; hDAT (query sequence) are aligned with the rDAT_robetta sequence.

As a result of this stage, a sequence identity of 93.1% has been detected between hDAT with rDAT_robetta .

In the second stage, the secondary structure modelling of the hDAT protein has been performed by making use of the rDAT_robetta structure template. As a result of this study, two different hDAT model structures have been obtained. The 3D structures of these two models and the rDAT_robetta is shown in Figure 4.11.

(A)   (B)   (C)
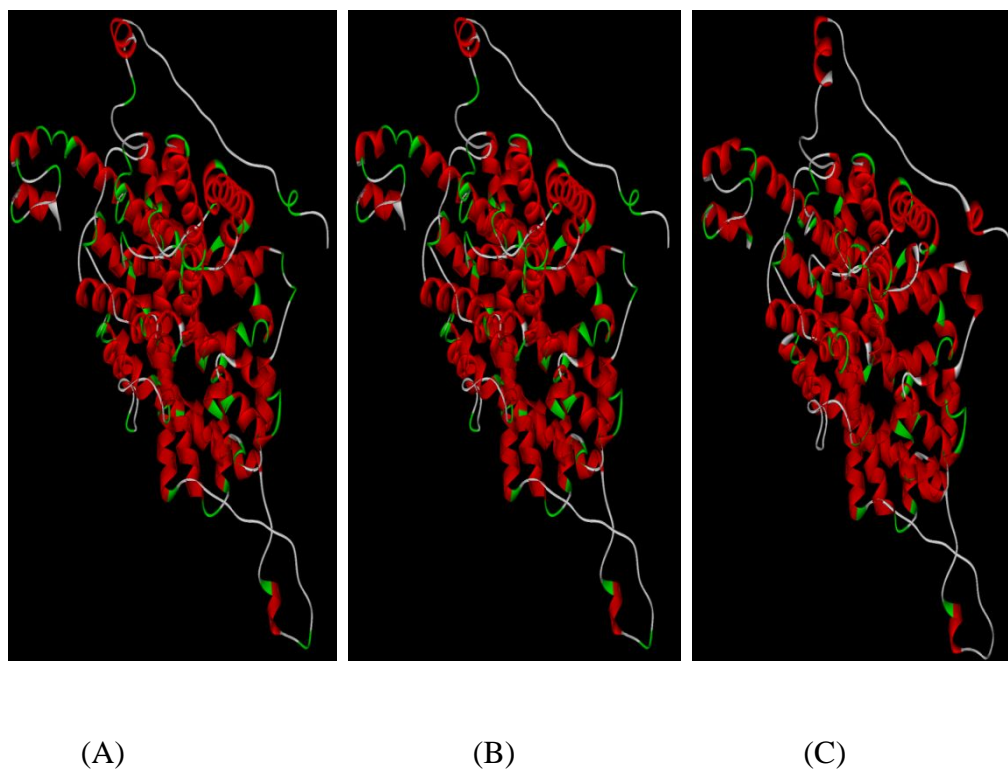
Figure 4.11: 3D structures of rDAT_robetta (A), model structure giM0001 (B) and giM0002 (C).

Table 4.9: The MODELER PDF energy data and DOPE Scores are stored in each model structure.

Model Scores			
Name	**PDF Total Energy**	**PDF Physical Energy**	**DOPE Score**
gi.M0001	3728.801	1890.910	-75957.132
gi.M0002	4087.605	1936.183	-75521.0937

    Model scores corresponding to these two models are given in Table 4.9. According to these results, the PDF Total Energy of gi.M0001 is lower than that of gi.M0002. As previously mentioned, lower energy level is an indication of better alignment. Therefore, we used only gi.M0001 model sequence in the further steps of the modelling.

And finally, a validation procedure should have been applied. For this reason, the required verification scores are given in Table 4.10. According to these values, the Verify Score result of the gi.M0001 model protein is higher than Verify Expected Low Score value, therefore this model is verified to be acceptable.

Table 4.10: The Verify Protein (Profiles-3D) Verify Score, Verify Expected High Score and Verify Expected Low Score values are shown in each model structure and rDAT_robetta.

Name	Verify Score	Verify Expected High Score	Verify Expected Low Score
rDAT_robetta	180.23	283.307	127.488
gi.M0001	131.43	283.307	127.488

In Figure 4.12, the 3D structure of superimposed rDAT_robetta and the model hDAT is depicted. The high similarity also in structure can be seen from the figure conforming the amino acid sequence similarity.
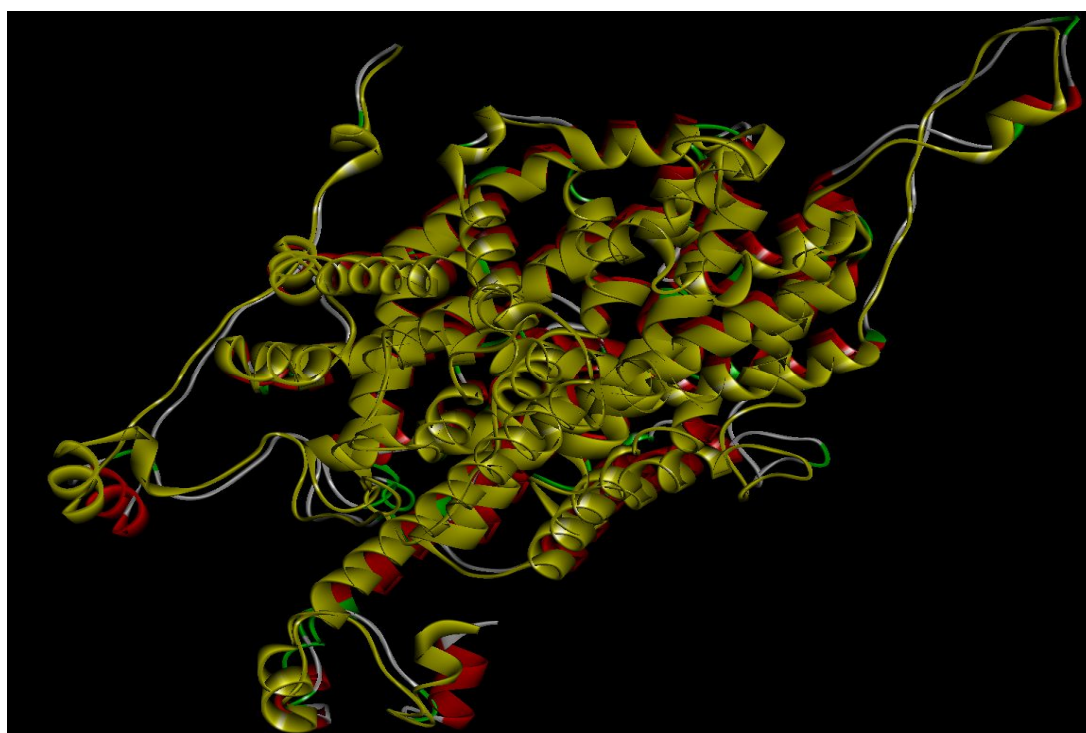
Figure 4.12: The tertiary structure prediction results rDAT_robetta (template protein) red color and giM0001 (model protein) yellow color.

In the table below, the amino acid residue intervals for the alpha helices belonging to the 3D model obtained here are given.

Table 4.11: The amino acid sequence positions of the TMH (alpha helix) regions pertaining to the giM0001 model structure by VMD.

TMH Number	Aminoacid seq.position	TMH Number	Aminoacid seq.position
1	5-11	13	328-335
2	27-34	14	342-373
3	65-77	15	376-384
4	79-93	16	387-400
5	96-124	17	404-438
6	137-172	18	442-457
7	194-199	19	467-479
8	206-221	20	481-497
9	240-253	21	501-509
10	258-282	22	518-550
11	290-298	23	553-594
12	308-322	24	611-619

As seen in the table, 24 alpha helix regions have been identified on the model protein obtained from this study. As previously mentioned, 12 alpha helix regions had been identified as a result of the experimental study. The helix regions identified in both computational studies are similar with some extra alpha helix regions identified.

Consequently; for enlightening the 3D structure of the hDAT protein, two different hDAT model structures have been built using different template structures pertaining to different template proteins. After the structure alignment performed in

VMD for these two resultant models. The RMSD value of C_α atoms pertaining to these two different models have been calculated as 4.792 A^0 for $C\alpha$ atoms. This number is an indication of two similar but some how different structure models of hDAT protein. The variations in the structure can be examined from the Figure 4.13 given below, which is the 3D structures of hDAT protein obtained by using different templates.
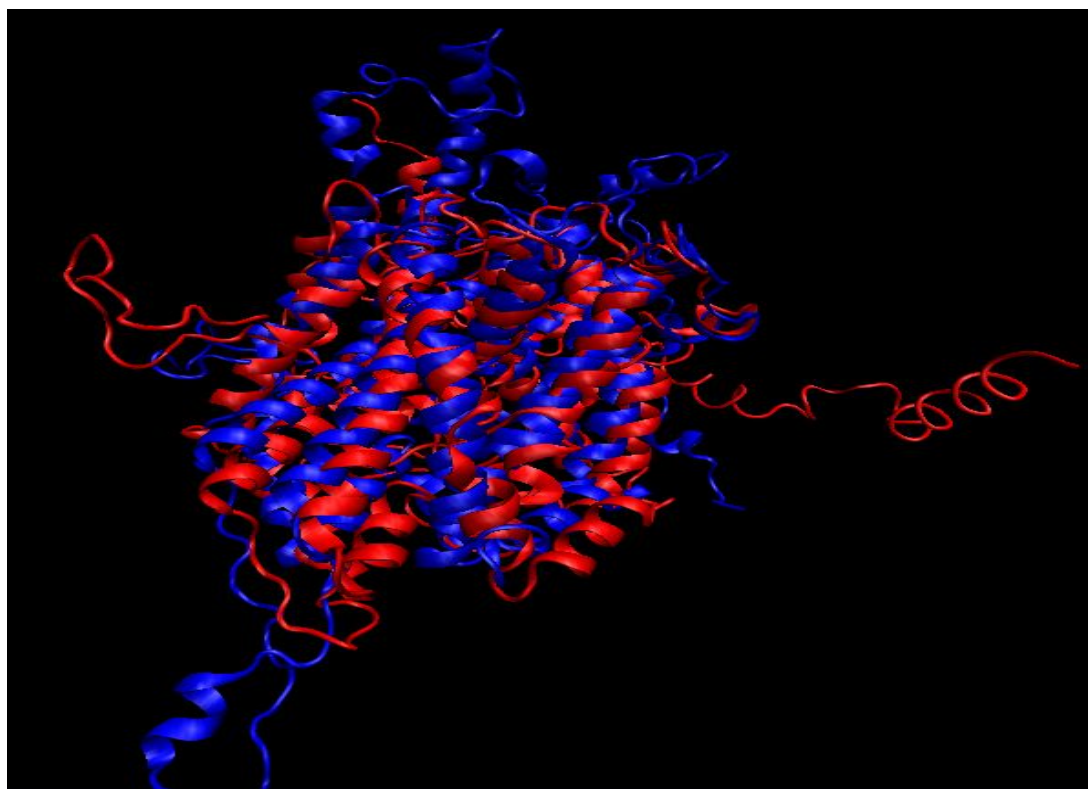
Figure 4.13: The hDAT model structure (red) created initially by using multiple templates from the database and the hDAT model structure (blue) created as the second which based on the rDAT_robetta as a template.

In this study, we must concern the fact that LeuT$_{Aa}$ protein had been used as a template for the homology modelling of the rDAT_robetta protein, which was itself used as the template in the latter part. LeuT$_{Aa}$ protein is also one of the proteins that we have used as a template during the homology modelling we have performed in the first part throughout the multiple sequence alignment. As a result, the 3D structure pertaining to LeuT$_{Aa}$ has been in a way involved in as a template both in the first and second modelling hDAT protein. Similarities observed despite the slight

differences identified in some regions in the 3D model structures pertaining to the hDAT protein can be a consequence related to this fact.

And further investigation of these two model structures were performed in order to see the effects on ligand binding to this protein. This further analysis are given in the next Chapter.

4.4 Conclusion

The molecular modelling of the hDAT protein pertaining to the NSS family whose 3D structure isn't yet available, has been performed through two different comparative homology modelling techniques. Whereas experimental multiple data has been preferred as template in the first part, a single template structure pertaining to the rDAT_robetta has been used in the second.

Three different template proteins (2A65, 3FZEA and 1SWWB) has been used in the first part. The sequence similarities between each of these proteins and the hDAT protein has been calculated less than 40%. 23 alpha helix regions have been identified in the hDAT model (hDAT_MS) created through the use of these multiple template proteins. In the latter part, the sequence similarity between the hDAT and the rDAT_robetta which has been used as template has been calculated as 93.1%. The human DAT model (hDAT_robetta) obtained through this way has 24 alpha helix regions. Even though the sequences of the hDAT models obtained from the application of both methods were the same, differences in the 3D structures of hDAT models has been identified in some regions.

In order to enlighten the functionality of the hDAT models built through the application of these two different techniques, and to identify the changes affecting in the mechanism as a result of the structure differences between the two different molecular ligand docking analysis was performed in the following stage of our thesis, and this studies are given in the next Chapter.

Chapter 5

Calculation of binding energies of DAT with some important ligand molecules by docking

5.1 Introduction

Ligand molecules that binds to dopamine transporter can be categorized as agonists and antagonists. DAT agonists lessen the hypokinesia of Parkinson's disease. And DAT antagonists are used for the prevention of hallucinations and delusion from which schizophrenic patient suffer. Despite the known curing features of both, in some cases adverse effects resulted from dopamine depletion similar to those caused by the restriction of dopamine receptors. For example, dopamine agonists in high doses may result in psychoses. So, in other words, dopamine imbalances may causes serious adverse effects. Inventing selective dopaminergic drugs that do not cause adverse effects is a difficult task in the development of pharmaceutical agents. This task can be achieved by comprehending how the agonists of subtypes of dopamine receptors involve in the mechanism and how they influence each other. Therefore, understanding of the dynamics is an important task and the interactions should be foreseen. Making such predictions directly with the receptor sequences or trial and error methods is tiresome. Hence, in the development of agonists and antagonists, predicting the three dimensional structure and function of DAT and binding sites of dopamine is an indispensable work **[34].**

In the past, there had been some experimental attempts made, in order to investigate the DAT molecule. The cloning of DAT cDNA was worked by Shimada et al. and Vandenbergh et al. **[35, 36].** DAT's structure, function and drug interaction were further investigated as a result of these studies.

However, the crystal structure of the hDAT hasn't been developed yet. Kitayama et al., Lien et al**. [37, 38]**, had also performed some experimental studies in which particular DAT amino acids were mutated. Experimental mutation results on DAT have been used by the group members for the rationalization of the experimental ligand binding measurements by this group member**.**

Furthermore, a more recent crystallization experiment has proven that the bacterial leucine transporter (LeuT$_{Aa}$) consists of 12 TM domains with intracellular N- and C-termini **[10]**. These LeuT$_{Aa}$ TM domains are reported to be merely overlapping in part with the domains suggested for the eukaryotic transporters **[39].**

An inhibitor-binding site located in an extracellular vestibule above the substrate-binding site has been discovered by the LeuT$_{Aa}$ structure. Binding of the inhibitor evidently stabilizes the external gate in a closed conformation at this substrate-binding site. Therefore, this function results in non-competitive inhibition of substrate transport. Along with these discoveries, the question of whether this is a binding mode that can be generalized to the mammalian transporters and to different inhibitor classes emerges **[40].**

Following the studies on LeuT$_{Aa}$, a 3D structural model of DAT was generated computationally on the basis of the x-ray crystal structure of LeuT$_{Aa}$ by the group members Xiaqin Huang and Chang-Guo Zhan. They made use of the structural model of DAT, by using the AutoDock 3.0.5 and the DOCK 5.4 programs, for the inspection of the binding mode of dopamine with DAT through molecular docking **[1].**

In another computational study, Martin Indarte, Jeffry D. Madura and their colleagues also made use of the LeuT$_{Aa}$ crystal structure as a template and targeted rDAT protein sequence. To build three rDAT models, they used three different comparative modelling approaches. In addition to this, to identify potential binding sites for the substrates dopamine and psychostimulant d-amphetamine, three docking methods (MOE-Dock 2005.06, ASEDock and MOE-DOCK 2004.03 GA) were applied to the three DAT by the same group **[31].**

Considering the researches mentioned above, here in this thesis, we have constructed a new 3D structural model for hDAT through homology modelling based on various reasonable templates. Furthermore, we have used the resultant DAT 3D models for the binding analysis of DAT ligands computationally with a docking program, AutoDock 4.0. In addition, based on the developed structural model of DAT, the substrate-binding modes have been determined through the calculation of binding free energies. The overall agreement between the computational results and available experimental data demonstrated some important structural features of DAT and its binding properties with several substrates, providing a valuable insight into the molecular working mechanism for DAT.

5.2 Methods

AutoDock programme makes use of a semiempirical free energy force field and Lamarckian Genetic Algorithm, which is a docking simulation method that provides flexible docking. Lamarckian Genetic Algorithm is mostly useful for systems with about 10 rotatable bonds in the ligand. Lamarckian Genetic Algorithm is more effective than the Simulation Annealing, though it can be useful in applications where search starting from a given point is desired.

5.2.1 The AutoDock Protocol:

The program AutoDock was developed to provide a procedure for predicting the interaction of small molecules (ligand) with macromolecular targets and calculating their binding energies. The steps of AutoDock4 used here were as follows;

1. First, we created the ligand and receptor coordinate files to use as input information for AutoGrid and AutoDock. It included polar hydrogen atoms, partial charges and atom types and information on the articulation of flexible molecules.

2. Second, AutoGrid created grid maps for each atom type of the ligand being docked. This helps to make the docking calculation fast. AutoGrid requires a grid parameter file (gpf) to specify the input files and parameters used in the calculation. The parameters used in our study are as such; the grid size was set to be 70x70x70 and grid space was chosen as 0.375 A^{o}

3. Next we created docking parameter file (*.dpf) including the search parameters and the docking parameters. We have chosen the genetic algorithm for docking. The system runs 100 genetic algorithm with 2500000 number of evaluations.

4. At the end of AutoDock used these parameters in the *.dpf file and calculated the binding free energy of the protein-ligand complex and computed the coordinates for each docked conformation in the output file (*.dlg).

5.3 Results and Discussion

The theory behind the study of the interactions of small molecules (ligands) with macromolecules by means of docking is described in detail in Chapter 3.

The tasks performed here in this thesis were predicting the binding sites and the calculation of the binding energies using AutoDock for eight different ligands using the three-dimensional structure of DAT predicted in our previous steps.

Ligand molecules

5- hydroxydopamine   6- hydroxydopamine   Dopamine
(1)   (2)   (3)

Amphetamine   Methamphetamine
(4)   (5)

Cocaine   Benzene-1.3-diol   Tyramine
(6)   (benzenediol)   (8)
(7)
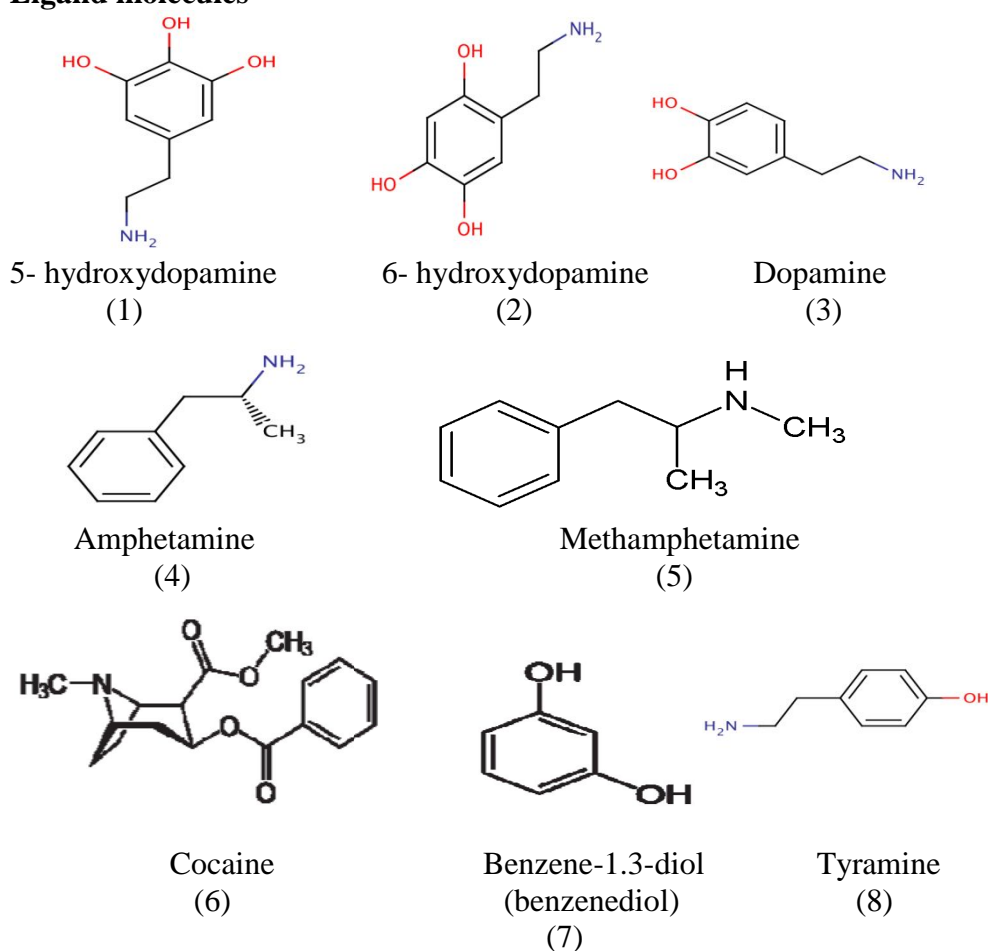
Figure 5.1: Structures of the eight ligands studies as the DAT receptor. 5-hydroxydopamine(1), 6-hydroxydopamine(2), Dopamine(3), Amphetamine(4), Methamphetamine(5), Cocaine(6), Benzene-1.3-diol(benzenediol)(7), Tyramine(8)

When we analyse the chemical structures of the above ligands; we see that all the ligands have an unsaturated benzene ring known as aromatic hydrocarbon. Also,

NH2 (amine) group is common in all the 5-hydroxydopamine, 6-hydroxydopamine, Dopamine and Tyramine ligands, but the number and orientation of their OH (hydroxyl) groups are different. If we look at the organic structures of Methamphetamine and Amphetamine ligands; different from the Amphetamine, an extra CH3 (methyl) group is present in the Methamphetamine ligand. Cocaine ligand, however being quite different than the other ones, is a bigger compound containing an aromatic benzene ring and a piperidine bi-ring. Furthermore, it contains more than one carboxyl oxygens and its free electron number is apparently higher than other ligands, when simply compared.

DAT Models;

(A) hDAT_MS   (B) hDAT_robetta   (C) rDAT_robetta
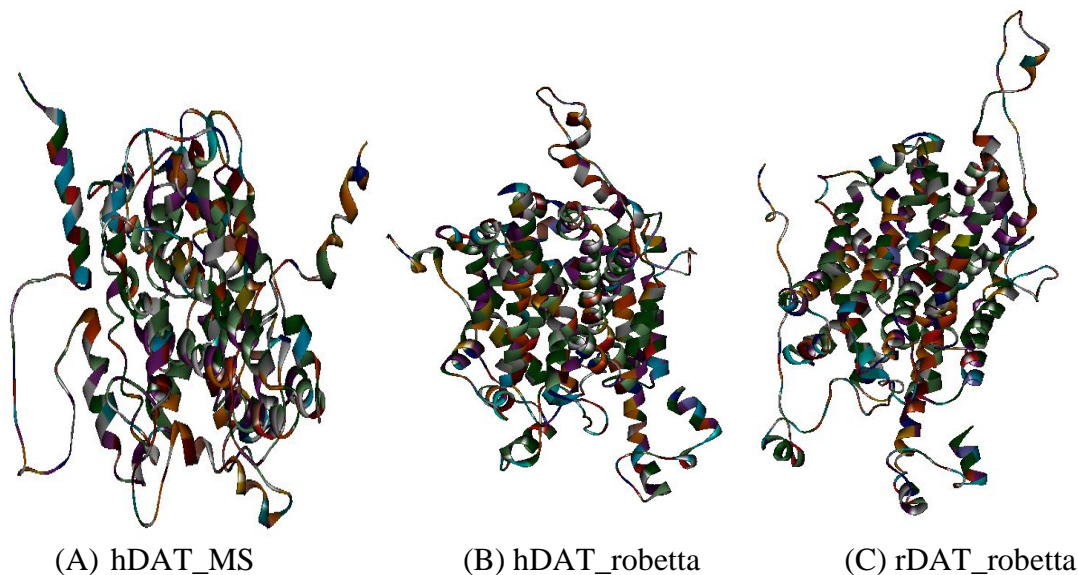
Figure 5.2: (A) The structural energy minimized model of the 3D hDAT protein which has been obtained using multiple experimental data as template. (B) The structural model of the 3D hDAT created using by taking rDAT_robetta protein as template. (C) The 3D structural model pertaining to the rDAT_robetta protein.

5.3.1 Binding Energies of the eight ligand molecules in DAT

The lowest binding energies for the DAT-ligand complex were computed by using the AutoDock4.0. The 3D hDAT's structure models obtained in our previous steps and for DAT of rDAT_robetta were used here. Computational results were given in Table 5.1, Table 5.2 and Table 5.3 for the first model of DAT (hDAT_MS), the second one (hDAT_robetta), and finally rDAT_robetta respectively.

Table 5.1: The calculated binding energies of the eight ligands with 3D model of hDAT_MS.

LIGANDS	LOWEST BINDING ENERGY(kcal/mol)
5-hydroxydopamine	**-4.81**
6-hydroxydopamine	**-6.33**
Amphetamine	**-5.91**
5-(2-aminoethyl) benzene-1.3-diol (benzenediol)	**-5.45**
Cocaine	**-6.33**
Dopamine	**-6.03**
Methamphetamine	**-5.05**
Tyramine	**-6.08**

Among these lowest binding energy scores in the Table 5.1, two best binding ligands for hDAT_MS are found to be cocaine and 6-hydroxydopamine. However, when compared with the other energy values, it is seen that there is not much difference among all.

71

Table 5.2: The calculated binding energies of the eight ligands with 3D model of hDAT_robetta

LIGANDS	LOWEST BINDING ENERGY(kcal/mol)
5-hydroxydopamine	-7.54
6-hydroxydopamine	-7.51
Amphetamine	-7.78
5-(2-aminoethyl) benzene-1.3-diol (benzenediol)	-7.06
Cocaine	-9.28
Dopamine	-7.43
Methamphetamine	-7.29
Tyramine	-7.25

Table 5.2 shows that the best lowest binding energy score pertains to the hDAT_robetta-cocaine complex, throughout the all ligands.

Table 5.3: The calculated binding energies of the eight ligands with 3D model of rDAT_robetta.

LIGANDS	LOWEST BINDING ENERGY(kcal/mol)
5-hydroxydopamine	-7.01
6-hydroxydopamine	-7.35
Amphetamine	-7.23
5-(2-aminoethyl) benzene-1.3-diol (benzenediol)	-7.10
Cocaine	-4.26
Dopamine	-7.12
Methamphetamine	-6.84
Tyramine	-6.71

Looking Table 5.3, the best lowest binding energy scores for all ligand-DAT complexes in rat are observed to be quite similar to each other, expect for cocaine.

According to the experimental studies conducted previously with the DAT-dopamine complex, the binding energy score was calculated as -7.4 kcal/mol **[41].** The scores obtained from our computational research show that the binding energy score pertaining to the DAT-dopamine complex is equal to the experimental score as shown on Table 5.2, and the scores on Table 5.1 and Table 5.3 are quite close confirming a good agreement.

So, our computational results and the experimental ones were compared. However, the binding energy scores are not sufficient for comparing the complex forming tendencies. Therefore, the binding site regions of the DAT-ligand complexes and the atomic interactions between the protein-ligand in these regions have been further investigated by docking analysis in the following sections.

5.3.2 Binding sites of the eight ligand molecules in DAT

Next, we will analyze the predicted binding sites to determine how each residue in the binding pocket contributed.

5.3.2.1 Binding sites in complexes of hDAT_MS

Table 5.4: The residues in the binding sites pertaining to the hDAT_MS-ligand complexes. Common residues found in these regions are shown in dark.

LIGAND	BINDING SITES			LIGAND	BINDING SITES		
5-hydroxy Dopamine	TRP162 ILE159 PHE155 PHE472	THR473 ASP476 PHE320 ARG85	LEU80	Cocaine	ASP476 THR473 PHE472 PHE155	ILE159 TRP162 VAL364 TRP84	VAL158 TYR88 ARG85
6-hydroxy Dopamine	PHE332 **VAL73** SER72 **VAL328** **LEU329**	**ALA77** PHE76 **LEU47** **GLU117** PHE114	LEU322 SER44 **SER45** **GLY323** THR48	Dopamine	**SER45** PHE114 **LEU47** **GLU117** SER72	**LEU329** **VAL328** VAL324 LEU322 PHE332	**GLY323** **ALA77** PHE76 **VAL73**
Amphetamine	**GLU117** PHE76 **VAL328** **LEU329**	**SER45** **LEU47** **GLY323** LEU322	**VAL73** VAL324 **ALA77**	Meth-amphetamine	PHE472 THR473 ASP476 PHE155	ILE159 ARG85 PHE320 LEU80	
5-(2-amino ethyl) Benzene -1.3-diol (benzenediol)	GLY75 ILE74 SER72 **VAL73** **ALA77** PHE114	ALA76 LEU322 **GLY323** VAL324 THR48 **GLU117**	**SER45** **LEU47** **VAL328** **LEU329** PHE332	Tyramine	PHE332 SER72 GLY75 ILE74 **VAL73**	PHE76 **ALA77** **LEU47** **GLU117** **LEU329**	**SER45** PHE114 **GLY323** LEU322 **VAL328**

As seen in the above table, SER45, LEU47, VAL328, LEU329, GLY323, VAL73, GLU117 and ALA77 are the common residues in interaction with all the ligands except for 5-hyroxydopamine, methamphetamine and cocaine. Considering the commonality of binding residues, 6-hyroxydopamine, amphetamine, benzenediol, dopamine and tyramine were expected to bind to the same volumetric cavity in the hDAT_MS. Corresponding, visual analysis were shown in the next figure.

Figure 5.3: Binding site regions pertaining to the hDAT_MS-ligand complexes. Protein is shown in line, and ligands are shown in CPK representation. Orange color cocaine, white color methamphetamine, purple color 5-hyroxydopamine, yellow color benzenediol, red color tyramine, green color amphetamine, blue color 6-hyroxydopamine and pink color dopamine.

As seen in the Figure 5.3, 5-hyroxydopamine, cocaine and methamphetamine molecules are found to bind with the DAT in different regions from other molecules.

5.3.2.2 Binding sites in complexes of hDAT_robetta

Table 5.5: The residues in the binding sites pertaining to the hDAT_robetta-ligand complexes. Common residues found in these regions are shown in dark.

LIGAND	BINDING SITES	LIGAND	BINDING SITES
5-hydroxy Dopamine	VAL152 **ASP79 GLY426 PHE326 SER321 SER422 GLY323** ALA81 ALA423 **LEU322** ALA77 **TYR156 PHE320**	Cocaine	ALA77 **PHE326 GLY426** PHE76 **PHE320** SER149 **SER321** ALA423 GLY153 **LEU322** LEU329 VAL152 **ASP79** VAL328 **GLY323 TYR156**
6-hydroxy Dopamine	ALA423 ALA77 VAL328 **SER422** ALA81 **PHE326** PHE76 **TYR156 PHE320 ASP79 GLY426 GLY323 LEU322 SER321**	Dopamine	**PHE326** ALA77 **TYR156 GLY323 ASP79** ALA81 **GLY426** VAL152 **SER321 PHE320 LEU322** PHE76 **SER422**
Amphetamine	**SER321 PHE326** PHE76 **LEU322** ALA81 **SER422 PHE320** ALA77 **TYR156 GLY323 ASP79** VAL152	Meth-amphetamine	**ASP79** PHE76 VAL152 ALA81 **TYR156 PHE326 ALA77 GLY426 GLY323 SER422** GLY153 **PHE320 SER321**
5-(2-amino ethyl) Benzene -1.3-diol (benzenediol)	ASP421 VAL328 **ASP79 SER422** SER149 ALA81 ALA423 **PHE326 SER321** GLY425 **GLY323 PHE320 GLY426 ALA77 LEU322**	Tyramine	**LEU322 ASP79 SER422 SER321** TYR79 **GLY426 PHE320 TYR156** ALA423 ALA81 GLY425 MET427

As seen in the above table, ASP79, PHE326, LEU322, GLY323, SER321, SER422, TYR156, PHE320 and GLY426 are the common residues that are commonly in interaction with all other ligands. Especially, ASP79 residue is in interaction with all the ligands. ASP79, which is a negatively charged aminoacid present in the side chain of the protein, is an important residue in electrostatic interaction with the dopamine. As a conclusion, the ASP79 residue is present in binding site region of all the ligands for this model.

Figure 5.4: Binding site regions pertaining to the hDAT_robetta-ligand complexes. Protein is shown in line, ligands are shown in CPK. Orange color cocaine, white color methamphetamine, black color 5-hyroxydopamine, yellow color benzenediol, red color tyramine, green color amphetamine, blue color 6-hyroxydopamine and pink color dopamine.

As seen in the Figure 5.4; all ligands are computed to bind to the same place in the DAT.

5.3.2.3 Binding sites in complexes of rDAT_robetta

Table 5.6: The residues in the binding sites pertaining to the rDAT_robetta-ligand complexes. Common residues found in these regions are shown in dark.

LIGAND	BINDING SITES			LIGAND	BINDING SITES		
5-hydroxy Dopamine	**GLY322** SER320 ASP79 PHE319 ALA81	**PHE76** LEU321 **ALA77** **TYR156**	VAL327 GLY424 VAL152 **SER421**	Cocaine	ALA479 ASP476 LEU475 LEU89	ARG85 GLY480 THR481 LEU484	ILE311 THR315 SER538 THR472
6-hydroxy Dopamine	VAL152 **SER421** LEU321 PHE325 **PHE76**	**GLY322** **ASP79** ALA81 **PHE319**	**ALA77** **SER320** **TYR156** VAL327	Dopamine	GLY425 GLY424 VAL327 **SER421** **PHE76**	LEU321 PHE325 **ASP79** **PHE319** **ALA77**	ALA81 **SER320** **TYR156** **GLY322** VAL152
Amphetamine	**PHE76** PHE325 **SER421** VAL327 **TYR156**	**PHE319** **SER320** **ASP79** GLY425	**ALA77** LEU321 **GLY322** ALA81	Meth-amphetamine	GLY425 VAL327 GLY153 VAL152 SER149	PHE325 **TYR156** **PHE76** **ASP79** **SER320**	**ALA77** **GLY322** **PHE319** **SER421**
5-(2-amino ethyl) Benzene -1.3-diol (benzenediol)	**PHE319** **ALA77** **ASP79** **TYR156** **PHE76**	**SER320** **GLY322** PHE325 LEU321 VAL152	**SER421** GLY425 GLY424 VAL327	Tyramine	ALA422 GLY424 GLY425 **SER421** **GLY322** ALA81	**PHE76** PHE325 ALA77 **TYR156** **PHE319**	**ALA77** **ASP79** VAL152 LEU321 **SER320**

According to the above table, SER421, GLY322, SER320, ASP79, PHE319, PHE76, ALA77 and TYR156 are the common residues that are present in every protein-ligand binding site region except for cocaine. These binding sites results are depicted visually in Figure 5.5.
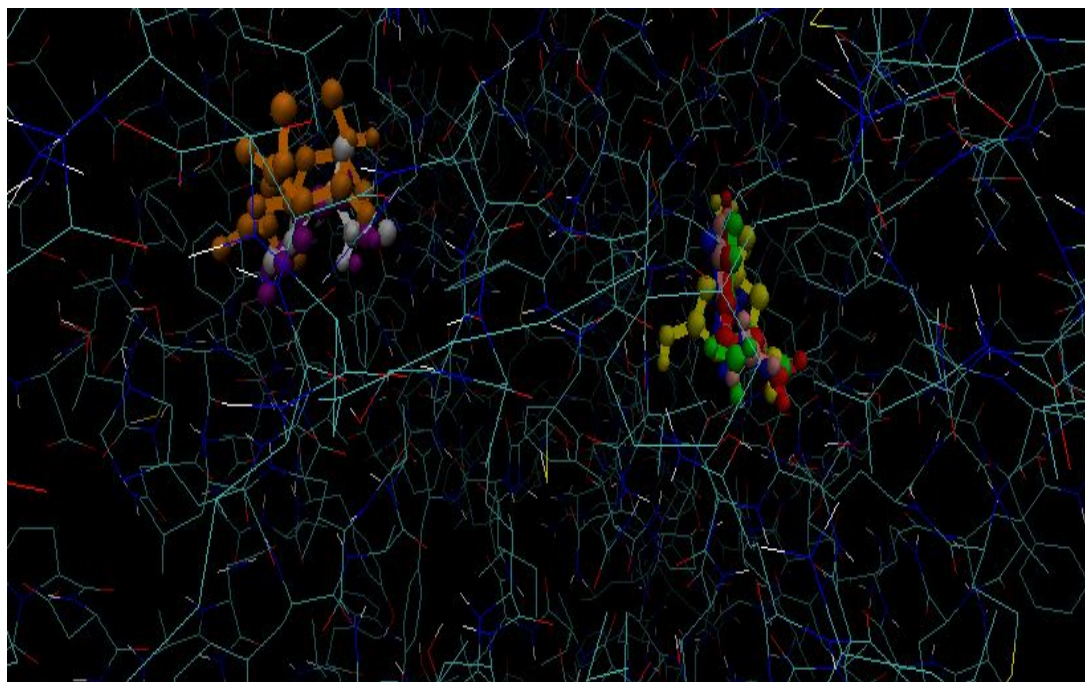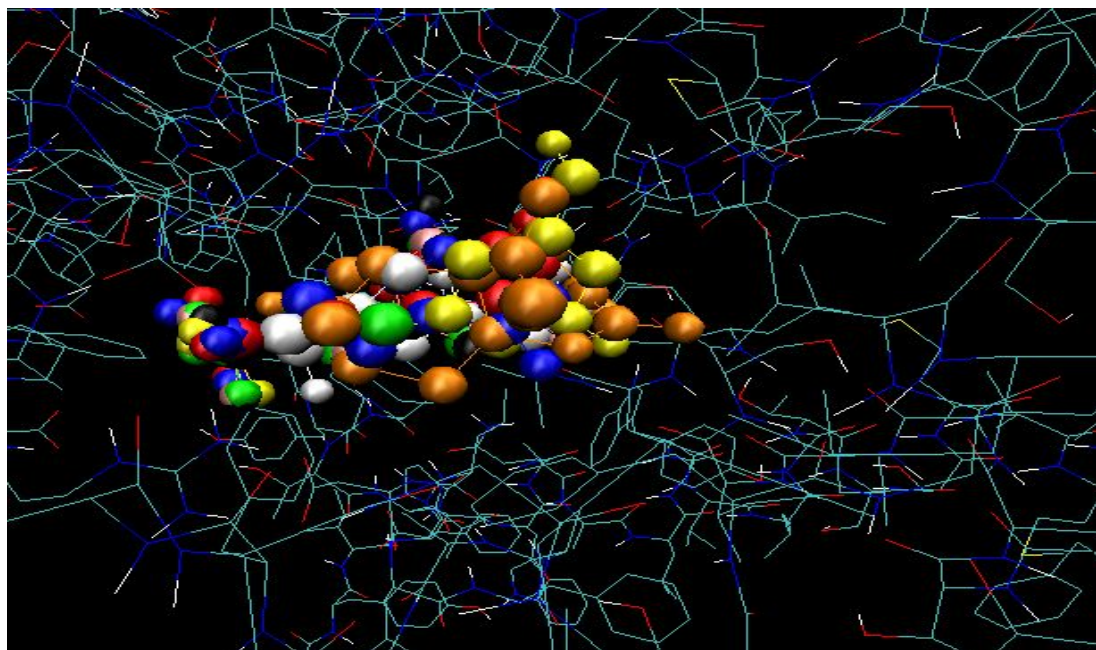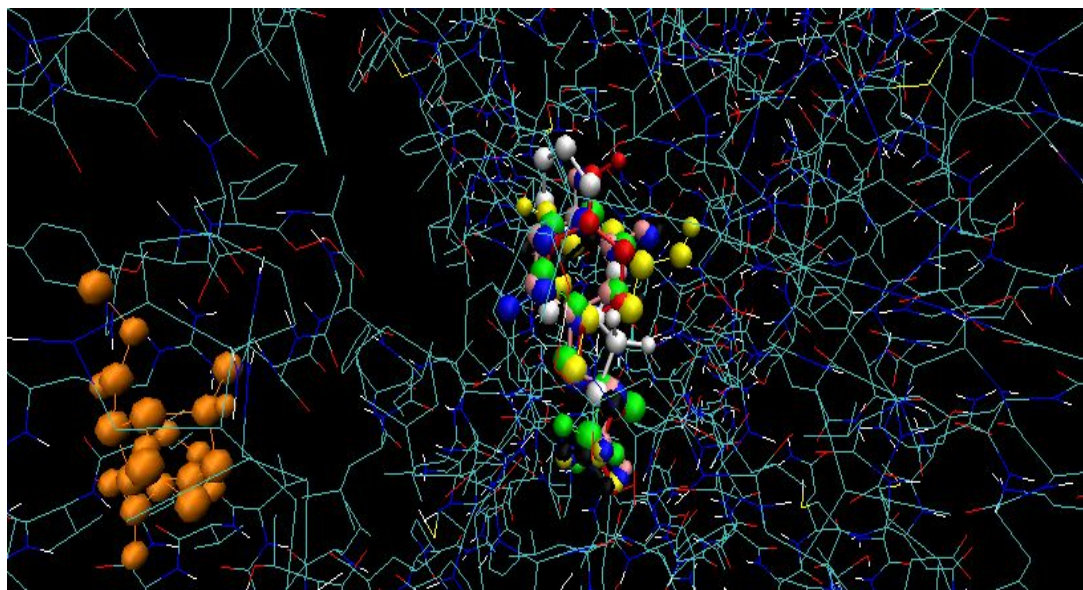
Figure 5.5: Binding site regions pertaining to the rDAT_robetta-ligand complexes Protein is shown in line, ligands are shown in CPK. Orange color cocaine, white color methamphetamine, black color 5-hyroxydopamine, yellow color benzenediol, red color tyramine, green color amphetamine, blue color 6-hyroxydopamine and pink color dopamine.

In the Figure 5.5, all ligand molecules except for cocaine can apparently be seen to bind with the DAT in the same region.

According to the docking analysis performed with three different DAT (two of them belonging to human and one for rat) and eight different ligands in this section, binding site regions pertaining to the hDAT_robetta-ligand and rDAT_robetta-ligand complexes have been found to be in harmony with each other, and it has been detected that only the binding site region pertaining to the rDAT_robetta-cocaine complex is different from the other ligands. Docking analysis pertaining to hDAT_MS-ligand complexes has been found incompatible with these studies.

ASP79, GLY323, SER422, SER321, TYR156, PHE320 residues are the residues that are commonly present in the binding site regions of both hDAT_robetta-ligand and rDAT_robetta-ligand complexes.

As a result of our overall binding site studies, the residues present in the binding site regions of the hDAT_robetta-ligand and rDAT_robetta-ligand complexes were detected to be similar.

(A)   (B)
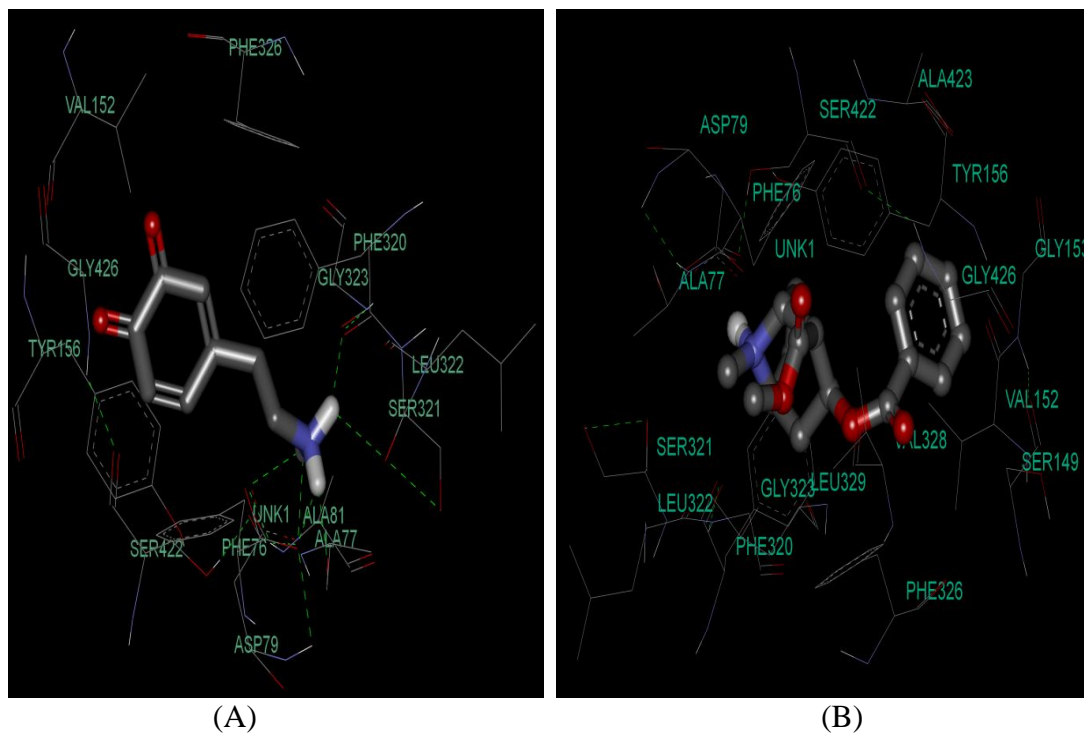
Figure 5.6: Molecular interactions between hDAT_robetta and dopamine (A) hDAT_robetta and cocaine (B) in their binding sites. Dopamine and Cocaine are shown in ball-and-stick and protein in lines.

As seen in the above figure, dopamine and cocaine molecules bind with the hDAT_robetta protein through similar regions. The molecular orientation of ligand molecules are also observed to be similar. For example, the aromatic group is found to exist close to the residues GLY426 and TYR156. Furthermore, the amine group of the ligand molecules are found to be in interaction with ALA77 residue.
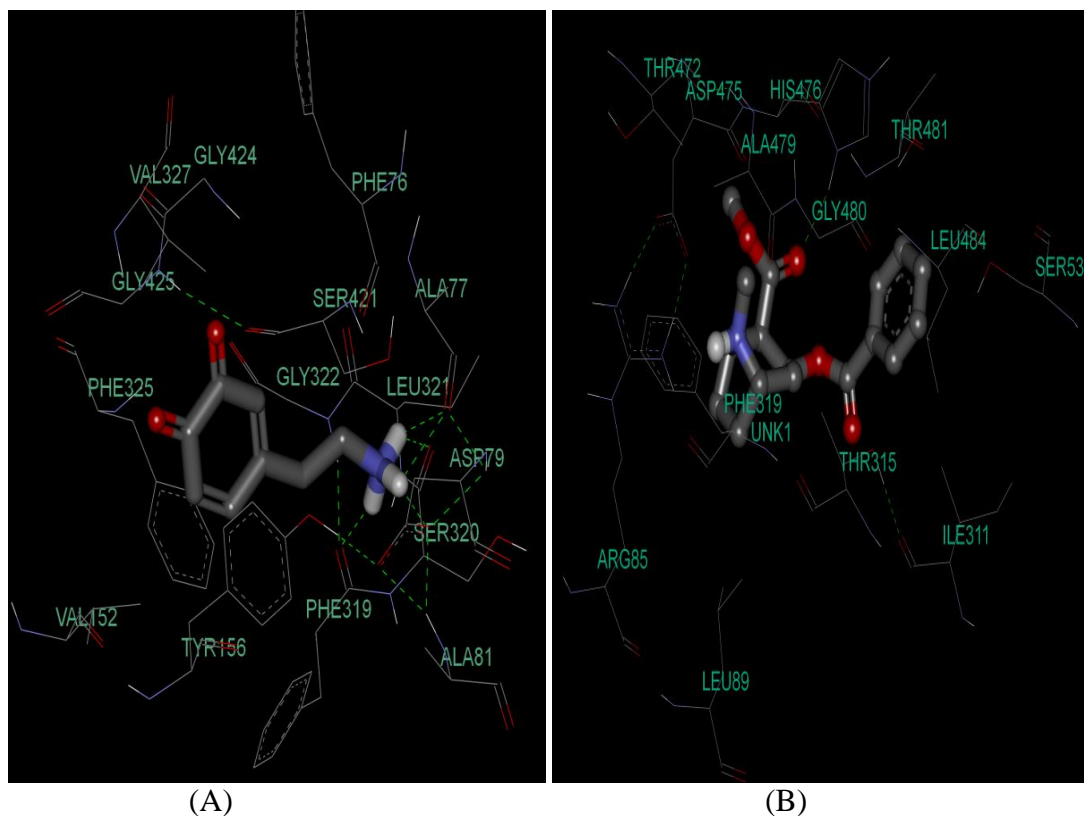
Figure 5.7: Molecular interactions between rDAT_robetta and dopamine (A) rDAT_robetta and cocaine (B). Dopamine and Cocaine are shown in ball-and-stick and protein in line

In Figure 5.7, the binding sites for rDAT_robetta are shown together with molecular interaction details.

On the contrary to the results of binding site analysis of hDAT_robetta, Figure5.7 shows that the cocaine binds with the rDAT_robetta protein in a region different from the binding region of the dopamine molecule.
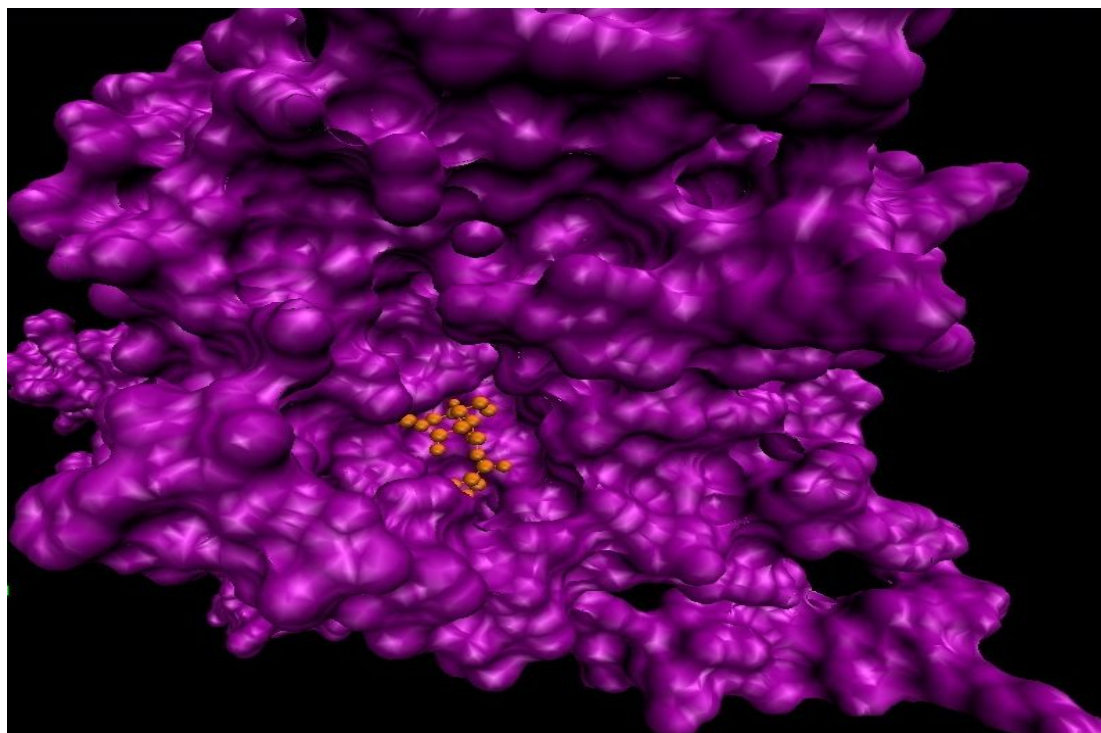
Figure 5.8: Binding region of cocaine on the surface of rDAT_robetta. The rat DAT protein, rDAT_robetta, indicated as opaque surface and cocaine indicated with CPK representation.

As seen in the above figure, cocaine had interaction with the external surface of the rDAT_robetta protein and the binding energy score has been calculated as -4.26 kcal/mol for this position as seen in the section 5.3.1. However, cocaine and hDAT_robetta protein bind in an internal cavity and the binding energy scores for this pose has been calculated as -9.28 kcal/mol. This internal binding cavity is shown in Figure 5.9.
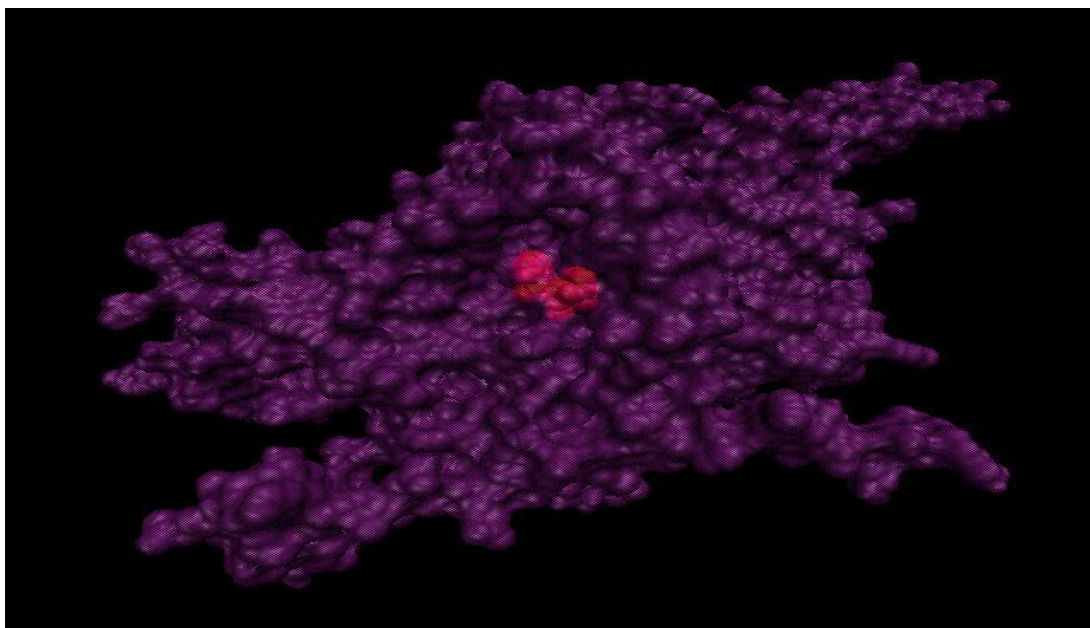
Figure 5.9: Binding region pertaining to hDAT_robetta-cocaine. hDAT_robetta indicated as transparent and cocaine indicated as VDW.

As seen in the figure of hDAT_robetta complex, cocaine binds in an internal cavity, which overlaps with the binding site of dopamine. Since, cocaine bounds in a way to the residues ASP79-TYR156, and thereby formed a closed binding pocket, when both ligands exist at the same time dopamine can not bind with DAT. As a result of this study, cocaine can be identified as a competitive inhibitor of hDAT_robetta for the transport of dopamine.

However, as seen in the study regarding the rDAT_robetta, whose similarity sequence with the hDAT_robetta is 93.2%, cocaine didn't bind in the DAT internal cavity. This might have resulted from the 6.8% difference in sequence between the two structural models. Because, even a small difference occurring in the structure also affects the binding affinity between cocaine and the protein.

Furthermore, a set of experimental studies have been conducted for analysing the cocaine's effect on mice **[42].** According to these experiments, the differences between the responses of human and mice to cocaine were determined. The relation between the two molecular structures and these differences in responses should be examined, to see whether there exist any correlation or not.

5.3.3 Atomic interactions between DAT and ligands

5.3.3.1 Interactions of hDAT_MS with ligands

Table 5.7: The hydrogen-bonding (HB) interactions between DAT and eight ligands in the hDAT_MS-ligand complex.

Ligand Name	Residue	Donor	Acceptor	Hb dist(A^o)
5-hydroxy dopamine	ASP476	N-HA N-HC	OD1 OD2	2.2 2.2
6-hydroxy Dopamine	ASP79	N-HA N-HC	OD2 OD2	2.7 2.4
Benz-1,3-diol	THR48 GLY323 VAL73 SER72 SER45 GLU117	O-HA O-HA N-HE N-HC O-HB O-HB	OG1 O=C O=C O=C OG OE2	2.2 2.0 2.0 2.3 2.2 2.0
Dopamine	SER44 SER45 LEU322	N-HB N-HC N-HB N-HA	O=C O=C OG O=C	2.2 1.8 2.0 1.8
Meth-Amphetamine	ASP476	N-HA N-HB	OD1 OD1	2.2 1.8
Tyramine	GLU117 SER45 SER44 VAL73	N-HC N-HD N-HC N-HD N-HC O-HA	OE2 OE2 OG OG O=C O=C	2.1 2.4 2.4 2.3 2.1 1.7

As seen in the above table, Table 5.7, a strong hydrogen bonding interaction has been observed between the hydroxyl group of DAT's SER45 residue and cationic head of all the ligands except for three; 5-hydroxydopamine, 6-hydroxydopamine, and methamphetamine. And looking at the table, we see that the hydrogen bond distances between the ligand and protein functional groups are varying from 1.7 A^o to 2.4 A^o.

To illustrate the 2D and 3D representations of interactions are depicted for Tyramine in Figure 5.10 and Figure 5.11.
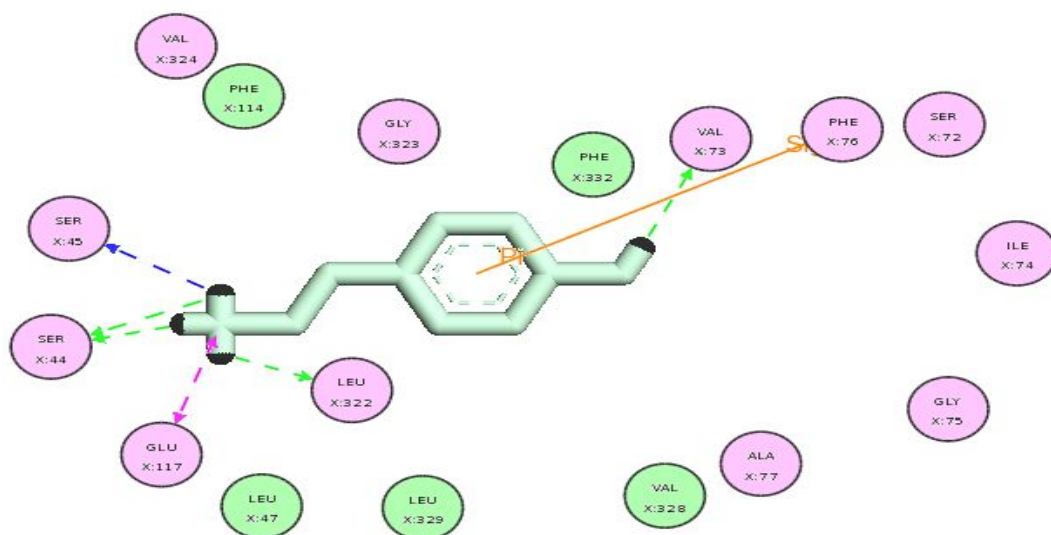
Figure 5.10: 2D representation of the atomic interactions between the hDAT_MS-Tyramine. Tyramine is shown with ball-and-stick model.
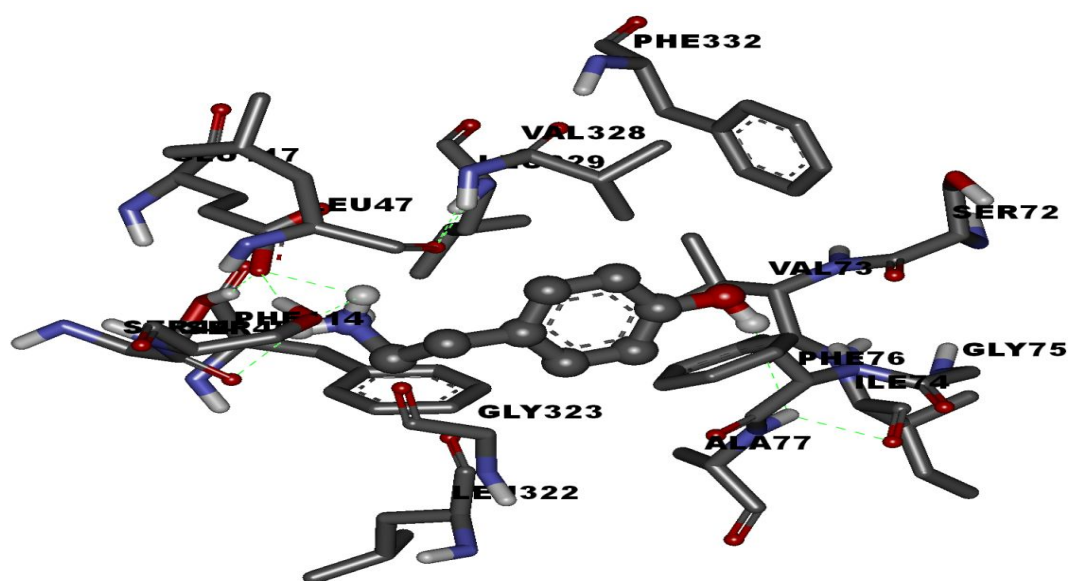
Figure 5.11: 3D representation of the atomic interactions between the hDAT_MS-Tyramine. Tyramine is shown with ball-and-stick model.

From the figures, it is seen that the nearest hydrogen atom on the aromatic ring of Tyramine interacts with the aromatic side chain of PHE76, indicating a cation-π interaction. And it is also observed in the figures that GLU117, SER45, SER44 and VAL73 hydrogen-bonded to the Tyramine. Similarly, other ligands can be visualized at their binding sites to show the interacting groups.

5.3.3.2 Interaction of hDAT_robetta with ligands

Table 5.8: The hydrogen-bonding (HB) interactions between DAT and eight ligands in the hDAT_robetta-ligand complex.

Ligand Name	Residue	Donor	Acceptor	HB dist (A^o)
5-hydroxy dopamine	SER321	N-HB	O=C	2.4
		N-HA	O=C	2.5
	ASP79	N-HC	OD1	2.1
		N-HC	OD2	2.7
	ALA77	N-HA	O=C	1.9
	PHE320	N-HA	O=C	1.8
6-hydroxy Dopamine	ALA77	N-HB	O=C	2.7
		N-HC	O=C	2.7
	ASP79	N-HC	OD1	2.1
	SER321	N-HB	O=C	1.9
	PHE320	N-HA	O=C	2.0
Amphetamine	ASP79	N-HC	OD2	2.4
		N-HB	OD1	2.7
		N-HC	OD1	2.1
	ALA77	N-HB	O=C	1.9
	PHE320	N-HA	O=C	1.9
	SER321	N-HA	O=C	2.4
Benz-1,3-diol	ASP79	N-HE	O=C	2.0
		N-HE	O=C	2.7
	ALA77	N-HD	O=C	2.0
	PHE320	N-HC	O=C	1.7
	SER321	N-HC	O=C	2.2
	SER422	O=HB	O=C	1.9
Dopamine	ASP79	N-HA	OD2	2.4
		N-HC	OD1	2.7
		N-HA	OD1	2.1
	SER321	N-HB	O=C	2.4
	ALA77	N-HC	O=C	1.9
	PHE320	N-HB	O=C	1.8
Methamphetamine	ASP79	N-HB	OD2	2.8
	PHE320	N-HA	O=C	2.1
Tyramine	ALA77	N-HC	O=C	2.2
		N-HD	O=C	2.1
	ASP79	N-HD	OD1	1.9
	PHE320	N-HB	O=C	2.1
	SER422	O-HA	O=C	2.0
	SER321	N-HC	O=C	1.8

As seen in the above table; hydrogen bonding interactions occurred between the cationic heads of the ligands and the carbonyl and side chain oxygens (OD1 and OD2) of the ASP79, SER321, ALA77 and PHE320 residues pertaining to the DAT protein. Also, hydrogen bonding interactions between the Benz-1,3-diol, Tyramine ligands and SER422, SER321 residues were observed.

As an example, the interactions for the ligand Benz-1,3-diol is depicted in Figure 5.12 and Figure 5.13 in 2D and 3D respectively.
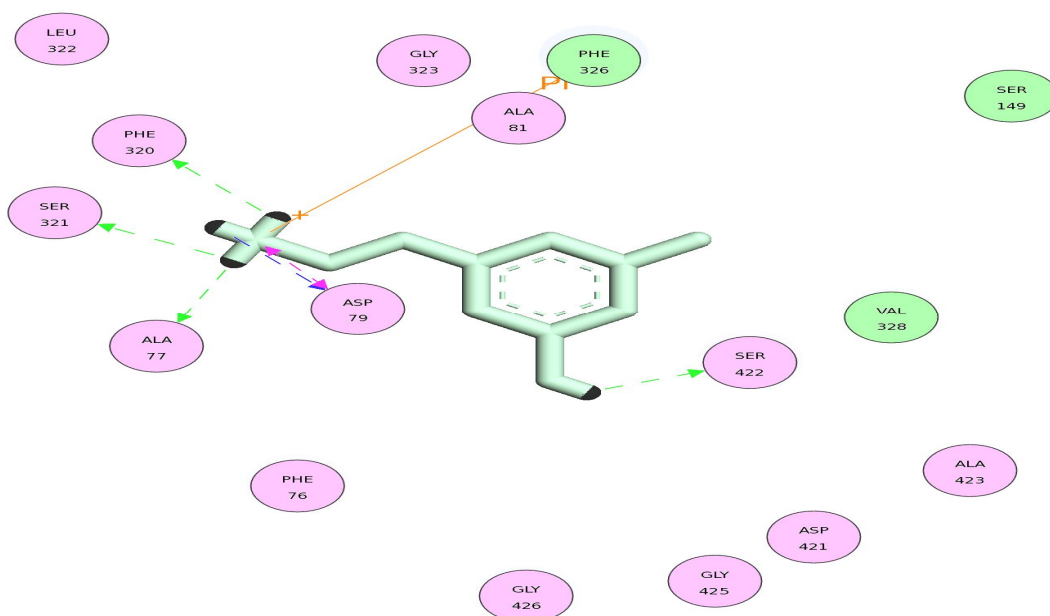
Figure 5.12: 2D representation of the atomic interactions between the hDAT_robetta-Benz-1,3-diol. Benz-1,3-diol is shown with ball-and-stick model.
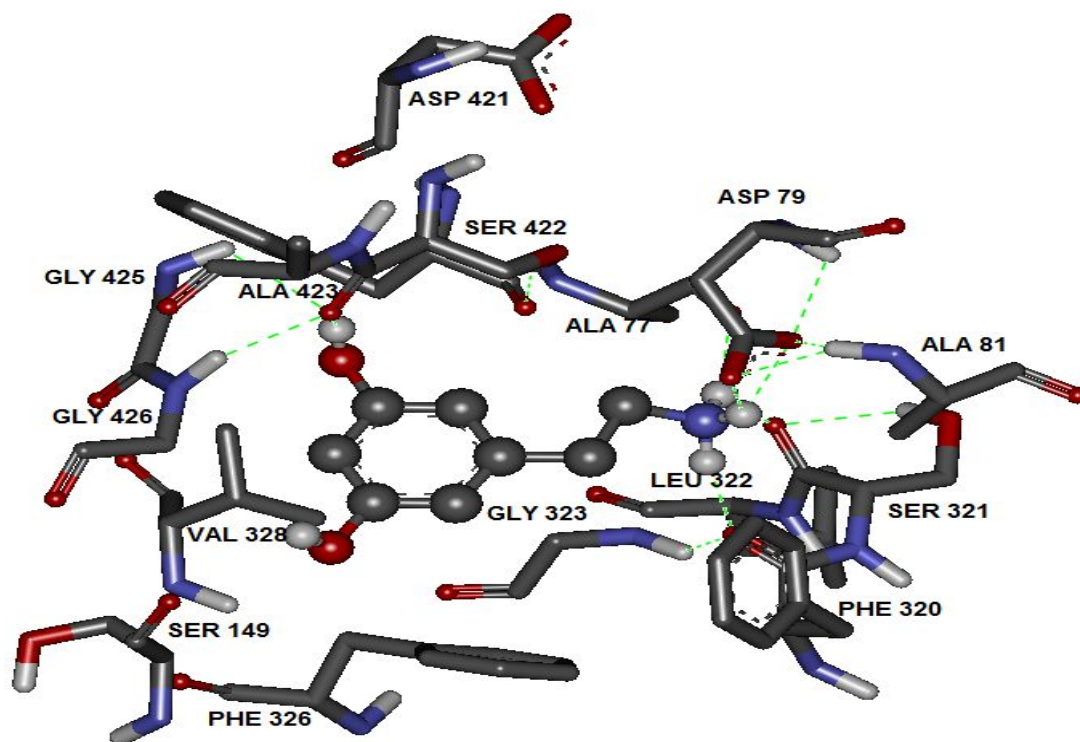
Figure 5.13: 3D representation of the atomic interactions between the hDAT_robetta-Benz-1,3-diol. Benz-1,3-diol is shown with ball-and-stick model.

The nearest hydrogen atom on the cationic head of Benz-1,3-diol interacts with the aromatic side chain of PHE326, indicating a cation-π interaction. Side chains of ALA77, ALA81, and PHE320 interact with the alkyl chain of Benz-1,3-diol from its cationic head. GLY426, GLY425 and side chains of VAL328 make contacts additionally with the aromatic ring of Benz-1,3-diol.

5.3.3.3 Interactıon of rDAT_robetta interaction with ligands

Table 5.9: The hydrogen-bonding (HB) interactions between DAT and eight ligands in the rDAT_robetta-ligand complex.

Ligand Name	Residue	Donor	Acceptor	Hb dist(A°)
5hydroxydopamine	ASP79	N-HA	OD1	2.5
		N-HC	OD1	1.9
	PHE319	N-HA	O=C	1.7
	SER320	N-HB	O=C	2.0
	ALA77	N-HA	O=C	2.1
6ydroxydopamine	ALA77	N-HB	O=C	2.2
		N-HC	O=C	2.1
	ASP79	N-HC	OD1	1.7
	SER320	N-HB	O=C	1.8
	PHE319	N-HA	O=C	1.9
Amphetamine	ASP79	N-HA	OD1	2.1
		N-HB	OD1	2.2
	SER320	N-HC	O=C	2.4
	PHE319	N-HC	O=C	2.0
	ALA77	N-HA	O=C	1.9
Benz-1,3-diol	ALA77	N-HD	O=C	2.2
		N-HE	O=C	2.2
	ASP79	N-HE	OD1	1.7
	PHE319	N-HC	O=C	1.9
	SER320	N-HD	O=C	1.7
	SER421	O-HB	O=C	1.9
Cocaine	HIS476	O=C	N-HD	2.2
Dopamine	ALA77	N-HB	O=C	2.2
		N-HC	O=C	2.0
	ASP79	N-HC	OD1	1.7
	SER320	N-HB	O=C	2.0
	PHE319	N-HA	O=C	1.9
Methamphetamine	PHE76	N-HB	O=C	1.9
	TYR156	N-HB	O-H	1.9
Tyramine	ALA77	N-HC	O=C	2.2
		N-HD	O=C	2.2
	ASP79	N-HD	OD1	1.7
	SER320	N-HC	O=C	1.6
	PHE319	N-HB	O=C	1.9
	SER421	O-HA	O=C	2.4

As seen in the above table; hydrogen bonding interactions occurred between the cationic heads of all the ligands, except for cocaine, and the carbonyl and side chain oxygens of the ASP79, SER320, ALA77 and PHE319 residues pertaining to the DAT protein.

And finally, for this rDAT_robetta structure, 2D and 3D representation of amphetamine interactions were given as an example in Figure 5.14 and 5.15.
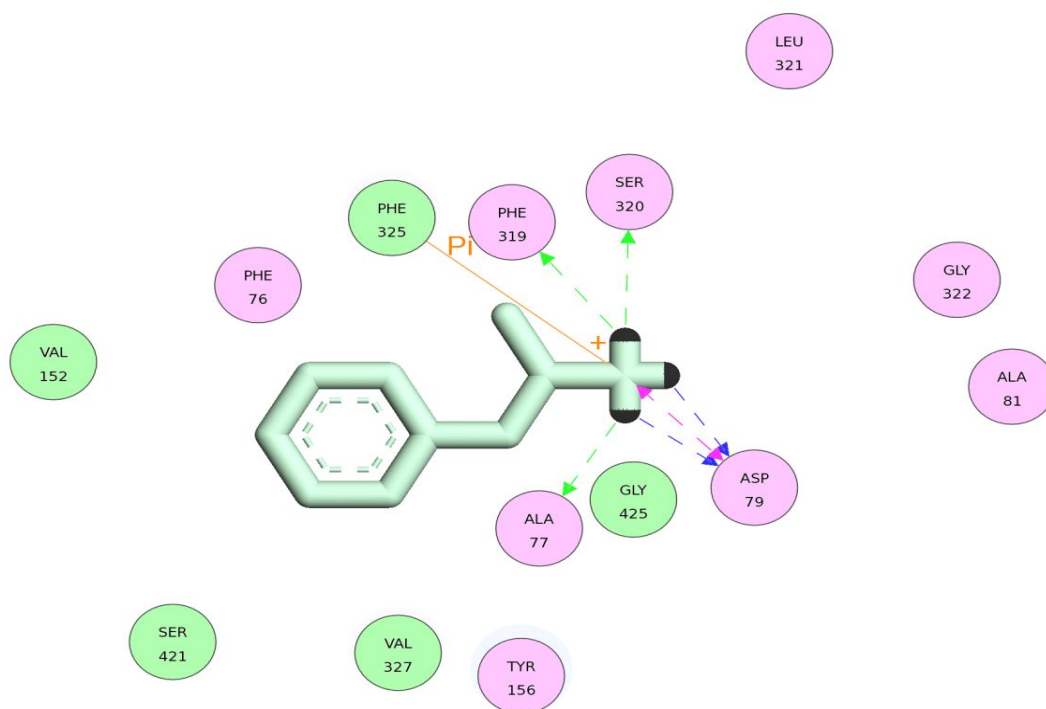
Figure 5.14: 2D representation of the atomic interactions between the rDAT_robetta-Amphetamine. Amphetamine is shown with ball-and-stick model.
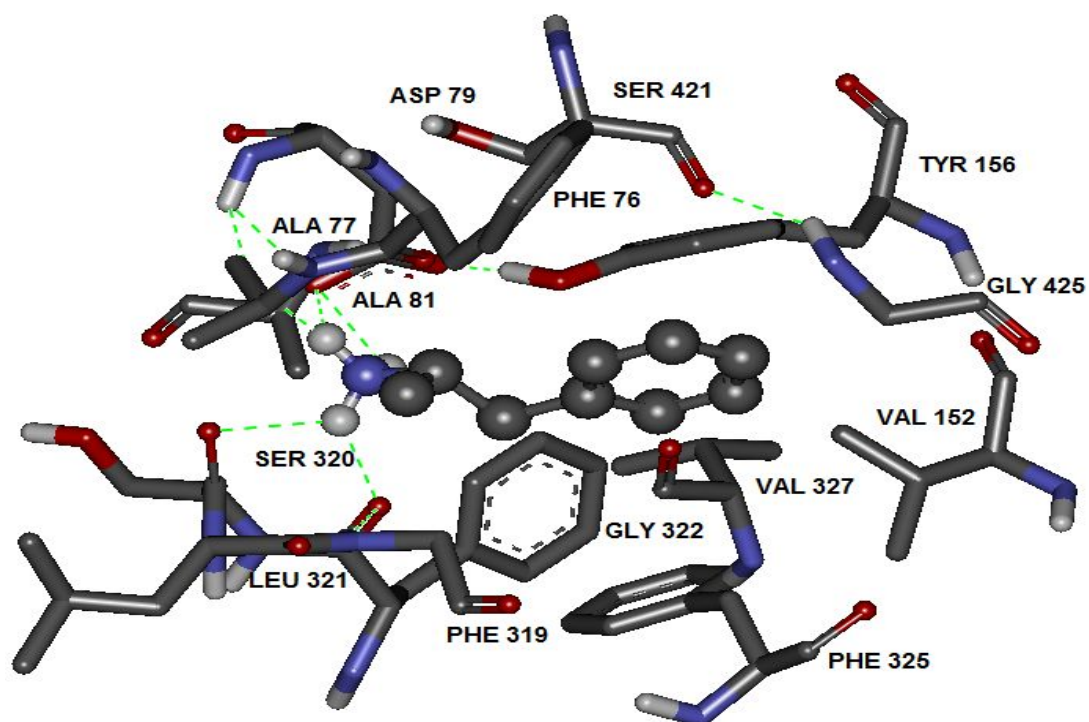
Figure 5.15: 3D representation of the atomic interactions between the rDAT_robetta-Amphetamine. Amphetamine is shown with ball-and-stick model.

The nearest hydrogen atom on the cationic head of Amphetamine interacts with the aromatic side chain of PHE325, indicating a cation-π interaction. Side chains of ALA77, ALA81, and PHE319 interact with the alkyl chain of Amphetamine between its cationic head. And also, GLY425 and side chains of VAL327 make additional contacts with the aromatic ring of Amphetamine. And interactions between Amphetamine and the side chains of ASP79, TYR156 and SER421 should be the major binding forces anchoring Amphetamine to its binding site.

5.4 Conclusion

The differences in binding site prediction between hDAT_MS model and the hDAT_robetta model are most likely due to the differences in packing in these protein structure. Considering the hDAT_MS-ligand complex data, binding site for ligands SER45, LEU47, VAL328, LEU329, GLY323, VAL73, GLU117, and ALA77 are the common residues in interaction with all the ligands except for 5-hydroxydopamine, methamphetamine and cocaine. And according to the hDAT_robetta-ligand complex data, binding site for ligands are buried deep in ASP79, PHE326, LEU322, GLY323, SER321, SER422, TYR156, PHE320, and GLY426.

However, in the experimental studies conducted previously by a group on the hDAT-cocaine and hDAT-dopamine, the residue ASP79, VAL152 and TYR156 residues had been identified as important residues charged in the interaction between cocaine, dopamine and DAT. According to the computational docking study of the same group, PHE76, ALA77, ASP79, SER149, VAL 152, TYR156, PHE320, SER321, LEU 322, GLY323, PHE326, VAL328, SER422, ALA423 and GLY426 are the residues that are present in the binding site regions pertaining to the DAT-cocaine and DAT-dopamine complexes [37].

In addition, in an another experimental study it is reported that; TYR156 is a strictly conserved residue between all members in the NSS family and SER422 is also conservative. In our studies here we also saw these residues takes place in the binding sites as well. And, the mutual interactions between the side chain of TYR156 and ASP79 are pronounced to possibly act as a latch to stabilize the irregular structure around the unwound region of helix 1. The involvement of ASP79 in the vital local interactions with both dopamine and Na^+ can explain why mutations ASP79ALA, ASP79GLY, and ASP79GLU dramatically reduced dopamine reuptake as demonstrated in previously reported experimental studies on DAT [1].

Consequently, through the comparison of our data with the previous studies, for hDAT_robetta-ligand complex, binding site for ligands overlaps extensively with that of dopamine reported in older experimental studies and computational studies.

Chapter 6

Conclusion

Comparative modelling methodologies allow us to develop a reasonable three dimensional structural model of DAT. And further applied molecular docking methods enable us to investigate its binding interactions with the ligands of our interest. The ligand investigated in our study are 5-hydroxydopamine, 6-hydroxydopamine, Benz-1,3-diol, Dopamine, Amphetamine, Methamphetamine, Cocaine, and Tyramine.

As a result of the comparative (homology) modelling of hDAT using different templates two DAT models, hDAT_MS and hDAT_robetta were obtained. The RMSD differences between these two models 16.585 A^0 for Cα atoms. Even though, alpha helices cover the similar regions in both structure models, a small difference in the orientation of these helices brings about this RMSD difference, indicating variations in 3D structure.

The differences in binding site prediction between the hDAT_MS model and the hDAT_robetta model are most likely due to the differences in packing in these protein structures.

The hDAT_robetta model created by use of rDAT templates was through to have the correct packing, since binding cavity for all the ligands found to be in the interior part of DAT. Our hDAT_robetta-ligand complex data suggest a common binding site for all the ligands buried deep in the protein between the residues ASP79, PHE326, LEU322, GLY323, SER321, SER422, TYR156, PHE320, and GLY426. This binding site overlaps extensively with that of dopamine reported in experimental studies. In addition the calculated binding energies of the ligands to hDAT_robetta model correlate well with experimental results in hand.

However, using multiple data from protein databank as templates to create hDAT_MS model doesn't seem to give so accurate results. When we compare the experimental and our computational results with this model, we see that they are not consistent as much as for the hDAT_robetta model.

The structural differences between these two models are resulted from differences between the templates used. For the hDAT_robetta a template of rat DAT model with 93.2% sequence similarity is used, whereas the sequence similarity of the templates used for the creation of hDAT_MS model structure was less than 30%. This means that more structural information is used for hDAT_robetta coming from its template.

By the way, the overall results also show the accuracy of the rat DAT template as well, since the accuracy of our hDAT_robetta model is dependent also to the template used. The rat DAT template was coming from computational study with de novo method of Robetta server together with additional Monte Carlo (MC) simulations having been performed on that.

As a conclusion when we consider our results for hDAT_MS and hDAT_robetta model, we end up with idea that further Molecular Dynamics simulations are required to obtain a better equilibrated structure for hDAT_MS models. The MC simulations having been performed for hDAT_robetta seem to compensate for this imperfection of our studies.

REFERENCES

[1] Xiaoqin Huang and Chang-Guo Zhan, How Dopamine Transporter Interact with Dopamine: Insights from Molecular Modeling and Simulation, *Biophysical Society*, **2007**; 0006-3495/07/11/3627/13

[2] Nianhang Chen, Maarten E.A. Reith, Structure and function of the dopamine transporter, *European Journal of Pharmacology* **2000**; 329-339

[3] Amara SG, Kuhar MJ, Neurotransmitter transporters: recent process. *AnnuRev Neurosci*, **1993**; 16: 73-93

[4] Krueger BK, Kinetics and block of dopamine uptake in synaptosomes from rat caudate nucleus. *J Neurochem,* **1990**; 55: 260-267

[5] Hitri A, Hurd YL, Wyatt RJ, Deutsch SI, Molecular, functional and biochemical characteristics of the dopamine transporter: regional differences and clinical relevance, *Clin Neuropharmacol,* **1994**; 17: 1-22

[6] Nianhang Chen, Maarten E.A. Reith, Na^+ and the substrate permeation pathway in dopamine transporters, *European Journal of Pharmacology,* **2003;** 213-221

[7] http://en.wikipedia.org/wiki/File:DAT1regulation.jpg

[8] Xiaoqin Huang, Howard H. Gu, and Chang-Guo Zhan, Mechanism for cocaine blocking the transport of dopamine: Insights from molecular modeling and dynamics simulation, *J. Phys. Chem. B* **2009**; 113, 15057-15066

[9] Giros, B., Mestikawy, S.E., Bertrand, L., Caron, M.G, Cloning and functional characterization of a cocaine-sensitive dopamine transporter, *FEBS* **1991**; 295, 149-154

[10] Yamashita, A., Singh, S.K., Kawate, T., Jin, Y., Gouaux, E. Crystal structure of a bacterial homologue of Na$^+$/Cl$^-$ dependent neurotransmitter transporters. *Nature,* **2005**; 437, 215-223.

[11] Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A., Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, **2000**; 29: 291-325

[12] Marketa Zvelebil, Jeremy O. Baum, Understanding Bioinformatics, *Garland Science, Taylor&Francis Group,* LLC,**2008**

[13] Srinivas Aluru ; Handbook of Computational Molecular Biology, *Chapman&Hall/CRC,Taylor&Francis Group*, **2006**

[14] Jonathan Pevsner; Bioinformatics and Functional Genomics Second Edition, *Wiley Blackwell, A John Wiley&Sons, Inc, Publication*, **2009**

[15] Steven Henikoff and Jorja G. Henikoff ; Amino acid substitution matrices from protein blocks; *Proc. Natl. Acad. Sci. USA,***1992;** 10915-10919

[16] Albert Bray Hopkin Johnson Lewis Raff Roberts Walter; Essential Cell Biology Third Edition; *Garland Science, Taylor&Francis Group;* **2009**

[17] A.Storch, A.C.Wdolph and J.Schwarz; Dopamine transporter: Involvement in selective dopaminergic neurotoxicity and degeneration, *Journal of Neural Transmission,* **2004;** 111: 1267-1286

[18] Philip E. Bourne and Helge Weissig; Structural Bioinformatics, *Wiley Liss, A John Wiley&Sons Publication,* **2003**

[19] Discovery Studio Tutoriol

[20] Sergio Filipe Sousa, Pedro Alexandrino Fernandes, and Maria Joao Ramos, Protein-Ligand Docking: Current Status and Future Challenges, *Proteins*: *Structure, Function, and Bioinformatics*, **2006**; 65:15-26

[21] David E. Clark, Evolutionary Algorithms in Molecular Design,*Volume 8, Wiley Vch*;**2000**

[22] H.-J. Böhm, Prediction of Non-bonded Interaction in Drug Design, *Wiley-Vch,* **2003**

[23] N Moitessier, P Englebienne, D Lee, J Lawandi and CR Corbeil, Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go, *British Journal of Pharmacology,* **2008**; 153, S7-S26

[24] Hernan Alonso, Andrey A. Bliznyuk, Jill E. Gready, Combining Docking and Molecular Dynamic Simulations in Drug Design, *Medicinal Research Reviews,* **2006**; Vol. 26, No. 5, 531^568

[25] Ruth Huey, Garrett M. Morris, Arthur J. Olson, Davıd S. Goodsell, A semiempiricial free energy force field with charge-based desolvation*, J Comput Chem,* **2007**; 28: 1145-1152

[26] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W.E.; Belew, R. K.; Olson, *A. J. J Comp Chem* **1998**; 19, 1639

[27] Garrett M. Morris, David S. Goodsell, Michael E. Pique, William "Lindy" Lindstrom, Ruth Huey, Stefano Forli, William E. Hart, Scott Halliday, Rik Belew and Arthur J.Olson, AutoDock Version 4.2 Automated Docking of Flexible Ligands To Flexible Receptors, **2009**

[28] Stanley Fahh and David Sulzer, Neurodegeneration and Neuroprotection in Parkinson Disease, *The journal of the American Society for Experimental NeuroTherapeutics,* **2004**; 139-154

[29] Abramson J., Smirnova I., Kasho V., Verner G., Kaback H.R, Iwata S., Structure and mechanism of the lactose permease of Escherichia coli, *Science,* **2003;** 301:610-615

[30] Huang Y., Lemieux M.J., Song J., Aver M., Wang D.N., Structure and mechanism of the glycerol-3-phosphate transporters from Escherichia coli, *Science,* **2003;** 301:616-620

[31] Martin Inderte, Jeffry D. Madura, and Christopher K.Surratt, , Dopamine transporter comparative molecular modeling and binding site prediction using the LeuT$_{Aa}$ leucine transporter as a template, *Proteins,* **2008;** 70:1033-104632

[32] Steven Henikoff and Jorja G. Henikoff; Amino acid substitution matrices from protein blocks, *Proc.Natl.Acad.sci. US*, **1992;** 10915-10919

[33] Bruno Giros, Salah El Mestikawy, Nathalie Godninot, Cloning, Pharmacological Characterization, and Chromosome Assignment of the Human Dopamine Transporter, *Molecular Pharmacology,* **1992;** 42:383-390

[34] M. Yashar S. Kalani, Nagarajan Vaidehi, Peter L. Freddolino,Maziyar A. Kalani, and William A. Goddard III* (Caltech Team) and James Hendrix, Felix Sheinerman, Isabelle Morize, and Abdelazize Laoui (Aventis Pharma Team), The predicted 3D structure of the human D2 dopamine receptor and the binding site and binding affinities for agonist and antagonists, *PNAS*, **2004**; Vol 101 No 11, 3815-3820

[35] Shimada, S., Kitayama, S., Lin, C.C., Patel, A., Nanthakumar, E., Gregor, P., Kuhar, M., Uhl, G.R., Cloning and expression of cocaine-sensitive dopamine transporter complementary DNA. *Science,* **1991**; 254, 576-578

[36] Vandenbergh, D.J., Persico, A.M., Uhl, G.R.,. A human dopamine transporter cDNA predicts reduced glycosylation, displays a novel repetitive element and provides racially-dimorphic *TaqI* RFLPs. *Brain Res. Mol. Brain Res.* **1992;** 15, 161-166

[37] Kitayama, S., Shimada, S., Xu, M., Markham, L., Donovan, D.M., Uhl, G.R., Dopamine transporter site-directed mutations differentially alter substrate transport and cocaine binding. *Proc. Natl. Acad. Sci*. **1992;** 89,7782–7785.

[38] Lin, Z., Wang, W., Kopajtic, T., Revay, R.S., Uhl, G.R., Dopamine transporter: transmembrane phenylalanine mutations can selectively influence dopamine uptake and cocaine analog recognition. *Mol. Pharmacol*. **1999**; 56,434–447.

[39] Dalit E Dar, Thomas G. Metzger, David J. Vandenbergh, George R. Uhl, Dopamine uptake and cocaine binding mechanisms: The involvement of charged amino acids from the transmembrane domains of the human dopamine transporter, *European Journal of Pharmacology,* **2006;** 538, 43-47

[40] Thijs Beuming, Julie Kniazeff, Marianne L Bergmann; The Binding sites for cocaine and dopamine in the dopamine transporter overlap, *Nature Neuroscience*, **2008;** Volume 11, Number 7

[41] Dar, D. E., T. G. Metzger, D. J. Vandenbergh, and G. R. Uhl, Dopamine uptake and cocaine binding mechanisms: the involvement of charged amino acids from transmembrane domains of the human dopamine transporter, *Eur. J. Pharmacol.,***2006**; 538:43–47.

[42] Ken T. Wakabayashi, Mark J. Weiss, Kristen N. Pickup, and Terry E. Robinson, Rats Markedly Escalate Their Intake and Show a Persistent Susceptibility to Reinstatement Only when Cocaine is Injected Rapidly, *The journal of Neuroscience,***2010**; 30(34):11346-11355

Curriculum Vitae

Name: GİZEM TATAR

E-mail: gizemttr@hotmail.com

Date of Birth: 11.09.1986

Place of Birth: Gaziantep

Marital Status: Single

Gender: Female

Nationality: Turkish

EDUCATION

- B.S, Molecular Biology and Genetics (2009)

Haliç University, İstanbul, Turkey

RESEARCH AND TEACHİNG EXPERIENCE

- Alman Hospital Genetic Department ; 2009 Intern
- Heinrich Heine University of Dusseldorf Hematology, Oncology and Clinical Immunology; 2007 Intern
- Gaziantep University of Genetic Department; 2007 Intern

TECHINICAL AND LANGUAGE SKILLS

Instrument Skills

- DNA isolation, PCR , Cells selection with MACS columns and Robosep (Stem Cell Technologies), Cell culture work, RNA isolation, Light Cycler quantitative Real Time PCR, Nested rt-PCR, FACS analyses, Western blot analysis

Computer Skills

- Windows XP-Vista, MS-word, Excel, Powerpoint, Linux, MATLAB, Visual Molecular Dynamic (VMD), Autodock, NAMD, Pymol, Vina, Discovery Studio (Accelrys Software

Languages Skills

- Turkish(Native), English(Upper-Intermediate)