# Extracting Parameters from a PDB File

## Anitha Kumari H.C[a] Gloriya Thomas[a] and Pallavi M.S[a]

[a]*Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Amrita University, Mysuru Campus Karnataka, India*

*E-mail: anithakumarihc@gmail.com, gloriyaglo@gmail.com, palls.ms@gmail.com*

*Abstract:* Protein structure resolution through extraction of parameters from PDB file is an important and stimulating task in molecular dynamics. Extracting parameters from PDB file with a computational approach reduces the effort of validation and assembly of protein structure. It plays a key role in the creation of 3D structure which is further helpful in ease analysis of PDB file that is simple view of PDB file, view the atoms as a coordinate, easy bonding of atoms. Extracting parameters from a PDB file is given in the work which uses a programming language to fetch, evaluate and represent the parameters. Extracting parameters and further these parameters are used in building the three dimensional structure which can be then applied with the energy minimization and the best pathway in molecular dynamics can be evaluated.

*Keywords:* PDB (Protein Data Bank), 3D structure, atoms, parameters, bonding, molecular dynamics.

## 1. INTRODUCTION

Computer technology has different fields and branches for the materialization of data. One among them is Bioinformatics, which deals with the understanding of biological aspects (like biological data such as amino acid sequences, protein structures and protein domains) through mathematical and computational analysis and approaches. The recognition and analysis of human diseases and the identification of molecular dynamics for drug invention held with the practice of genomic data can be executed by combining the molecular biology of Bioinformatics and Computer Science. Computational biology is the authentic process of analyzing and inferring data which includes:

1. Implementation and development of several types of information can enable the access and authentication with computer programs.

2. Algorithms or mathematical formulas and statistical analysis that evaluates the large dataset members and its relationships.

Structural Bioinformatics keeps track of three-dimensional structure of protein which helps to predict the interaction of protein-protein and its docking problem. The docking algorithm that is computational algorithm for molecular relations is employed with the principle of simulation of atoms in molecular dynamics.

The Protein Data Bank file format is a textual file format which gives a standard representation for macromolecular structure data derived from X-ray diffraction and NMR studies. X-ray Crystallography merges information into macromolecular crystallographic information file with various atomic coordinates and factors. NMR method determines the statistics from PDB file will be merged into macromolecular crystallographic information file with coordinate along with other constraint files. This extraction of parameters from an PDB file include statistical information such as force, kinetic energy required to bond, temperature etc. these information's will help to derive a 3D structure from it.

Objective of our work is to extract the parameters from the PDB file, convert those to a three dimensional structure later finding the accurate pathway for the analysis of Melanoma, a type of skin cancer.

## Atomic Coordinates: PDB Format

```
                            Chain name
                           /   Sequence Number
         Amino Acid       /   /
                   \     /   /
         Element    \   /   /     -----Coordinates-----
               \  \   /  /        X       Y       Z      (etc.)
   ATOM    1   N    ASP L   1     4.060   7.307   5.186   ...
   ATOM    2   CA   ASP L   1     4.042   7.776   6.553   ...
   ATOM    3   C    ASP L   1     2.668   8.426   6.644   ...
   ATOM    4   O    ASP L   1     1.987   8.438   5.606   ...
   ATOM    5   CB   ASP L   1     5.090   8.827   6.797   ...
   ATOM    6   CG   ASP L   1     6.338   8.761   5.929   ...
   ATOM    7   OD1  ASP L   1     6.576   9.758   5.241   ...
   ATOM    8   OD2  ASP L   1     7.065   7.759   5.948   ...
                   \\
          Element position within amino acid
```

**Figure 1: Sample PDB file**

Mathematical and biological approaches with computational assay of molecular pathway can be used in the medical field for the identification of type of cancer. Extracting the parameters from a PDB file is one of the mathematical and computational assay which can be used in identification of specific type of molecular pathway. The bonding of different elements in a perfect angle can be done with the help of three-dimensional view and then find out the energy released in bonding. Also, this three-dimensional structure leads us to the identification of protein and nucleic acid in PDB file. The differentiation of protein and nucleic acid can be easily done with the help of 3D structure of the PDB file. The best pathway can be identified by the low energy release while bonding. Decision of best pathway gives a raise to the generation of drugs for melanoma by the researchers and therapeutics. The developed software such as VEGA and SuMo gave a raise in the three-Dimensional structure in molecular dynamics can be applied for the conversion PDB file to its three-dimensional structure. It was limited to the interoperability and flexibility of its platform dependency and independency. So that both the features are included in the work. The methods for the extraction of parameters from the PDB file is being Proposed which further lead to conversion of a PDB file to 3D structure.

## 2. LITERATURE REVIEW

Pedretti et al [1] proposed VEGA in 2002 to produce a bond between the software packages of molecules. It features to the file format conversion, surface calculation, trajectory analysis and analysis, display and management of 3D molecular structure. Vega is user friendly and platform independent.

M S Pallavi et.al [2] has proposed a work on spam detection using support vector machine algorithm. The support vector machine algorithm determines the results fast and provide exact results for the problem. This algorithm can be used to find a quick and optimal solution for any kind of related works of optimization.

Sabitha Mangalathillam et.al [3] has proposed a work on curcumin loaded chitin nanogels to cure melanoma. The drug scatters in water and is less toxic. There is active transdermal penetration where occurs a vigorous elements transported through the skin for complete dissemination.

Gans et al [4] has presented a program Qmol in 2001 intended for inspecting the output of notional calculations and simulations of peptides and small molecules. To overcome the limitation of interoperability with windows platform they introduce Qmol with tentative visualization of data and read binary line data engendered by molecular simulations. It accept command line arguments and assist as a co-worker application to be executed by other applications.

Gouet et al [5] has designed a Program Easy Sequencing in PostScript (ESPript) in 1998 permits the hasty visualization through PostScript output of sequences associated with general programs such as CLUSTAL-W or GCG PILEUP. To produce a combination of both structural information and sequence it reads files of secondary structure. ESPript can also adhesive material on filtrate maintenance into coordinate files, for consequent visualization with a graphics program.

Hamelryck et al [6] has introduced a tool in 2003, Biopython is a set of tools in bioinformatics developed using python. It deals with the structure of macromolecule. Data structure and parser makes the tool working with flexibility and ease handles other packages such as residue disorder and atom ineffectively. To spot apparent errors parser accomplishes sanity checking.

Humphrey et al [7] has proposed a molecular graphics program in 1996, Visual Molecular Dynamics (VMD) for the study and exhibition of molecular gatherings in specific nucleic acid and proteins. Simulation trajectories from files expressed using colouring structure of precise subset of atom and specific interpretation method.

Jambon et al [8] has designed tool in 2003 for three-dimensional structure of proteins for analysis and detection of local spatial similarities of amino acids. A SuMo named software allowed the management of the chemical groups and selection of proteins. The evolutions, convergent and divergent are being focused to the applicability of the approach which differentiates functional and non-functional lectins.

Van et al [9] has established an approach of PDB based protein loop prediction in 1997 to acquire evocative statistics and to predict the correlation between the internal loop conformation and the conformation of the loop of adjacent stem bonds. This analysis is helpful for the energy minimization and the protein yield better results with non-bonded contacts so that study of the non-bonded contacts of the designated template loops is beneficial in defining the anticipated accuracy of a prediction.

Voss et al [10] has developed 3V in 2010 to identify and analyse clefts, channels and cavities internal volumes of a macromolecular structure by 3V web server which can analyse and extract automatically from structure of protein and RNA inputs. It analyses two rolling-probe solvent differences with large feasible radius of probe and radius of other solvent. The limitation is that the volumetric representation couldn't made as images and files which can be downloaded.

R I Ramachandran et.al [11] explains the basics of computational chemistry and molecular modelling which is the fast evolving zone used for modelling and simulation to determine the behaviour of biological systems at molecular level.

## 3. PROPOSED METHODOLOGY

The architecture diagram for the computational analysis of Melanoma is represented in the Fig. 2 where in which the first step to start with is getting the file from online using PDB, once we get the file the next step is to prepare the data for visualization which includes the steps like conversion of the PDB file to 3D structure, Evaluating the particular angle in which the atoms are to be bonded, Energy minimization, Docking.
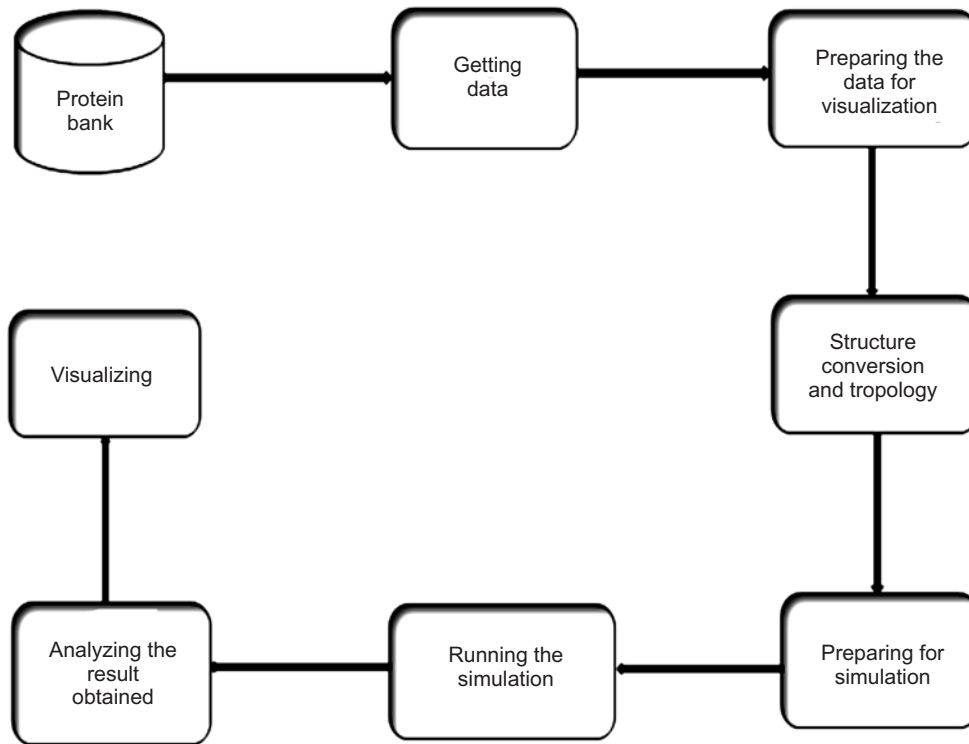
**Figure 2: Logical design for computational Analysis of Melanoma**

Then comes structure conversion which is to be done by using the structure which is already obtained, after this we need to prepare, simulate, analyze the structure respectively, once these steps are performed we have to visualize by using software such as the Chimera, flipbook etc.

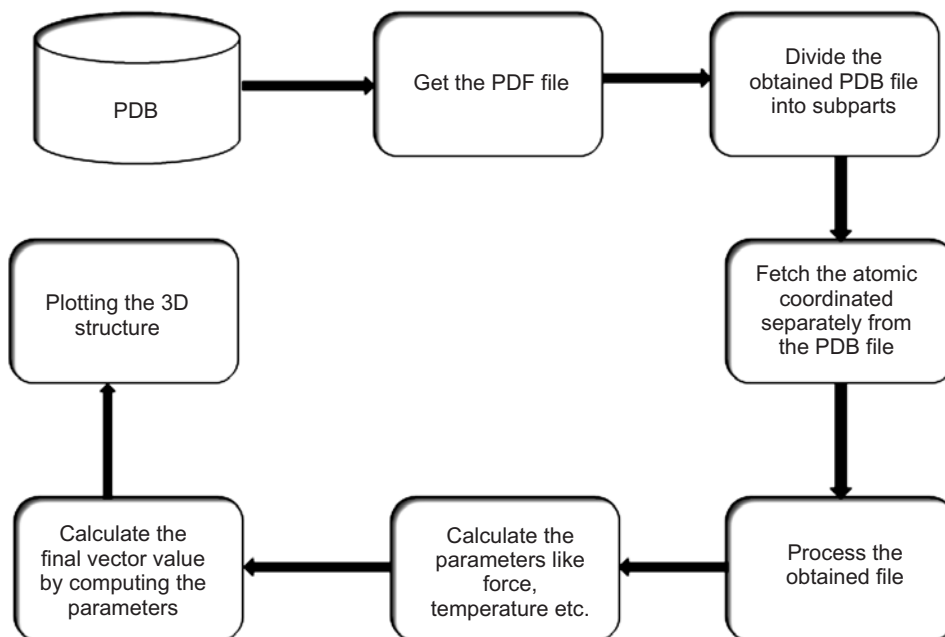The steps involved in the conversion of PDB file (.pdb) to three-dimensional structure is shown in the Fig. 3



**Figure 3: Design for the conversion of PDB to 3-Dimensional structure**

### 3.1. Steps for extracting the parameters from pdb file

**Step 1:** Create a class DataAtom.

**Step 2:** Create a method in DataAtom where it assigns name of the atom, mass, force, distance, maximum inter-atomic distance for bonds, force field option and cut-off distance.

**Step 3:** Create another class InOut, which is inherited by the base class DataAtom.

**Step 4:** Create a method in the InOut where it fetches the data from .dat file.

**Step 5:** Create another method in InOut class where it assigns the value which is fetched from the .dat file and assigns it to a variable.

**Step 6:** Another method is used to calculate, simplify and print the output in InOut class.

**Step 7:** Create another class called Init.

**Step 8:** Create a method where it calculates the points (x, y, z) by using various parameters.

**Step 9:** Another method is used in Init to calculate the heat, kinetic energy.

**Step 10:** Another method is used to calculate the value of (x, y, z) when it moves the atoms to their center of mass system

**Step 11:** One more method to find rescales the velocities such that the total kinetic energy amounts to initial total kinetic energy.

### 3.2. Classes used to implement the extraction of parameters from pdb file

**Program class:** Main class

**Inout class:** Class used to get input, print output, PDB input/output

**DataAtom class:** Atomic parameters

**Units class:** For the conversion factors between different systems of units

**InitConfig class:** For initial configuration

**Propag class:** For time integration algorithm

**Forces class:** For interaction forces

**Rescal class:** For position and velocity rescaling functions

**Measure class:** For measurements (bond statistics etc.)

**PlotCfg class:** To display of the geometric configuration

**Plots class:** To display plots of spatial and temporal dependencies

### 3.3. Standard representations of macromolecular structures (pdb document)

1.  Title Section
2.  Primary Structure Section
3.  Heterogen Section
4.  Secondary Structure Section
5.  Connectivity Annotation Section
6.  Miscellaneous Features Section
7.  Crystallographic and Coordinate Transformation Section
8.  Coordinate Section
9.  Connectivity Section
10. Bookkeeping Section

## 4.  EXPERIMENTAL RESULTS



```
file:///C:/Users/Anitha/documents/visual studio 2013/Projects...    —    □    ×

Enter the atom name
C
Name of the atom =C
 Mass of the atom =1.27150235407603E+22
 force of the atom =100
 distance of an atom =1.4
 interatomic distance =1.6
 cut-off distance=2
 force-field
1
type of run: 0
initialization: 0
no. of atoms is 60
chemical symbol is C
atomic mass is 12.011
force-like parameter is100
distance-like parameter is 1.4
cut-off distance is 2
kinetic temperature in K is 3481352
integration time step is 1E-05
no. of time steps is 10000
no. of time steps for output is 100
The value of x[] is 4.24651122508077E-09
 The value of y[] is -3.65666248628336E-16
 The value of z[] is 4.24651122508078E-09
```

**Figure 4: Extracting the parameters with respect to carbon atom**

Parameters which are required to evaluate the vector $(x, y, z)$ is represented for the carbon atom is as represented in the Fig. 4 where it is depicting the parameters like mass of the atom, force of the atom, distance of an atom from another, interatomic distance, cut-off distance, kinetic energy, time steps for input and output etc.

Parameters which is required to evaluate the vector $(x, y, z)$ is represented for the argon atom is as depicted in the Fig. 5 where it is depicting the parameters like mass of the atom, force of the atom, distance of an atom from another, interatomic distance, cut-off distance, kinetic energy, time steps for input and output etc.

## 5.  CONCLUSION

In this work we are extracting the various parameters from the PDB file. We examined methods to improve the computational exploration of the extraction of PDB file. We consider providing the atoms in the PDB file and molecular dynamic system. The approach to be executed by the therapeutics in the process of analysis of cancer and the structure of protein bond to be identified for the drug development process. The atoms formed with all its features to its bonding and unbonding possibilities. So that the atom with its features can be visualized in the form of graphics. This work deal with the extraction of the various parameters from the PDB file. The extraction of parameters from the PDB file further lead to conversion of a PDB file to 3D structure and for the minimization of energy calculation of protein bonding.

**Figure 5: Extracting the parameters with respect to Argon**

## 6. ACKNOWLEDGMENT

## REFERENCES

[1] A.Pedretti, L. Villa, and G. Vistoli, "VEGA: A versatile program to convert, handle and visualize molecular structure on Windows-based PCs," *J. Mol. Graph. Model.*, vol. 21, no. 1, pp. 47–49, 2002.

[2] M. S. Pallavi, Hegde, V., Anushadevi, H. G., and Ambika, V., "Automated spam detection in e-mail using SVM", International Journal of Applied Engineering Research, vol. 10, pp. 25219-25228, 2015.

[3] Mangalathillam, S., Rejinold, N. S., Nair, A., Lakshmanan, V. K., Nair, S. V., & Jayakumar, R. (2012). Curcumin loaded chitin nanogels for skin cancer treatment via the transdermal route. Nanoscale, 4(1), 239-250.

[4] Gans, J. D., & Shalloway, D. (2001). Qmol: a program for molecular visualization on Windows-based PCs. *Journal of Molecular Graphics and Modelling*, *19*(6), 557-559.

[5] Gouet, P., Courcelle, E., & Stuart, D. I. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, *15*(4), 305-308.

[6] Hamelryck, T., & Manderick, B. (2003). PDB file parser and structure class implemented in Python. *Bioinformatics*, *19*(17), 2308-2310.

[7] Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, *14*(1), 33-38.

[8] M. Jambon, A. Imberty, G. Deléage, and C. Geourjon, "A new bioinformatic approach to detect common 3D sites in protein structures," *Proteins Struct. Funct. Genet.*, vol. 52, no. 2, pp. 137–145, 2003.

[9] van Vlijmen, H. W., & Karplus, M. (1997). PDB-based protein loop prediction: parameters for selection and methods for optimization. *Journal of molecular biology*, *267*(4), 975-1001.

[10] Voss, N. R., & Gerstein, M. (2010). 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic acids research*, gkq395.

[11] Ramachandran, K. I., Deepa, G., & Namboori, K. (2008). *Computational chemistry and molecular modeling: principles and applications*. Springer Science & Business Media.