SIMULATION AND 3D VISUALIZATION OF COMPLEX MOLECULAR STRUCTURE FOR STUDY OF PROTEIN AND NANO MATERIALS

Himanshu S Mazumdar[1], Shivangi R Shah[2]

[1]*Professor and Head, Research and development Center, Dharmsinh Desai University, Gujarat, India*
[2]*Student, Information and Technology, Dharmsinh Desai University, Gujarat, India*

Abstract

Simulation and visualization of complex bio-molecules are gaining importance; this is because of the fact that functional properties of such molecules are more dependent on their 3D structures. One of the challenges of computational biology is prediction of structures of protein from amino sequence. Three dimensional visualization of molecular structure is also an important research goal in nano engineering. Such visualization, simulation and animation of reaction dynamics are essential for modern chemistry. Force field simulation of large number of atoms in complex molecular structure is computationally challenging. To find equilibrium of force fields and resultant dynamic structure of large atomic clusters which assembles to predictable molecules is one of the major goals of computational biology. In this paper we describe a simple but efficient algorithm (MoliSim3D) to simulate and dynamically visualize in 3D the assembly of atoms in presence of internal and external forces. It helps monitoring and measuring different bond angles, dimensions and displacements of selected portions from thousands of atoms forming target molecule. Pattern recognition, pattern error detection and reaction control are advance tools of virtual molecular assembly. The proposed system is basically intended as educational tool and to help researchers in understanding complex molecular dynamics and functionality with high degree of confidence in simulated environment.

Keywords: *Simulation, 3D Visualization, Structure, Dynamic*

----------***----------

1. INTRODUCTION

In the field of molecular biology, protein is the major structural constituent. It is the protein that represents most of the complex life functions. Major constituent of protein are amino acids which are most fundamental molecule in life present on the earth. Research has been carried out using complex searching algorithms for efficient usage of large Proteins Data-Bases (PDB) [5], similar protein search in large PDB like UniRef100 [6], protein structure prediction [9] and multiple protein sequence alignment [10]. Two major facts for protein are: (a) Different Protein structure is the cause for different protein function [3], (b) Different Protein function depends on different physical and chemical parameter, so it can be said that a given type of protein sequence can form multiple structures and those structures further can provide different functions. Enzymes are protein catalysts which can perform most complex chemical function by making or breaking the chemical bonds [1]. A protein needs to be positioned accurately in 3D space for efficient reaction in a given substrate. It is scaffolding that tethers other proteins to form stable complexes structure by implementing accurate docking [7]. For a large organic molecule the atom level representation using computer graphics is a complex task. From atom level representation of complex molecule, it is difficult to conceive the underlying pattern which needs to be identified for probable efficient tethering. Shape and structure of a protein causes different function. A function further uses 3D structure more than the sequence. Even though DNA molecule contains necessary information for life, the important process like maintenance, replication, defense and reproduction are carried out by proteins. It is the chemical and physical structures of a protein that determines its activity [4].

Each structure basically contains thousands of atoms per molecule. For this reason, it is often consider convenient to simplify the problem by using a hierarchical description of protein structure in which successive layer of the hierarchy describe increasingly more complex levels of organization[8]. Precision of structure is important along with the atomic detail as a small alteration in the atomic composition may dramatically change the functionality like hemoglobin to sequel shell structure which defers by Amino acid sequence change from glutamic acid to valine in 2 beta sheet out of 4 chains of hemoglobin [2]. Characterization of natural proteins and altering its function require not only the knowledge of its chemical composition but its physical structure and orientation in different dynamic conditions. Dynamic force field simulation of such molecules allows a researcher to visualize and plan desired modification strategies of controlled reaction. Such simulations will reduce the trial and error experimentations in determining structure of protein and design of targeted drugs. This process is further accelerated with availability of database of structure of known targeted drugs. Three dimensional patterns search for complementary structures in database automates the drug design process. This paper describes an educational or research tool that can perform the task of simple chemical lab experiments. One of the novel features of the proposed system is motion dynamics. This is implemented using an iterative algorithm of inter-atomic force equilibrium difference-equation. The algorithm is validated by measuring the resultant bond angles of synthesized well known molecules and standard amino acids as shown in figure-1.

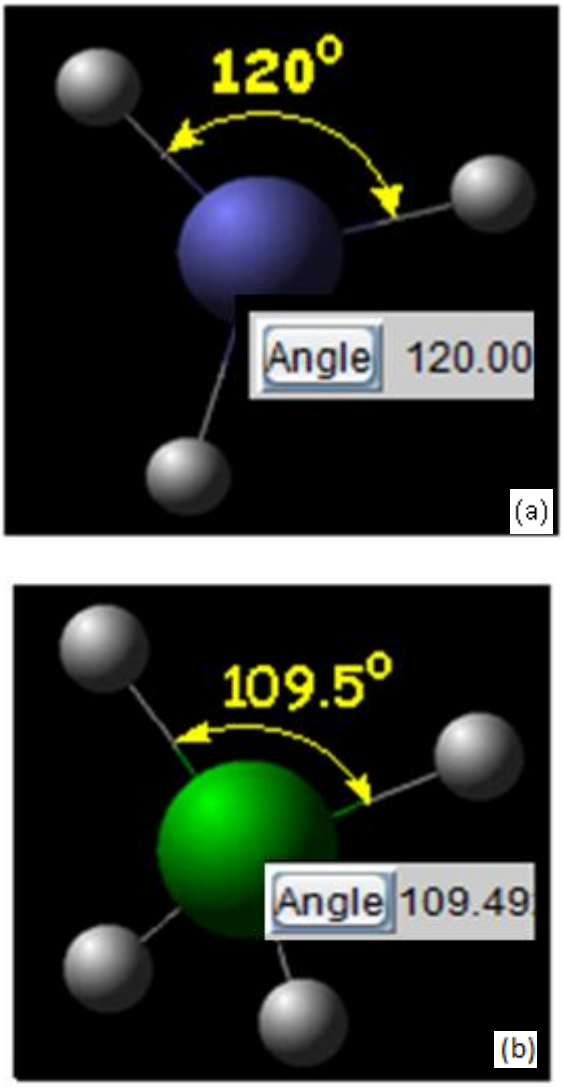

Fig-1: Results of dynamic force field equilibrium condition of trigonal amine NH_3 bonds in (a) and tetrahedron bonds of methane CH_4 in (b).

2. SYSTEM ARCHITECTURE

The architecture of MoliSim3D consists of a 3D graphics editor, inter-atomic force simulator, amino acid database and powerful utilities for 3D viewing of atoms, bonds and molecules. Most of the activities of the simulator are around the atom-bond 3D position database as shown in figure-2 and described in section-3. System is organized based on different menu driven utilities for composing, viewing, simulating reaction, database management and documentation. Important features of MoliSim3D are explained in section-4 and Section-5. The simulator generates a script file consisting of list of atoms with its attributes, list of bonds between atoms and current 3D position of all the atoms as shown in figure. During simulation atom-bonds and atom positions are continuously updated. This script file is saved as a library component of molecules. Number of these components can be loaded together and allowed to react to create bigger molecules like proteins.

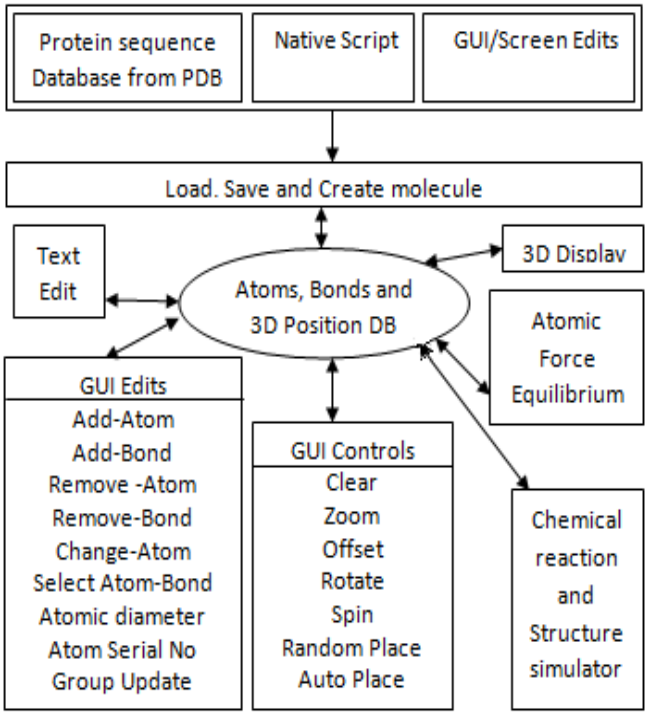

Fig-2: System architecture of MoliSim3D. All the activities of IO, GUI and force simulations are around the atoms, bonds and 3D position database of atom descriptor array.

3. DATA STRUCTURE

The heart of the MoliSim3D is atom-bond 3D position database as shown in figure-2. This database consists of array of atoms with properties and array of connections. Attributes associated with each atom are explained below.

3.1 Atom Structure

Most activities of this simulator architecture are around the array of atom structure which confirms acceptable nomenclatures of the molecular chemistry. Important elements of the atom structure are described below.

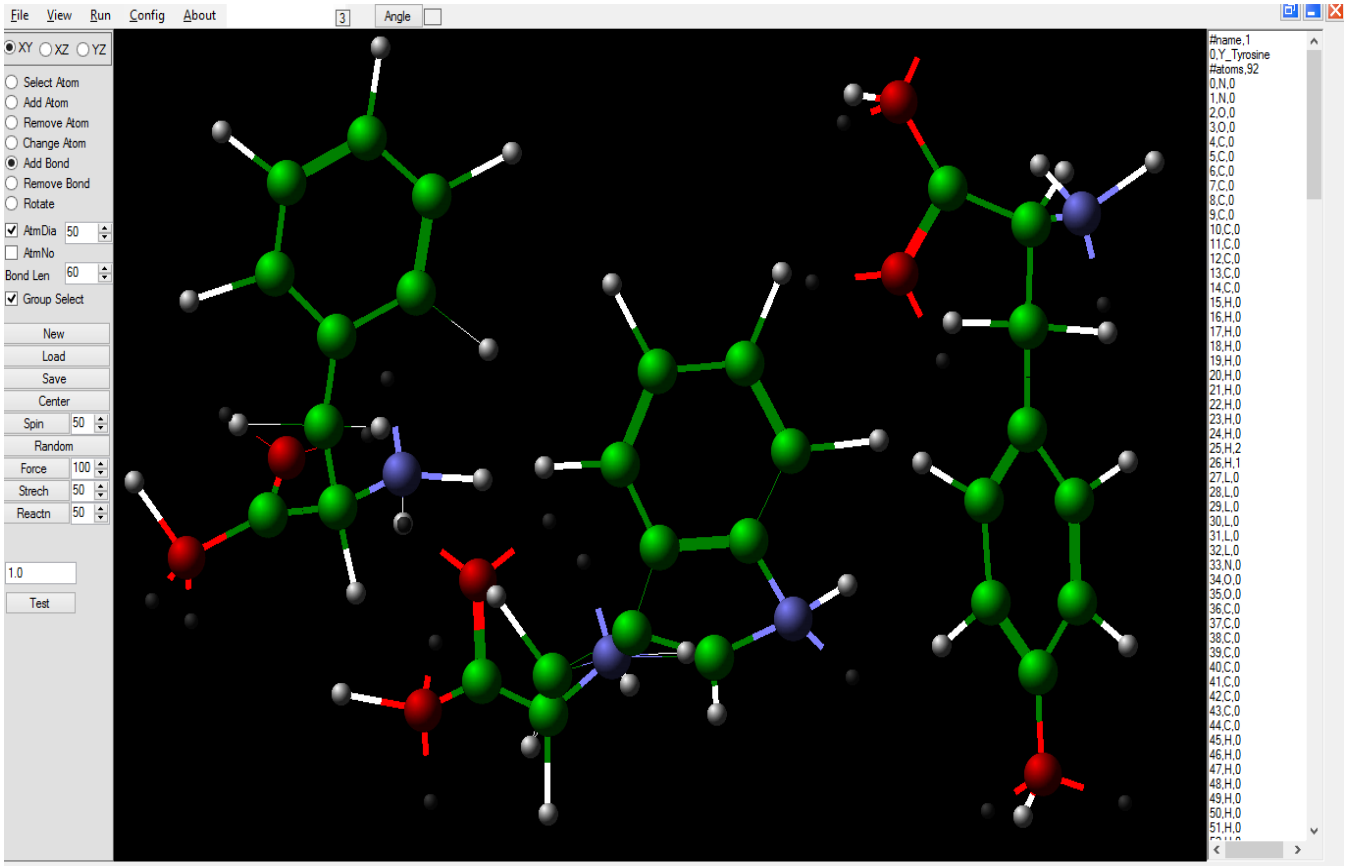

Fig-3: Dry Chemistry Lab application screenshot. On the left controls are provided for molecule composition and animation of force equilibrium. On the right script is generated describing the assembled molecules. The 3D display in middle shows animation of molecular dynamics and structure formation.

3.2 Atom Number

This field is used as unique atom ID. This ID is displayed on atom body and used in molecules descriptor list to define bonds and atom's 3D locations. Atom number is reorganized in run time after any new inter molecular bonding.

3.3 Atom Symbol

Atoms are labeled as 'H', 'O', 'C', 'N', 'S', and 'L'. These labels are associated with atom's color, diameter and valance used in molecules descriptor list.

3.4 Atom Color

Atom color is associated with the atom symbol and is automatically updated. This parameter is used for 3D atom and bond color display.

3.5 Atom Location

This field is used to hold the current position of the associated atom using normalized 3D coordinates. This is the key parameter for atomic and molecular reaction. Continuous updating of this parameter results in 3D animation of molecular reaction and structure visualization.

3.6 Atom Select

This is a Boolean variable used mostly for interactive editing and visualization of molecule. Multiple auto-selections such as group select help identifying and relocating connected atoms. Selected atoms are visually highlighted and used for adding, removing, changing or moving atoms.

3.7 Atom Source Link

This is an integer array field that contains list of atom IDs that are greater than this atom's ID and connected to this atom. Collection of atoms bonding list defines the bonding details of the molecule or connected atoms group. No bond is duplicated in these lists for a molecule.

3.8 Atom All Link

This is an integer array field that contains list of all atom IDs that are connected to this atom. All Link array of an atom contains all bonding details of source and destination atom IDs in a given molecule or connected atoms group. All bonds are duplicated in these lists. These arrays are used to calculate the dynamic force fields and continuously updated during reaction simulation.

3.9 Atom Diameter

This field provides the relative diameter value of an atom normalized with respect to diameter of Hydrogen atom. This parameter is used for realistic 3D projection of molecules.

4. COMPOSING AND EDITING FEATURES

As shown in figure-4 creation of a new 3D molecular structure or modifying an existing molecule is performed using a set of powerful but simple editing commands. The editing operation is performed in X-Y, X-Z and Y-Z projected plane of 3D space. Following are brief descriptions of important commands.

4.1 Add Atom

A new selected type of atom can be placed on any of the 3 plains using mouse double click. After placing it in a given plain it can be seen in other plane and dragged to desired position in 3D space. Any new placement updates the atom and location entries of source list of text window at the right.

4.2 Add Bond

Any two atoms can be bonded together using mouse click. Any editing actions are immediately reflected in the source list of the target molecule.

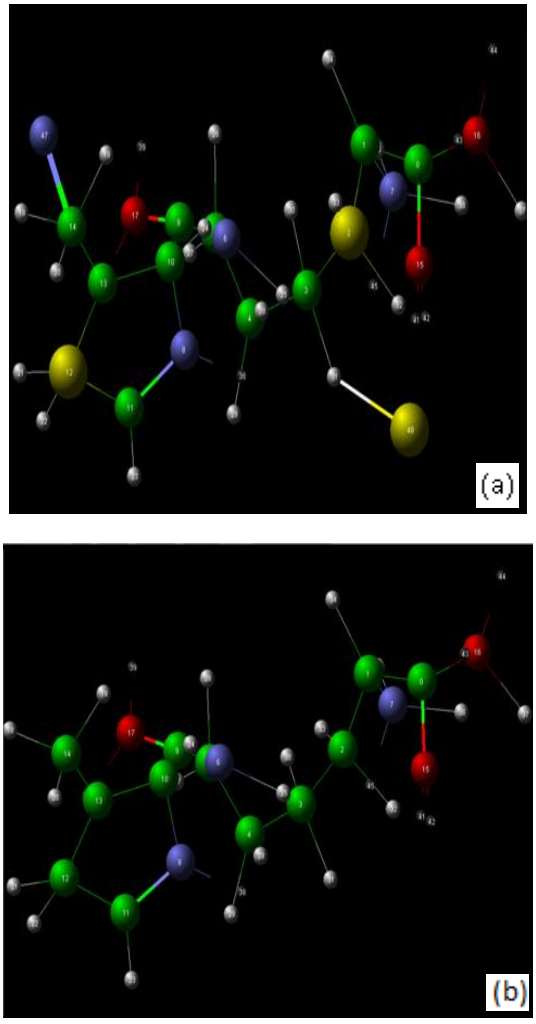

Fig-4: Screenshots of 3D molecule composition. (a) image shows 'Add Atom' and 'Add Bond' operations. On (b), editing utilities like 'Remove atom', 'Remove bond' and 'Change atom' are shown.

4.3 Remove Atom

Any atom can be removed from screen and list by double clicking on it if Remove Atom radio button is selected.

4.4 Remove Bond

Any bond can be removed from screen and list by double clicking on center of the desired bond if Remove Bond radio button is selected.

4.5 Change Atom

Type of any atom can be changed in screen and list by double clicking on it if Change Atom radio button is selected along with selected appropriate destination atom type.

4.6 Lone Pair

Insertion of Lone Pair is performed by adding 2 Lone pair elements and three bonds connecting them with associated atom. There is an external control to trim to calibrate the bond angles.

5. VISUALIZATION OPTIONS

Composed molecules are projected in 2D plane. A set of utilities are provided to enhance the viewing precipitation. Interactive measurement of 3D bond angles is an important feature of visualization. Descriptions of important viewing options are given below with relevant screenshots in figure-5.

5.1 Offset

Displacement of the 3D position of a selected atom, group of connected atoms or all atoms in reference to an origin or to a previous position is performed interactively using mouse drag. A complete 3D offset is a two step operation performed in two different planes.

5.2 Rotate

Rotation in 3D space of selected group in reference to group center or all atoms in reference to center of image are performed in X or Y axis using mouse drag.

5.3 Spin

This operation invokes continuous rotation of selected group in reference to group center or all atoms in reference to center of image in Z axis. Visualization of reaction or angular stabilization is enhanced in this mode.

5.4 Place at Center

This operation repositions average center of all atoms or selected atoms to the center of the screen. This operation is useful when atoms are drifted away from center under the influence of force fields.

5.5 Measure Bond Angles

Interactive measurement of 3D bond angles between any two bonds having a common node atom is performed by pointing connected atoms in sequence. Result in degrees is displayed in a text box as shown in figure-1.

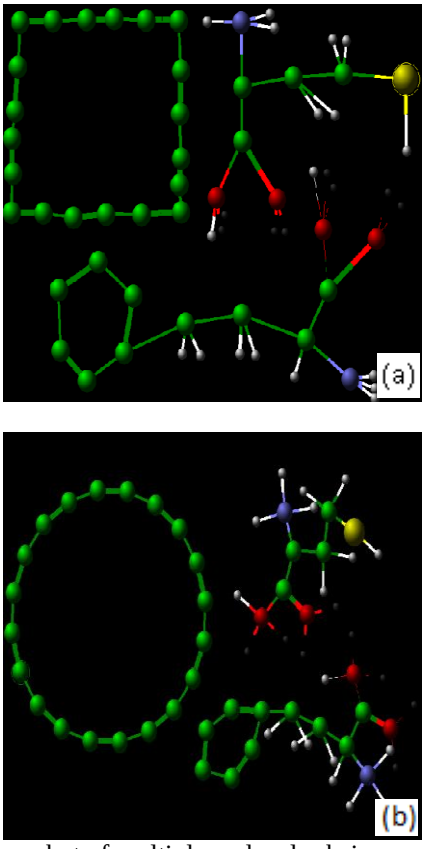

Fig-5: Screenshot of multiple molecules being composed. In (a), 3 molecules are assembled using manual editing and in (b), force field equilibrium is applied to get stable structures.

6. MOTION DYNAMICS

Simulation of motion dynamics is one of the important features of this simulator. Motion dynamics in MoliSim3D consist of computation and visualization of the temporal structure of atomic bonds forming molecule. Initially the atoms are placed using 3D graphics editor and are connected according to the chemical formula using bonds. Force equilibrium algorithm is applied to the assembled structure to compute the 3D co-ordinates of each atom in a given molecule under inter-atomic forces. The algorithm is iterative and uses difference equation to compute the atom co-ordinates. The intermediate co-ordinates of the atoms, when visualized produce effect of visual simulation of structure stabilization. There are 4 utilities associated with motion dynamics as described below.

6.1 Randomize Atom Position

The initial condition of atoms could be randomized using randomize utility. This utility helps to examine different possible stable states of any assembled molecule.

6.2 Stretch Amino and Carboxyl Ends

A special utility for amino acid or protein is included to generate an artificial repulsive force field between free ends of amino and carboxyl group atoms. This allows viewing of the entire amino acid chain of protein as a linear primary structure.

6.3 Force Equilibrium Algorithm

A computationally simple but effective algorithm is developed to organize large number of bonded atoms in 3D space very similar to their natural 3D structure in dynamic equilibrium for animation and final static position. This is further enhanced using chemical bonding rules to simulate molecular reaction like peptide bond formation from random initial states. A simplified flowchart of force equilibrium algorithm is shown in figure-6. It works on condition for attraction and repulsion forces where only end atoms are repulsive to each other. Bond angles in the stable or quasi-stable state are measured and are in agreement with experimental results as shown in table 1.

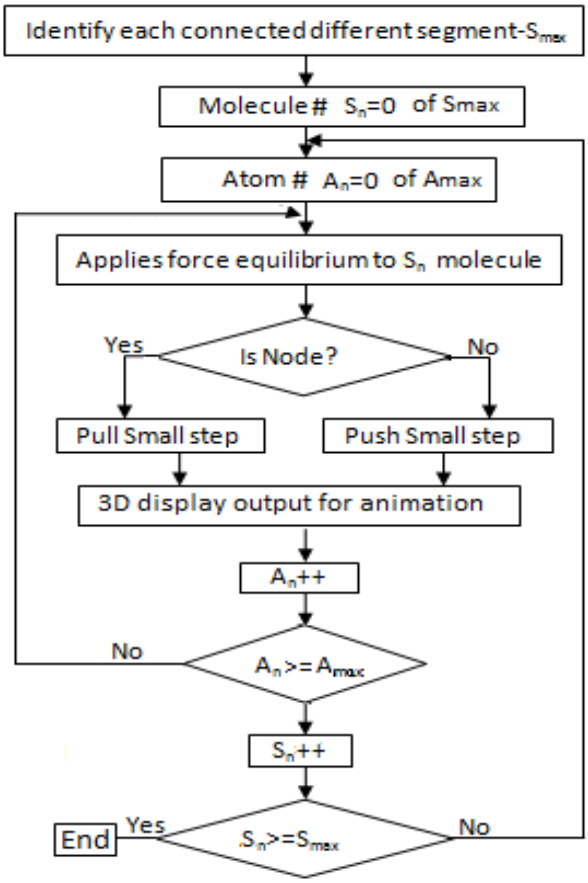

Fig-6: Flowchart of Force Equilibrium algorithm. The push and pull actions are implemented using difference equation of molecular dynamics.

6.4 Chemical Reaction Algorithm

Animation of amino acid reaction utility is implemented under force field equilibrium. This algorithm uses rule based interaction between amino and carboxyl group of nearest

molecules. This animated action produces peptide bond and releases H_2O molecule as illustrated in screenshots of figure 11 and 12. A simplified representation of the reaction algorithm is shown in figure-7.

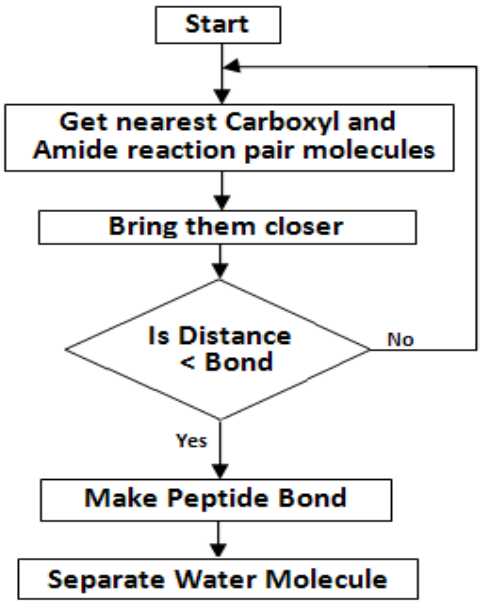

Fig-7: Flowchart of chemical reaction between standard amino acids.

7. PERFORMANCE ANALYSIS AND RESULTS

Different test are performed to validate the accuracy and performance of the proposed system. Actual angle and simulated angle between atomic bonds are compared after force field equilibrium. Average test results of 10 experiments are shown in table-1.

System is also tested with 100 randomly selected amino acids for dynamic equilibrium and reaction to make a single chain of protein. The operation consists of creation of 99 peptide bonds and separation of 99 H_2O molecules in pc based desktop environment. Result indicating total 10 numbers of experiments each for randomly selecting amino acids at constant force equilibrium of 100 is shown in table-2. These experiments are carried out using:

- Processor: I5, operating system: windows 8.1(64 bit), Ram: 8 GB and graphics card: 1GB.

Table 1: shows the comparison of actual angle and simulated angle between atomic bonds selected from a molecule. Result is calculated based on average of randomly selected bonds of 10 randomly selected molecules.

Total no of bonds	Molecular geometry	Actual bond angle	Measured average bond angles of 10 samples
2	Linear	180	180.0
3	Triagonal planer	120	120.0
4	Tetrahedral	109.5	109.4
5	Trigonal bipyramidal	90,120, 180	90.2,119.6, 179.5
6	Octahedral	90, 180	90.5, 180.0

Table 2: shows time taken and memory usage to create peptide bonds between randomly selected amino acids using force field equilibrium, stretch and reaction utilities.

Total no of amino acid	Total no of atoms present	Time taken to form peptide(s)	Memory taken to form PB (GB)	Maximum cpu usage(%)
2	46	3	1.1	25
5	210	5	1.1	25
10	563	21	1.2	25
20	1030	49	1.3	35
40	2060	200	1.4	35
50	1276	265	1.5	39
60	3090	595	1.5	35
100	2575	719	1.9	40
200	5150	3700	2.2	50

7.1 Display Type

Large number of different types of molecular assemblies are composed and structured under force field to test performance of the system. All the possible outputs of different compositions are found very close to the desired structure. Some of the samples are shown in figure- 8, 9 and 10

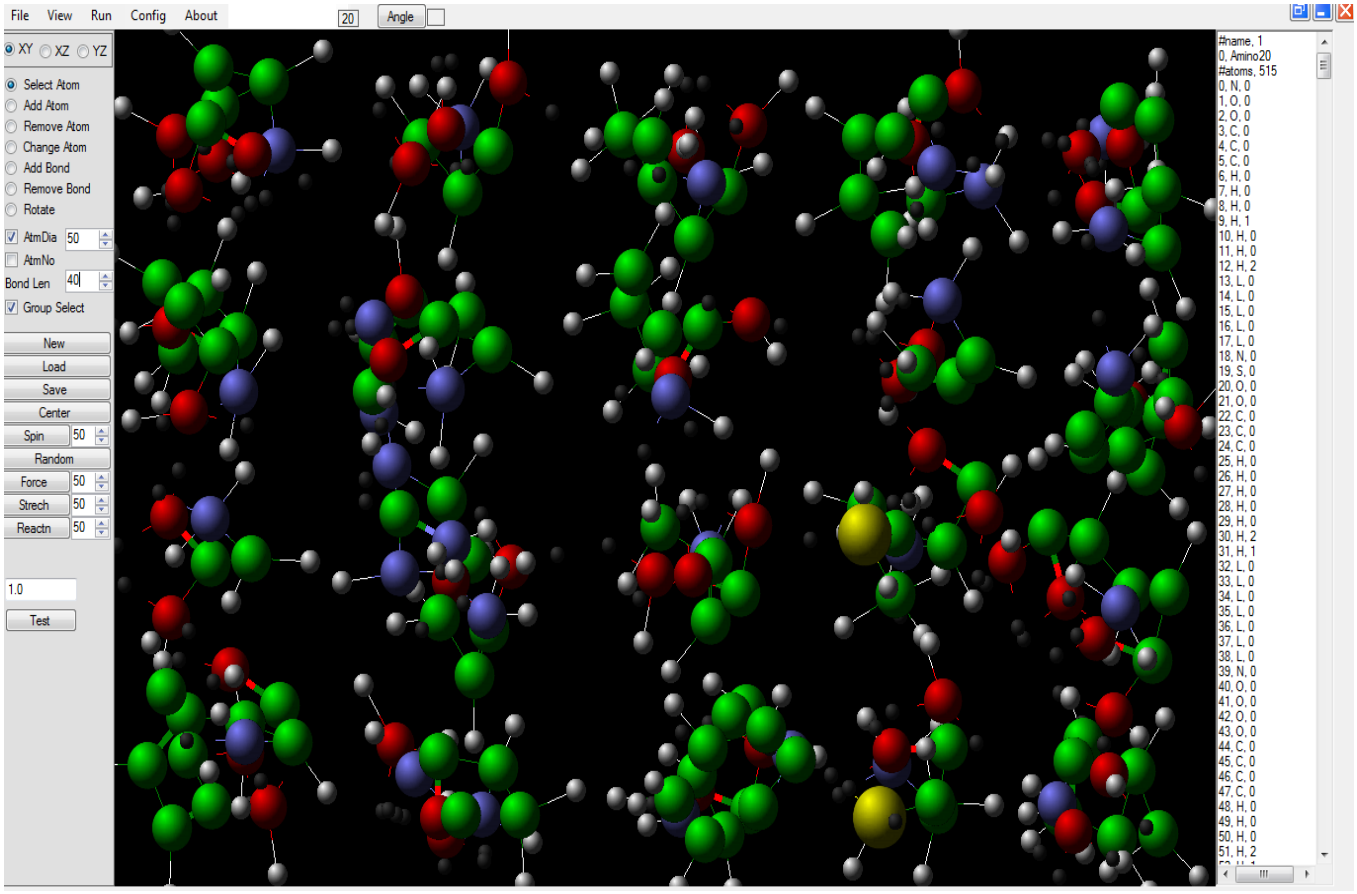

Fig-8: Screenshot of standard 20 Amino Acids

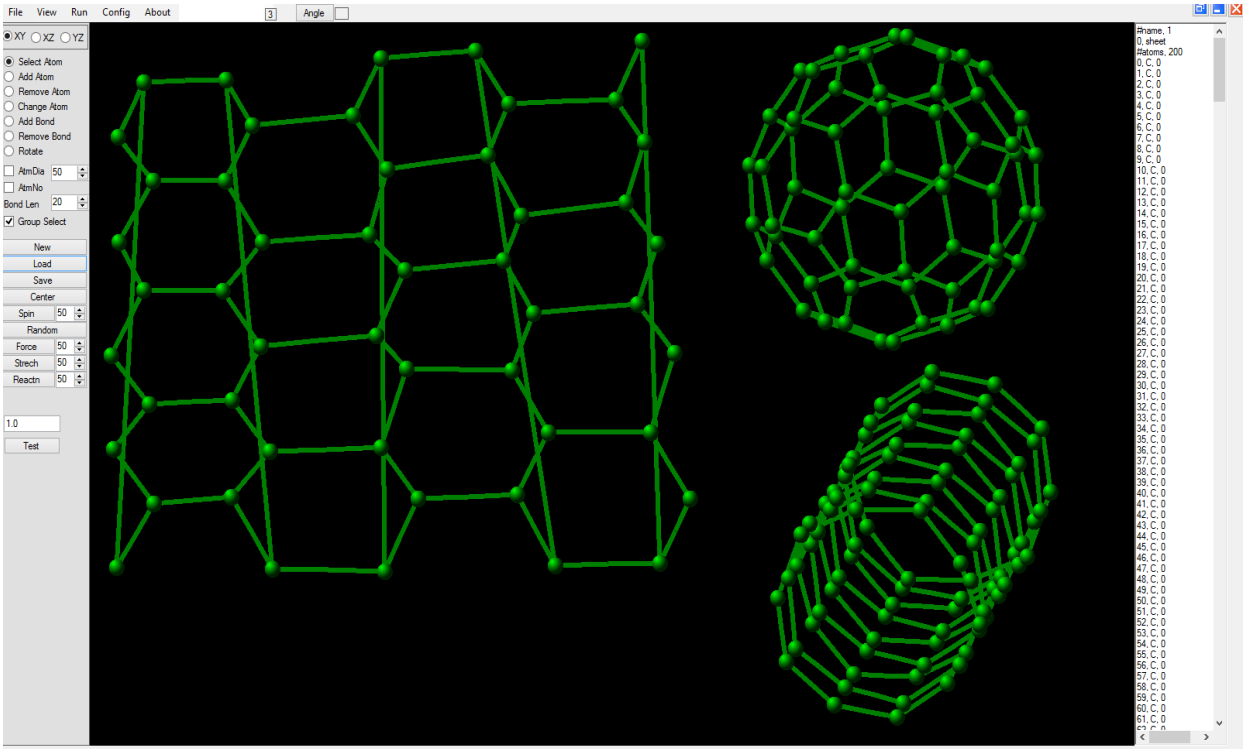

Fig-9: Assembly of three nano structures: Nano sheet, Bucky ball and Nano tube

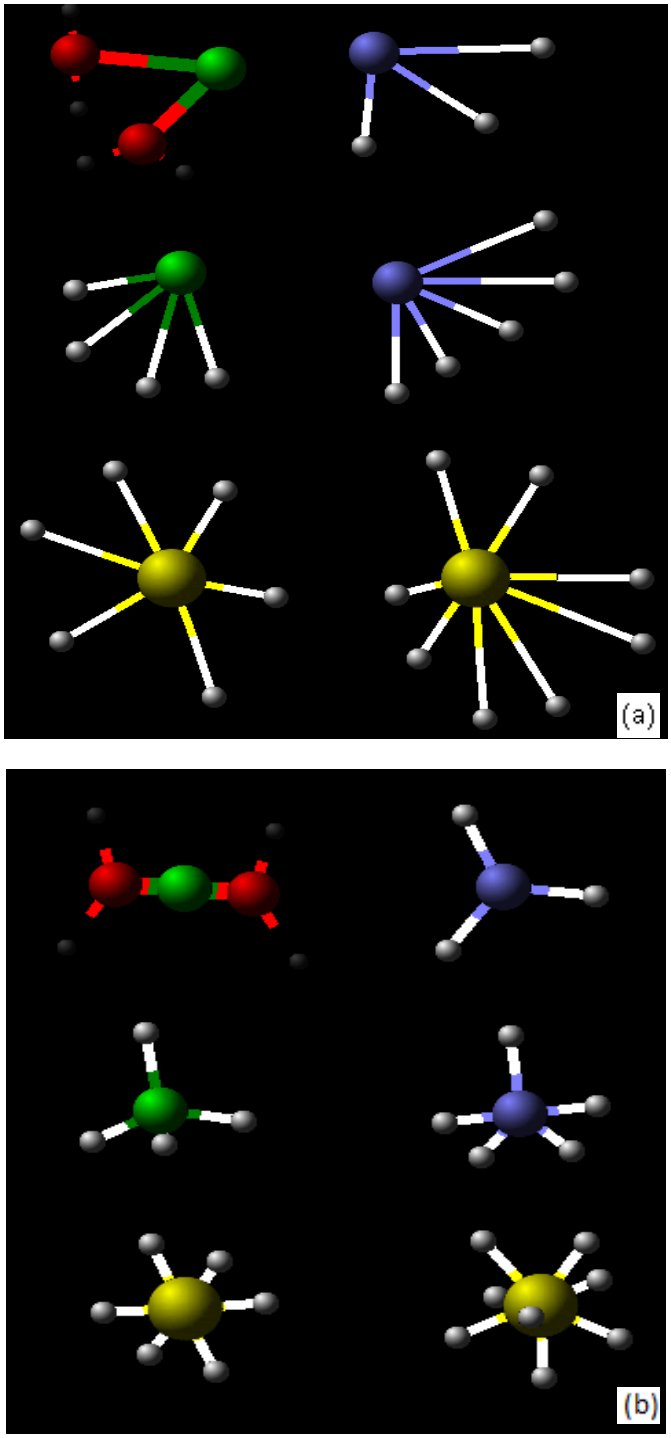

Fig-10: CO_2, NH_3, CH_4, NH_5, SH_6 and SH_8 molecules being composed at (a) and 3D view of each after equilibrium at (b) with appropriate bond angles.

7.2 Results of Molecular Reaction Simulation

Large number of amino acids are loaded and allowed to stabilize and react under force field and reaction utility. The system completes the reaction in few minutes using the hardware configuration as mentioned earlier. A screenshot of 100 amino acids before and after completion of reaction is shown in figure- 11 and 12. The operation took 8.5 minutes to compete**.**

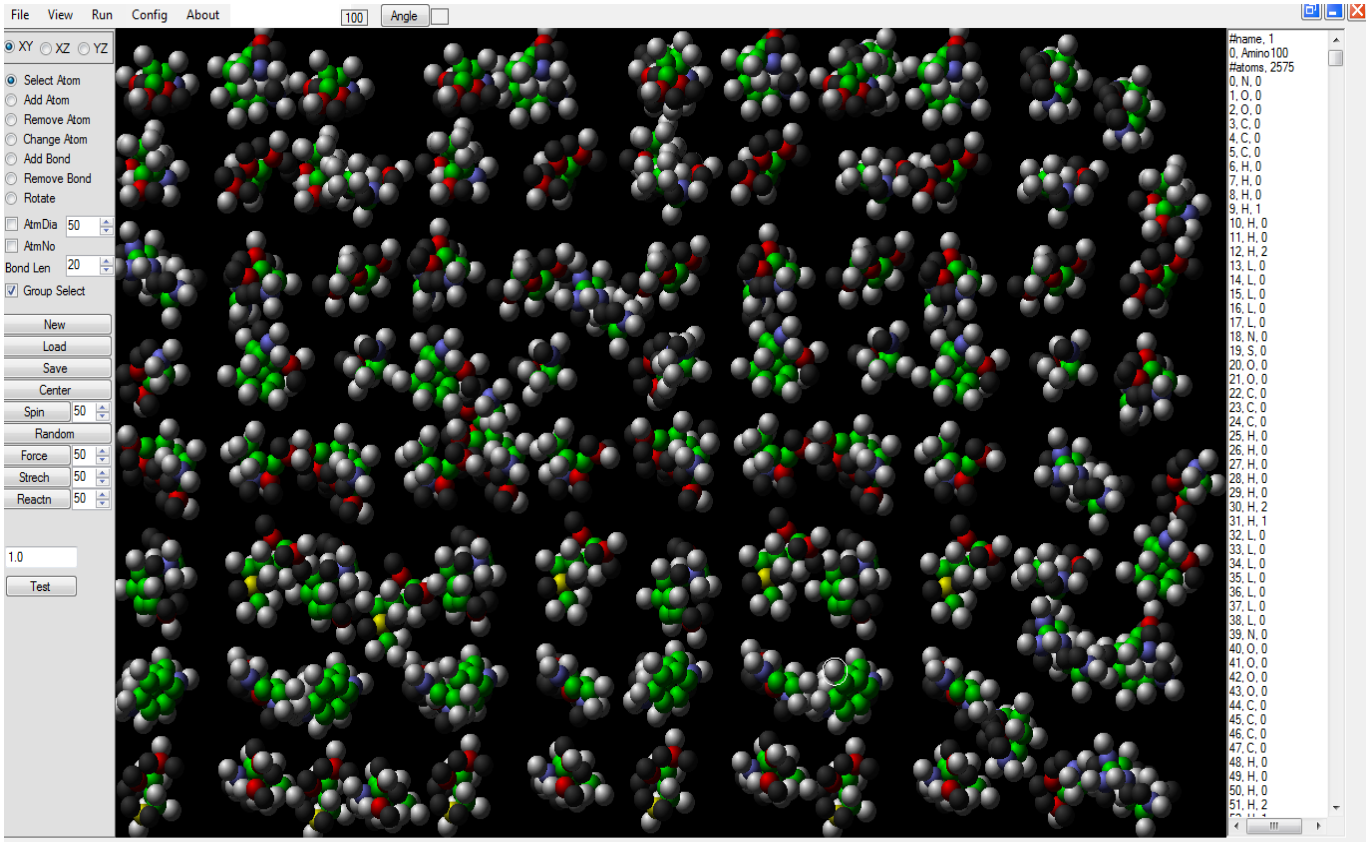

Fig-11: 100 amino acids consisting of randomly chosen from 20 standard amino acids are loaded from amino acid library.

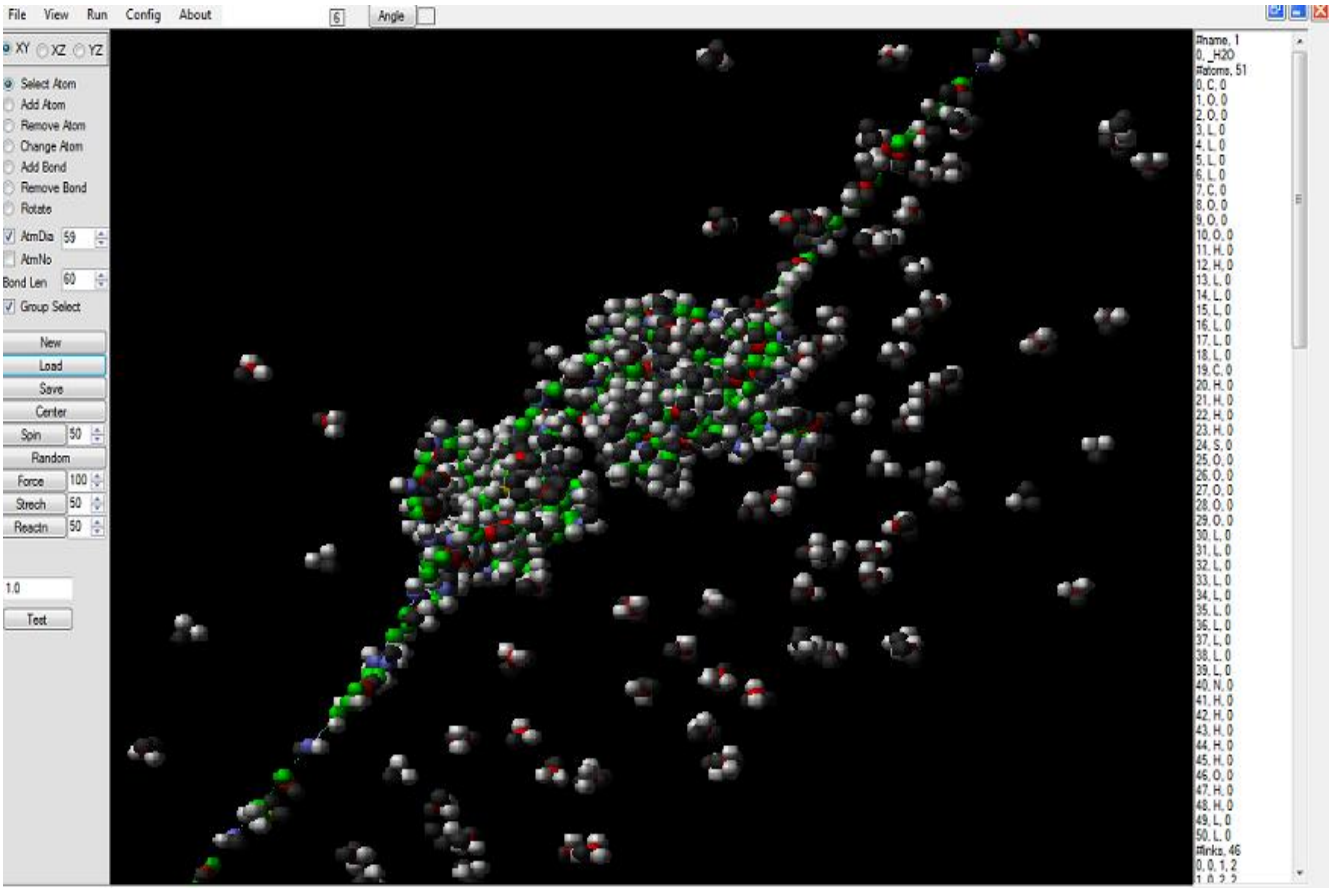

Fig-12: The view while reaction of 100 amino acids as shown in figure-11. New 100 molecules include 99 water molecules along with the linear protein is seen under stretched condition.

8. CONCLUSION

A desktop application for simulation and 3D visualization of molecular reaction is developed using C# Visual studio. The application is validated using reaction of 20 standard amino acids that are randomly placed and are allowed to interact chemically to produce a protein chain as shown in figure 11 and 12. The system is capable of interactively composing simple molecules like H_2O with electron lone pair in CAD like environment and supports measurement of the bond angles in equilibrium condition. It supports large external and internal database as input. It can load amino sequence from entire protein database of uniref100 or from native database of 20 amino acids. It can assemble both proteins and other inorganic molecules. Multiple molecules could be loaded at the same time to simulate and visualize intermolecular reaction dynamics. One of the useful features for portability is use of text representation of the current state of atoms, bonds and positions in 3D with details of bond strength and lone pairs. Its force field simulator with randomize initialization helps exploring number of possible stable states in equilibrium.

ACKNOWLEDGEMENTS

This work is carried out in R&D Center, Dharmsinh Desai University Nadiad and with resources of PLANEX, PRL project as M.Tech Dissertation work. We sincerely thank the staff of R&D Center for providing technical support.

REFERENCES

[1] Alberts B, Johnson A, Lewis J, et al. New York "Molecular Biology of the Cell. 4th edition" Garland Science; 2002.

[2] Julie Ann Buford. "Sickle Cell Hemoglobin and Malaria: An Adaptive Study of Natural Selection on an Infectious Disease." (2004).

[3] Charles W. O'Donnell, Mieszko Lis, Jérôme Waldispühl & Srinivas Devadas "Predicting Protein and RNA Structures and Interactions using Biophysically-Motivated Energy Functions and Machine Learning" Computer Science and Artificial Intelligence Laboratory (CSAIL)-2007

[4] Gruber, Arthur, Alan M. Durham, Chuong Huynh, and Hernando A. del Portillo. "Bioinformatics in tropical disease research." (2008).

[5] Maulika S Patel and Himanshu S Mazumdar, "Utilities for Efficient Usage of Large Biological Database," Phil. WCECS 2010 Vol II, October 20-22, 2010, San Francisco, USA.

[6] Maulika S Patel and Himanshu S Mazumdar, "Similarity Search Using Pre-Search in UniRef100 Database," International Journal of Hybrid Information Technology Vol. 4 No. 3, July, 2011

[7] Catherine Qiurong Pan, Marius Sudol, Michael Sheetz and Boon Chuan Low "Modularity and functional plasticity of scaffold proteins as p(l) acemakers in cell signaling" Cellular Signalling-2012

[8] Tverdislov, Vsevolod A. "Chirality as an Instrument of Stratification of Hierarchical Systems in Animate and Inanimate Nature." arXiv preprint arXiv:1212.1677 (2012).

[9] Maulika S Patel and Himanshu S Mazumdar, "Knowledge base and neural network approach for protein secondary structure prdiction," Journal of Theoritical Biology 2014

[10] Himanshu S Mazumdar Ankita C Baravaliya and Maulika S Patel, "Keyword based Iterative Approch to multiple Sequence Alignment," Int. J. Pure App. Biosci. 2 (3): 139-144 (2014)

BIOGRAPHIES

Himanshu S Mazumdar, Ph.D., Senior Member, IEEE is currently working as Professor, EC and Head of Research and Development Center at Dharmsinh Desai University. He has worked in important space missions in University College London, NASA's Space Shuttle and Indian Space Research Organization.

Shivangi Shah received the Bachelor degree in Computer engineering in 2013 and currently pursuing Masters in Information Technology and has carried out dissertation work in Research and Development Center during 2014-2015 from Dharmsinh Desai University