

Molecular Dynamics Simulation Studies of HYD1

Cephas Mawere¹, Suresh Kumar G²

^{1,2}Department of Bioinformatics, School of IT, Jawaharlal Nehru Technological University Hyderabad, India

Abstract: *The unique features of HYD1 protein and its primary involvement in fungal pathogenesis of Gibberella moniliformis have drawn attention to in silico study of potential inhibitors. This destructive mold is also on the rise of causing cancer related diseases in both humans and animals upon ingestion of the infected maize and/or rice grains. To date, currently used fungicides are ineffective against G. moniliformis as the fungus can grow and adapt to any environment. It has been realized that such behavior is attributed to class I hydrophobins where HYD1 belongs. These bio-surfactant macromolecules spontaneously self assemble at hydrophobic: hydrophilic interfaces into thin nanometric rodlet biofilms. For this reason, this paper focuses on refining a predicted HYD1 protein and studying its conformational and thermodynamic properties. Molecular dynamics simulation studies thus prove the protein to be thermally stable, compact, and results in a better quality model. This will help in understanding the self assembly mechanism of HYD1 and future docking studies.*

Keywords: HYD1, Gibberella moniliformis, Hydrophobins, Bio-surfactant, Self-assemble, Molecular dynamics simulation

1. Introduction

Hydrophobins are a group of small surface active protein uniquely found in filamentous fungi [1] like *G. moniliformis*. They constitute ten percent of the fungal cell wall as structural components and when secreted they are localized on the surface of aerial mycelia [3]. At hydrophilic: hydrophobic interfaces these proteins spontaneously self-assemble into different types of supramolecules. Based on this unique property, solubility and the pattern of Cys residues in their primary sequence, hydrophobins are divided into two groups- class I and class II [2]. Class I hydrophobins, where HYD1 belongs, can spontaneously self-assemble into a 10 nm wide amphipathic rodlet film in the presence of a hydrophilic: hydrophobic interface [5]. This biofilm lowers the surface tension of the medium or substratum in/on which *G. moniliformis* will grow [6]. Of interest, class I hydrophobins exists in three states; as monomers they are soluble in water, but self- assemble into the beta state at water-air interface. However, at the interface between water and a hydrophobic solid like Teflon, HYD1 attains an alpha helical structure [7]. Such a self assembly property of HYD1 causes it to be thermally resistant to heat treatments below 90°C. The protein also tends to be insoluble in aqueous solvents only to be dissociated by harsh treatment with strong acids such as concentrated TFA [9]. It has therefore been suggested by Wosten et al [1] that the eight conserved cysteine residues that form four disulphide bridges prevent self- assembly of the hydrophobin in the absence of a hydrophilic: hydrophobic interface, while the N-terminal part of hydrophobin, at least partly, determines wettability of the hydrophobic side of the assemblage. Experimental evidence of related class I hydrophobins also indicate that the alpha helical state is an intermediate while the end form of assembly is the beta state [8]. These two assembled forms have an amphipathic nature which confers water repellent properties to many conidia, hyphae and multicellular structures. The rodlet layer is covered by a mucilaginous extracellular matrix that helps the fungal conidia to bind to the substrate [6]. This makes class I hydrophobins, which includes HYD1, to be essential for the *G. moniliformis* to complete its biological life cycle. As such, class I

hydrophobins are actively involved in formation of hydrophobic aerial structures like spores and fruiting bodies, in fungal development, environmental adaptation and pathogenic interactions. During pathogenesis, for instance, class I hydrophobins mediate attachment of hyphae to each other [3] and to the hydrophobic surface of the host plant before penetration and infection commence [4]. For this cause *G. moniliformis* has been a destructive pathogen to more than 11,000 plant species including maize and rice [9]. This study therefore seeks to characterize the dynamic behaviour of HYD1 in water, prior to docking simulations in a quest to find a potential class I hydrophobin inhibitor for the aforementioned fungus.

2. Materials and Methods

A predicted HYD1 protein model from the M4T_HYD1 server [10] validated by PROCHECK and Verify3D was used for MD simulation studies. GROMACS 4.5.4 package [11] was thus chosen to explore the dynamic behavior of M4T_HYD1 protein with the empirical force-field GROMOS96 43a1 being used. The study on protein conformation, stability, compactness and refinement was therefore done in a four staged simulation process as indicated in Fig. 1 below.



Figure 1: The four staged molecular dynamics simulation

2.1 Energy Minimization

In order to understand the atomic scale mechanisms of HYD1, a topology file (.top), a GROMACS-compatible coordinate file (.gro) and a position restraints file (.itp) were created using the 'pdb2gmx' command. The .gro file of HYD1 was then used to prepare the system simulation cell. The resulting cubic box was defined by a 1.0 Å edge length. This was followed by centering the protein in the box. An SPC (a single three-point water model coordinate) file (spc216.gro) was then used to fill up the box, hence solvating the system. The system net charge was then neutralized by replacing six solvent molecules (SOL) in the .top file with Three Na⁺ and three Cl⁻ ions. In order to relax the molecular geometry of the system by getting rid of atomic clashes or any existing irregularities certain parameters were considered. These include a maximum force (emtol) was set at 100kJ/mol/nm, periodic boundary conditions in the directions x, y, z, 0.8 nm cut-offs for short range VDW and electrostatic interactions as well as a particle-mesh Ewald long range electrostatics. Once these factors were set, the system was minimized for 50,000 steps using the steepest descent algorithm.

2.2 Equilibration

System equilibration was characterized by two ensemble based simulations- NVT and NPT. Once minimization was complete the system was first simulated at 25,000 steps with a 2fs time step (a total of 50ps) to bring it to equilibrium. Thus, an NVT ensemble was used to create trajectories at a Langevin temperature of 298 K and a constant volume of 158 nm³ using Berendsen coupling algorithm. Position restraints were applied on the protein backbone to avoid any wild fluctuations. Also, all velocities were evenly distributed across all molecule types by coupling protein atoms and non-proteins in separate baths to keep the M4T_HYD1 rigid. Finally, the linear constraint solver (LINCS) algorithm was employed to constrain all bonds.

After this calibration step to a target temperature, the system was relaxed with an NPT ensemble. This allows the system to find the correct density. This time, the simulation was extended to 500ps with position restraints still being applied to the backbone of M4T_HYD1. A Nose-Hoover thermostat was then used in place of the Berendsen method to keep the kinetic energy of the system from fluctuating wildly. Lastly, pressure was isotropically (the same in all directions) coupled to a value of one bar.

2.3 Production MD

As soon as the system reached equilibrium and the correct density was found, a classical molecular dynamics simulation step was introduced. Here, the position restraints on the backbone of M4T_HYD1 were lifted. As such, all other parameters applied in the NPT equilibration were used for 2,500,000 MD steps with a 2fs timestep, totaling 5,000 ps (or 5ns) of sampling.

2.4 Analysis

A 'trjconv' command was used to convert trajectory files to .xtc and .pdb format for further analysis. A 'g_energy' was also used to output total energy, volume, potential energy, kinetic energy and pressure data results for the equilibration

stage and energy minimization stage. 'g_rms', 'g_rmsf', 'g_sas', and 'g_gyrate' commands were respectively employed to calculate the RMSD, RMSF, Solvent Accessible Area, and Radius of gyration of the simulated protein. Likewise 'g_mindist' and 'g_dist' were used to measure the distance between atomic residues. The resulting files (.xvg) were analyzed using xmgrace tool [12]. Moreover, the 'g_confirms' command was used to generate a pdb file of the original M4T_HYD1 protein superimposed to the simulated structure. RasMol [13] and PYMOL Viewer [14] programs were used to visualize and interpret the trajectories, the generated protein simulate, and the generated rmsf.pdb file. Finally, the secondary structure content of the simulated M4T_HYD1 as a function of time was analyzed against that of the original protein using the DSSP (Dictionary of Secondary Structure of Proteins) tool [15].

3. Results and Discussion

3.1 Energy minimization (EM)

EM of M4T_HYD1 foresaw steepest descent algorithm converging to Fmax (maximum force) in less than 100 in 6,153 steps (approx. 6ns). This resulted in the system potential energy being reduced from -1.8e+05 kJ/mol to global minimum value of -2.7129166e+05 kJ/mol (Fig. 2). Thus, energy minimization has ensured that we have a reasonable starting structure, in terms of geometry, solvent orientation and without sterical clashes.

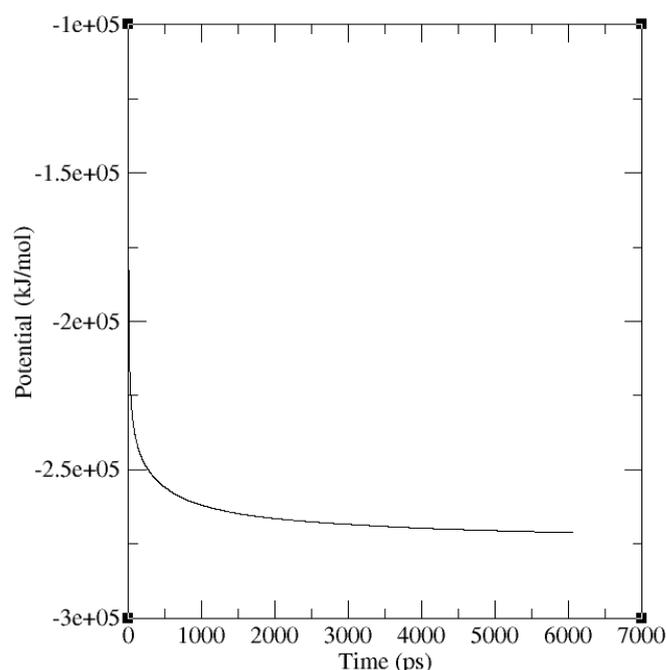


Figure 2: Potential energy graph for minimized M4T_HYD1

3.2 Equilibration

a) NVT Simulation

Equilibration using the NVT ensemble was validated as successful based on results obtained for energy attributes like temperature, potential energy, kinetic energy, and total energy. From Table 1.0 and temperature and total energy plots (Fig. 3 and Fig. 4) it is apparent that the system is well equilibrated around the target temperature (298 K) and total system energy in the first 50ps.

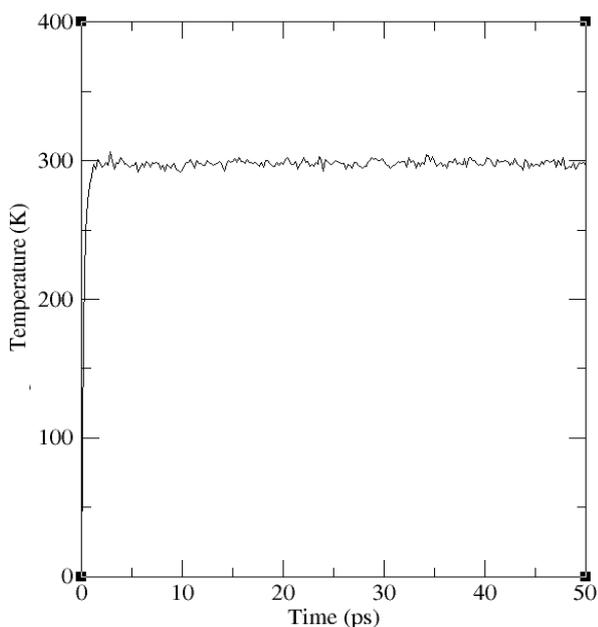


Figure 3: System temperature as a function of time during NVT equilibration

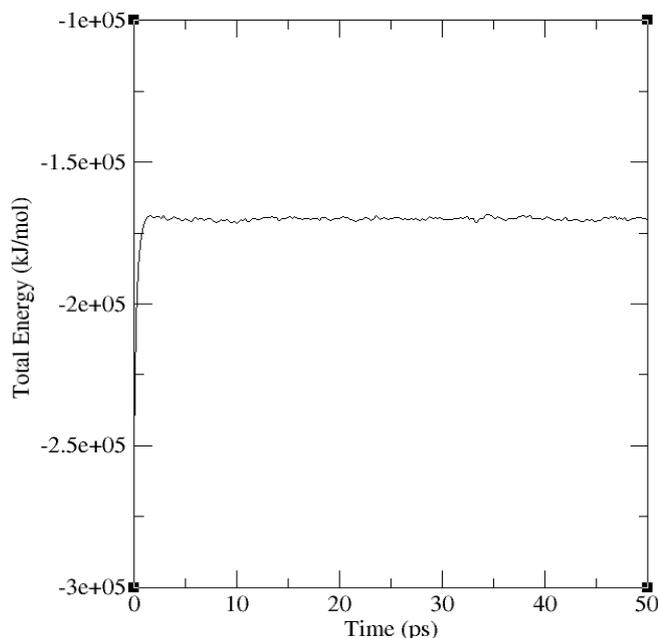


Figure 4: Total Energy of system under equilibration

Table 1: Attributes used to validate the equilibration stage

| Attribute | Average Score | Err. Est | RMSD | Total drift | S.I Unit |
|-------------|---------------|----------|---------|-------------|-------------------|
| Temperatur | 296.69 | 1.6 | 13.3317 | 8.8066 | K |
| Pot. Energy | -207614 | 270 | 2455.01 | 1462.4 | kJ/mol |
| Kin. Energy | 37238 | 200 | 1673.29 | 1105.3 | kJ/mol |
| Tot. Energy | -170093 | 44 | 1395.11 | 206.91 | kJ/mol |
| Pressure | 2.67034 | 2.3 | 186.631 | -12.82 | Bar |
| Volume | 157.371 | 0.031 | 0.18585 | -0.125 | nm ³ |
| Density | 994.646 | 0.2 | 1.17671 | 0.7904 | Kg/m ³ |

b) NPT Simulation

However, a correct density among other factors like pressure and volume (Table 1) are required to authenticate a

successful position restraint based simulation.

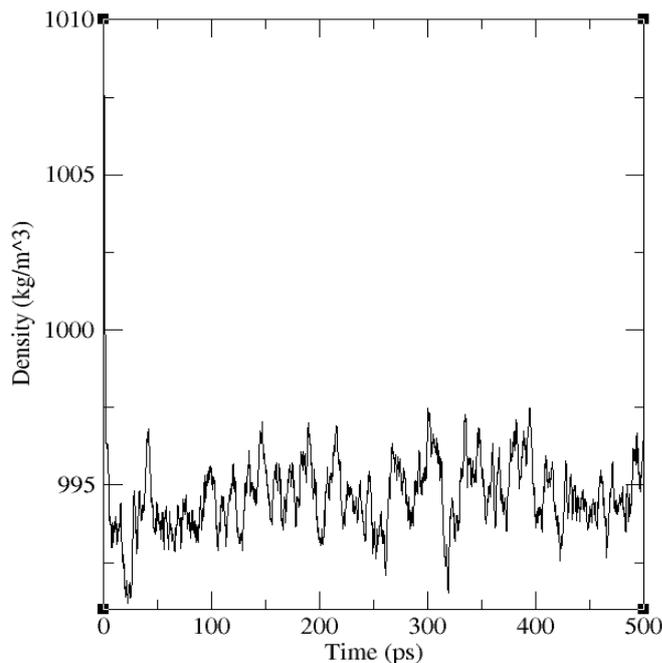


Figure 5: System density as a function of time during NVT equilibration

Accordingly, the density of the system appears to have quickly reached equilibrium quite near to a physiological value as shown by the density plot (Fig. 5).

3.3 Analysis

3.3.1 RMSD, RMSF and Radius of Gyration

The root means square deviation (RMSD) is a measure of how much the structure of a protein changes over the course of the simulation. For small, globular proteins like HYD1, changes on the order of 1-3 Angstroms are acceptable and expected [16]. Beyond this range the protein might be misbehaving. In this study, the backbone option was chosen for the least fit and for the RMSD measurement, and the output 'rmsd.xvg' file was visualized in Grace (xmgrace):

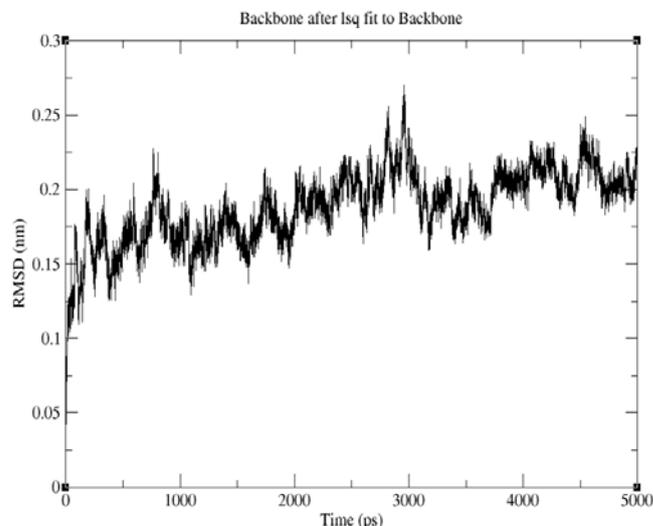


Figure 6: M4T_HYD1 backbone RMSD during a 5ns simulation

The black line in Fig. 6 thus represents the RMSD of the protein backbone during the simulation, which lies in the acceptable range. Despite this being a 5ns simulation, M4T_HYD1 is reasonably stable around 0.2 nm. Now, while RMSD measures the global backbone deviation, the local changes are measured using the root mean square fluctuation (RMSF). In this study much interest was in the entire residue fluctuation (Fig. 7) as well as the residues of the protein backbone (Fig. 9). Thus the 'g_rmsf' command resulted in two important output files; a data file of RMSF values plotted by residue number that can be viewed with Grace and a structure file (.pdb) that contains beta-factors equal to RMSF values.

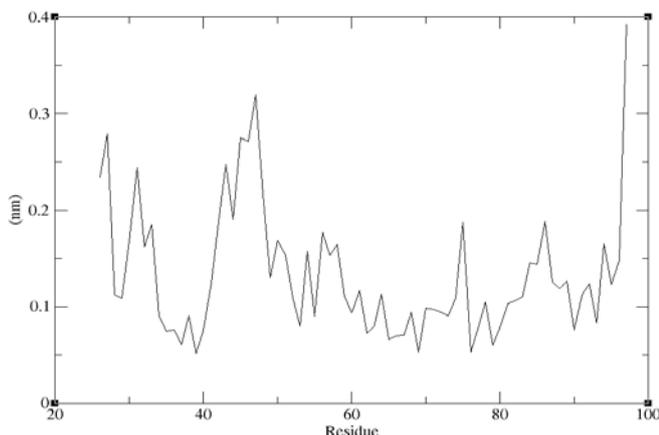


Figure 7: RMS fluctuations of protein residues during simulation

On this plot, peaks indicate areas of the protein that fluctuated the most during the simulation. The C-terminal tail here fluctuated much more than any other part of the HYD1 protein.

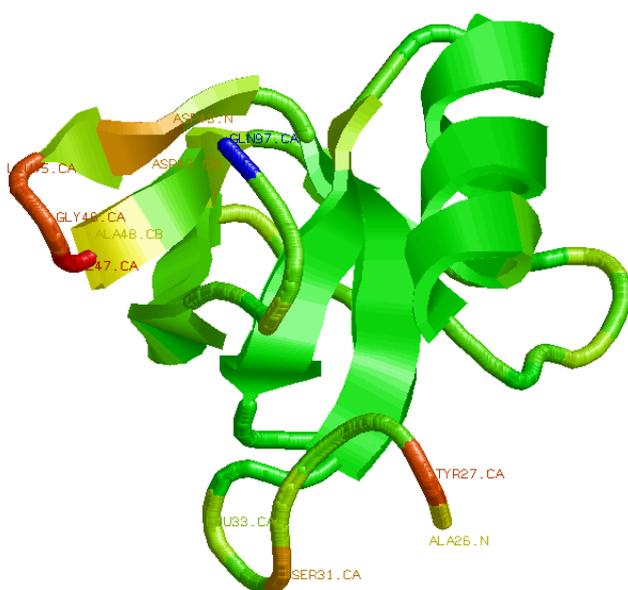


Figure 8: RMSF protein residues as viewed in RasMol

When the rmsf.pdb for the above plot was loaded in RasMol and visualized using the temperature colour scheme, the above structure was produced. Here, blue value indicates the most fluctuation which decreases in the colour order : red-

orange- yellow- yellowgreen- green. Information from RasMol also reveals that HYD1 has 6 strands, 12 turns and 1 helix. It therefore appears that the beta sheet is the most stable part of the protein. However, for rms fluctuation of backbone residues each residue is coded according to the anisotropic temperature (beta) value stored in the pdb file [13]. Typically, this gives the measure of the mobility of a given residue's position. The beta value thus increases in order from blue-cyan- green- yellow- orange then to red as shown in Fig. 10.

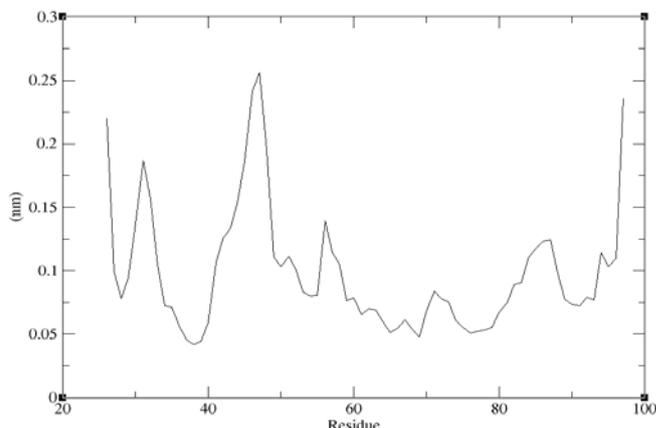


Figure 9: RMS fluctuations of backbone residues during simulation

Also, in plot for backbone residues, three residues that form two beta strands fluctuated the most as indicated by the red colour in Fig. 10.

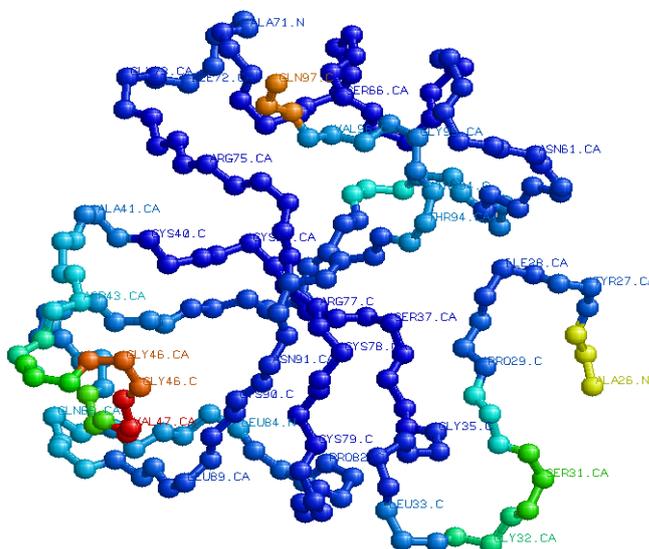


Figure 10: RMSF backbone residues as viewed in RasMol

These include Valine-47, Glycine-46 and Leucine-45. Fluctuations were also noticed with Alanine-26 and Glytamine-97 residues as they transition to form a fully solvated and dynamic HYD1 protein.

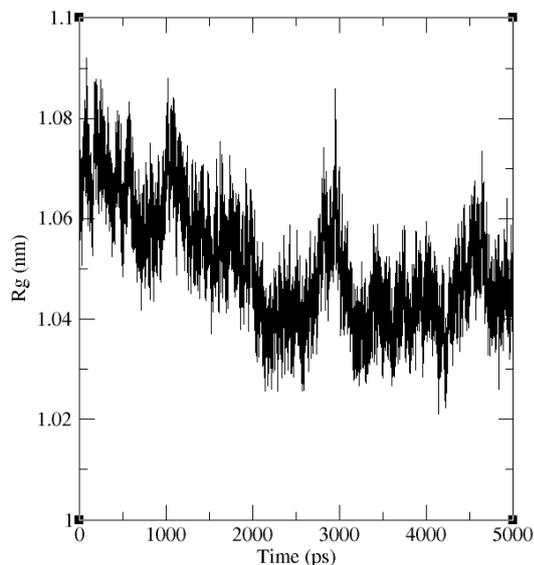


Figure 11: Radius of gyration of protein residues over 5ns

Upon compact assessment with `g_gyrate` which measures the radius of gyration, the structure of M4T_HYD1 seemed to collapse from 1.08 nm to 1.04 nm. Unusual peaks were observed at 1ns and 3ns as the protein tried to resolve into its original shape, before it finally attained a constant stable structure at 1.05nm towards the 5ns mark (Fig. 11). The collapse of the HYD1 protein was probably accompanied by changes to its secondary structure, as revealed by backbone fluctuations in the rmsf plot (Fig. 8, 10) and DSSP Table 2.

3.3.2 Trajectory Analysis

Trajectory data generated from the classical MD simulation step and converted to pdb format was viewed using PYMOL tool. The protein file thus contained 2,500 frames which caused the protein to rotate around all three axes. Six frames were thus selected at time intervals shown in Fig. 12 below.

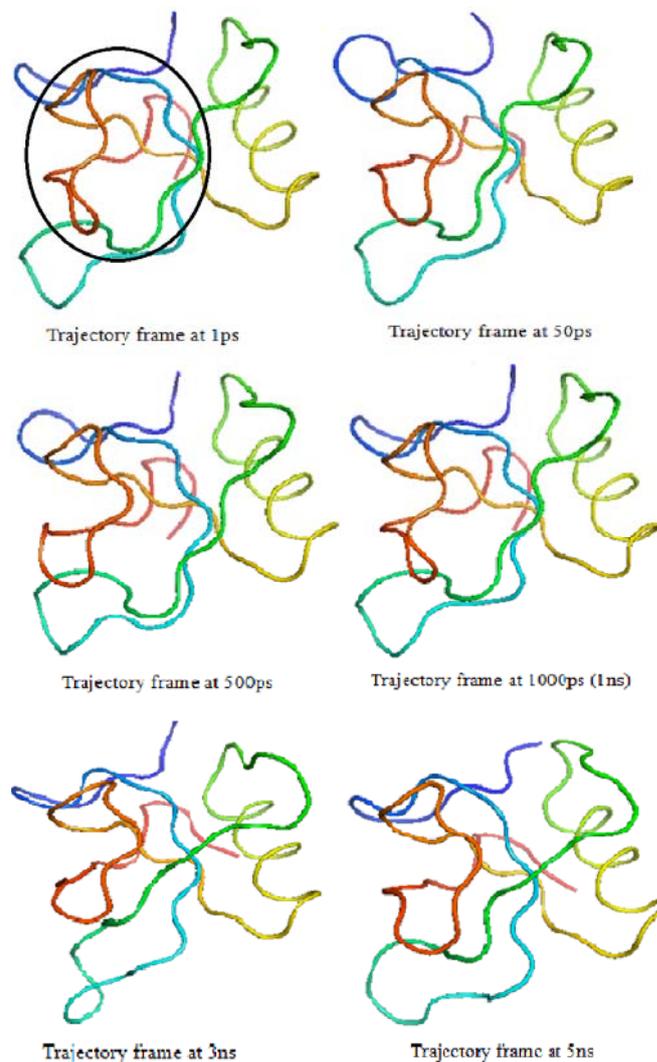


Figure 12: Snapshots of protein conformations in cartoon representation during the classical MD step

The circled area is a region of beta strands and turns where most fluctuations were observed (Fig. 7-10). Consistent to RMSD and Rg plots significant residue/ atomic fluctuations were observed at 1ns, and 3ns before HYD1 conformed to a stable dynamic structure between 4ns and 5ns.

Of interest, the simulated protein seem to have slowly exposed and partly exposed its binding cavity in the hydrophobic region of strands resulting in an increase in the surface accessible area for inhibitors to bind to the exposed pocket. Fig. 13 reveals a gradual reduction of the hydrophobic area with a sudden drop being observed towards 1ns (which is 1000ps) from 27 nm² to 21.6 nm². However, from 2ns till the end of the simulation M4T_HYD1 keeps an average hydrophobic area of 24 nm².

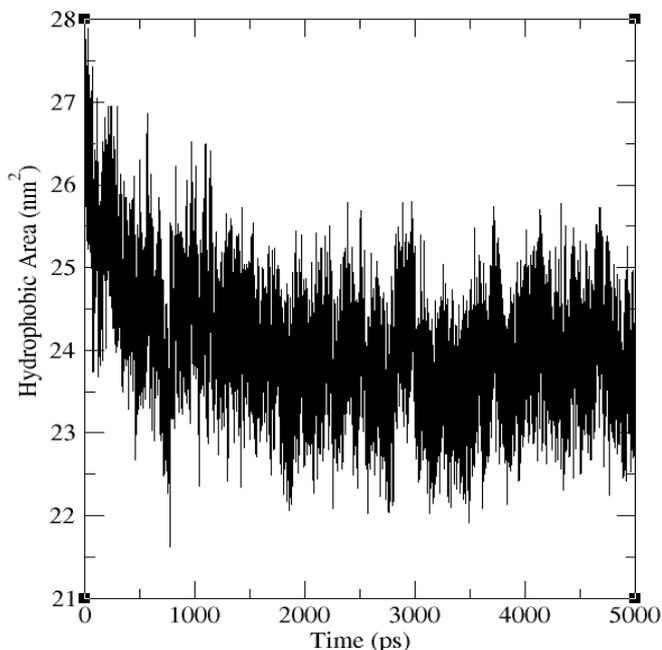


Figure 13: Hydrophobicity measurement as a function of time during simulation of M4T_HYD1

3.3.3 Overall Conformation Analysis

When the solvated protein now in dynamic form is analyzed against the original M4T_HYD1 protein, the conformation shows an RMSD value score of 0.291738 nm. Again the deviation lies in the acceptable range and can be credited to the secondary structure changes observed by DSSP (Table 2).

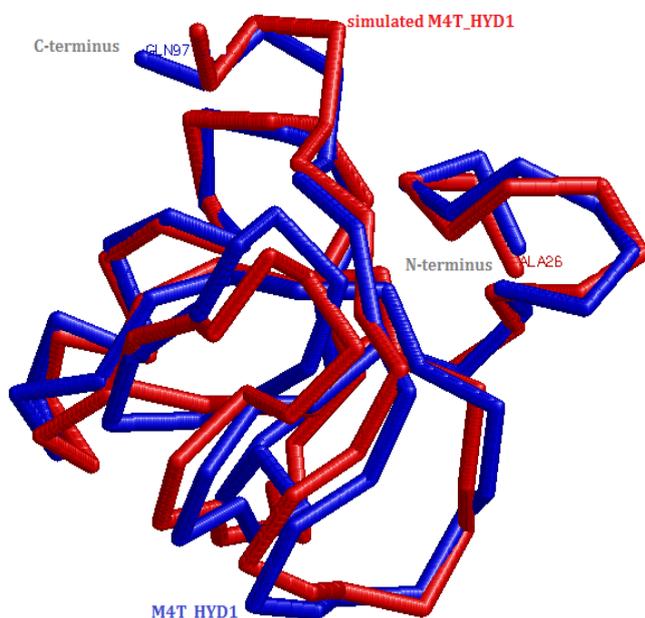


Figure 14: Superimposed backbone structures of original and simulated HYD1 protein

Comparison of these two structurally aligned proteins with DSSP reveals that as HYD1 conformed to a less compact structure, there were significant loss of beta sheets, and less of the helix to other secondary structure elements like bends, hydrogen bond turns and amino acids not in atom records:

Table 2: DSSP three state composition of HYD1

| Method | DSSP value (%) | |
|--------|----------------|----------------|
| | M4T_HYD1 | Simulated HYD1 |
| Helix | 15.3 | 13.9 |
| Sheet | 37.5 | 23.6 |
| Others | 47.2 | 62.5 |

However, the total content of sheet against that of helix in the secondary structure of the simulated protein still remains greater by approximately double that of helix. This can prove the theoretical fact that class hydrophobins attain a beta state in a water solvent.

4. Conclusion and Future Scope

MD simulations were used to compute atomic trajectories by solving equations of motion numerically using the GROMOS96 43a1 empirical force field. From a dynamic point of view the HYD1 protein proved to be thermally stable in the presence of an explicit water environment at 298K and constant pressure conditions. The protein also attained a compact structure which was credited to the low energy fluctuating beta strands. These findings are in agreement with previous theoretical and experimental evidence on dynamics of class I hydrophobins in related filamentous fungi. The authors therefore look forward to use the simulated M4T_HYD1 structure for docking studies in pursuit of a potential inhibitor molecule against *G. moniliformis*.

Acknowledgements

The authors would like to thank Dr. A. Govardhan, Director, School of Information Technology (SIT), Jawaharlal Nehru Technological University Hyderabad (JNTUH) for his support and encouragement in this research.

References

- [1] H. A. B. Wosten, "Hydrophobins, the fungal coat unraveled", *Biochimica et Biophysica Acta* 1469, pp.79-86, 2000.
- [2] J. G. Wessels, J. G., "Hydrophobins: proteins that change the nature of the fungal surface", *Adv Microb Physiol.* 38, pp.1-45, 1997.
- [3] M. Stubner, G. Lutterschmid, R. F. Vogel, and L. Niessen, "Heterologous expression of the hydrophobin FcHyd5p from *Fusarium culmorum* in *pastorisa* and evaluation of its surface activity and contribution to gushing of carbonated beverages", *Int. J. Food Microbiol.* 141, pp.110-115, 2010.
- [4] N. J. Talbot, M. J. Kershaw, E. G. Wakley, O. De Vries, and J. Wessels, "MPG1 Encodes a fungal hydrophobin involved in surface interactions during infection-related development of *Magnaporthe grisea*", *Plant Cell* 8, pp.985-999, 1996.
- [5] B. M. Linder, "Hydrophobins: proteins that self assemble at interfaces", *Curr. Opin. Colloid Interface Sci.* 14, pp.356-363, 2009.
- [6] X. Wang, H. P. Permentier, R. Rink, J. Kruijtzter, R. M. Liskamp, H. A. Wösten, B. Poolman, and G. T. Robillard, "Probing the self-assembly and the

- accompanying structural changes of hydrophobin SC3 on a hydrophobic surface by mass spectrometry”, *Biophys J* Vol.87, pp.1919-28, 2004.
- [7] J. Bayry, V. Aimanianda, I. Guijarro, M. Sunde, and J. P. Latge, “Hydrophobins—Unique Fungal Proteins”, *PLoS Pathog* Vol.8 Issue 5, e1002700, 2012.
- [8] S. Askolin, M. Linder, K. Scholtmeijer, M. Tenkanen, M. Penttilä, M. L. de Vocht, H. A. Wösten, “Interaction and comparison of a class I hydrophobin from *Schizophyllum commune* and class II hydrophobins from *Trichoderma reesei*”, *Biomacromolecules* Vol.7, pp.1295-301, 2006.
- [9] C. W. Bacon, I. E. Yates, W. P. Norred, and J. F. Leslie, “Production of fusaric acid by *Fusarium* species”, *Applied and Environmental Microbiology* Vol.62, pp.41, 39-41, 43, 1996.
- [10] C. Mawere., S. Kumar. G, and S. Muzondo, “Comparative Modelling, Quality Assessment and Validation of HYD1”, *International Journal of Innovative Research in Computer and Comm. Eng*, IV (5), pp. 4242-4250, 2014.
- [11] B. Hess and C. Kutzner and D. van der Spoel and E. Lindahl, “GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation “, *J. Chem. Theory Comput.* 4 pp. 435-447, 2008.
- [12] W. Kabsch, and C. Sander, "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers* 22 (12) pp. 2577-637, 1983.
- [13] R. Sayle and E. J. Milner-White. "RasMol: Bimolecular graphics for all", *Trends in Biochemical Sciences (TIBS)*, 20 (9), p. 374, 1995.
- [14] The PyMOL Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC. [Online]. Available: <http://www.pymol.org/>
- [15] R. C. Rizzo, “MD Simulation: Protein in Water”, para. 13, Aug. 15, 2013. [Online]. Available: http://ringo.ams.sunysb.edu/index.php/MD_Simulation:_Protein_in_Water_Pt2. [Accessed: Aug. 1, 2014].

Author Profile



Cephas Mawere is an M.Tech Student in the field of Bioinformatics currently underway at the SIT, JNTUH, India. He attained his B. Tech. Degree in Biotechnology from CUT, Zimbabwe in 2010. He is a Harare Institute of Technology staff development research fellow. His research interests are in the area of Microbiology, Molecular Dynamics Simulation, Computer- Aided Drug Design, High Performance Computing and Proteomics.



Suresh Kumar G is a Lecturer in Bioinformatics at School of Information Technology (SIT), Jawaharlal Nehru Technological University Hyderabad (JNTUH), Hyderabad, India. He has received Masters Degree (M.Tech) in Bioinformatics from University of Hyderabad, Hyderabad, India. His research interests include Computer- Aided Drug Design Molecular Modeling, Docking, 3D-QSAR studies, Genomics and Proteomics, Computational Systems Biology.