ISSN: 2755-0168

Journal of Biotechnology & Bioinformatics Research

SCIENTIFIC Research and Community

Research Article

Open Access

Computational Structural and Functional Characterization of *Azorhizobium caulinodans* NifH Protein and Associated Cluster Analysis

Pradeep Upadhyay[1], Vinay Kumar Singh[2*] and Rudra P Ojha[1]

[1]Dept of Botany, Nehru Gram Bharti Deemed University, Kotwa Jamunipur, Prayagraj, Uttar Pradesh

[2]Centre for Bioinformatics, School of Biotechnology, Institute of Science, Banaras Hindu University, Varanasi, Uttar Pradesh

ABSTRACT

Nitrogenase iron protein, a key enzyme in biological nitrogen fixation, is required for turning atmospheric nitrogen into a physiologically usable form. In the present work, nitrogenase iron protein (encoded by the nifH gene; WP_012169572.1) of *Azorhizobium caulinodans*, a nitrogen-fixing bacteria recognized for its symbiotic relationship with leguminous plants, was protein of interest. The nitrogenase iron proteins cluster with NifH were identified computationally and characterized using various tools. An in *silico* model prediction was also done for functional and structural classification including disulphide bonds analysis, molecular docking and other structural bioinformatics analysis. NifH or frxC family signatures (PS00746: ESGGPEPGVGCAG and PS00692: DVLGDVVCGGFAMP). The formation of disulphide bonds between cysteine residues Cys^{99} and Cys^{133}were recognized to play an important role in protein stability. Potential Fe-S cluster binding site within the nitrogenase iron amino acid sequences Cys^{99}, Cys^{133} and Gly^{135} were identified. In docking analysis the residues Gly^{14}-Thr^{20}, Asp^{41}, Lys^{43}, Asp^{45}, Arg^{48}, Asn^{186}, Arg^{188}, Val^{212}-Asp^{215}, Val^{218}, Gln^{219}, Glu^{222}, Gln^{237} and Tyr^{241} were interacted with Adenosine-5'-Diphosphate (DrugBank ID: DB16833/ Pharos: CHEMBL14830/ PubChem: 6022/ ChEMBL: CHEMBL14830/ ChEBI: CHEBI:16761). In this interaction, Gly^{14}, Ile^{15}, Gly^{16}, Lys^{17}, Ser^{18} and Lys^{43} residues were identified as active site residues. Furthermore, molecular dynamics simulation was performed to evaluate the protein's conformational stability and compactness with simulation time 300 ns. Also the impact of disulphide bond formation of NifH model on its overall structural integrity was also evaluated. Total 30 potential disulphide bonds were identified using BIOVIA Discovery Studio 2019. The *in silico* findings shed light on the stabilizing impact of disulphide bonds in the nitrogenase iron protein.

***Corresponding author**

Vinay Kumar Singh, Centre for Bioinformatics, School of Biotechnology, Institute of Science, Banaras Hindu University, Varanasi, 221005, Uttar Pradesh.

Received: December 02, 2023; **Accepted:** December 11, 2023; **Published:** January 05, 2024

Introduction

Azorhizobium a genus of nitrogen-fixing bacteria is well recognized for its symbiotic connection with leguminous plants. A fascinating enzyme called nitrogenase found among these bacteria plays a crucial role in biological nitrogen fixation, a process critical for maintaining soil fertility and supporting plant growth. The nitrogenase enzyme complex is made up of two parts: nitrogenase reductase, commonly known as the iron protein, and nitrogenase molybdenum-iron protein. In this paper, we will concentrate on the nitrogenase iron protein present in Azorhizobium and its critical role in converting atmospheric nitrogen into physiologically accessible forms [1].

The nitrogenase iron protein, which is expressed by the nifH gene, is an electron carrier and regulatory component of the nitrogenase enzyme complex. This metalloprotein comprises a distinct iron-sulfur cluster that acts as an electron transfer agent, transferring electrons from a reducing agent like ferredoxin to the nitrogenase molybdenum-iron protein. This electron transfer process is essential for the reduction of dinitrogen (N2) to ammonia (NH3), a process known as nitrogen fixation, which is the principal method by which atmospheric nitrogen is rendered accessible for both plants and other microbes [2].

Azorhizobium species have symbiotic relationships with leguminous plants, generating root nodules that fix atmospheric nitrogen. The nitrogenase iron protein works with the molybdenum-iron protein within these specialized structures to catalyze the complex and energy-demanding reduction of nitrogen gas into ammonia. This ammonia is subsequently digested by the host plant, providing a necessary source of nitrogen for important biological activities including protein synthesis and nucleic acid production [3].

Understanding the nitrogenase iron protein in Azorhizobium and its interactions with the host leguminous plants is of great significance in agricultural and ecological contexts. Harnessing the natural nitrogen-fixing capabilities of Azorhizobium and optimizing their symbiosis with legumes can lead to sustainable agricultural practices, reducing the reliance on synthetic nitrogen fertilizers and promoting environmental friendly farming methods. Moreover, unraveling the molecular mechanisms of nitrogen fixation in Azorhizobium contributes to our broader knowledge of nitrogen cycling in natural ecosystems and helps address the global challenge of sustainable food production and environmental conservation [4-6].

Based on above review, bioinformatics tool and techniques were used for classification of potential nif genes. Further total study was based on functional and structural characterization of NifH protein of *Azorhizobium caulinodans*.

Materials and Methods

Identification, Cluster and Network Analysis

The nif genes and their associated cluster was identified and retrieved from the genome of *Azorhizobium caulinodans* using *in silico* methods. The full genome of azorhizobium caulinodans was used for retrieval of total nif genes available in the genome. Using bioinformatics tools and sequence analysis algorithms, the specific regions containing the nif genes were identified and visualized from the extracted genomic sequence using GView Server. (https://server.gview.ca) and SnapGene software (http://www.snapgene.com) free trail [7]. STRING and STITCH Server were used for network analysis [8,9].

Physico-Chemical Characterization

The physiochemical properties were calculated by Prot Param server. ProtParam calculates many of the protein parameters including MW, Theoretical pI, Amino acid composition, Atomic composition, Extinction coefficient, estimated half-life, instability index, aliphatic index, and Grand average of hydropathicity [10].

Phylogenetic Tree Classification

16S RNA based phylogenetic tree was prepared as it was taken as one of the marker gene to consider *Azorhizobium caulinodans* as Bacteria. The phylogenetic tree of *Azorhizobium caulinodans* ORS571 was created using the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method of MEGA 7 software [11].

Structure Prediction and Evaluation

SWISS-MODEL is a widely used homology modeling server that employs a range of algorithms and techniques to develop three-dimensional models for protein sequences based on their homologous templates. Swiss-Model utilizes homology modeling to predict the 3D structure of a protein from its sequence. It identifies homologous templates, aligns sequences, and employs modeling algorithms like MODELLER for initial structure generation. The model is refined, assessed for quality, and presented for user analysis, with ongoing updates to algorithms for improved accuracy [12]. To check the quality and stability of NifH protein,web based server PROCHECK and ERRAT were used. The Ramachandran plot is a tool used to analyze the conformation of protein backbone dihedral angles (phi and psi angles) in a three-dimensional space. It provides information about the allowed and disallowed regions of the protein's conformational space [13,14]. A web tool Disulfide by Design 2.0 was used for disulfide engineering [http://cptweb.cpt.wayne.edu/DbD2/; 15]

Active Site Analysis, Molecular Docking Calculation and Interaction Studies

The Active site amino acids of protein Nif H were determined by conserved domain database for interactive domain family analysis CDD Database. The active site amino acids in the query were analyzed. Ligand (ADP) ADP is a purine ribonucleoside 5'-diphosphate was used for docking analysis. It has a role as a human metabolite and a fundamental metabolite [16].

Molecular Dynamics Simulation and Protein-Ligand Contacts Through Trajectory Analysis

DESRES software package DESMOND was used for molecular dynamics simulation. Desmond is free to use by faculty, researchers, and students at academic and non-profit institutions. Further trajectory analysis was done using Simulation Interactions Diagram [17].

Results and Discussion

Identification, Cluster Analysis and Sequence Retrieval

The genome of *Azorhizobium caulinodans* ORS 571 has a total size of 5,369,772 base pairs with GenBank ID: AP009384.1.Based on evaluation of *Azorhizobium caulinodans* genome (size: 5369772 bp), total 18 protein coding nif genes were successfully identified and based on literature survey functional details were also given in table below.(Table 1). STRING server was used for functional protein association networks preparation and all 17 were functionally associated except AZC_1035.

Table 1: Total 18 Nif Genes were Identified in *Azorhizobium caulinodans* Genome

S.N.	Protein ID	Old Locus Tag	Start	End	Preferred Name	Details
1.	WP_012169565.1	AZC_1034	1161899	1162195	fdxB	Ferredoxin.
2.	WP_012169566.1	AZC_1035	1162529	1162993	AZC_103	Protein of unknown function; DUF269.
3.	WP_012169567.1	AZC_1036	1163003	1163398	nifX	Nitrogenase molybdenum-iron NifX protein.
4.	WP_012169568.1	AZC_1037	1163424	1164791	nifN	Nitrogenase molybdenum-iron cofactor biosynthesis protein; Belongs to the NifD/NifK/NifE/NifN family.
5.	WP_012169569.1	AZC_1038	1164797	1166479	nifE	Nitrogenase MoFe cofactor biosynthesis NifE protein; Belongs to the NifD/NifK/NifE/NifN family.
6.	WP_012169570.1	AZC_1039	1166617	1168176	nifK	Nitrogenase molybdenum-iron protein beta chain; This molybdenum-iron protein is part of the nitrogenase complex that catalyzes the key enzymatic reactions in nitrogen fixation; Belongs to the NifD/NifK/NifE/NifN family.
7.	WP_012169571.1	AZC_1040	1168256	1169758	nifD	Nitrogenase molybdenum-iron protein alpha chain.

8.	WP_012169572.1	AZC_1041	1169826	1170716	nifH	Nitrogenase iron protein 2; The key enzymatic reactions in nitrogen fixation are catalyzed by the nitrogenase complex, which has 2 components: the iron protein and the molybdenum-iron protein.
9.	WP_043878932.1	AZC_1049	1178139	1179872	nif A	nif-specific transcriptional activator NifA
10.	WP_012171933.1	AZC_3410	3920412	3921611	nifU	Iron sulphur cluster assembly protein
11.	WP_244421877.1	AZC_3411	3938256	3938540	nif S	Mobilization of the sulfur required for metallocluster formation.
12.	WP_012171935.1	AZC_3412	3938614	3939705	nifT	Its precise role unclear but might assist in regulation or process
13.	WP_012171937.1	AZC_3414	3939853	3940071	nifB	Nitrogenase cofactor biosynthesis protein NifB
14.	WP_012171942.1	AZC_3420	3941201	3942760	nifZ	Accassary role in Maturation of nitrogenase MoFe protein
15.	WP_012171966.1	AZC_3443	3944696	3945001	nifH	Nitrogenase component 2 and also involved in nitrogen fixation
16.	WP_012171967.1	AZC_3444	3960536	3961252	nifQ	Donates molybdenum to NifEN/NifH for synthesis of MoFe cofactor
17.	WP_012171969.1	AZC_3446	3961980	3962309	nifW	Nitrogenase stabilizing protective protein
18.	WP_012173222.1	AZC_4703	5349726	5350286	nifU	Involved in iron-sulfur cluster assembly

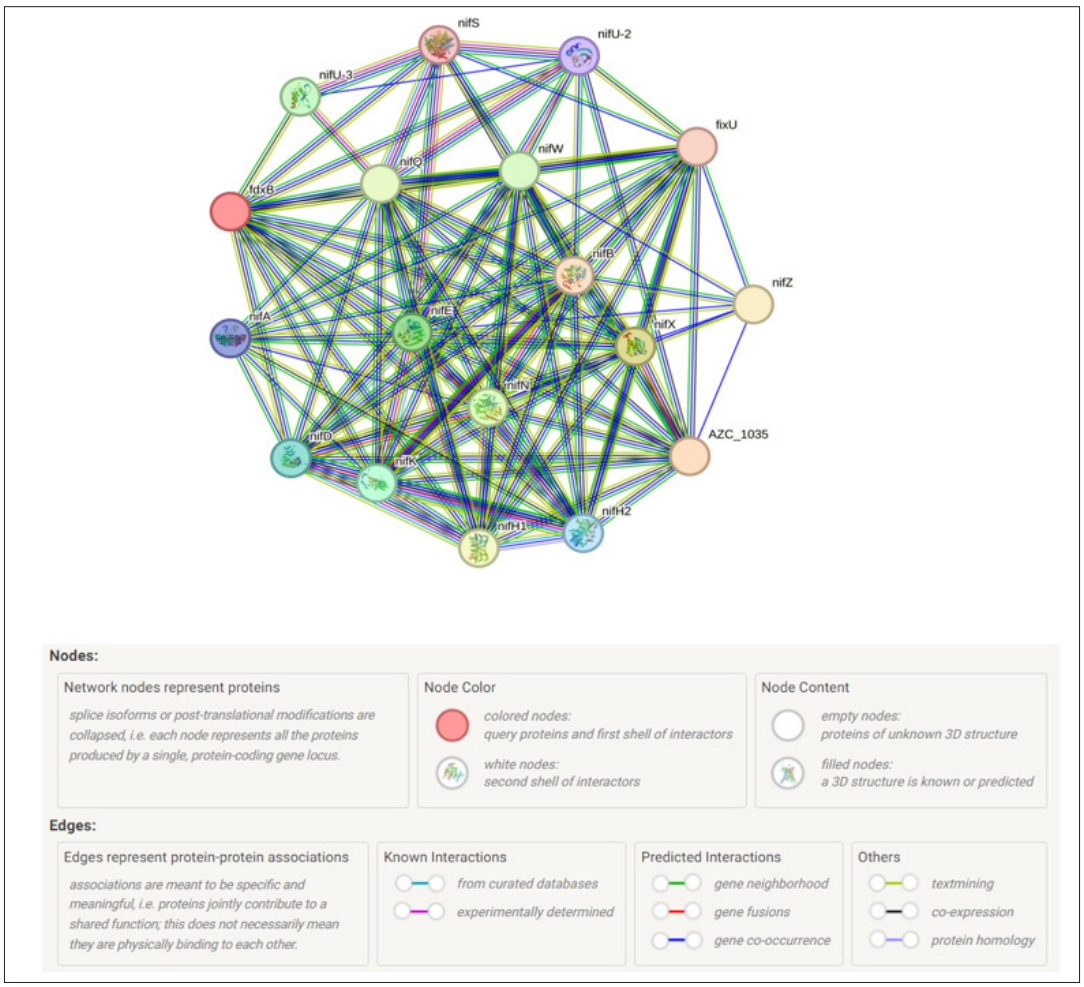


Figure 1: Protein-Protein Functional Association Network of Selected 18 Nif Proteins of *Azorhizobium Caulinodans*

Major stretch features were taken as input for cluster preparation using SnapGene software. Three major cluster were designed using SnapGene. Figure 1 illustrated all identified clusters (Figure 2).

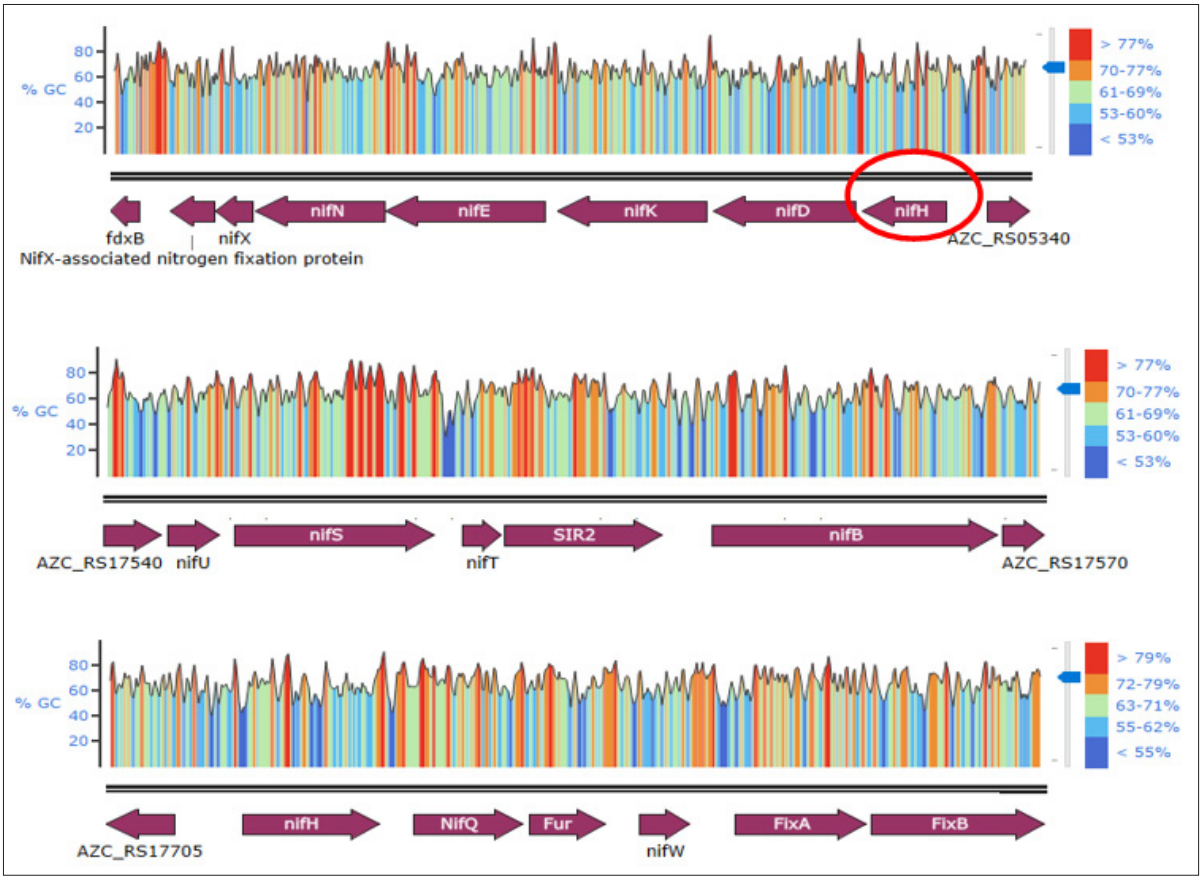


Figure 2: Three Clusters Mapped Using Snapgene Software (Free Trial)

Further, Cluster one was used for comprehensive analysis (Table 2). All genes of this cluster were functionally associated with each other (Figure 3). Gene ontology details were reported in Table 3. Protein-Chemical functional association of cluster1 proteins were given in figure 4.

Table 2: Comprehensive Analysis Details of Cluster 1

S.N.	Protein ID	Old Locus Tag	Preferred Name	Biological Process (Gene Ontology)	Details
1.	WP_012169565.1	AZC_1034	fdxB	GO:0009399 (Nitrogen fixation)	Ferredoxin.
2.	WP_012169566.1	AZC_1035	AZC_1035	-	Protein of unknown function; DUF269.
3.	WP_012169567.1	AZC_1036	nifX	GO:0009399 (Nitrogen fixation)	Nitrogenase molybdenum-iron NifX protein.
4.	WP_012169568.1	AZC_1037	nifN	GO:0009399 (Nitrogen fixation)	Nitrogenase molybdenum-iron cofactor biosynthesis protein; Belongs to the NifD/ NifK/NifE/NifN family.
5.	WP_012169569.1	AZC_1038	nifE	GO:0009399 (Nitrogen fixation)	Nitrogenase MoFe cofactor biosynthesis NifE protein; Belongs to the NifD/NifK/NifE/NifN family.
6.	WP_012169570.1	AZC_1039	nifK	GO:0009399 (Nitrogen fixation)	Nitrogenase molybdenum-iron protein beta chain; This molybdenum-iron protein is part of the nitrogenase complex that catalyzes the key enzymatic reactions in nitrogen fixation; Belongs to the NifD/NifK/NifE/NifN family.
7.	WP_012169571.1	AZC_1040	nifD	GO:0009399 (Nitrogen fixation)	Nitrogenase molybdenum-iron protein alpha chain.
8.	WP_012169572.1	AZC_1041	nifH	GO:0009399 (Nitrogen fixation)	Nitrogenase iron protein 2; The key enzymatic reactions in nitrogen fixation are catalyzed by the nitrogenase complex, which has 2 components: the iron protein and the molybdenum-iron protein.

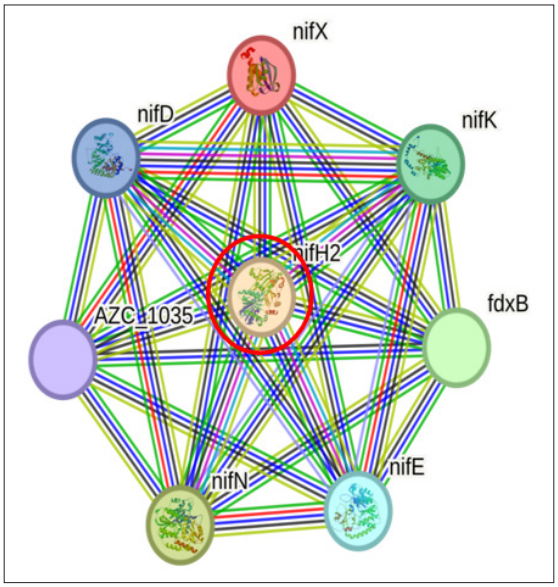


Figure 3: Protein-Protein Functional Association of Cluster1 Genes

Table 3: Interaction Analysis Details of Cluster 1

S.N.	Term ID	Functional Term description	Observed gene count	Matching proteins in your network (IDs)	Matching proteins in your network (labels)
1.	GO:0016163	Nitrogenase activity	5	438753. AZC_1037,438753. AZC_1038,438753. AZC_1039,438753. AZC_1040,438753. AZC_1041	nifN,nifE,nifK,nifD,nifH
2.	GO:0018697	Carbonyl sulfide Nitrogenase activity	3	438753. AZC_1039,438753. AZC_1040,438753. AZC_1041	nifK,nifD,nifH
3.	GO:0051540	Metal cluster binding	5	438753. AZC_1034,438753. AZC_1036,438753. AZC_1039,438753. AZC_1040,438753. AZC_1041	fdxB,nifX,nifK,nifD,nifH
4.	GO:0051536	Iron-sulfur cluster binding	4	438753. AZC_1034,438753. AZC_1039,438753. AZC_1040,438753. AZC_1041	fdxB,nifK,nifD,nifH

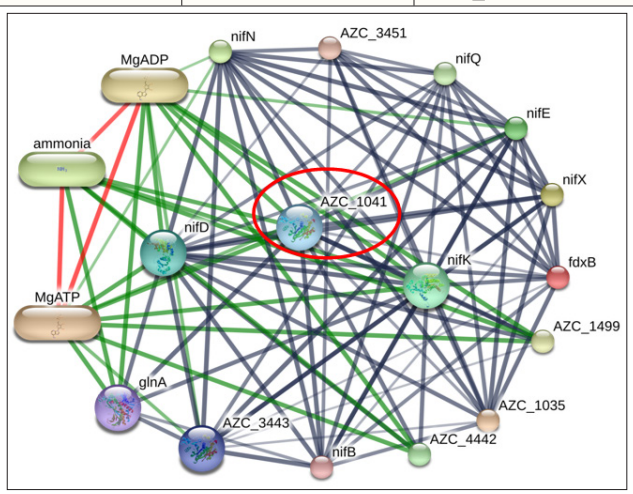


Figure 4: Protein-Chemical Functional Association of Cluster1 Genes

Phylogenetic Tree Classification

This tree represents the evolutionary history and relationships among the nucleotide sequences of this organism. The branch lengths in the tree correspond to the genetic distances between sequences, and the sum of branch length, which is reported as 0.05704444, represents the total evolutionary distance covered by the tree. To assess the reliability of the tree, a bootstrap test was conducted with 1000 replicates. The percentage values displayed next to the branches indicate the frequency at which the associated taxa clustered together in these replicates.

The analysis involved a dataset of 7 FASTA protein sequences, specifically focusing on *Azorhizobium caulinodans* ORS571. All the evolutionary analyses, including tree construction and distance calculations, were performed using MEGA7, a specialized software tool for molecular evolutionary analysis. Two major clusters were obtained (Figure 5).

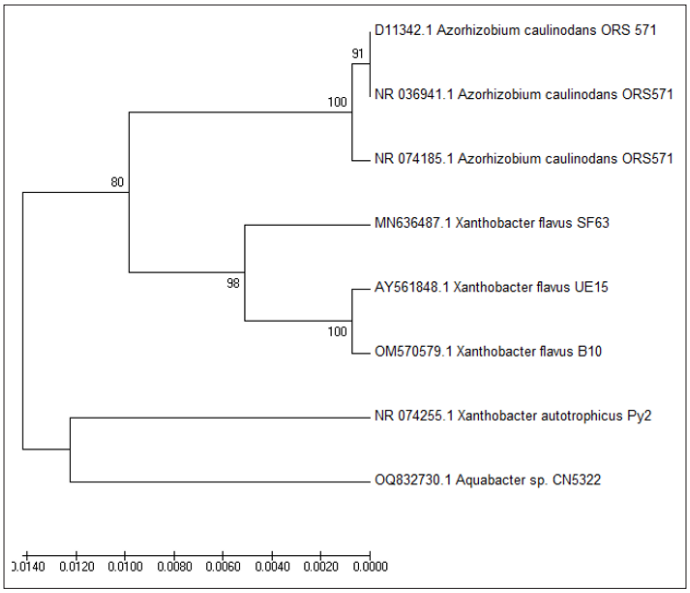


Figure 5: Phylogenetic Analysis of Selected *Azorhizobium Caulinodans* Nifh Gene with Other Closely Related Organisms

Physico-Chemical Characterization

The analyzed protein has 296 amino acids and a molecular weight of 31987.55 (Figure 6). Its theoretical isoelectric point (pI) is calculated to be 5.03. Looking at the amino acid composition, the protein consists of various amino acids in different quantities. Alanine (Ala) and glycine (Gly) each make up 10.1% of the mino acids present. Other amino acids like glutamate (Glu), leucine (Leu), and isoleucine (Ile) also contribute significantly to the composition.

The protein contains a total of 40 negatively charged residues, which include aspartate (Asp) and glutamate (Glu), and 28 positively charged residues, which consist of arginine (Arg) and lysine (Lys). This balance of charged residues plays a role in the protein's overall charge and interactions with other molecules. The protein was classified as stable based on its calculated instability index of 36.99. Additionally, the aliphatic index, which measures the relative volume occupied by aliphatic side chains, was determined to be 94.29. This index provides information about the protein's thermo stability.(4) Finally, the grand average of hydropathicity (GRAVY) calculated to be -0.100, indicating a slightly hydrophilic nature of the protein. The combination of these parameters and indices helps in understanding the physicochemical properties, stability, and composition of the analyzed protein.

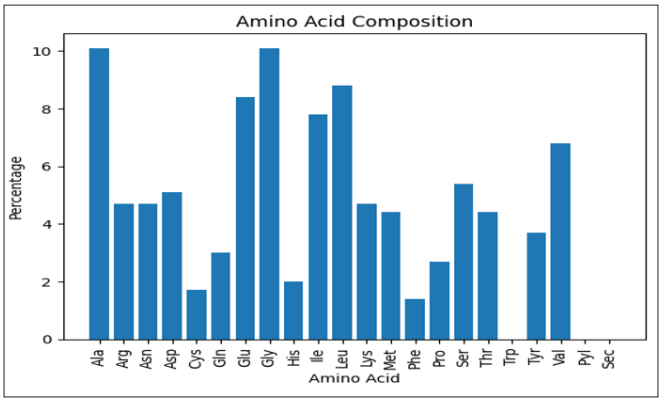


Figure 6: Graph Representing Amino Acids Composition Details in Percentage

Structure Prediction and Evaluation

Prediction of three dimensional structure of NifH protein was done using SWISS-MODEL (Figure 7). The overall quality was 95.849 using ERRAT (Figure 8). The statistics also mention the percentage of the protein for which the calculated error value falls below the 95% rejection limit. This indicates the proportion of the protein structure that was within an acceptable range of error. Good high-resolution structures typically have values around 95% or higher, indicating a high level of agreement with the reference structures (Figure 8).

Overall, the statistics from the ERRAT tool provide insights into the quality and stability of the Nif H protein structure. A higher percentage below the rejection limit indicates a more accurate and reliable structure, while lower percentages may suggest some deviations or discrepancies from the reference structures, particularly in lower resolution models.

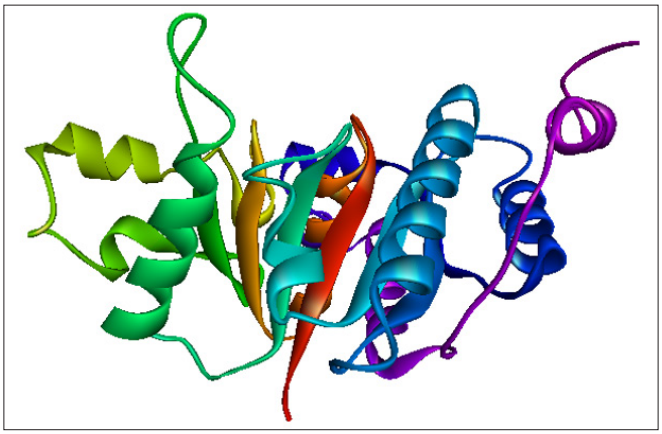


Figure 7: Visualization of Predicted Three Dimensional Structure of Nifh Protein

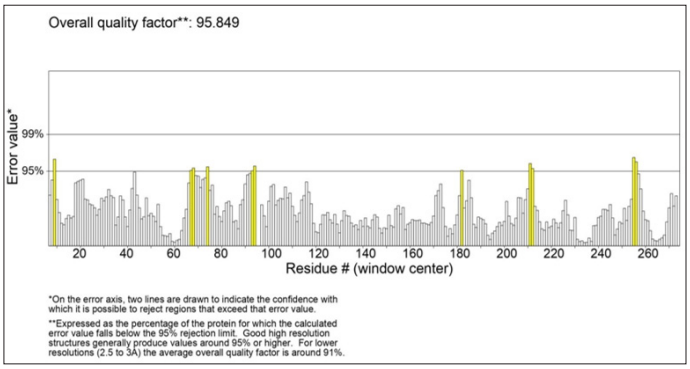


Figure 8: Structure Quality Analysis Using Errat Server

Further, PROCHECK (stereo chemical quality server) was used for analysis of predicted structure. According to the statistics, the majority of residues in the nifH protein (92.8%) were located in the most favored region of the Ramachandran plot (Figure 9). These regions correspond to the energetically most favorable conformations for the phi and psi angles of the protein backbone. It suggests that the protein adopts stable and preferred structural arrangements for the majority of its residues. A smaller portion of residues (6.8%) were found in the additional allowed regions [a, b, l, p] of the Ramachandran plot. These regions represent conformations that are also energetically favorable, although to a slightly lesser extent compared to the most favored regions. Residues in these regions may exhibit some flexibility or variations in their backbone conformation. Overall, the statistics suggest that the nif H protein predominantly adopts energetically favorable conformations, as indicated by the high percentage of residues in the most favored regions of the Ramachandran plot. This information is valuable in understanding the protein's structure and stability and provides insights into its functional properties.

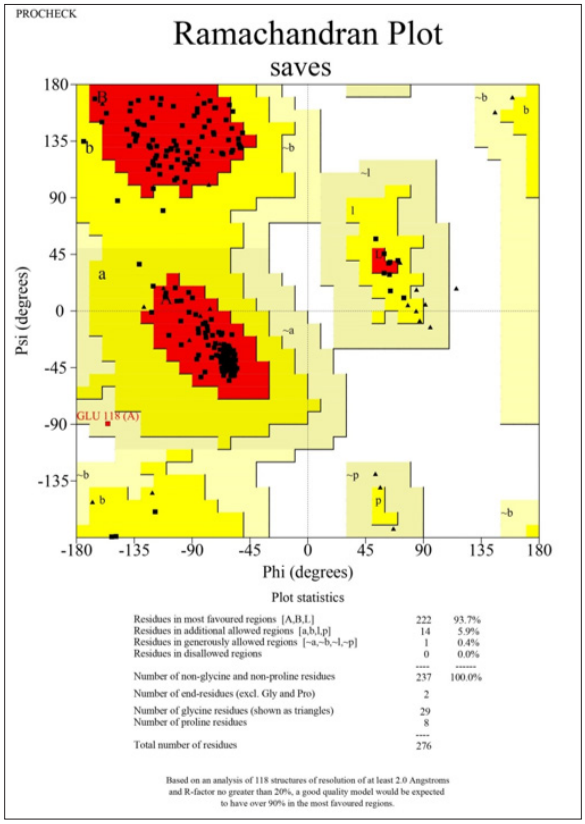


Figure 9: Ramachandran Plot Statistics Details Using Procheck Analysis

Bond and Conformational Stability Analysis

The Disulfide by Design server was used to analyze the protein NifH, and the results obtained are as follows. The Disulfide by Design Score, which indicates the quality of the predicted disulfide bonds, was determined to be 2.13. The analysis utilized the refined. pdb input file, specifically Model 1 of the protein structure. NifH consists of 276 complete residues, representing the total number of amino acids in the protein sequence [15]. The optimal Chi3 angle for the predicted disulfide bonds was identified as +97/-87, with a tolerance of 30. This angle is crucial for proper disulfide bond formation and plays a role in maintaining the protein's structural stability. The Ca-Cb-S angle, which characterizes the geometry of the Cα-Cβ-Sγ atoms involved in disulfide bond formation, was found to be 114.60 degrees, with a tolerance of 10. The analysis also involved checking for the presence of both INTER-chain and INTRA-chain disulfides.

The score is a numerical value that evaluates the quality of the protein structure, with higher scores indicating better agreement with experimental data. Energy, on the other hand, reflects the overall energy of the protein structure. Lower energy values suggest a more stable conformation. The number of bad contacts reveals the presence of unfavorable or steric clashes between atoms, which need to be minimized for a realistic and stable protein model. Thermal mobility provides insights into the flexibility and movement of atoms within the protein structure at a specific temperature, indicating the protein's dynamics and potential conformational changes (Figure 10). Sequence separation measures the distance or number of amino acid residues between specific residues or structural elements, aiding in the understanding of spatial arrangement and potential interactions.

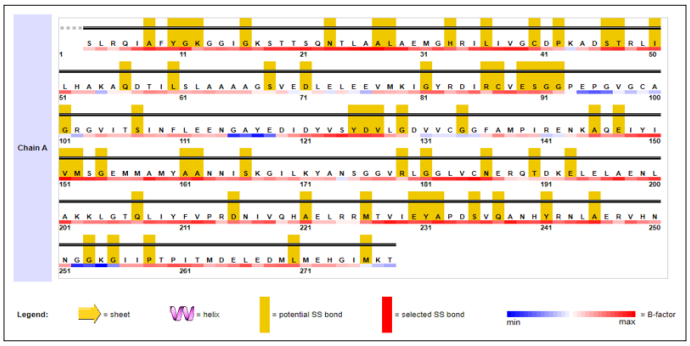


Figure 10: Disulphide Bond Analysis by Dbd2 (Disulphide by Design) Server

BIOVIA Discovery Studio 2019 was also used for protein stability analysis (Table 4). Evaluating protein stability involves diverse structural and energy criteria. Lower energies (<= 5.0) and absence of bad contacts signify stability, while high thermal mobility (>= 20.0) and close sequence separation (< 5) may suggest instability. Optimal residue depth (between >= 3 and < 8) and moderate volume changes (between -50.0 and 25.0) likely maintain stability, while certain secondary structure interactions, like coil-coil, promote stability, contrasting with destabilizing sheet-helix interactions. Consistency with minimum chi3 angles and maintaining proximity to structural minima often corresponds to stability, whereas deviations might indicate potential destabilization. These parameters collectively offer insights into protein stability, although individual cases may exhibit exceptions due to the intricate nature of protein structures. The table presents different properties of the protein and assigns them to different color-coded categories: Green (Stabilizing), Orange, and Red (Destabilizing).

The Energy property represents the overall energy of the protein structure. Energies less than or equal to 5.0 are categorized as Green, indicating a stabilizing effect. Energies between 5.0 and 10.0 fall into the Orange category, while energies greater than 10.0 are classified as Red, suggesting a destabilizing effect. Thermal Mobility measures the flexibility and movement of atoms within the protein structure. Proteins with a thermal mobility of 20.0 or higher are assigned to the Green category, indicating higher flexibility. Proteins with a thermal mobility between 5.0 and 20.0 fall into the Orange category, while those with a thermal mobility below 5.0 are categorized as Red, indicating lower flexibility.

Sequence Separation evaluates the distance between specific residues or structural elements within the protein. Proteins with a sequence separation of 25 or higher or involving interactions between different chains are considered Green, indicating a stabilizing effect. Proteins with a sequence separation of at least 5 fall into the Orange category, while those with a sequence separation less than 5 are classified as Red, suggesting a destabilizing effect.

Table 4: Conformational Stability Analysis Details

Property	Green (Stabilising)	Orange	Red (Destabilising)
Energy	<= 5.0	<= 10.0	> 10.0
Bad Contacts	0	1	> 1
Thermal Mobility	>= 20.0	>= 5.0, < 20.0	< 5.0
Sequence Separation	>= 25 or Inter Chain	>= 5	<5
Residue Depth	>= 3 and < 8	>=2 or >=8	< 2
Volume Change	>= -50.0 and < 25.0	>= -125 or >=25.0	< -125.0
Environment	Coil-Coil	Other combinations	Sheet-Helix
Chi3 (minima -87, +97)	<= 20.0 difference from minimum	<=30.0 difference from minimum	>30 from minimum

Residue Depth measures the burial or depth of specific amino acid residues within the protein structure. Proteins with a residue depth between 3 and 8 are assigned to the Green category, indicating a stabilizing effect. Proteins with a residue depth of 2 or 8, or greater, fall into the Orange category, while those with a residue depth below 2 are classified as Red, indicating a destabilizing effect.

Volume Change describes the structural rearrangements or conformational changes that occurred during the simulation. Proteins with a volume change between -50.0 and 25.0 are categorized as Green, indicating a stabilizing effect. Proteins with a volume change of -125 or greater, or at least 25.0, fall into the Orange category, while those with a volume change below -125.0 are classified as Red, suggesting a destabilizing effect.

Finally, Chi3 evaluates the chi-angle of the third side chain torsion angle of specific amino acid residues. Proteins with a difference from the minimum angle of 20.0 or less were categorized as Green, indicating a stabilizing effect. Proteins with a difference between 20.0 and 30.0 fall into the Orange category, while those with a difference greater than 30.0 classified as Red, suggesting a destabilizing effect.

Active Site Analysis, Molecular Docking Calculation and Interaction Studies

The residues Lys^{12}, Gly^{13}, Gly^{14}, Ile^{15}, Gly^{16}, Lys^{17}, Ser^{18}, Lys^{43} and Ala^{44} were identified as prominent site for docking calculation (Figure 11). The active site residues were identified based on structural alignment with template structure. For structural alignment crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii* (PDB ID: 1NIP) was used as reference model (Figure 12a). In docking analysis the residues Gly^{14}-Thr^{20}, Asp^{41}, Lys^{43}, Asp^{45}, Arg^{48}, Asn^{186}, Arg^{188}, Val^{212}-Asp^{215}, Val^{218}, Gln^{219}, Glu^{222}, Gln^{237} and Tyr^{241} were interacted with Adenosine-5'-Diphosphate (DrugBank ID: DB16833/ Pharos: CHEMBL14830/ PubChem: 6022/ ChEMBL: CHEMBL14830/ ChEBI: CHEBI:16761; Figure 12b, Figure 13-14).

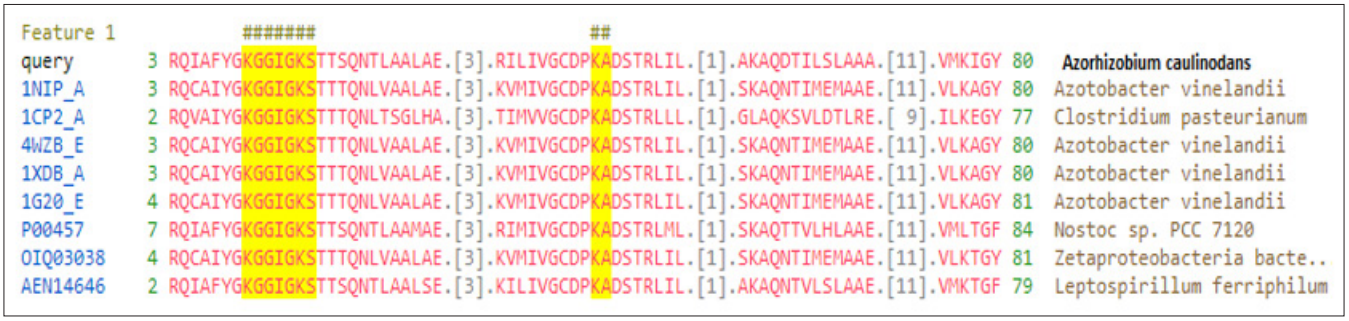


Figure 11: CDD BLAST Analysis for Active Site Residue Identification. Active Sites were Highlighted in Yellow Colour with #

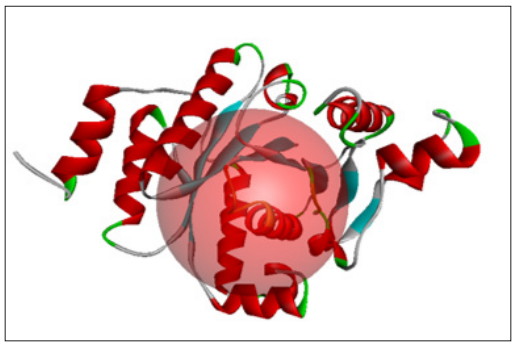


Figure 12a: Active Site Analysis

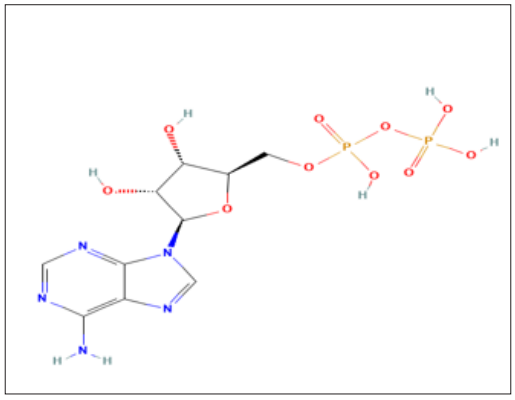


Figure 12b: ADP (CID: 6022)

It is an adenosine 5'-phosphate and a purine ribonucleoside 5'-diphosphate. The molecular formula, C10H15N5O10P2, illustrates the precise arrangement of atoms in the compound. Its molecular weight is determined to be 427.20 g/mol, indicating the mass of one mole of the substance. Identified by the IUPAC nomenclature as 5'-diphospho-adenosine, the compound is named systematically based on its structure and functional groups. The Isomeric SMILES notation, C1=NC(=C2C(=N1)N(C=N2)[C@H]3C@@HO)N, represents a specific linear notation of its structure, providing a visual representation of its interconnected atoms. This comprehensive table serves as a quick reference guide, outlining key characteristics and identifiers for the compound. The binding energy between ligand and protein was estimated good interaction score - 14.4355 Kcal/mol.

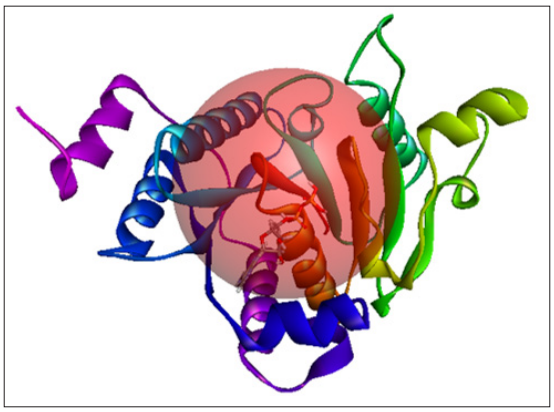


Figure 13: Docking Analysis of NifH with ADP(CID-6022) on Reported Active Site Residues

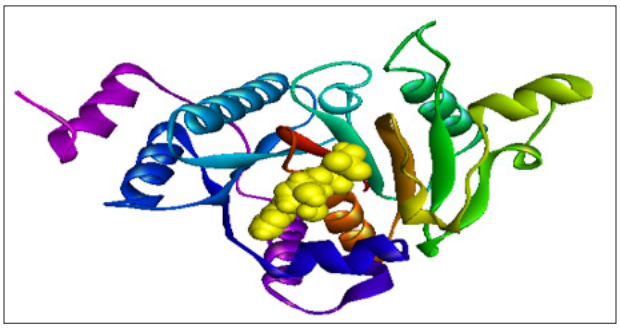


Figure 14: Representation of Nif H2 Protein and ADP Complex, Ligand Highlighted in Yellow Color

Molecular Dynamics Simulation and Protein-Ligand Contacts Analysis

Molecular dynamics simulation was performed using DESMOND software at 300ns. Protein RMSD and Ligand RMSD were found stable after 150 ns time scale.

Protein-Ligand RMSD

RMSD analysis (left Y-axis) indicating simulation equilibrated and no fluctuations towards the end of the simulation. Changes of the order of 1-3 Å were perfectly acceptable. Protein was undergoing a minimal conformational change during the simulation and RMSD values stabilized (Figure 15). Ligand RMSD (right Y-axis) results indicated how stable the ligand was with respect to the protein and its binding pocket (Figure 15).

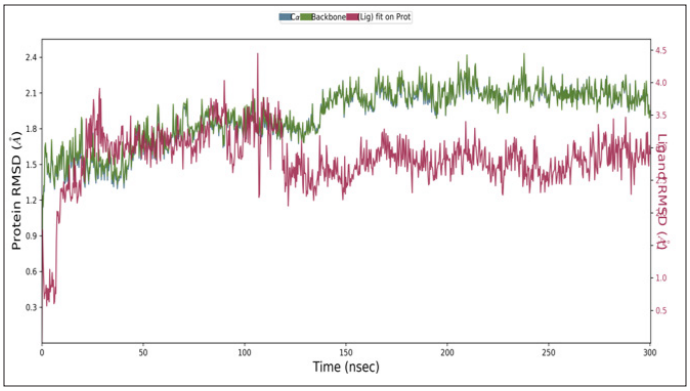


Figure 15: Protein-Ligand Root Mean Square Deviation of Simulated Protein Nif H2 in Complex with Adp During 300 ns Md Simulation

Protein RMSF

The Protein Root Mean Square Fluctuation is useful for characterization along the protein chain. Plot peaks indicated Gly96-98 residues of the protein showing fluctuation during simulation, no fluctuation observed in ligand binding residues. This indicated that protein active site residues were stability interacted with the ligand (Figure 16a).

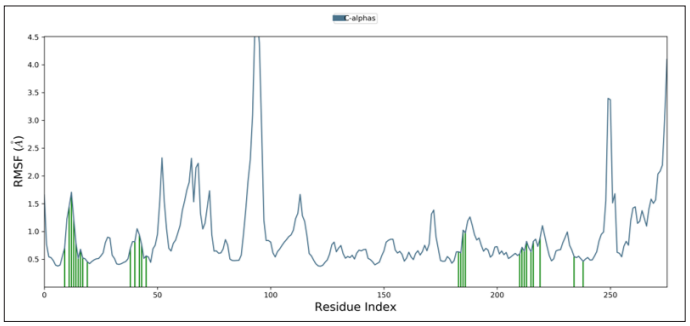
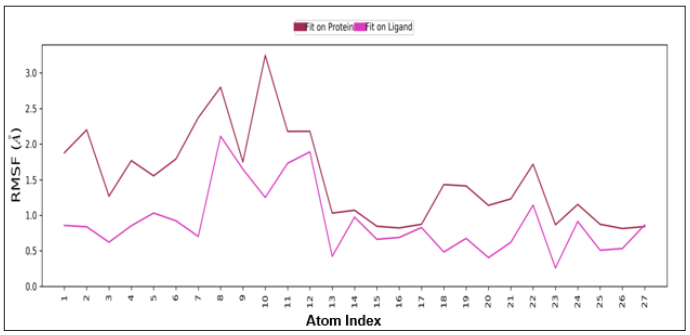


Figure 16a: Protein Root Mean Square Fluctuation with Ligand Contacts. Ligand Contacts' Highlighted in Green Color

Ligand RMSF

The Ligand Root Mean Square Fluctuation is useful for characterization changes in the ligand atom positions. The ligand RMSF indicated how ligand fragment interact with the protein and their entropic role in the binding event (Figure 16b).



Atom Index

Figure 16b: Ligand Root Mean Square Fluctuation

Protein-Ligand Contacts

Protein interaction with the ligand was monitored throughout the simulation. Protein-ligand interactions were categorized into four types: Hydrogen Bonds, Hydrophobic, Ionic and Water Bridges (Figure 17). The residues involved in hydrogen bonding represented in Figure 17b.

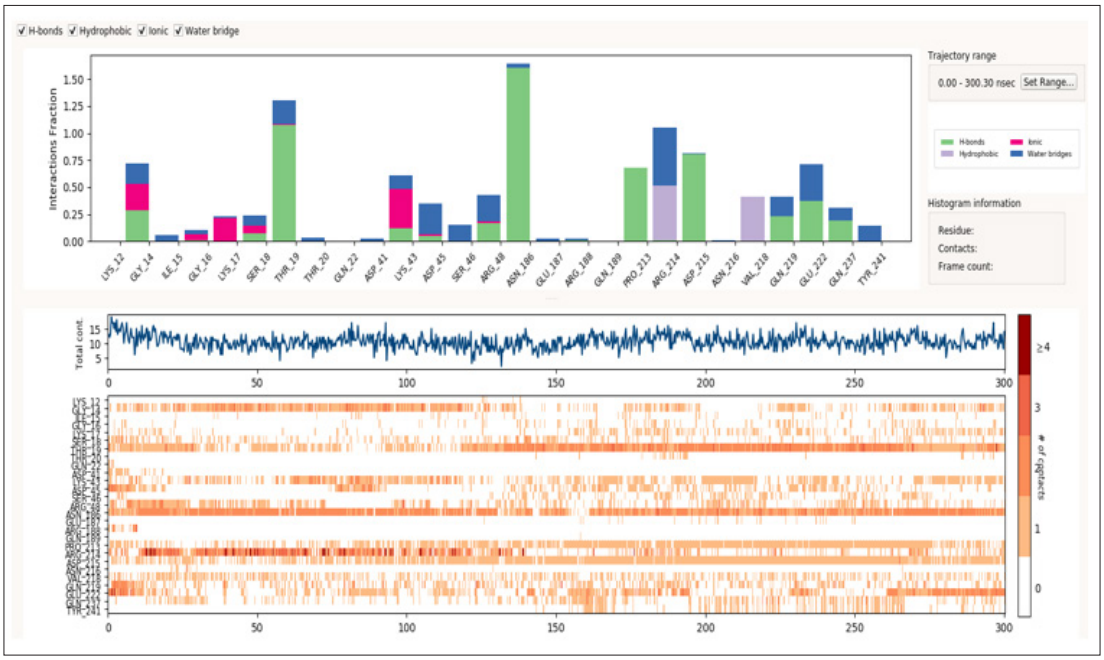


Figure 17a: Protein-Ligand Contacts Representing Hydrogen Bonds, Hydrophobic Interaction, Ionic Bond and Water Bridges

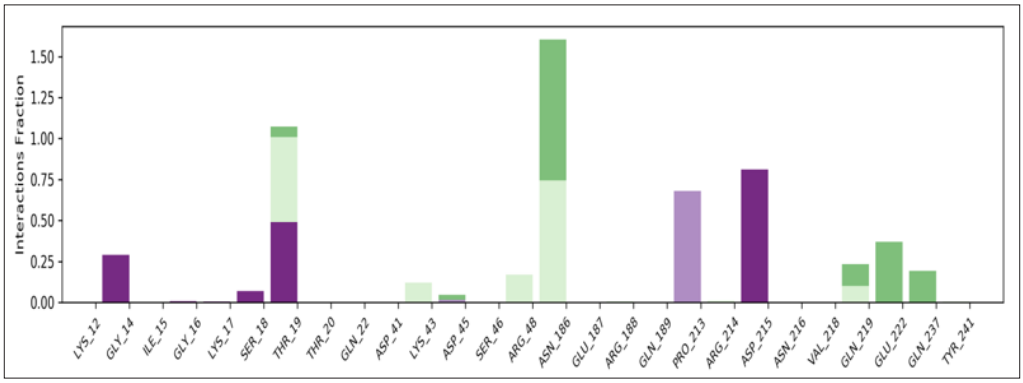


Figure 17b: Protein-Ligand Contacts Representing Hydrogen Bonds Details

Ligand-Protein Contacts

Ligand atom interactions with the protein active site residues were analyzed after simulation. Interactions that occur more than 30.0% of the simulation time in the selected trajectory (0.00 through 300.30 nsec), were shown in Figure 18.

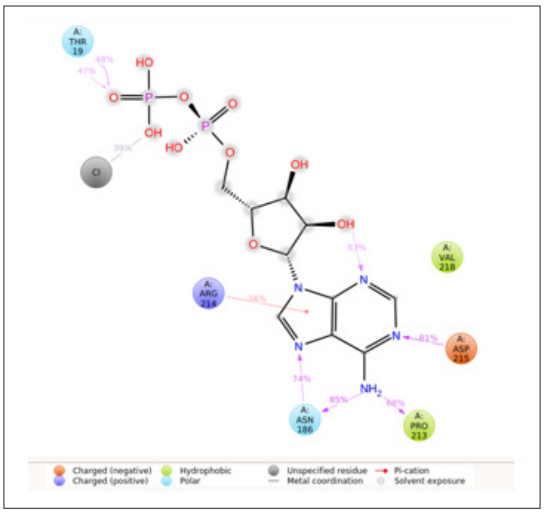


Figure 18: Protein-Ligand Contacts Details after 300 ns Simulation

Root mean square deviation, Radius of Gyration (rGyr), Intramolecular Hydrogen Bonds (intraHB), Molecular Surface Area (MolSA), Solvent Accessible Surface Area (SASA) and Polar Surface Area (PSA) were represented in Figure 19.

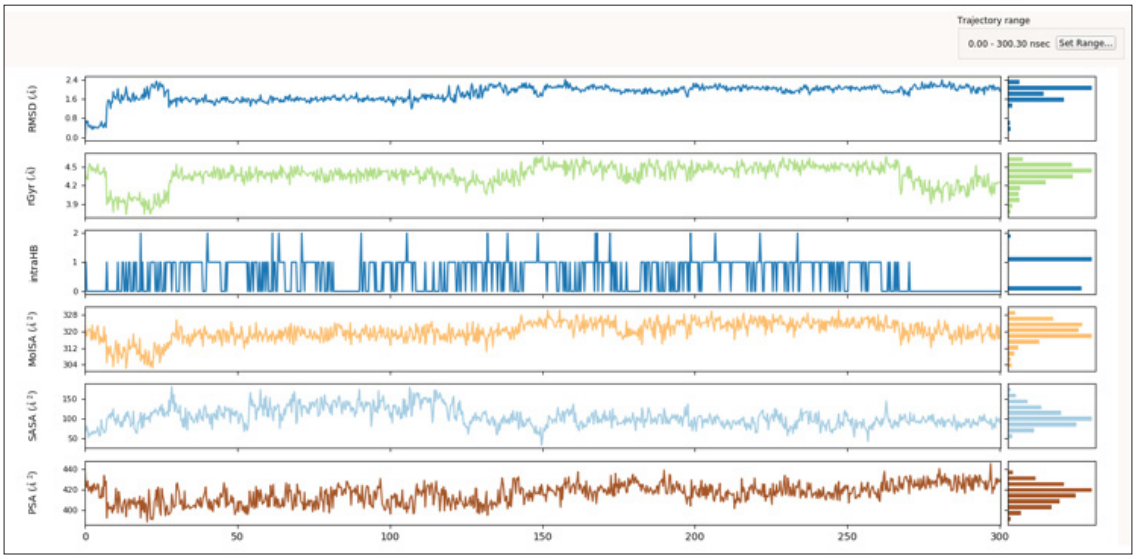


Figure 19: Ligand Rmsd, Radius Of Gyration, Intra Molecular Hydrogen Bonds, Molecular Surface Area, Solvent Accessible Surface Area, Polar Surface Area with 300 ns Time Scale

The ligand torsions plot (Figure 20) summarized the conformational evolution of every rotatable bond (RB) in the ligand throughout the simulation trajectory (0.00 through 300.30 nsec).

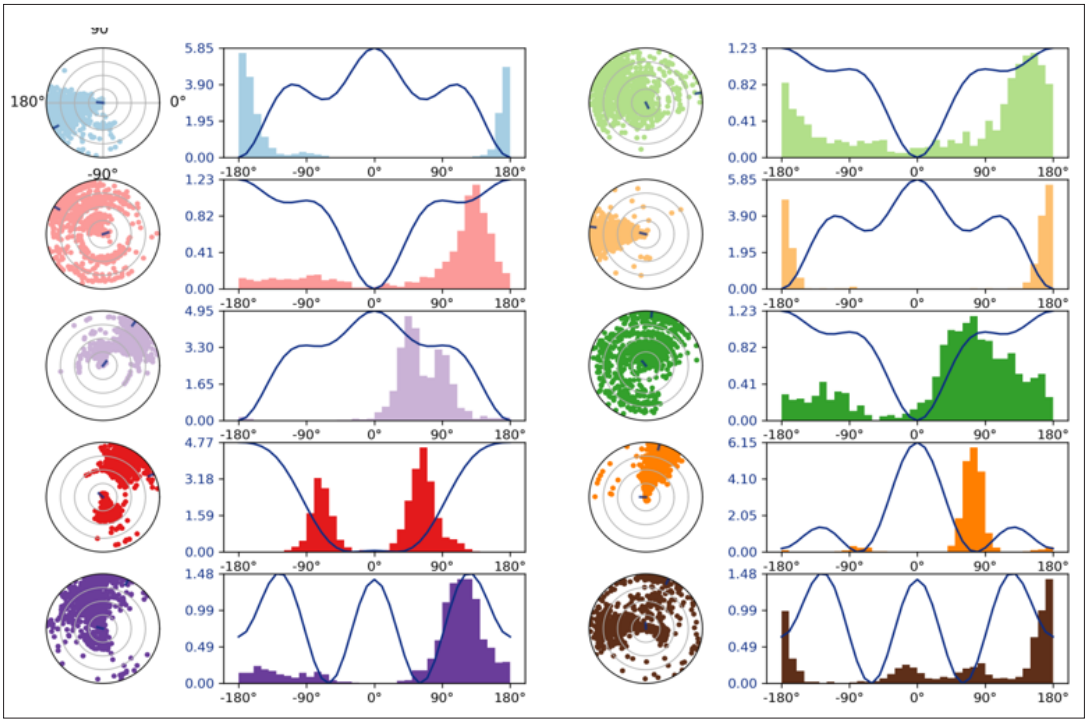


Figure 20: The Ligand Torsions Plot

Conclusion

Total 18 nif genes were sucessfully identified from *Azorhizobium caulinodans* genome and they are functionally associated with each other. *Azorhizobium caulinodans* NifH protein (WP_012169572.1: AZC_1041 from 1169826 to 1170716 was successfully characterized functionally associated with Nitrogen fixation (GO: 0009399). Total 93.7% residues in most favoured region and no residues in disallowed region indicated that the NifH model having good quality. *In silico* interaction of ADP (ADENOSINE-5'-DIPHOSPHATE) with nitrogenase iron protein of *Azorhizobium caulinodans* was done for functional characterization. In the conclusion, identified active site residues Gly14, Ile15, Gly16, Lys17, Ser18 and Lys43 of NifH were involved in interaction with ADP. To study of dynamic behavior of the protein-ligand complex, 300 ns time scale molecular dynamics simulation studies concluded that ATP interacts with the NifH and forming stable and compact complex over time.

Acknowledgement

Authors are thankful to Biotechnology Information System Network Distributed Information Sub-Centre, Centre for Bioinformatics, School of Biotechnology, Institute of Science, Banaras Hindu University and PARAM ShivaΨ, Supercomputing Facility IIT (BHU), Varanasi Uttar Pradesh 221005.

References

1. De Bruijn FJ, (1989) The unusual symbiosis between the diazotrophic stem-nodulating bacterium Azorhizobium caulinodans ORS571 and its host, the tropical legume Sesbania rostrata, In T. Kosuge and E. W. Nester (ed.) Plant-microbe interactions, molecular and genetic perspectives McGraw-Hill Publishing Co., New York, NY 3: 457-504.
2. Bergman B, EJ Carpenter (1991) Nitrogenase confined to randomly distributed trichomes in the marme cyanobacterium Trichode.smium thiebuutii J. Phycol 27: 158-165.
3. Soumare A, Diedhiou A G, Thuita M, Hafidi M, Ouhdouch Y,et al. (2020) Exploiting Biological Nitrogen Fixation: A Route Towards a Sustainable Agriculture. Plants (Basel, Switzerland) 9: 1011.
4. Elmerich C, Dreyfus BL, Reysset G, Aubert JP (1982) Genetic analysis of nitrogen fixation in a tropical fast-growing Rhizobium. The EMBO journal 1: 499-503.
5. Mahmud K, Makaju S, Ibrahim R, Missaoui A (2020) Current Progress in Nitrogen Fixing Plants and Microbiome Research. Plants (Basel, Switzerland) 9: 97. Rosenblueth M, Ormeño-Orrillo E, López-López A, Rogel MA, Reyes-Hernández BJ, et al. (2018) Nitrogen Fixation in Cereals. Frontiers in microbiology 9: 1794.
6. Hara S, Morikawa T, Wasai S, Kasahara Y, Koshiba T, et al. (2019) Identification of Nitrogen-Fixing Bradyrhizobium Associated With Roots of Field-Grown Sorghum by Metagenome and Proteome Analyses. Frontiers in microbiology 10: 40.
7. Stothard P, Grant J R, Van Domselaar G (2019) Visualizing and comparing circular genomes using the CGView family of tools. Briefings in bioinformatics 20: 1576-1582.
8. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, (2015) STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic acids research 28: 447-452.
9. Kuhn M, von Mering C, Campillos M, Jensen LJ, Bork P (2008) STITCH: interaction networks of chemicals and proteins, Nucleic acids research 36:684-688.
10. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel R D, Bairoch A (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic acids research, 31: 3784-3788.
11. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular

Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular biology and evolution, 33: 1870-1874.

12. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. Nucleic acids research, 31: 3381-3385.
13. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst 26: 283-291.
14. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein science: a publication of the Protein Society 2: 1511-1519.
15. Craig DB, Dombkowski,A.A (2013) Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. BMC bioinformatics 14: 346.
16. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, et al. (2005) CDD: a Conserved Domain Database for protein classification. Nucleic acids research 1: 192-196.
17. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, et al. (2022) Database resources of the national center for biotechnology information. Nucleic acids research 50: 20-26.