Int J Pharm Bio Sci 2012 July; 3(3): B 22 - 32

Research Article Bioinformatics

International Journal of Pharma and Bio Sciences

ISSN
0975-6299

IN SILICO STUDIES OF NifH PROTEIN STRUCTURE AND ITS POST-TRANSLATIONAL MODIFICATION IN *Bradyrhizobium* sp. ORS278.

SUBARNA THAKUR[1], ASIM K BOTHRA[2] AND ARNAB SEN[1]*

[1]*NBU Bioinformatics Facility, Department of Botany, University of North Bengal, Siliguri-734013, India*
[2]*Bioinformatics Chemoinformatics Laboratory, Department of Chemistry, Raiganj University College, Raiganj, India*

ABSTRACT

Bradyrhizobium is a gram negative, symbiotic nitrogen fixer that can form nodules in the roots of leguminous plants. The enzyme nitrogenase is the utmost important component of nitrogen fixing machinery, comprising two proteins, a molydenum-iron protein (MoFe protein) and an iron protein (Fe protein). Since, there is no comprehensive data on the tertiary structure of the nitrogenase iron protein (NifH) of *Bradyrhizobium* sp. ORS278 in the public domain, we decided to construct a three dimensional structure of the protein based on homology modeling technique. The model was constructed using the crystal structure of the NifH protein (1NIP) from *Azotobacter vinelandii* as a template. The functionally important regions of the protein and position of 4Fe-4S cluster in the protein were investigated. The 3D model was further utilized to highlight the importance of covalent modification site (Arg-102) in post-translational regulation of the nitrogenase enzyme through *in-silico* docking experiment.

KEYWORDS : Bradyrhizobium, symbiotic, homology modeling, template, docking.

ARNAB SEN
NBU Bioinformatics Facility, Department of Botany University of North Bengal,
Siliguri 734013, India

.

INTRODUCTION

Biological Nitrogen fixation is reaction catalyzed by a number of enzymes and carried out by a number of free-living and symbiotic bacteria by which atmospheric inert nitrogen (N_2) is reduced to ammonia. Nitrogen in the fixed form is of utmost importance for agriculture and industry. The enzyme nitrogenase, catalyzing the process of nitrogen fixation consists of nitrogenase iron protein and nitrogenase molybdo-ferro protein working in collaboration to fix atmospheric nitrogen[1, 2]. The nitrogenase iron protein (NifH) is a highly conserved dimer consisting of two equal subunits. This protein has a single 4Fe:4S cluster that helps in the transfer of electrons to the molybdo-ferro protein involving a series of associations and dissociations [3]. Besides, it also participates in the FeMo-cofactor biosynthesis and maturation of the molybdo-ferro protein [4].

The leguminous plants are associated with a number of nitrogen fixing soil bacteria collectively known as rhizobia. The rhizobia are housed in the different subclasses of Proteobacteria and consist of *Rhizobium, Bradyrhizobium, Mesorhizobium, Sinorhizobium, Azorhizobium* etc. *Bradyrhizobium* sp. ORS278 is a photosynthetic, nitrogen fixing bacteria entering into symbiotic association with *Aeschynomene sensitiva* by the formation of stem nodules[5]. The nitrogen fixing machinery of *Bradyrhizobium sp*. ORS278 comprises a nitrogenase enzyme system typical of diazotrophs. The nitrogenase iron protein (NifH) is an important component of this nitrogenase system. Although a lot of information has been there in the public domain regarding the crystallographic structure of nitrogenase iron protein of *Azotobacter vinelandii* and *Clostridium pasteurianum*, no crystal structure data of nitrogenase iron protein of *Bradyrhizobium* sp. ORS278 is currently available in Protein Data Bank. We therefore decided to construct a theoretical three dimensional model of the protein to garner insight into its structural and functional properties. The deficit of information of its three dimensional structure weakens a comprehensive knowledge on its conformational properties. Since, it is difficult to obtain accurate experimental structures from methods such as X-ray crystallography and NMR for the protein of interest; homology modeling can provide useful structural models for understanding a protein's function as well as conformational changes during the post-translational regulation mechanism.

Post translational modification is an important factor in regulation of nitrogenase activity. As seen in many photosynthetic nitrogen-fixing bacteria like *Rhodosprillium* and *Azospirillum* this involves reversible ADP-ribosylation of dinitrogenase reductase at Arginine residue which remains in close proximity of the Fe4S4 cluster in presence of the enzyme dinitrogenase reductase ADP-ribosyltransferase (DRAT)[6]. The ADP-ribosylation prevents the productive association of dinitrogenase reductase with dinitrogenase, such that neither electron transfer nor ATP hydrolysis can occur. In these circumstances, the nitrogenase activity is said to be switched off. The enzyme is reactivated upon removal of ADP-ribose by dinitrogenase reductase-activating glycohydrolase (DRAG). The modification site i.e Gly-Arg (ADR-ribose)-Gly-Val-Ile-Thr is highly conserved in nitrogenase iron protein of other nitrogen-fixing organisms. Here we have tried to look into the role of the Arg102 in ADP-ribosylation of dinitrogenase reductase and subsequent regulation in *Bradyrhizobium* sp. ORS 278 through *in-silico* docking experiment.

MATERIALS AND METHODS

i) Model Building:

The amino-acid sequence of the nitrogenase iron protein (NifH) from

Bradyrhizobium sp. ORS278 (Accession number NC_009445) was retrieved from the sequence database of Integrated Microbial Genomes (img.jgi.doe.gov)[7]. The amino acid sequence of the nitrogenase protein of *Bradyrhizobium* sp. ORS278 was subjected to PSI-BLAST (position specific iterative blast) [8] for finding out the suitable template. The search was restricted to PDB structures only. Features like quality of the template structure, environmental likeness and phylogenetic similarity were considered while choosing the desired template. Alignment of the template and target sequences was carried out with the help of CLUSTAL W [9]. A three dimensional all-atom model was constructed using Modeller 9v7 [10] from the sequence alignment between the template sequence and the target sequence with the parameters of energy minimization value.

Secondary structure of the target protein was predicted using the web interfaces of HNN (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pi)[11] and JUFO (http://www.meilerlab.org/view.php)[12].

ii) Model refinement and evaluation:

This model was then subjected to constraint energy minimization applied to all the protein atoms, using the steepest descent and conjugate gradient process to eradicate existing bad sectors between the protein atoms. All these computational analyses were performed using the Swiss-PDB Viewer package (http://www.expasy.ch/spdbv) [13] *in vacuo* with the GROMOS96 43B1 parameters set. Hydrogen bonds were not considered for the final model. The model was further analyzed with a series of checks for ensuring its internal stability and reliability. The Auto Deposition Input Tool (ADIT) (http://deposit.pdb.org/validate) was used to construct Ramachandran Plot to look into the possible conformations of φ and ψ angles for our modelled protein. The quality of the model was checked using ProSA [14] (https://prosa.services.came.sbg.ac.at/prosa.php) ERRAT analysis [15](http://nihserver.mbi.ucla.edu/ERRATv2/) and VERIFY3D [16] (http://nihserver.mbi.ucla.edu/Verify_3D/). Our refined model was then submitted to ProFunc server (http://www.ebi.ac.uk/thronton-srv/databases/ProFunc) [17] to provide functional insight of the protein and identify functionally important regions**.**

iii) Conformational dynamics study:

Conformational dynamics study of our modelled protein was done with the aid of Normal Mode Analysis (NMA).NMA provides an appropriate time-independent approach to study the dynamics of relatively large proteins like ours. Results from the NMAs of the protein was obtained from an online sources like the WEBnm@ [18] (http://www.bioinfo.no/tools/normalmodes).This servers provides online calculation of normal modes at the residue-level .The overall dynamics of the protein can be described by a superposition of a number of linearly independent normal modes.

iv) Docking studies:

The three dimensional co-ordinates of the ADP-Ribose molecule were retrieved from NCBI-PubChem Compound database (http://pubchem.ncbi.nlm.nih.gov/). The molecular docking simulation was carried out using the AutoDock 4.0 suite [19]. In this docking simulation, we used semi-flexible docking protocols in which the target protein NifH was kept as rigid. The ligand being docked was kept flexible, in order to explore an arbitrary number of torsional degrees of freedom. The Autodock tool program was used to prepare, run, and analyze the docking simulation. All the rotable dihedrals in the ligand were assigned with the help of auxiliary program Auto-Tors and were allowed to rotate freely. The AutoGrid 4.0 Program, supplied with AutoDock, was used to produce a grid map for each atom type present

in the ligand. This grid must surround the region of interest in the macromolecule. In the present study, the binding site was selected based on the amino acid residues which are involved in post-translational modification. Therefore, the grid was centered in the region Gly-Arg-Gly-Val-Ile-Thr. The grid box was set at 82, 60, and 50 A°(x,y, and z). The Lamarckian genetic algorithm was chosen to search for the best conformers. The docking parameters were as follows: 50 runs, population size 150, 25,000,000 energy evaluations, 27,000 generations and RMSD-tolerance of 2.0 Å. The docking results were analyzed in Autodock tools.

RESULTS AND DISCUSSION

For prediction of the three dimensional structure of a protein, at first a suitable template must be identified and compared with the target protein. The target protein and template protein may be structurally similar if both the sequences are alike. Position specific iterative blast (PSI-BLAST) results of the nitrogenase iron protein (NifH) of *Bradyrhizobium* sp. ORS278 revealed 73% identity with the X-ray crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii* (PDB ID- 1NIP) [20]. This protein is a homodimer consisting of 289 residues in each chain. The alignment between the template and target sequence is shown in Figure 1.

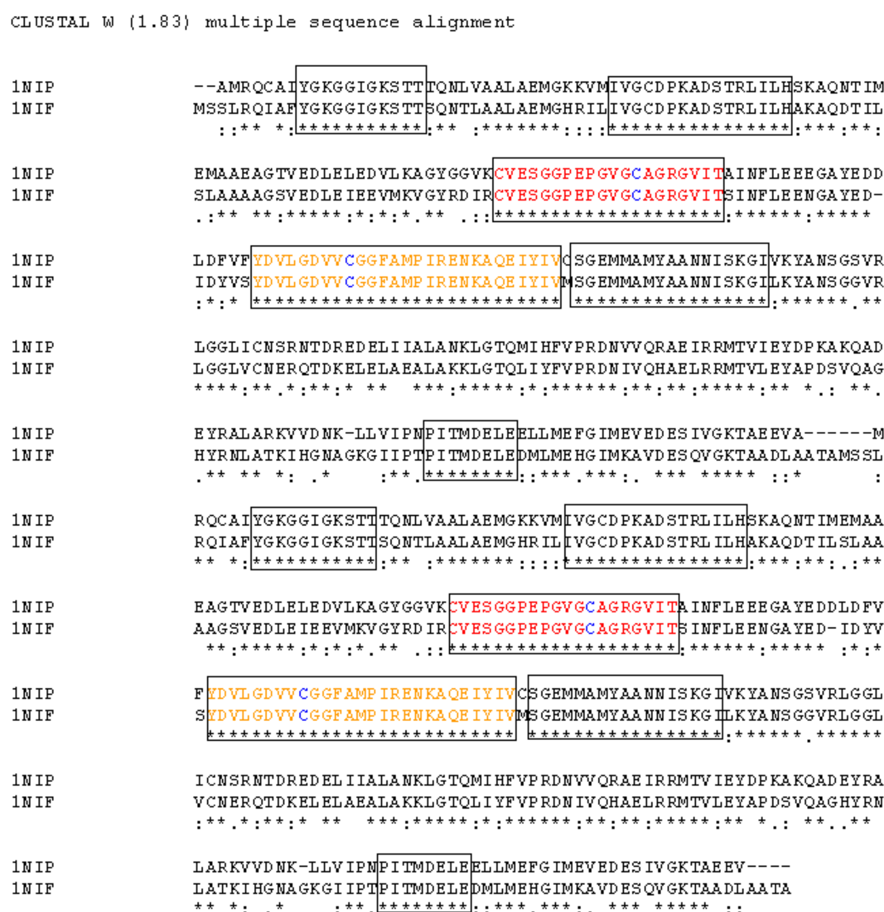

Figure 1
Alignment of target protein and template protein, with conserved regions shown in boxes. The regions highlighted in red shows the conserved regions around the metal binding sites

The major conserved regions in the NifH protein are amino acid residues: 10-20, 37-52, 87-106 (with a metal binding site), 125-151(with another metal binding site), 153-169,260-267,305-315,332-347,382-401 (with a metal binding site), 420-446 (with a metal binding site), 448-464, 555-562. The secondary structure prediction servers revealed that the modelled structure has alpha helix which constitutes 256 residues (43.39%), extended strand with 100 residues (16.95%) and the random coil with 234 residues (39.66%) of the protein. All these are interspersed between the protein residues throughout the structure.

The modelled structure of the nitrogenase iron protein revealed its existence as homodimer with 295 amino acid residues in each chain (Figure 2) connected at one surface by the 4Fe-4S cluster. The model has 4428 number of atoms and 4488 bonds. The exact molecular weight of the target protein is 63,575.3 Da and the molecular volume 37,826. The molecular formula of the protein is $C_{2754}H_{4436}N_{762}O_{868}S_{40}Fe4$. The net partial charge is -22. Hydrophilic amino acids occupied the surface whereas the hydrophobic amino acids occupied the core. The core functional region is represented by the 4Fe-4S cluster. The cluster is connected by the thiol groups of cysteine residues 99 and 133 in each sub-unit. The metal binding sites of the protein exhibit highly conserved nature. The nucleotide binding section of the nitrogenase iron protein comprises the residues in the Walker A motif, Switch 1 and Switch 2 regions. The Walker A motif lies between 11-18 residues, Switch I region between 38-44 residues and Switch II region between 124-128 residues in the modelled protein. These regions are associated by polar contacts and indispensable for maintaining the stability of the protein

.

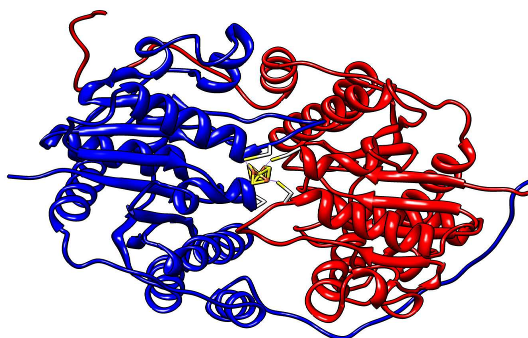

Figure2

3D structure of the nitrogenase iron protein from Bradyrhizobium ORS278. The structure comprises of two subunits connected at one surface by the 4Fe-4S cluster.

Results of ProFunc analysis revealed that the interacting interface of the nitrogenase iron protein from *Bradyrhzoblum sp.*ORS278 consist of two chains. There are 36 interface residues in Chain A with an interface area of 2129 angstroms whereas Chain B consists of 42 interface residues with interface area of 2019 angstroms. No disulphide bonds and salt bridges were recognized in the interface region of the protein. The analysis also confirmed the presence of nests in each chain of the protein. Eight such nests have been located in Chain A whereas Chain B consists of 9 nests. These nests are structurally crucial motifs forming a concave depression which can serve as a binding site for an atom or a group of atoms [21]. Analysis of the binding sites showed the presence of clefts and cavities in the surface of

the proteins (Figure 3). In enzymes, the active site is often characterized by a particularly large and deep cleft [22]. The presence of 4Fe-4S cluster in a large cleft further substantiates the functionality of the modelled protein .

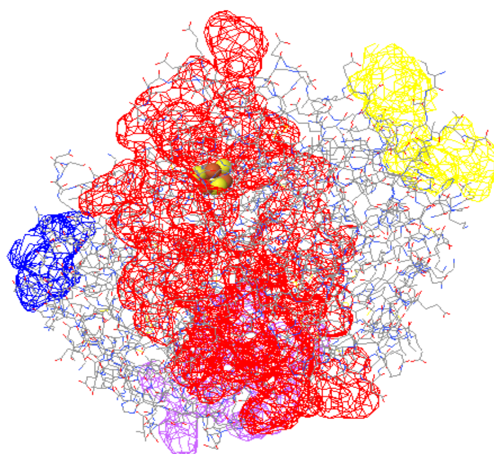

Figure 3
Various clefts and cavities present on the surface of the protein. The cleft are highlighted by various color.

RMSD calculations established that the nitrogenase iron proteins (NifH) of *Bradyrhizobium sp*.ORS278 had a deviation of 0.52 Å in the Cα residues and 0.55 Å in the backbone residues, from the template protein. This suggests that there isn't much significant deviation between the template protein and the modelled protein. The Ramachandran Plot (Figure 4a) illustrating the backbone conformation of the modelled protein revealed that the number of non-glycine and non-proline residues was 508. Out of these, 438 (86.2%) residues were in the most favoured regions. The refined model was analyzed by different analysis programs for the evaluation of the model quality. The overall quality score calculated by ProSA for our structure is displayed in a plot (Figure 4b) that shows the scores of all experimentally determined protein chains currently available in the Protein Data Bank (PDB) and has a z-score of −8.05. This result indicates that the z-score of our model is very much within the range of scores normally found for proteins of comparable size. ERRAT evaluation of the protein structure revealed a quality factor of 87.34.The normally accepted range of high quality model is <50 [15]. In the current case, the ERRAT score is well within the range of a high quality model. The model was further evaluated by VERIFY 3D Structure Evaluation Server and found to be stable and consistent.

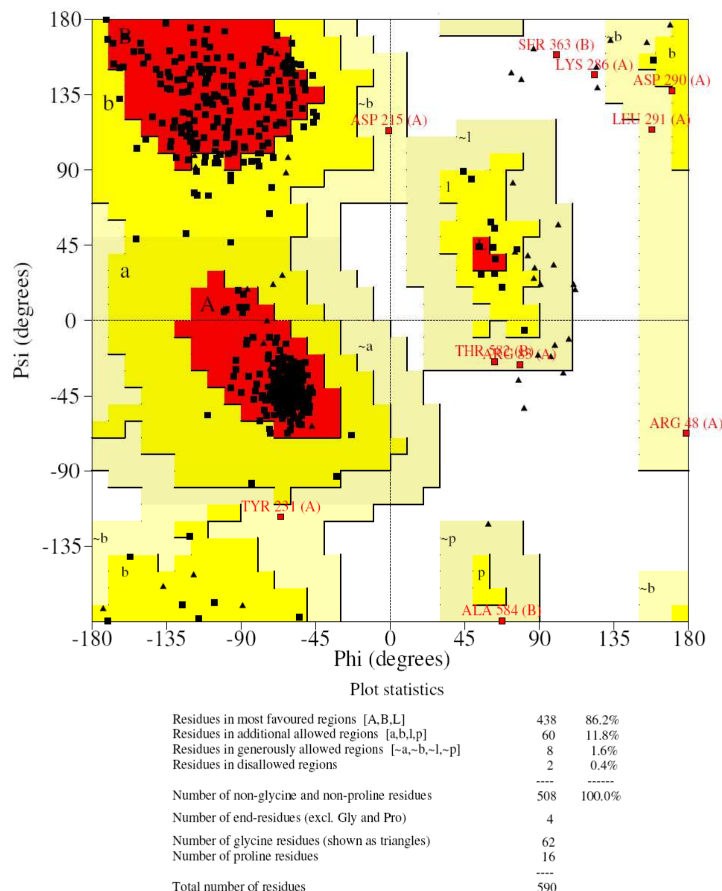

Figure 4(a)

Ramachandran plot of the nitrogenase iron protein of Bradyrhizobium sp. ORS278.The plot calculation on the 3D model of the NifH were computed with the PROCHECK program

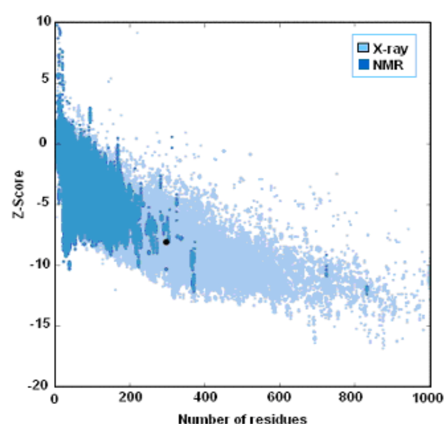

Figure4(b)

The plot shows z-scores of all protein chains in PDB determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to their length. The plot shows only chains with less than 1000 residues and a z-score -10. The z-scores of NifH protein are highlighted as large dots.

The intrinsic dynamics study of our modelled protein was carried out using the Normal mode analysis. NMA is a powerful technique for studying the vibrational and thermal properties of proteins at the atomic level. The WEBnm@ server offers fast calculation of the 200 lowest frequency modes. The first six modes which are referred to as zero frequency modes correspond to global rotation and translation of the system and are ignored in corresponding analysis[18]. Therefore the lowest frequency mode of interest is mode 7. The deformation energy of the seventh mode of our protein was 1231.61 which were lowest among the other modes. The deformation energy is a measure of the collectivity of the motion associated with the mode. Low deformation energy of the seventh mode indicates that the mode is with large rigid regions which have a good chance of describing domain motions. Normalized atomic displacements analyses were performed for modes 7 to12 (Figure 5).The plot thus obtained shows the displacement of each amino acid residue of the modelled protein for a particular mode. The highest correspond to the most displaced regions

.

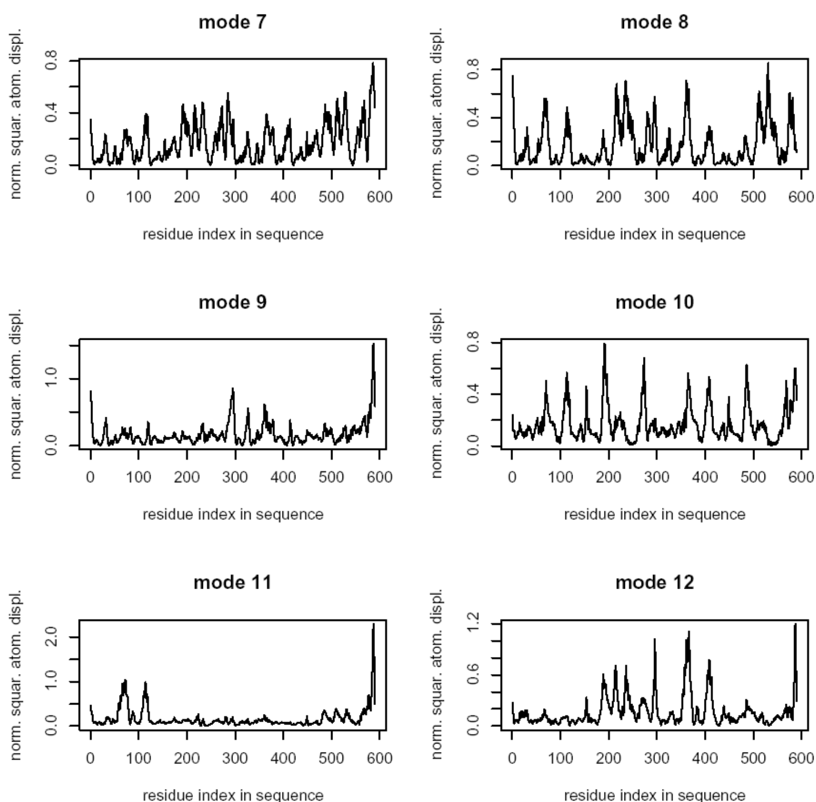

Figure 5
Displays the plot obtained by calculating the normalized atomic squared displacements

The region around the Fe4S4 cluster is highly conserved in the Fe proteins of various nitrogen fixing organisms. A close inspection of the structural details of the NifH protein model of the *Bradyrhizobium* ORS278 reveals that probable site of the post translational modification (Arg102) falls within a region of the protein predicted to possess p-

sheet or a-helical structure and the arginine residue is only three amino acids removed from one of the proposed cysteine ligands of the iron-sulfur center. To look into the possible aspects of ADP-ribosylation event, *in-silico* docking experiment was carried out. The docking of the substrates (ADP-Ribose) against whole structures of NifH protein of *Bradyrhizobium* sp. ORS278 revealed that the substrate binds the Fe protein in their cavities with the lowest docking energies of -2.78 Kcal/mol (Figure 6)

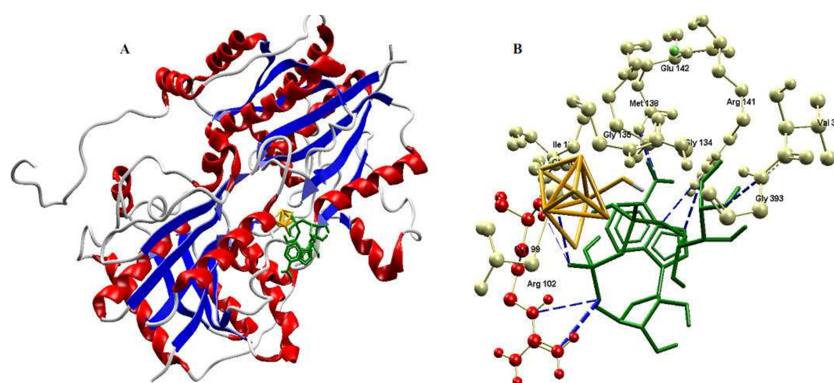

Figure 6 a)
The ADP-ribose docked into the three dimensional structure of NifH protein of Bradyrhizobium ORS278 ; b) The docked conformation of the substare ADP-ribose (green). All the amino acid residues which involved in molecular interaction are shown in ball and stick drawing. The arginine residue 102 is shown in Red. All the hydrogen bonds are shown in blue dashed lines.

Other than Arg 102, the amino acid residues found to be involved in molecular interaction with substrate include Cys 99, Arg 141, Met 138 and Gly 393. The possible hydrogen bond of the substrate ADP-Ribose with the amino acid is shown in the figure with blue dashed lines. ADP-Rib is linked to the arginine residue 102 through the guanidino group. Proximity of the docked substrate to the Fe4S4 active center leads to the speculation that this kind of modification may lead to blocking of the binding site for the MoFe protein or form a conformational change that inhibits the protein-protein interaction. As a result, the whole nitrogenase complex may be inactivated. Therefore, ADP-ribosylation of Arg102 may serve as important regulatory system in nitrogenase activity of *Bradyrhizobium* ORS278.

CONCLUSION

The 3D model of the NifH protein of *Bradyrhizobium* sp. ORS278 provides an insight into its conformational properties and structure-function relationship. The NMA analysis of the model was successful in determining the structural dynamics of the protein and describing its relative motion. All the dynamics analysis points towards higher degree of collectivity and probably refers to large regions of the protein being displaced. These results are essential for describing the functioning of the protein. *In-silico* docking results points to the fact that positioning of the arginine residue in NifH favours the ADP ribosylation which further hampers its interaction with other component of nitrogenase complex and it may lead to inactivation of nitrogenase system as a whole. The outcome of this study is

expected to benefit researchers investigating nitrogen fixing machinery of symbiotic nitrogen fixer like *Bradyrhzobium* sp. However, availability of site-directed mutagenesis experimental data of this protein may provide further light on the details of the role of this protein.

.

REFERENCES

1. Howard, J.B. and D.C. Rees, Structural Basis of Biological Nitrogen Fixation. Chem Rev, 96(7): 2965-2982, (1996).
2. Kim, J., D. Woo, and D.C. Rees, X-ray crystal structure of the nitrogenase molybdenum-iron protein from Clostridium pasteurianum at 3.0-A resolution. Biochemistry, 32(28): 7104-15, (1993).
3. Strop, P., P.M. Takahara, et al., Crystal structure of the all-ferrous [4Fe-4S]0 form of the nitrogenase iron protein from Azotobacter vinelandii. Biochemistry, 40(3): 651-6, (2001).
4. Allen, R.M., R. Chatterjee, et al., Biosynthesis of the iron-molybdenum cofactor of nitrogenase. Crit Rev Biotechnol, 14(3): 225-49, (1994).
5. Giraud, E., L. Hannibal, et al., Effect of Bradyrhizobium photosynthesis on stem nodulation of Aeschynomene sensitiva. Proc Natl Acad Sci U S A, 97(26): 14795-800, (2000).
6. Lowery, R.G. and P.W. Ludden, Effect of nucleotides on the activity of dinitrogenase reductase ADP-ribosyltransferase from Rhodospirillum rubrum. Biochemistry, 28(12): 4956-61, (1989).
7. Markowitz, V.M., N. Ivanova, et al., An experimental metagenome data management and analysis system. Bioinformatics, 22(14): e359-67, (2006).
8. Altschul, S.F., T.L. Madden, et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res, 25(17): 3389-402, (1997).
9. Thompson, J.D., D.G. Higgins, and T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res, 22(22): 4673-80, (1994).
10. Eswar, N., B. John, et al., Tools for comparative protein structure modeling and analysis. Nucleic Acids Res, 31(13): 3375-80, (2003).
11. Guermeur, Y. and P. Gallinari, Combinaison de classifieurs statistiques, application à la prédiction de la structure secondaire des protéines. PhD Thesis (1997).
12. Meiler, J., M. Müller, A. Zeidler, and F. Schmäschke, Generation and evaluation of dimension-reduced amino acid parameter representations by artificial neural networks. Journal of Molecular Modeling, 7(9): 360-369, (2001).
13. Kaplan, W. and T.G. Littlejohn, Swiss-PDB Viewer (Deep View). Brief Bioinform, 2(2): 195-7, (2001).
14. Wiederstein, M. and M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res, 35(Web Server issue): W407-10, (2007).
15. Colovos, C. and T.O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci, 2(9): 1511-9, (1993).
16. Luthy, R., J.U. Bowie, and D. Eisenberg, Assessment of protein models with three-dimensional profiles. Nature, 356(6364): 83-5, (1992).
17. Laskowski, R.A., J.D. Watson, and J.M. Thornton, ProFunc: a server for predicting

protein function from 3D structure. Nucleic Acids Res, 33(Web Server issue): W89-93, (2005).
18. Hollup, S.M., G. Salensminde, and N. Reuter, WEBnm@: a web application for normal mode analyses of proteins. BMC Bioinformatics, 6: 52, (2005).
19. Morris, G.M., D.S. Goodsell, et al., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry, 19(14): 1639-1662, (1998).
20. Georgiadis, M.M., H. Komiya, et al., Crystallographic structure of the nitrogenase iron protein from Azotobacter vinelandii. Science, 257(5077): 1653-9, (1992).
21. Watson, J.D. and E.J. Milner-White, The conformations of polypeptide chains where the main-chain parts of successive residues are enantiomeric. Their occurrence in cation and anion-binding regions of proteins. J Mol Biol, 315(2): 183-91, (2002).
22. Laskowski, R.A., N.M. Luscombe, M.B. Swindells, and J.M. Thornton, Protein clefts in molecular recognition and function. Protein Sci, 5(12): 2438-52, (1996).

This article can be downloaded from www.ijpbs.net