In-silico analysis suggests alterations in the function of XisA protein as a possible mechanism of butachlor toxicity in the nitrogen fixing cyanobacterium *Anabaena* sp. PCC 7120

Shilpi Singh[1]* & Prem Pal Singh[2]

[1]Molecular Biology Section, Laboratory of Algal Biology, Center of Advanced Study in Botany, Banaras Hindu University, Varanasi-221005; [2]Centre of Bioinformatics, Shobhit University, Meerut-250110; Shilpi Singh – Email: shilpi.0805@gmail.com; Phone: +91- 542-6701110; Fax No: +91- 542-2368174; *Corresponding author

Abstract:
Butachlor, a commonly used herbicide adversely affects the nitrogen fixing capability of *Anabaena*, an acclaimed nitrogen fixer in the Indian paddy fields. The nitrogen fixation in *Anabaena* is triggered by the excision of nifD element by xisA gene leading to rearrangement of nifD forming nifHDK operon in the heterocyst of *Anabaena* sp. PCC7120. Functional elucidation adjudged through in-silico analysis revealed that xisA belongs to integrase family of tyrosine recombinase. The predicted functional partners with XisA protein that have shown cooccurence with this protein in a network are mainly hypothetical proteins with unknown functions except psaK1 whose exact function in photosystem I is not yet known. The focus of this study was to find out the relation between XisA and butachlor using in-silico approaches. The XisA protein was modeled and its active sites were identified. Docking studies revealed that butachlor binds at the active site of XisA protein hampering its excision ability vis-à-vis nif genes in *Anabaena* sp. PCC7120. This study reveals that butachlor is not directly involved in hampering the nitrogen fixing ability of *Anabaena* sp. PCC7120 but by arresting the excision ability of XisA protein necessary for the functioning of nif gene and nitrogen fixation.

Background:
Butachlor a commonly used herbicide is frequently used in the rice paddy fields to block the growth of undesirable weeds. This butachlor toxicity also harms the cyanobacterium strains which are very important agriculturally as they convert atmospheric nitrogen into the form available to the plants for various purposes**.** *Anabaena sp. PCC7120* is a completely sequenced **[1]** heterocystous,gram-negative photoautotroph, endowed with two agriculturally important traits of carbon and atmospheric nitrogen fixation within the heterocyst cells. It contributes to the global nitrogen economy of soil and supports rice paddy production in tropical countries including India. The heterocyst and vegetative cells have division of labour, where heterocysts provide nitrogen to vegetative cells which in turn provide photosynthate to heterocysts. It is now known that the nitrogen fixing genes (nif) are wide spread in the genome of *Anabaena* which harbours an 11,278 kb nifD element, a 59,428 kb fdxN element and a 9419kb hupL element **[2]**. All these three elements are excised by site-specific recombination process during the late stages of heterocyst differentiation **[3, 4]**. XisA, required for the excision of the nifD is located near the nifK-proximal end of the nifD element **[5]**. XisA protein shows sequence homology with the integrase family of tyrosine recombinase **[6]**. The loss of excision may

BIOINFORMATION

occur in case of any mutation or absence of xisA recombinase, which is absolutely essential for excision. During heterocyst formation substantial changes occur in the cell. In addition to transcriptional gene regulation, two site-specific chromosomal rearrangements are tightly coupled to heterocyst differentiation: (i) excision of an 11kb from the nifD gene in the nifHDK operon, and (ii) excision of a 55kb element from the fdxN gene in the nifB-fdxN-nifS-nifU operon. The rearrangement thus obtained produces continuous coding sequences and functional operons essential for nitrogen fixation. The nifHDK operon is thought to encode the structural proteins of nitrogenase, the absence of nitrogenase activity in the strain of *Anabaena* sp. PCC7120 reiterates the necessity of nifD element excision for nitrogenase expression. Furthermore, the nifB-fdxN-nifS-nifU is thought to encode for the proteins required for the maturation of nitrogenase [7]. The site- directed inactivation of the *Anabaena* xisA gene can block the rearrangement of the 11kb element and also the nitrogen fixation. It is worth mentioning that *Anabaena variabilis* also contains a xisA gene which may complement the defective xisA gene of *Anabaena* sp PCC7120 though they are quite different in many ways [5].

XisA is a soluble, cytoplasmic protein with Gene ID: 1105037 in lineage: Bacteria; Cyanobacteria; Nostocales; Nostocaceae; *Nostoc*. It is composed of 472 aminoacids with positive strand. With the synonym of alr1442, xisA is found to be located at 1700741 to 1702159 in the *Anabaena* genome. It requires mention that reduction in nitrogenase activity at high concentrations of herbicide may be due to inhibition of photosynthesis [8, 9], which provides reductant and ATP required for nitrogenase activity and carbon skeleton for heterocyst formation [10], a entity harbouring nitrogenase. The present study is an attempt to ascertain at sequence level if xisA is a type II restriction endonuclease having a recombinase activity using in-sillico approaches. Efforts have also been made to construct the homology model of XisA protein for its proper functional elucidation and its docking with butachlor, a rice field herbicide to trace out the possible mechanism of butachlor toxicity on nitogen fixing ability of Anabaene sp. PCC7120 if it shows any interaction with the XisA protein, thereby regulating the function of nitrogenase.

Methodology:

Sequence Retrieval of xisA gene
The nucleotide sequence of the xisA gene was retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/) [11] with the accession NC_003272.1. The corresponding amino acid sequence of the xisA was also retrieved from the same database.

Physico-chemical characterization
This study was done to determine the physiochemical properties of xisA gene in *Anabaena* sp. PCC7120. The isoelectric point (pI), molecular weight, number of atoms present, aliphatic index and grand average of hydropathicity (GRAVY) were computed using the Expasy's protparam server [12].

Study of Protein-Protein function association network
The protein- protein interaction was done using STRING [13], a web-server to retrieve and display the repeatedly occurring

neighborhoods of XisA protein which are functionally associated with XisA protein.

Homology modelling for three dimentional structure prediction of XisA protein
Homology modelling (HM) also known as comparative modelling was used for three dimensional structure prediction. In view of the fact that the three-dimensional structure of the XisA protein from *Anabaena* sp. PCC7120 was not available in the Protein Data Bank, hence an attempt was made to construct the 3D model using Discovery studio 3.5 [14] and also to find out a suitable template protein for the modelling of the target protein. The template protein was searched through Brookhaven Protein Data Bank (PDB) [15] advance Blast to find out the most identical and positively similar model as a template. From the homology searching 1Aop (the site-specific recombinase) was selected as template protein model. The modelled structures outcomes were ranked on the basis of an internal scoring function and those with the least internal scores were utilized for model validation. Validation of the modelled structure was done to assess the reliability of the structure of XisA protein. The backbone conformation of the structure was calculated by analyzing the phi (Φ) and psi (ψ) torsion angles using PROCHECK, as determined by Ramachandran plot statistics using PDBsum [16]. Finally, the quality of the consistency between the template and the modelled xisA was evaluated using ProSA [17] during which the energy criteria for the modelled structure were compared with the potential mean force obtained from a large set of known protein structures.

Probable active site prediction for butachlor
Q-site finder [18] was used for prediction of active sites in XisA protein structure. Q-site finder uses the interaction energy between the protein and a simple Van der Waals probe. This was done to find out the binding site and the interacting residues for the ligand butachlor.

Cofactor-Ligand docking study with XisA protein
Docking of the XisA protein structure was done using Discovery Studio 3.5 [14]. The study of interaction of the XisA protein was done to trace the interacting residues with ligand butachlor herbicide. This study provides useful information regarding protein and the ligand.

Result & discussion:
Homology search showed *Anabaena* sp. PCC7120 circular DNA, with complete genome length of 6413771 bp and xisA gene possesses the accession no.NC_003272.1 and location starting at 1700741 and ending at 1702159 with 1419nt. The protein length was found to be 472 aa long with the accession NP_485485.1

Physico-chemical characterization
XisA protein was found to have a molecular weight of 55227.6 g mol^{-1}. The computed isoelectric point (Pi) of 8.88 indicates that the protein will precipitate in acidic buffer. The relatively high Ai 78.41 value indicates that the cyanobacterial XisA protein will be stable over a wide range of temperature. The grand average of hydropathicity (GRAVY) value 0.749, suggests its favourable water solubility. The amino acid

composition of nifH protein shows the abundance of Leu (9.7%) and Lys (9.1%).
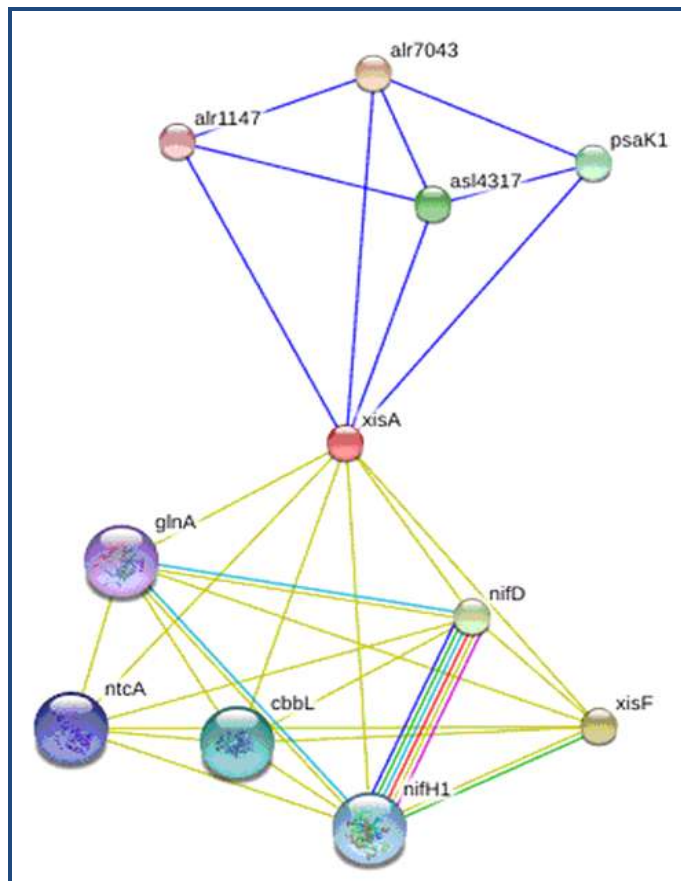
Figure 1: Protein-protein interaction network of the predicted funtional parters of XisA protein.

Study of protein- protein interaction

Protein–protein association has emerged as a useful concept for organizing all protein-coding genes in a genome**.** This study was done to find out the network of proteins **(Figure 1)** that might be interacting with XisA protein in one or the other way. The protein association network revealed the functional modularity and interconnectivity of XisA protein in the cell. Each different coloured lines represent the manner in which these predicted functional partners are associated with the XisA protein **Table 1 (see supplementary material)** such as the blue line indicates cooccurence of alr7043, alr1147, asl4317 and psaK1 with XisA as shown in the **(Table 1).** The predicted functional properties of the participating partners are catalogued in **(Table 1)** where nifD nitrogenase molybdenum-iron protein alpha chain is a part of the nitrogenase complex that catalyzes the key enzymatic reactions in nitrogen fixation, the exact function of psaK in photosystem I (PSI) is not yet known; cbbL catalyzes two reactions: the carboxylation of D-ribulose 1,5-bisphosphate, the primary event in carbon dioxide fixation, as well as the oxidative fragmentation of the pentose substrate in the photorespiration process, NifH1 is involved in key enzymatic reactions in nitrogen fixation catalyzed by the nitrogenase complex, which has 2 components: the iron protein and the molybdenum-iron protein, ntcA required for full expression of proteins subject to ammonium repression and glnA glutamate--ammonia ligase.
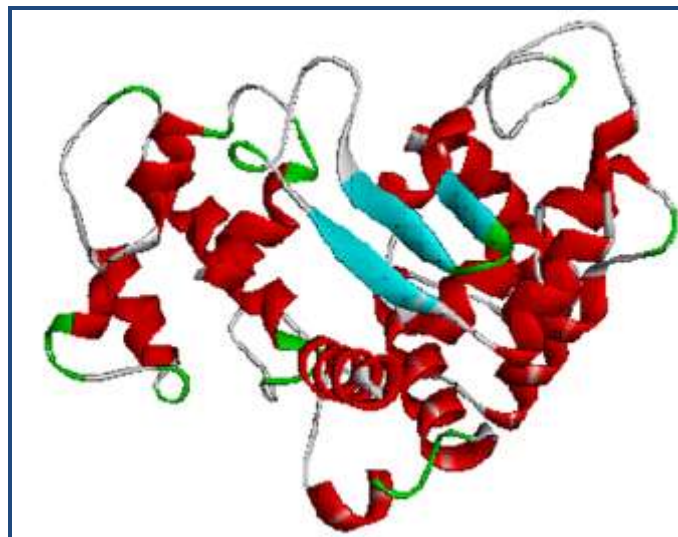

**Figure 2:** Modeled structure of XisA protein by using Discovery Studio 3.5.
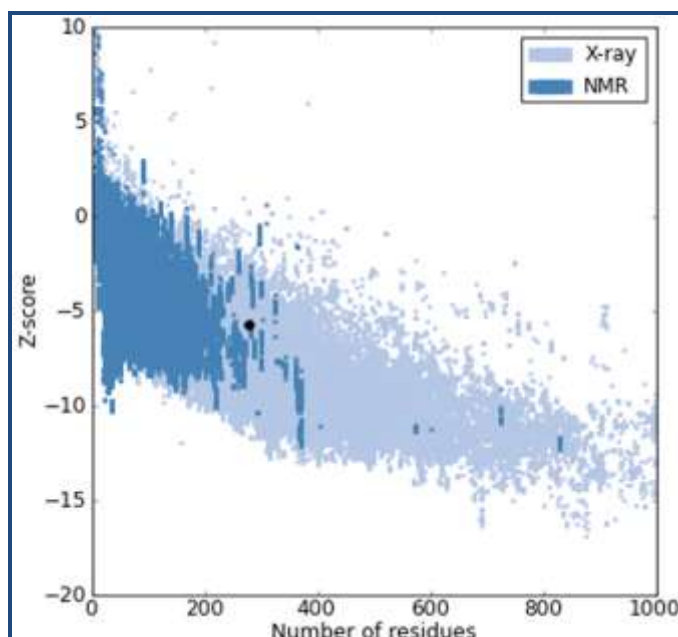
Figure 3: Prosa graph plot analysis of XisA protein structure using NMR and X-ray crystallography.

Homology modeling and assessment

The XisA protein (target) was modeled for structural classification and functional analysis taking 1Aop (crystal structure of the site-specific recombinase, XerD) from *E.coli* showing 19% identity with XisA protein as a template **(Figure 2).** The three-dimensional structural classification helps in studying the ligand-protein binding. The modeled protein structure was then validated through Prosa (**Figure 3)** and Rampage **(Figure 4)**. The rampage results showed 3.3% residues **Table 2 (see supplementary material**) in the outlier region. The Prosa **(Figure 3)** used for overall quality assessment through NMR and X-ray crystallography gave the z-score as -5.7. The results obtained through these assessment tools showed the overall satisfactory model quality. Thus the modeled structure was deposited to PMDB with PMDBID as XisA_PM0078645.
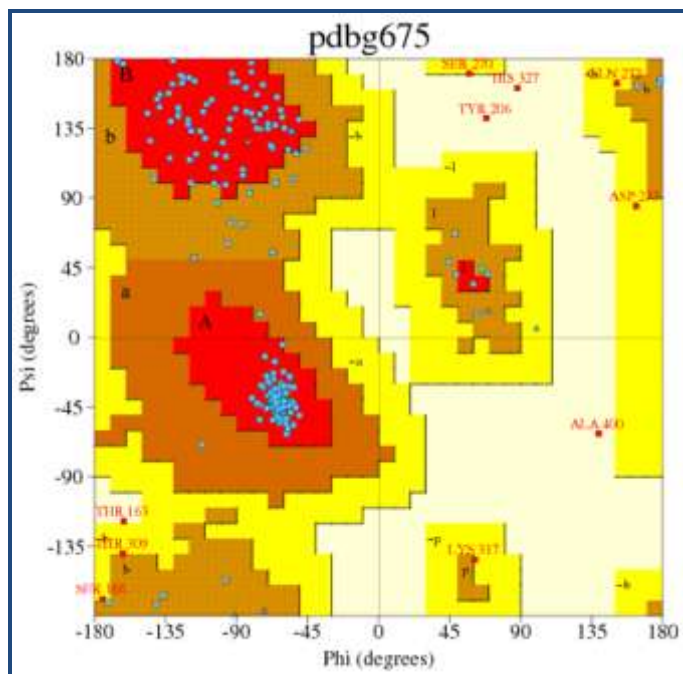
**Figure 4:** Ramachandran Plot for relibilityassessment of the modelled protein Xis**.**
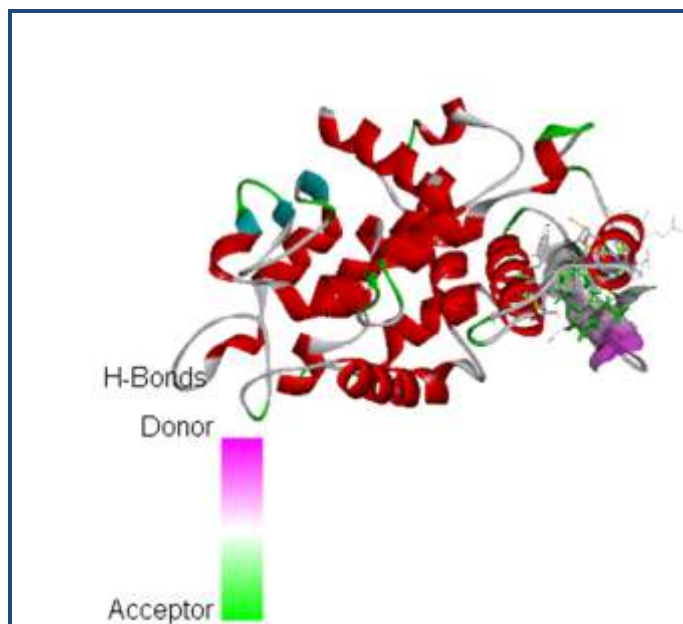
**Figure 5:** Surface view of the ligand butachlor and protein XisA docking

Butachlor (herbicide) docking study with XisA protein
Active site binding region and prediction of functional sites in modeled protein structure are the computational methods continuously attracting the attention. The protein surface was analysed for pockets through ligand binding site prediction **(Figure 4).** The major interacting cavity has the major active binding region and these cavities are defined by the energy criteria (**Figure 5**). The interacting energies of the ligand with the protein were calculated by Van-der Waals interactions. The ligand butachlor interacted with the modeled protein XisA with the libdock score of 88.1706 confirming the interaction between the ligand and the protein **Table 3 (see supplementary material).** The ligand butachlor and the

protein XisA possessed the interacting residues ALA[189], LEU[192],ILE[19 3],SER[195] ,ILE[196], ALA[211], ILE[212], PHE[215], CYS[216], MET[231], PRO[232], ASP[233], ARG[234] between them (**Figure 6)**. The Q-site finder predicted 10 active binding sites in XisA protein of *Anabaena* sp. PCC7120 **(Figure 7).** Most favourable binding sites contain amino acids with high conservation residue scores. The q-site finder also gave the information about the volume of each of the predicted site and the overall protein volume **Table 4 (see supplementary material).** When traced with Q-site finder the site 3 with a volume of 320 cubic angstroms was found to be the actual binding site of the ligand butachlor with XisA protein structure **(Table 4).** It was observed that ALA[211], PHE[215] (site 2), ILE[217] (site 4), and, LEU[192], SER[195], PHE[215] (site 10) were present at more than one site. Furthermore, the MET[231] was present solely at the ligand binding site and not found at any other site and location **(Table 4)**. This amino-acid is coded by the initiation codon AUG which indicates mRNA's coding region required for initiation of translation into protein.
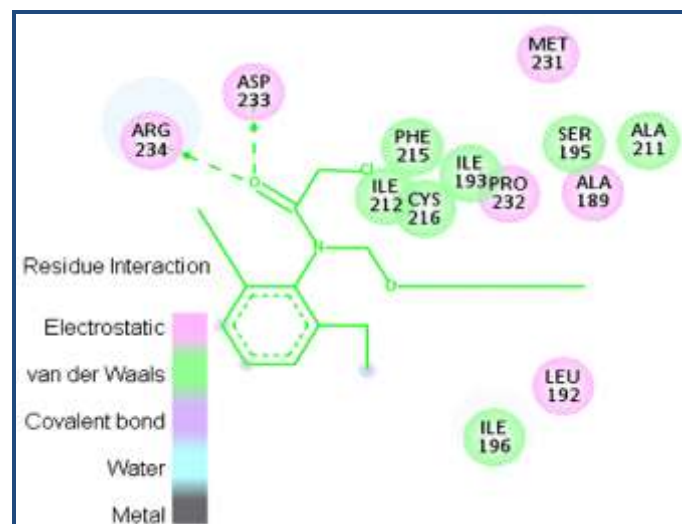
Figure 6: 2D view of the ligand and the protein docking showing different types of interactions with each residues**.**
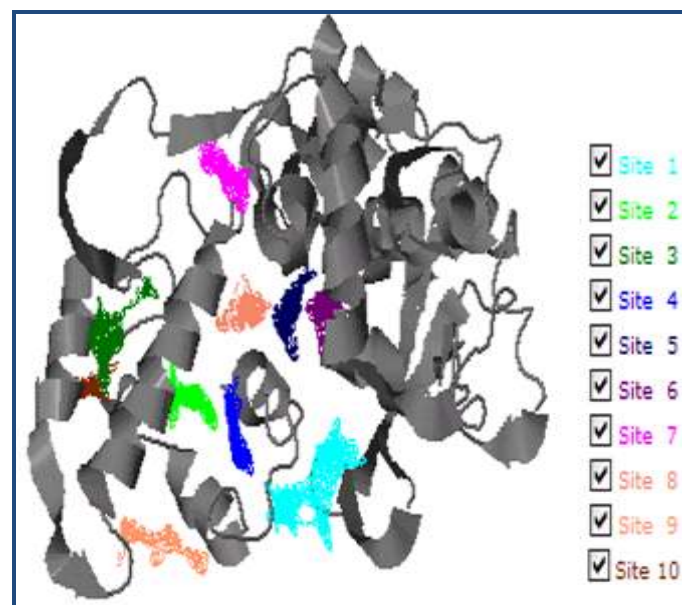
Figure 7: Active binding site in the modelled protein

Conclusion:

Using in-silico approaches, an attempt has been made for the first time to elucidate that binding of butachlor at the active site of XisA protein hampers the nitrogen fixing mechanism of *Anabaena* sp. PCC7120. The XisA protein on predicting its functionally associated partners in a protein interaction network showed co-occurrence with hypothetical proteins of unknown function and psaK1 whose exact function is not yet known. Through docking studies it was revealed that butachor binds at site 3 of the modelled XisA protein and MET[231] was present solely at the ligand binding site. It is well-known that this amino-acid is coded by the initiation codon AUG which indicates mRNA's coding region required for initiation of translation into protein. Thus binding of butachlor at this site may hamper the regulation mechanism XisA protein thereby altering the excision of nifD element hence disrupting the nitrogen fixing capacity of *Anabaena* sp. PCC7120. These findings suggest that wet lab approaches may be applied to verify the adverse affects of butachlor binding at the active site of XisA protein and inhibition of nitrogen fixation in cyanobacteria

References:

[1] Kaneko T *et al. DNA Res*. 2001 **8**: 205 [PMID: 11759840]
[2] Apte SK & Prabavathi N. *J Biol Sci*. 1994 **19**:579
[3] Haselkorn R, *Annu Rev Genet*. 1992 **26**: 113 [PMID: 1482108]
[4] Brusca JS *et al. J Bacteriol*. 1989 **171**: 4138
[5] Shah GR *et al*. *Res Microbiol*. 2007 **158**: 679 [PMID: 18023966]
[6] Wei TF *et al. J Bacteriol*. 1993 **4025**: 35 [PMID: 8391534]
[7] Orme-Johnson WH, *Annu Rev Biophys Chem*. 1985 **14**: 419 [PMID: 3890886]
[8] Jo K & Topal MD, *Science.* 1995 **267**: 1817 [PMID: 7892605]
[9] MacWfiams MP *et al. Genetics*. 1996 **143**: 1069
[10] http://www.ncbi.nlm.nih.gov/
[11] http://web.expasy.org/protparam/
[12] http://stringdb.org/newstring_cgi
[13] Shahi SK et al. *J Mol Model* 2013 19:973-983 [PMID:23097003]
[14] http://www.rcsb.org/pdb/home/home.do
[15] http://www.ebi.ac.uk/pdbsum/
[16] https://prosa.services.came.sbg.ac.at/prosa.php
[17] http://www.modelling.leeds.ac.uk/qsitefinder/

BIOINFORMATION

Supplementry material:

Table 1: Predicted functional partners of XisA protein with their interaction score and colours indicating the kind of association

Table 2: Rampage and Pdbsum results of model assessment

Gene Name	Number of residues in favoured region	Number of residues in allowed region	Number of residues in outlier region	Number of residues in favoured region	Number of residues in allowed region	Number of residues in generously allowed region	Number of residues in disallowed region
XisA	90.9%	5.8%	3.3%	222	29	6	4

Table 3: Results of Butachlor docking with protein

Protein Name	libDock Score	Relative energy	Pose_ number	Pubchem Shape_vol.	Pubchem MMFF94_ Energy	Interacting residues
Nifxis	88.1706	16.7613	1	250.8	64.6364	ALA^{189},LEU^{192},ILE^{193},SER^{195},ILE^{196}, ALA^{211},ILE^{212},PHE^{215},CYS^{216},MET^{231}, PRO^{232},ASP^{233},ARG^{234}

Table 4: Ten predicted active sites of the modelled protein structure with overall protein volume and volume of individual sites. Protein Volume: 26787

Site 1: 639 Cubic Angstroms	GLU^{154}, TYR^{156}, PHE^{157}, LYS^{158}, THR^{159}, HIS^{160}, LYS^{161}, ARG^{162,} THR^{163}, LYS^{165}, SER^{166}, GLU^{167}, PHE^{170}, ARG^{289}, GLU ^{290}, PHE^{292}, ILE^{293}, THR^{309} , TRP^{310}, LYS^{311}, ASP^{313}, LYS^{314}, GLU^{315}, CYS^{316}, LYS^{317}, THR^{318}
Site 2: 348 Cubic Angstroms	LEU^{149}, GLU^{150}, PHE^{152}, ALA^{153}, PHE^{170}, TYR^{173}, PHE^{174}, SER^{175}, THR^{177}, GLN^{178}, ASN^{182}, SER^{183}, LYS^{184}, ALA^{211}, ALA^{214}, PHE^{215}, THR^{218}, PHE^{219}
Site 3: 320 Cubic Angstroms	ALA^{189}, LEU^{192}, ILE^{193}, SER^{195}, ILE^{196}, ALA^{211}, ILE^{212},PHE^{215}, CYS^{216}, ILE^{217}, ASN^{220}, LYS^{230}, MET^{231}, PRO^{232}, ASP^{233}, ARG^{234}
Site 4: 192 Cubic Angstroms	PHE^{152}, ALA^{153}, GLU^{154}, GLU^{155}, TYR^{156}, PHE^{157}, LYS^{161}, LYS^{165}, SER^{166}, GLU^{167},THR^{169}, PHE^{170}, TYR^{173}, ILE^{217}, THR^{218}, ILE^{221}
Site 5: 225 Cubic Angstroms	GLU^{155}, TYR^{156}, LYS^{158}, ILE^{221}, LEU^{281}, GLY^{285}, LEU^{286}, ARG^{287}, PRO^{288}, ARG^{289}, PRO^{380}, TYR^{381}, LEU^{383}, ARG^{384}, TRP^{387}
Site 6: 256 Cubic Angstroms	GLU^{155}, LYS^{158}, PHE^{284}, GLY^{285}, LEU^{286}, ARG^{287}, GLU^{290}, LEU^{308}, GLN^{322}, ARG^{384}, HIS^{412}, THR^{413}, GLN^{414}, TYR^{416}, GLN^{417}, PHE^{420}
Site 7: 226 Cubic Angstroms	ILE^{223}, ASP^{224}, LEU^{225}, GLN^{227}, TYR^{228}, ALA^{243}, GLU^{244} ILE^{245}, LEU^{246}, TYR^{381}, ASP^{382}, HIS^{385}, HIS^{412}
Site 8: 231 Cubic Angstroms	LEU^{149}, LYS^{151}, PHE^{152}, GLU^{155}, LYS^{184}, ILE^{221}, GLU^{222} ASP^{224}, SER^{226}, ARG^{287}, ARG^{384}, VAL^{411}, THR^{413}
Site 9: 286 Cubic Angstroms	THR^{163}, THR^{164}, LYS^{165}, SER^{166}, HIS^{168}, THR^{169} ,TYR^{172}, ASP^{200}, TRP^{203}, ALA^{204}, ARG^{205}, TYR^{206}, ASN^{207}, ARG^{210}
Site10:221CubicAngstroms	SER^{175}, ARG^{176}, THR^{177}, TYR^{180}, ASN^{182}, ALA^{187}, ASN^{191}, LEU^{192}, ASN^{194}, SER^{195}, GLN^{198}, PHE^{215}