Docking analysis and catalytic site prediction of azoreductase in *E. coli,* with a wide range of industrially important azodyes.

Samrat Adhikari*, Samudra Prasad Sutradhar, Nangkyntiewbor Jungai, Bikash Thakuria and Gopesh Paul.

Bioinformatics Infrastructure Facility, Department of Biotechnology, St Edmund's College, Shillong- 793003, Meghalaya, India, Phone- 91-0364-2220808; 91-9862041757

**Corresponding author Email:* stedmundc.btisnet@nic.in

Abstract

Azoreductase is an FMN-dependent and NADH dependent enzyme of *E. coli* responsible for degrading azo dyes. The crystal structure of the enzyme and commercially important azo dyes were retrieved from PDB and NCBI PubChem. These azo dyes were then docked with the FMN-dependent NADH-azoreductase enzyme to analyze the binding affinity of the azo dyes with the enzyme and to predict the catalytic sites. In this approach, the catalytic residues of FMN-dependent NADH-azoreductase of *E. coli* was identified which were then analyzed based on its function, conservation, hydrogen bonding, B-factor and flexibility. The results predicted that the Phe-172, Glu-174, Lys-145, Asp-146 and Lys-169 contributes an important role as catalytic site residues in the enzyme. This study mainly focuses on better understanding of the biodegradation of commercially important and carcinogenic azodyes by azoreductase from *E.coli*. Furthermore, the catalytic site study is essential for understanding and altering the substrate specificity and for the design of a harmless azodye.

Keywords Azoreductase, NADH, FMN, EDO, IPA, azo dyes.

Abbreviation

NADH Nicotinamide Aminidine dinucleotide

FMN Flavin mononucleotide

EDO 1, 2 Ethanediol

IPA Isopropyl alcohol

PDB Protein Data Bank

NCBI National Centre for Biotechnology Information

CDART Conserved Domain Architecture Retrieval Tool

Introduction

Azo dyes are widely used for industrial, printing, and clinical purposes as well as textile dyeing because of its chemical stability, ease of synthesis, and versatility. Their durability, however, causes pollution once the dyes are released into the environment as effluent [1]. In addition, azo dye compounds are also the most commonly used drugs in the treatment of inflammatory bowel disease. However, the release of these dyes into the environment is undesirable, not only because of their colour, but also because many azo dyes and their by products are toxic and mutagenic. Biological treatment of azo dyes by the use of bacteria has been studied widely. Enzymes that catalyze the reduction of azo groups are termed azoreductases (EC 1.7.1.6). Utilizing NADH/FMN and/or NADPH as an electron donor, azoreductase can decolorize azo dyes into corresponding aromatic amines by reductive cleavage of azo bonds. The decolourization is a rate-limiting step, which was followed by the aerobic mineralization of the colourless aromatic amines. Furthermore, azoreductase is also involved in the site-specific delivery of azo prodrugs, which are therapeutically inactive in their intact form and rely on azo reduction by azoreductases of intestinal microflora for activation. Proteins with azoreductase activity have been identified and characterized from a wide variety of bacteria [2, 3].

There are at least two different types of bacterial azoreductases: Flavin dependepent and Flavin independent [4]. Flavin-dependent azoreductases can be further classified into two families according to their amino acid sequences. Azoreductases from *E. coli* and *Bacillus sp*. strain OY1-2 are representative of the two flavin-dependent azoreductase families, respectively. Although they are effective for an *in vitro* enzyme assay, overexpression of these azoreductases *in vivo* results in little or no increase of bacterial decolorization activity. Thus, the physiological role of azoreductase has recently been a subject of debate. As the introduction of azo Dyes into the environment is due mainly to human activities, reduction of azo dyes may not be the primary roles of these enzymes [6].

Azo dyes such as Sudan dyes are not legal for use as colourants in food, however, recently these dyes have been detected as contaminants in the food supply [7]. The human health impact of exposure to azo dyes used in certain food products has caused concern since they may have genotoxic properties. The environmental fate and subsequent heath effects of the azo dyes released in textile and paper industry wastewater are increasing being studied by the scientific community [8].

While azo dyes are generally considered to be persistent pollutants because they are typically recalcitrant to aerobic biotransformation [9], they might be metabolized by azoreductases from commensal microorganisms, mammalian liver cells, and soil microorganisms [10]. A variety of microorganisms, including bacteria and fungi, are capable of decolorizing a diverse range of azo dyes. Some bacteria have the ability to degrade azo

dyes both aerobically and anaerobically [11]. Bacterial degradation of azo dyes is often initiated by cleavage of azo bonds by azoreductases which are followed by the aerobic degradation of the resulting amines [4].

To understand the biodegradation of the azo dyes mediated by the enzyme and to design harmless azo dyes, it is essential to discover the catalytic site residues within the azoreductase enzyme and to perform flexibility analysis of the catalytic residues [5]. In the present study, docking study of the FMN-dependent NADH-azoreductase of *E.coli* and the industrially important dyes is carried out to understand the mechanism of azoreductase catalyzed enzymatic reactions, the catalytic site residues and the binding affinities of the azodyes and the enzyme. The industrially important dyes based on the related health hazards are chosen to study. In this, an FMN-dependent NADH-azoreductase enzyme of *Escherichia coli* is chosen as the target protein or protein of interest for the theoretical studies of biodegradation of azodye compounds. This enzyme was chosen to view the interaction with azodyes, which are pollutant to the nature. This study is applicable for understanding and altering the substrate specificity and for the design of a harmless azodye.

Materials and Methods

Data set

The enzyme molecule, FMN-dependent NADH-azoreductase of *E.coli* was obtained from Protein Data Bank (PDB) [12]. The three-dimensional (3D) structure of this protein has been solved by X-ray crystallographic techniques with 1.80 A^{o} resolution. Commercially important azodyes with serious health hazards [13, 14] were selected (table1) and retrieved from the NCBI PubChem Compound (15) for the docking study.

Multiple sequence alignment

Multiple sequence alignment is an essential tool for protein structure and function prediction. It illustrates the relationships between two or more sequences. When the sequences involved are diverse, the conserved residues are often key residues associated with maintenance of structural stability or biological function. Multiple sequence alignment was carried out using CLUSTAL W [16].

Computation of docking score between the ligands and the enzyme

Protein or ligand modification was carried out by removing water molecules from the protein and energy minimization was carried out using Swiss PDB Viewer (SPDBV) [17] tool. Autodock Vina [18] was used for docking studies of azoreductase (protein) and azodyes (ligands). The interactions between the different ligands

with the protein of interest were viewed in PyMOL [19] and the analysis of the H-bond and its positions were also done with PyMOL.

Functional site location

The catalytic or functionally important residues of a protein are known to exist in evolutionary constrained regions. However, the patterns of residue conservation alone are sometimes not very informative, depending on the homologous sequences available for a given query protein. Hence, the prediction of functional sites in newly solved protein structures is a challenge for computational structural biology. Most methods for functional site identification utilize measures of amino acid sequence conservation in homologous sequences, based on the assumption that functional sites are relatively conserved during evolution. Protein structural information has also been used to help identify protein functional sites. Ligand binding site prediction of the azoreductase enzyme was carried out using Q-SiteFinder [20]. It works by binding hydrophobic (CH3) probes to the protein, and finding clusters of probes with the most favourable binding energy. These clusters are placed in rank order of the likelihood of being a binding site according to the sum total binding energies for each cluster.

Hydrogen bonding and B factor

The hydrogen bonding and the B-factor are the two vital parameters in the catalytic residue prediction. Catalytic residue hydrogen bonding was investigated using the result obtained from the docking study. Catalytic residue B-factors is a measure of residue flexibility. Catalytic residues tend to have lower B-factors than all residues. Catalytic residues are very precisely positioned and held in place, as shown by their low B-factors and hydrogen bonding. The hydrogen bonding information is obtained from the docking studies. The B-factors of the residues interacting with the ligands and the active sites were taken from PDB file for each atom in a residue, and then averaged over the whole residue.

Results and Discussion

PDB structure

The enzyme, FMN-dependent NADH-azoreductase of *E.coli* has a crystallized structure characterized by three ligands FMN, EDO and IPA. The structure was obtained from PDB with the PDB id, 1V4B.

Multiple sequence alignment

Catalytic residues are more conserved than the average residues [21]. Hence the conserved residues were investigated to have a relationship with the catalytic residues. A BLAST search along with a multiple sequence alignment with the BLAST result showed the conservation of almost all the residues.

Proceedings IWBBIO 2014. Granada 7-9 April, 2014

Computation of docking score between the ligands and the enzyme

Proteins are the basis of life processes at the molecular level. The protein interaction is either with another protein or with small molecules. Many biological studies will benefit from interaction predictions. Docking study showed the binding affinity, number of hydrogen bonds and the binding residues. It is interesting to note that the binding affinities have negative values. This reveals the high feasibility of this reaction. The docked complexes were analyzed with the molecular visualization tools, PyMOL as shown in fig 2. The docking analysis showed that three dyes viz. (**a**) Sudan black B (61336), (**b**) Kayaku acid blueblack 10B (5473482), (**c**) FD & C Yellow no.5 (25245842) formed H-bonds with the enzyme residues Asp-109, Arg-59 and Ala-112, and Arg-59 of Site 7 respectively. Moreover, six dyes showed docking conformation with two of the natural ligands of the target protein; 3.3'dichlorobenzidine (7070) with IPA, Methyl red (10303) with IPA and Glu-174, HABA (5357439) IPA and Lys-145, Amaranth (FD & C red no. 2) (5473445) with IPA, Glu-174 and Phe-172, Grasal orange (5858445) with IPA and Antipyrylazo III (9573878) with FMN and His-144. The variation in the docking score indirectly gives the idea about the rate of decolourization. The rate of colour removal for trypan blue is higher than any other azo dye. The amino acids, Phe-172;Glu-174;Asp-189;Asp-146;Lys-145;Asp-109;Arg-59;Ala-112;Ala-200;Asp-146;Asp-150;Gly-174;Lys-169;Phe-170;Asp-167;Thr-166;Gly-131;Gly-164;Lys-132;His-144;Arg-59 interact with different ligands and are taken into account for comparison with the active sites obtained from active site prediction.

Functional site location

Active sites of the target protein were predicted using Q-SiteFinder and the output file was viewed under Chimera. Ten active sites were obtained from the study along with the corresponding amino acid residues present in each active site (Fig 1b). Each of the sites were analysed and compared with the amino acids interacting with the ligands in the docking study. The amino acid residues were found to have lower B-factor than the other residues. The 7^{th} active site consists of amino acids which are interacting with ligands in the docking study. The docking result shows that the amino acids such as Phe-172, Glu-174, Lys-145, Asp-146 and Lys-169 are very much repeated in the interaction with more than one ligand. This reveals that these amino acids are catalytic residues. The active site variations suggest that the enzyme can decolourize a wide range of azo dyes [22].

Hydrogen bonding and B factor

Both the Hydrogen bonding and the B-factors in the crystal structures were used as a measure of residue flexibility. Analysis shows that the amino acids interacting with the ligands are involved in hydrogen interaction

be it as a donor or as an acceptor. This shows that catalytic residues have a limited conformational freedom. The docking result shows that the ligands have hydrogen bonding with amino acids and it is illustrated in table 2, column 4. Catalytic residues tend to have lower B-factors than all residues, suggesting that they have to be more rigidly held in place than the averaged over the whole residue [23]. B-factors were taken from the PDB file for each atom in a residue, and then averaged over the whole residue. The B-factor for the residues, Phe-172 is 16.45, Glu-174 is 16.53, Lys-145 is 25.16, Asp-146 is 25.77 and Lys-169 is 31.8.

Conclusion

A major challenge in azo dye degradation and a harmless azodye design is the elucidation of biochemical and biological properties of enzymes, including the determination of catalytic residues that belong to the ligand-substrate binding site. Comparing the results from Q-Site Finder and docking studies, it indicates that the amino acid residues Phe-172, Glu-174, Lys-145, Asp-146 and Lys-169 play an important role as catalytic site residues in the azoreductase enzyme of *E.coli*. This docking study also provides information on the binding affinity of the ligands with azoreductase enzyme. The rate of colour removal for trypan blue is higher than any other azo dye. It is hoped that this information would provide a better understanding of the molecular mechanisms involved in catalysis and a heuristic basis for predicting the catalytic residues in enzymes of unknown function. The natural ligands (FMN, EDO and IPA) were also found to interact with some of the dyes and hence definitely binding site for the dyes. In this work the catalytic residues are reported as well as the binding affinities for some commercially important azodyes are also predicted. The study made in this project would facilitate researchers a better understanding of enzyme mechanisms and also used to improve the designing strategies of less harmful azodyes.

ACKNOWLEDGEMENT

We take this opportunity to acknowledge the funding received from Department of Biotechnology, Govt. of India for setting up BIF facility under BTISNET programme in Department of Biotechnology, St. Edmund's College, Shillong. We also express our heartfelt gratitude to Dr Sylvanus Lamare, Principal, St. Edmund's College, Shillong for his support throughout the work.

References

1. Nakanishi, M., Yatome, C., Ishida, N., and Kitade, Y. "Putative ACP Phosphodiesterase Gene (acpD) Encodes an Azoreductase." *J. of Biological Chemistry.*2001, 276(49):46394-46399.

2. Hanauer, S. B. *"*Inflammatory bowel disease*." N. Engl. J. Med.,* 1996, 334:841–848.

3. Wang, C. J., C. Hagemeier, N. Rahman, E. Lowe, M. Noble, M. Coughtrie, E. Sim, and I. Westwood. 2007. "Molecular cloning, characterization and ligand-bound structure of an azoreductase from *Pseudomonas aeruginosa*". *J.Mol. Biol*., 2007, 373(5)**:**1213–1228.

4. Chen, H. **"**Recent advances in azo dye degrading enzyme research*."Curr. Protein Pept. Sci*., 2006, **7**(2):101–111.

5. RamanthanK, ShanthiV, SethumadhavanR.
"In silico identification of catalytic residues in azobenzene reductase from Bacillus subtilis and its docking studies with azo dyes." Interdiscip Sci.,2009 ,1(4):290-7.

6. Liu, G., Zhou, J., Fu, Q, S., 1, 2, 3 and Wang, J. "The Escherichia coli Azoreductase AzoR Is Involved in Resistance to Thiol-Specific Stress Caused by Electrophilic Quinones*." J. Bacteriol.,* 2009, 191(20):6394-6300.

7. Xu, H., Heinze, T. M., Chen, S. Cerniglia, C. E., and Chen, H. "Anaerobic metabolism of 1-amino-2-naphthol-based azo dyes (Sudan dyes) by human intestinal microflora*." Appl Environ Microbiol,* 2007, 73(23):7759-7762.

8. McMullan, G., Meehan, C., Conneely, A., Kirby, N., Robinson, T., Nigam, P., Banat, I. M., Marchant, R., Smyth, W. F. "Microbial decolourisation and degradation of textile dyes. *Appl Microbiol Biotechnology*," 2001, 56(1-2):81-87.

9. Stolz, A. "Basic and applied aspects in the microbial degradation of azo dyes." *Appl Microbiol Biotechno,l* 2001, 56(1-2):69-80.

10. Levine, W. G. "Metabolism of azo dyes: implication for detoxication and activation." *Drug Metab Rev,* 1991, 23(3-4):253-309.

11. Prasad, N, K., Vindal, V., Narayana, S, L., V, R., Kunal, S, P., M, S. **"**In silico analysis of Pycnoporus cinnabarinus laccase active site with toxic industrial dyes*." J. Mol. Model,* 2011.

12. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. "The Protein Data Bank: a computer-based archival file for macromolecular structures." *J. Mol. Biol.*, (1977), 112, 535-542.

13. Fonovich TM. "Sudan dyes: are they dangerous for human health?",Drug Chemcal Toxicology. 2012 Sep 5. [Epub ahead of print]

14. Hilgemann DW, Delay MJ, Langer GA. "Activation-dependent cumulative depletions of extracellular free calcium in guinea pig atrium measured with antipyrylazo III and tetramethylmurexide." Circulation Research. 1983 Dec;53(6):779-93.

15. Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J., Bryant, S.H. "PubChem: a public information system for analyzing bioactivities of small molecules, 2009, 37(Web Server issue)": W623–W633

16. J.D. Thompson, D.G. Higgins, 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.* 1994;11 ;22(22):4673-80.

17. Guex, N., and Peitsch, M.C. "SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling"**.** *Electrophoresis,* 1997, 18, 2714-2723.

18. Trott, O., Olson, A. "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading". *J Comp Chem,* 2010, 31:455–461.

19. DeLano, W.L. "The PyMOL molecular graphics system." 2002. http://www.pymol.org.

20. Laurie, A.T.R., Jackson, R.M. "Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites." *Bioinformatics*, 2005, 21:1908–1916.

21. Bartlett GJ, Porter CT, Borkakoti N, Thornton JM. "Analysis of catalytic residues in enzyme active sites." J Mol Biol. 2002, 324(1):105-2.

22. Philem & Adhikari, Bioinformation. "Homology modeling, docking studies and functional analysis of various azoreductase accessory interacting proteins of *Nostoc sp.PCC7120". Bioinformation,2012,*8(7): 296-300.

23. Guo Y, Yue Q, Gao B. "Molecular docking study investigating the possible mode of binding of C.I. Acid Red 73 with DNA." Int J Biol Macromol. 2011 Jul 1;49(1):55-61. Epub 2011 Apr 1.

Table 1: Detailed docking analysis result showing interaction of azodyes and azoreductase enzyme. Listed below is the binding affinity, number of hydrogen bond formation of each azodye with azoreductase. The amino acid residue position of azoreductase interacting with each azodye is also provided along with the azodye information.

Azo Dyes	Common Names	Uses	Health Hazards
7070	3.3'DICHLOROBENZIDINE; Dichlorobenzidine base	Used in the production of yellow, and some red and orange pigments for the printing ink, textile, paper, paint, rubber, plastic, and related industries.	Carcinogenic
7340	O-Aminoazotoluene; C.I. Solvent Yellow 3	Used to color oils, fats, and waxes.	There is limited evidence for the carcinogenicity of Disperse Yellow 3 in experimental animals.
8411	O-Dianisidine;3.3'-DIMETHOXYBENZIDINE	Used in iso- cyanate-based adhesive systems	Carcinogenic
10303	Methyl red; O-METHYL RED	Used as indicator dye	Carcinogenic to animals
11490	4-(Phenylamino)-3'-Sulfoazobenzene	Used in automobile exhaust	Induces immune response to restricted antigens
61336	Sudan black B; Solvent black 3	Used in staining system	Carcinogenic in rat
5357439	HABA;2-(4-Hydroxyphenylazo) benzoic acid	Used in manufacturing laboratory chemicals	Skin, eye irritation
5473445	UNII-37RBV3X49K; Amaranth (FD & C red no. 2)	Used as food dye	Carcinogenic in rat
5473482	Amidoschwarz; Kayaku acid blueblack 10B	used to stain proteins in electrophoretic techniques	skin, eye, and respiratory tract irritant
5483782	D.C. Red no. 9; Pigment Red 53-1	Used for coloring in cosmetics	Ala-200
5809667	Sudan (1); Solvent red 23	Mainly oils, fats and waxes also in alcoholic, ester and hydrocarbon solvents. Occasionally in	The quantity of Solvent Red 23 that may be released to the environment is below the level expected to cause harm to organisms.

		polystyrene also used in cosmetics.	
58584 45	1-Phenylazo-2-napthol; Grasal orange	Used to colour waxes, oils, petrol solvents, polishes and foodstuffs	There has been concern that this material can cause cancer or mutations but there is not enough data to make an assessment.
58765 71	Scarlet red; Sudan IV	Used for the staining of lipids,triglycerides and lipoproteins on frozen paraffin sections.	Have potential carcinogenic effect.
59042 46	Dianil blue; TRYPAN BLUE	Used to selectively colour dead tissues or cells blue	The effects of trypan blue on the expression of apoptosis related and cell cycle arrest gene expressions.
60686 28	hexa-hydrchloride salt		Lys-145(2),169, Phe-170(4), Asp-146(2)
64533 40	Dianisidine sulfate; 119-90-4 (parent)	Used in the manufacture of azo dyes	Highly toxic and has oncogenic effect under experimental conditions (PMID: 969345)
95738 78	AntipyrylazoIII; EINEC3238988-2	Metallochromic indicator	Detrimental effect on muscle
25245842	FD & C Yellow no.5	Used in food, drugs, cosmetics and medical devices	Prolonged use leads to hypersensitivity

Table 2: Target protein Docked With Natural Ligands.

Ligands	Affinity(kcal/mol)	H-Bonds	Positions(residues)
EDO	-3.3	6	Ser-15,17(2),139(2), Gln-16
FMN	-10.5	15	His-144, Gly-142,141, Phe-98, Asn-97, Tyr-90, Met-95, Ser-9,15,17(2),139(2), Gln-16(2)
IPA	-3.2	5	Ser-15,17(2),139, Gln-16

Table 2: Table showing the amino acid composition of the different active sites predicted. The amino acid positions of the sites are also listed below.

Active Sites	Residues
Site 1	ILE 143
Site 2	SER 9, LEU 11, TYR 14, SER 15, GLN 16, SER 17, PRO 94, MET 95, TYR 96, SER 139, ARG 140, GLY 141
Site 3	PHE 24, TRP 28, VAL 88, LYS 133, ALA 134, ILE 135
Site 4	MET 95, TYR 96, ASN 97, PHE 98, SER 139, ARG 140, GLY 141, GLY 142
Site 5	PRO 48, LEU 54, ARG 68,GLU 71, ALA 72, LEU 75
Site 6	-
Site 7	LEU 50, VAL 55, LEU 58, ARG 59, ASP 109, ALA 112, ARG 113, ALA 114, PHE 118
Site 8	GLN 19, LEU 20, ASP 22, TYR 23, GLU 26
Site 9	ILE 10, TYR 96, ASN 97, ASN 99, ILE 100, SER 101,THR 102, GLN 103
Site 10	PHE 98, ASN 99, ILE 100

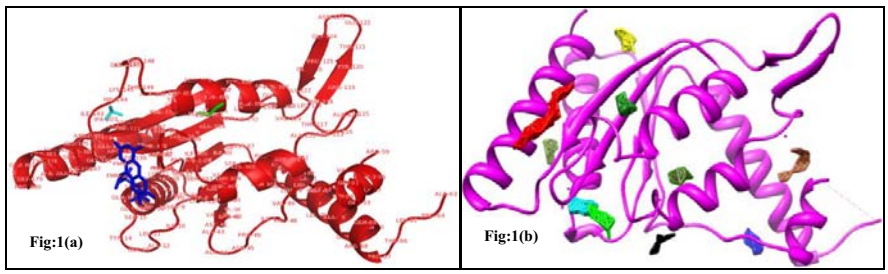

Figure 1: Figure 1(a) shows the three dimensional structure of the target protein (1V4B) obtained from PDB and figure 1(b) shows the active sites displayed in the three dimensional structure of the target protein predicted with the aid of Q-SiteFinder and viewed under Chimera.

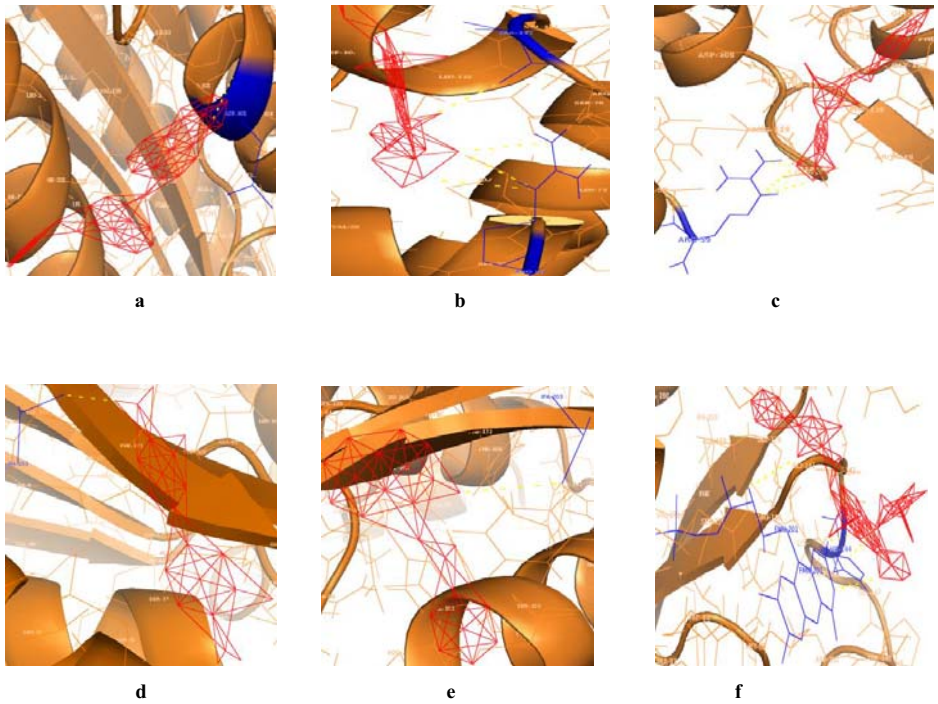

Fig 2: Docked conformations of (a) Sudan black B (61336), (b) Kayaku acid blueblack 10B (5473482), (c) FD & C Yellow no.5 (25245842) in the active site. Fig 4: Docked conformations of (d) 3.3'DICHLOROBENZIDINE with IPA, (e) Grasal orange with IPA and (f) Antipyrylazo III with FMN and His-144.