BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Prediction model

Volume 8(5)

Comparative modeling of DszC, an enzyme in biodesulfurization, and performing *in silico* point mutation for increasing tendency to oil

Ibrahim Torktaz[1], Zahra Etemadifar[2]* & Peyman Derikvand[2]

[1]Department of Biotechnology, Faculty of Advanced science and technologies, University of Isfahan, Hezarjarib St, 81746-73441, Isfahan, Iran, and National Institute of Genetic Engineering and Biotechnology, Tehran, Iran; [2]Department of Biology, Faculty of Sciences, University of Isfahan, Hezarjarib St, 81746-73441, Isfahan, I.R. Iran; Zahra Etemadifar - E mail: z.etemadifar@sci.ui.ac.ir; Phone: (98) 311 7932367; Fax: (98) 311 7932456; *Corresponding author

Received December 06, 2011; Accepted January 01, 2012; Published March 17, 2012

Abstract:
Desulfurization protein named DszC from *Rhodococcus erythropolis* is the key enzyme for biodesulforization of dibenzothiophene (DBT) in 4S pathway, which is a pathway with four enzymes. DszC enzyme biodesulfurizes DBT and its derivatives in oil components and biphasic systems. It functions well at the oil- water interface. In this study point mutation performed in DszC enzyme regarding to increase protein hydrophobicity and stability for application in immobilized form. 3D model of DszC predicted using Phyre2, SAM-T08 and M4t servers. I-Mutant 2 server used to determine potential spots for point mutation, and Molegro Virtual Docker (MVD) used for performing point mutation on 3D model. Hydrophobicity plots generated by Bioedit version 7.0.8.0 in Kyte-Doolittle scale indicated that protein hydrophobicity is increased after mutation. Also protein stability increased 26.11 units in scale of DDC2.

Background:
Sulfur concentrations in crude oil supplies are increasing, resulting in high sulfur concentrations in finished petroleum products. Hydrodesulfurization (HDS) used to achieve lower sulfur concentrations in refiners that need to high temperatures and pressure, requiring high operating costs **[1].** Benzothiophene (BT), DBT, and alkilated derivatives are recalcitrant organic sulfur and more resistant to HDS treatment than other sulfur compounds such as mercaptans and sulfides **[2].** Application of HDS also elicits undesirable effects on fuel quality; other chemical compounds in the oil reduced due to high temperature and pressure needed to achieve low sulfur levels. Biodesulfurization is used as a more selective method for lowering the sulfur content of petroleum products. DBT has been used as a model aromatic sulfur compound for the isolation and characterization of bacteria capable of biodesulfurize organosulfur compounds found in a variety of fossil fuels. The 4S pathway is a specific desulfurization pathway in which DBT is desulfurized and converted to 2-HBP. Through this pathway the carbon skeleton of DBT is released intact and thus the calorific value of the fuel is not lost. Consequently, development of biocatalytic desulfurization for the selective removal of aromatic sulfur compound from petroleum products has focused on the 4S pathway. Several bacterial species have been identified, including *Rhodococcus erythropolis*, *Gordonia*, *Arthrobacter*, *Brevibacterium*, and *Pseudomonas,* which are capable of either biotransforming DBT or growing with it as a sole sulfur source. The DBT desulfurization (*dsz*) operon from *Rhodococcus erythropolis IGTS8* encodes three proteins, DszC, DszA, and DszB. These proteins have been isolated, cloned, mutated, and overexpressed **[3, 4].** DBT is oxidized by DszC, first to DBT-5-oxide (DBTO) and then to DBT-5, 5'-oxide (DBTO2). DszA catalyzes transformation of DBTO2 to 2-(2'-hydroxyphenyl) benzene sulfinate (HPBS), which opens the thiophenic ring. HPBS is then desulfinated by DszB to produce 2-hydroxybiphenyl (2-HBP) **(Figure 1) [3].**

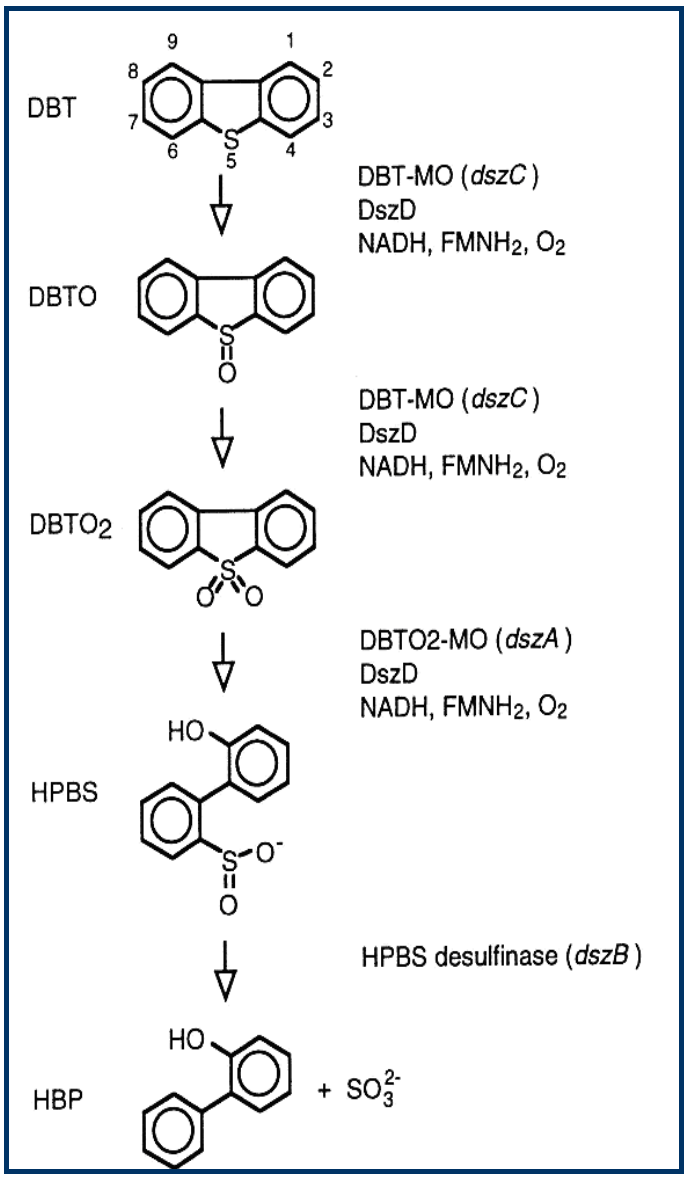

Figure 1: Proposed Dsz (4S) pathway for the biodesulfurization of DBT by R. erythropolis IGTS8. DBTO, DBT-5-oxide; DBTO2, DBT-5, 5′-dioxide; HPBS, 2-(2'-hydroxyphenyl)-benzene sulfinate; HBP, 2-hydroxybiphenyl; DBT-MO, DBT mono-oxygenase; DBTO2-MO, DBTO2 mono-oxygenase; FMN, flavin mononucleotide.

The transport of substrates and products might also contribute to desulfurization activity, as demonstrated by the fact that cell-free lysates of desulfurization cultures can exhibit a broader substrate range than the intact cell culture **[5].** There is no evidence that Dsz enzymes are excreted from IGTS8 cells, but the size of substrates metabolized show that desulfurization does not occur intracellular; desulfurization occurs in association with the external surface of cells **[6].** Immobilized-cell or free-cell biocatalyst (Dsz enzymes) was used to remove sulfur from oil in two-liquid phase systems in researches **[7].** To increase the rate of biodesulfurization by cultured or immobilized cells and Dsz enzymes, more researches are needed in areas such as increasing specific desulfurization activity, enzymes stability and hydrophobicity. The purpose of this study was the changes in the structure of DszC enzyme to increase stability and improve its performance in the oil phase in biphasic system. The sequence of key desulfurization pathway enzyme DszC, retrieved from NCBI database with accession number 32363475, was used to optimization of the protein stability and increasing hydrophobicity.

Methodology:

Prediction of 3D model of DszC

Phyre2 **[8],** SAM-T08 **[9]** and m4t **[10]** servers used for predicting 3D structure of DszC. M4t model choosed best based on prosa score **[11].**

Optimization of model

Modeller v9.8 **[12]** used for optimization of predicted model. Also molegro virtual docker 2011.4.3.0 (MVD) used for final check of model. For this purpose charge calculated by MVD and added to model. Explicit hydrogen bonds created and explicit bonds assigned. Energy minimization performed using UCSF chimera candidate version1.5.3. During minimization step update interval was 10 and step size was 0.02. Minimization performed in 100 steps.

Finding ligand binding sites

3DLigand Site server **[11]** used for prediction of potentially binding sites of model. This server searches query structure against Uniprot **[13]** library using matching molecular models obtained from theory (MAMMOTH) **[14]** method. In the server output 129 ASN, 131 SER, 132 SER, 137 HIS, 138 VAL, 160 HIS, 161 PHE, 162 CYS, 163 SER, 215 SER, 391 HIS, 394 VAL, 396TYR predicted as present in binding site. Also as an alternative approach MVD used for finding cavities of model. For this purpose probe size was 1.2, max number of ray checks was 16, minimum number of ray hits was 12 and Grid Resolution was 0.8. Five cavities found by MVD **(Figure 2).**

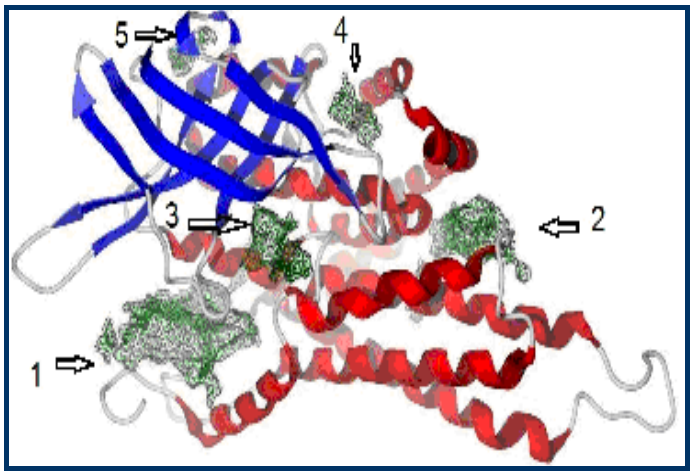

Figure 2: Cavities of DszC predicted model. MVD used for cavity detection. Detecting parameters: probe size 1.2, max number of ray checks was 16, minimum number of ray hits 12 and Grid Resolution 0.8

Point mutation

Point mutation performed in positions that had not any significant effect in 3D configuration of DszC model. To do this we selected amino acids which were not present in cavities and their side chain did not affect 3D structure of binding sites, as candidate for point mutation. Selected positions submitted to I-Mutant2 server **[15]** and in output data energy level of protein for changing a specific position to any other 19 possible amino

acids received. This was done for all positions that had potential to perform mutation. Based on increasing stability of the protein, hydrophobic amino acids according to the energy business were selected. The best amino acid substitution was not selected in terms of energy but the selection only was taken in hydrophobic amino acids.

Based on the output of the server, best spots for performing point mutation in a manner that the protein become more hydrophobic and enhance its stability were in positions including:9,11,12,13,14,51,119,122,178,179,180,203,204,221,222,278,295 and 338. Since conditions of DszC enzyme activity are at 30°C and neutral pH, thus protein stability upon point mutation was measured regarding to the same temperature and pH.

Mutation in three-dimensional model

MVD software was used for doing mutations in 3D model. In Selected locations on three-dimensional model, first, native amino acid was replaced with a new amino acid. Replaced residues in new position may stand with it's beside residues at a smaller distance than the threshold. In this case residue side-chain charges are in interference. To avoid interference of charges, optimization of neighborhood residues performed. During this step the charge of replaced amino acid root with lateral-roots of neighbor amino acids became in optimal state, because of the minimum force of mutated amino acid to its neighbors. Mutation and optimization of all selected positions were done to hydrophobic amino acids **Table 1 (see supplementary material)**. The final model for analysis, again submitted to the prosa server. The Z-score for this model was -5.08

Analysis of mutated model in aspect of affinity to water and oil

BioEdit version 7.0.8.0 used for generating hydrophobicity plot of native and mutant sequence of query **(Figure 3-A)**. Kyte-Doolittle scale used for depiction of hydrophobic plot.

Discussion:

Predicting 3D structure of query

Phyre2 server uses comparative modeling and any twenty identical amino acids are hits for this server. In phyre2, 3D model of query generated based on template of chain A from acyl CoA dehydrogenase (C2Z1QA). 394 residues (94% coverage) with 100% confidence have been modeled by this server. M4t and SAM-T08 servers used as alternative for predicting 3D structure of query. M4t uses comparative modeling for predicting query model. In this server five known structures used as template. (2rfq_12_367_A, 2vig_10_374_E, 2vig_10_368_C, 3mpi_5_378_C and 1ukw_4_379_A). SAM-T08 uses Hidden Markov Model algorithm to predict 3D model of Protein. In SAM-T08 serve misaligned residues in query, that have not any identical structure in database, are modeled by HMM algorithm and 3D model of misaligned residues conjectures by this server. Generated models checked in prosa server and Z-score of phyre2 was-7.53, SAM-T08 reached score-6.83and score of m4t model was -7.66. M4t output use for further study.

Prediction of protein stability

I-Mutant 2 which used in this research is a Support Vector Machine –based server that is trained based on a dataset from ProTherm. Trained SVM in this server can predict stability of protein upon point mutation based on both protein sequence and structure. When protein sequence is used, overall accuracy of prediction is 0.77 **Table 2 (see supplementary material)**. Output of server is energy level of protein in DDG value. DDG is defined as unfolding Gibbs free energy of mutated protein minus unfolding Gibbs free energy of wild type protein. DDG unit is Kcal/mol.

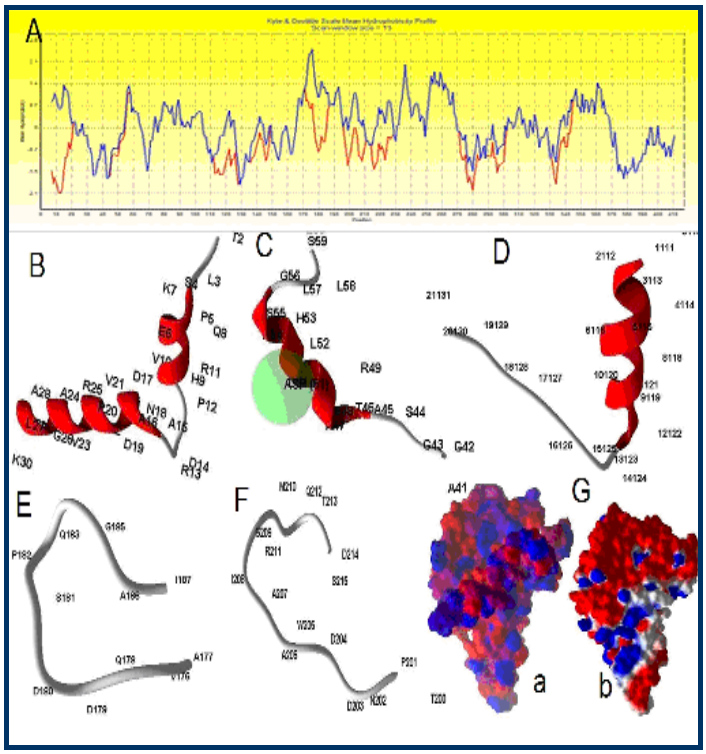

Figure 3: (A): Hydrophobicity plot of mutated (red) and native (blue) sequence of DsZc; **(B):** Mutation in positions 9, 11, 12, 13, 14; **(C):** Mutation in position 51; **(D):** Mutation in positions 119 and 122; **(E):** Mutation in positions 178, 179 and 180; **(F):** Mutation in positions 203 and 204; **(G):** hydrophobicity based coloring of native and mutated model, (a: after mutation, b: before mutation), Blue color shows hydrophilic area.

Point mutation

Mutation in positions 9, 11, 12, 13, 14 changes structure of first loop of protein **(Figure 3-B)**. This fold has a loop. Because amino acids that present in this loop were not detected as present in binding site by 3DLigandSite and MVD, it is probable that structure changing which caused by mutating 5 amino acids has not significant effect in structure of protein pockets. Mutation in position 51 **(Figure 3-C)** changes structure of a separate and surface exposed fold of protein. This fold has not any effect in configuration of detected binding sites of enzyme. Positions 119 and 122 are located in an alpha helix and mutating these positions do not change helix form **(Figure 3-D)**. Positions 142, 278, 295 and 338 are in surface of protein and mutation in these positions does not affect structure of fold and cavities of protein significantly. Mutation in 178, 179 and 180 do not change 3D configuration of protein **(Figure 3-E)**, because neighbor tern is made by prolin 182. Also as we see in (**Figure 3-F**), mutating positions 203 and 204 do not have any significant effect on protein folding. His 221 and Asn 222 are two surface amino acids that are located away from cavities and turns. After

performing mutation, protein stability increased 6.24 units in DDG scale.

Generating hydrophobicity plot of query

BioEdit version 7.0.8.0 was used for generating hydrophobicity plot and hydrophobicity scale was Kyte-Doolittle **[16].** This scale is widely used for detecting hydrophobic regions in proteins. In this scale, regions which reach positive value are hydrophob. Depending on the window size, this scale can be used for surface exposed and transmembrane regions.Window sizes of 5-7 is used for surface-exposed regions and window sizes of 19-21 is used for finding transmembrane domains if the values calculated are above 1.6 . Because DszC is not transmembrane protein, default window size of 13 used for detecting surface exposed regions of query.

Conclusion:

In native DszC enzyme affinity to water is more than oil, because of the side-chain root of amino acids present in surface of protein. Causing the mutation in surface amino acids of DszC to hydrophobe residues changed hydrophobicity properties of this protein. After performing mutation, the DszC enzyme had more tendencies to oil surface in a manner that its stability increased. This mutated enzyme can work better in biphasic system by increasing hydrophobicity, and can maintain in active structure in longer time by increasing stability. Also differences in coloring of native model and mutant model **(Figure 3-G)** showed the more water escaping in mutant model. Since the stability of this enzyme increased with mutation, so it is suggesting that mutated enzyme can be used by high performance for biodesulfurization in biphasic system as immobilized enzyme.

References:

[1] Monticello DJ & WR, *Finnerty Annu Rev Microbiol.* 1985 **39**: 371[PMID:3904605]

[2] Kabe T *et al. Ind Eng Chem Res.* 1992 **31**: 1577

[3] Oldfield C *et al. Microbiology.* 1997 **143**: 2961 [PMID: 9308179]

[4] Piddington C *et al. Appl Environ Microbiol.* 1995 **61**: 468 [PMID: 7574582]

[5] Kilbane JJ, *Curr Opin Biotechnol.* 2006 **17**: 305 [PMID: 16678400]

[6] Kilbane JJ & Le Borgne S, *Developments and Perspectives.* 2004. pp. 29–65.

[7] Tao F *et al. Appl Environ Microbiol.* 2006 **72**:4604 [PMID: 16820450]

[8] Kelley LA & Sternberg MJE, *Nature Protocols.* 2009 **4**: 363 [PMID: 19247286]

[9] Karplus K, *Nucleic Acids Research.* 2009 37 [PMID: 19483096]

[10] Rykunov D *et al. Struct Funct Genomics.* 2009 **10**: 95

[11] Wiederstein M & Sippl MJ, *Nucleic Acids Research.* 2007 **35**: W407 [PMID: 17517781]

[12] Eswar N *et al. Current Protocols in Bioinformatics, John Wiley & Sons, Inc.* 2006 **15**: 561

[13] The UniProt Consortium. *Nucleic Acids Res.* 2009. **37**: D169 [PMID:18836194]

[14] Ortiz AR *et al. Protein Sci.* 2002 **11**: 2606[PMID: 12381844]

[15] Capriotti E *et al. Nucleic Acids Res.* 2005 W**306**: 10 [PMID: 15980478].

[16] Kyte J & Doolittle RJ, *Mol Biol* 1982 **157**: 105[PMID: 7108955]

Edited by P Kangueane

Citation: **Torktaz** *et al.* Bioinformation 8(5): 246-250 (2012)

BIOINFORMATION open access

Supplementary material:

Table 1: Performed point mutations in DszC enzyme. Change in DDG scale indicates changing protein stability upon point mutation

Position	Native amino acid	Mutated amino acid	Energy in DDG scale
9	Histidine	Leucine	1.25
11	Arginine	Isoleucine	-0.58
12	Proline	phenylalanine	-0.26
13	Arginine	Leucine	-0.28
14	aspartic acid	phenylalanine	-0.09
51	aspartic acid	Serine	0.03
119	Glutamine	Isoleucine	0.99
122	Glutamine	Isoleucine	0.88
142	Lysine	phenylalanine	0.92
178	Glutamine	Isoleucine	0.00
179	aspartic acid	Leucine	-0.59
180	aspartic acid	Leucine	-0.19
203	aspartic acid	Leucine	-0.31
204	aspartic acid	Leucine	-0.31
221	Histidine	Leucine	1.20
222	Asparagines	Isoleucine	1.95
278	Arginine	Leucine	1.25
295	aspartic acid	Leucine	0.63
338	Arginine	Isoleucine	-0.25

Table 2: Accuracy of trained SVM for prediction protein stability upon point mutation

Prediction method	Q2	P(+)	Q(+)	P(-)	Q(-)	C
I-Mutant2.0-Seq	0.77	0.69	0.46	0.79	0.91	0.42

Q2 is defined overall accuracy: Q2=p/N
P: number of correctly predicted residues. N: total number of residues
C is correlation coefficient and defined as: C(s) = [p(s) n(s)-u(s) o(s))] / D
And D is normalization factor: D =[(p(s)+u(s))(p(s)+o(s))(n(s)+u(s))(n(s)+o(s))]1/2