BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 9(19)

Insilico Characterization and Homology Modeling of Arabitol Dehydrogenase (ArDH) from *Candida albican*

Muhammad Waseem Sarwar*, Irum Baddisha Saleem, Asif Ali & Farhat Abbas

Department of Bioinformatics and Biotechnology, Government College University (GCU), Faisalabad, Pakistan; Muhammad Waseem Sarwar - Email: waseemsarwar@live.com; Phone: +92-03007277872; *Corresponding author

Received November 11, 2013; Accepted November 19, 2013; Published December 06, 2013

Abstract:
Background: Arabitol dehydrogenase (ArDH) is involved in the production of different sugar alcohols like arabitol, sorbitol, mannitol, erythritol and xylitol by using five carbon sugars as substrate. Arabinose, d-ribose, d-ribulose, xylose and d-xylulose are known substrate of this enzyme. ArDH is mainly produced by osmophilic fungi for the conversion of ribulose to arabitol under stress conditions. Recently this enzyme has been used by various industries for the production of pharmaceutically important sugar alcohols form cheap source than glucose. But the information at structure level as well as its binding energy analysis with different substrates was missing. **Results:** The present study was focused on sequence analysis, *insilico* characterization and substrate binding analysis of ArDH from a fungus specie *candida albican*. Sequence analysis and physicochemical properties showed that this protein is highly stable, negatively charged and having more hydrophilic regions, these properties made this enzyme to bind with number of five carbon sugars as substrate. The predicted 3D model will helpful for further structure based studies. Docking analysis provided free energies of binding of each substrate from a best pose as arabinose -9.8224calK/mol, d-ribose -11.3701Kcal/mol, d-ribulose -8.9230Kcal/mol, xylose -9.7007Kcal/mol and d-xylulose 9.7802Kcal/mol. **Conclusion:** Our study provided insight information of structure and interactions of ArDH with its substrate. These results obtained from this study clearly indicate that d-ribose is best substrate for ArDH for the production of sugar alcohols. This information will be helpful for better usage of this enzyme for hyper-production of sugar alcohols by different industries.

Keywords: Arabitol dehydrogenase, Homology Modeling, *Insilico*, Charaterization, Molecular Docking.

Background:
Arabitol dehydrogenase is involved in the reduction of a five carbon sugar ribulose to form arabitol in pentose phosphate pathway **[1].** This enzyme is produced in various osmophilic fungus species under osmotic stress, to produce compatible solute arabitol to balance the osmotic pressure across the cell membrane. The D-arabitol dehydrogenase belongs to the short chain hydrogenases/reductases superfamily. It is a group of enzymes with a broad spectrum of substrates and is produced by all domains of life. It catalyzes the oxidation of D-arabitol to D-ribulose and therefore belongs to the class of D-arabitol 2-dehydrogenases, which are typically produced by yeast. Similar to other polyol oxidizing or reducing enzymes, ArDH requires NAD/NADP(H) as a co-substrate. Polyols matching the Bertrand–Hudson rules (such as D-arabitol, D-sorbitol, D-mannitol, meso-erythritol, etc.) are all good substrates of ArDH, while those that do not match the rule (L -arabitol, xylitol, etc) are either poor substrates or not accepted at all **[2].** If desirable, arabitol can also be converted to xylitol by arabitol dehydrogenase. Arabitol dehydrogenases have been produced by bacteria also. A membrane-bound arabitol dehydrogenase was purified from acetic acid bacteria that could oxidize

polyhydroxy alcohols. The ArDH has the potential to be a useful biocatalyst for the production of D-ribulose starting from the inexpensive D-arabitol due to its regiospecificity **[3].** Arabitol dehydrogenase gene can be used as selectable marker as plants especially edible crops can grow on d-xylulose, so if a plant cell could express arabitol dehydrogenase, then such a cell would be able to grow in a medium containing d-arabitol, whereas an untransformed plant cell would not proliferate **[4].** Since most plants cannot metabolize most sugar alcohols, including d-arabitol, there is an opportunity to develop positive selection systems based on sugar alcohols. ArDH also cloned for the production of xylitol from arabitol in different arabitol producing species. The d-arabitol dehydrogenase converts arabitol into xylulose that is an intermediate of the oxidative pentose phosphate pathway. The arabitol can maintain cell turgor and thus it is the driving gradient for water uptake under stress conditions **[5, 6].**

Arabitol dehydrogenase is well known the conversion of many rare sugars & sugar alcohols like arabitol to xylulose, manitol to fructose, sorbitol to fructose, erythritol to erythrose, xylulose to arabitol, fructose to manitol and ribulose to arabitol in bacteria and other organisms. Rare sugar derivatives are used as effective antivirals, for example against hepatitis-B virus and human immunodeficiency virus, and antitumor agents **[7, 8].** The demand for rare sugars has been rising recently; global production was in excess of 133 million tons in 2002. These sugars are useful in the sweetener and pharmaceutical industries. Five-carbon rare sugars are also used as sweeteners and normally do not have an unpleasant after taste that is common in artificial sweeteners, such as saccharin and cyclamates. Moreover, D-ribulose produced from D-Arabitol dehydrogenase is an important precursor for the synthesis of oligosaccharides, amino sugars, and glycosides and is used for the industrial synthesis of the alternative sweetener xylitol **[9].** Keeping in view the its importance and applications we have done structural as well as functional annotation of ArDH from *Candida albican*s using bioinformatics tools, for the very first time. The study presents novel insights into the structural features and substrate binding of ArDH.

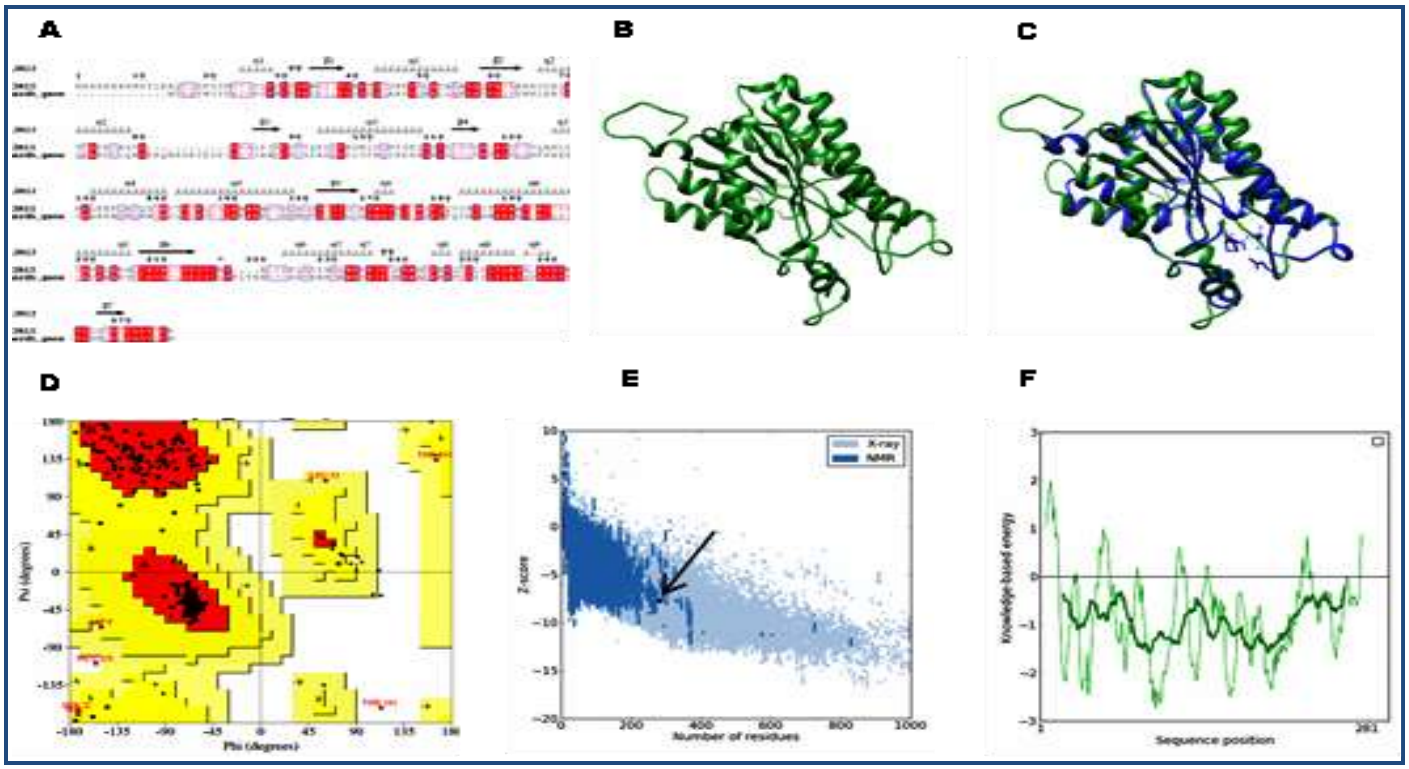

Figure 1: A) Query and template (3R1I) alignment with secondary structure elements showing alpha helixes and beta sheets; **B)** Predicted Model of ArDH; **C)** Template (Blue) model (Green) superimposition showing very high similarity as done by Chimera v1.7.3 showed a RMS value of 0.241 A;º **D)** Ramachndran plot analysis for Predicted Model of ArDH. The plot statistics are: Total number of residues are 281 with 91% residues in the most favored region (red); 6.1 % residues in allowed (yellow) and 2% in generously allowed (light yellow) region. Number of glycine residues (labelled as G) is 24 and Number of Pro residues is 11; **E)** Z-score plot from ProSA-web server showing the quality of predicted model in NMR region (dark blue); **F)** Energy plot showing all residues of predicted model at very stable position (dark green line).

Methodology:

The study was conducted using Intel(R) Core(TM) i3-2370M CPU@ 2.40 GHz, 4 Core(s) processor and 64-bit Operating System.

Sequence Retrieval

The 281 amino acid long sequence of ArDH from *Candida Albican* was obtained from the Protein sequence database of EBI (Id: JN578089) and subjected to BLASTP on NCBI against Protein Data Bank (PDB) entries to find similar sequences. The default parameter values of BLASTP were used for search.

Characterization of Target Sequence

The physicochemical properties such as molecular mass, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index,

aliphatic index, and grand average of hydropathicity (GRAVY) were calculated by ProtParam from EXPASy server (www.expasy.ch/tools) **[10].** The CELLO v.2.5 **[11]** was used to analyze the subcellular localization of ArDH. MEMPACK from UCL **[12]** was used to study the transmembrane topology. In order to know the key residues responsible for catalytic activity of the enzyme, multiple sequence alignment of ArDH with other closely related enzymes of dehydrogenase family having known crystallographic structures was done on ClastalW2 server **[13].**

Secondary structure prediction

The secondary structure prediction was done by SOPMA **[14]** and PsiPred **[15]**, server. The predicted secondary structural information of the enzyme was considered to improve the target-template alignment and for building conformations for 3D model of the ArDH.

Model building by Homology Modeling

Homology modeling was used to predict the 3D structure of the ArDH being the most suitable method for protein model building **[16].** The template was selected from CPH Model Server **[17].** The template and query alignment was done using Modellerv9.11 **[18]** program using align2d command, the output file (PIR format) was used to build ten models by Modeller. The constructed models were then solvated and subjected to energy minimization with the GROMOS 96 43B1 parameters set through Swiss-PDB Viewer program **[19].** Model evaluation, model quality and stereochemical analyses were performed by Errat, verify 3D. ProSA-web Z-score, **[20]** Qmean plot **[21]** and PROCHECK Ramachandran plot **[22].** Furthermore, Root Mean Squared Deviation (RMSD), superimposition of query and template structure, performed using UCSF Chimera 1.7.3 workbench **[23].**

Molecular docking analysis

Docking studies were carried out using the MOE (Molecular Operating Environment) software **[24].**

Ligand preparation: Chemical structures of substrate molecules were downloaded from pubchem (pubchem. ncbi.nlm.nih. gov/). These structures were optimized by adding hydrogens and energies were minimized with parameters (gradient: 0.05, Force Field: MMFF94X).

Preparation of Receptor Structure: ArDH model was predicted through homology modeling, best model was selected for docking analysis. This model is subjected to 3D protonation and energy minimization using parameters (gradient: 0.05, Force Field: MMFF94X+Solvation). The minimized structure was used as the receptor protein for Docking.

Docking Run: MOE docking program with default parameters was used to bind the selected ligands with receptor protein and to find the correct conformation of substrate. Free energy of binding of the ligand from a given pose was estimated by MOE London dG scoring function.

Results & Discussion:

Characterization of ArDH

Physiochemical properties of ArDH by ProtParam tools is presented in **Table 1 (see supplementary material).** Results show that ArDH protein has a molecular weight of 30642.8 Daltons and an isoelectric point of 5.59. The computed pI value revealed that ArDH is an acidic protein. The computed isoelectric point will be useful for separating the protein on a polyacrylamide gel by isoelectric focusing. The extinction coefficient can be used to calculate the concentration of a protein in solution. Stability of ArDH was studied by analyzing the values for instability index, aliphatic index and Grand average of hydropathicity (GRAVY) index **[25].** The value of instability index was 44.16 hence it could be safely predicted stable protein. The aliphatic index refers to the relative volume of a protein that is occupied by aliphatic side chains and contributes to the increased thermo stability of protein. Aliphatic index of ArDH was 85.77. GRAVY index indicates the solubility of proteins, GRAVY index of ArDH was -0.118. A negative GRAVY value for ArDH describes it to be hydrophilic in nature. Subcellular localization was predicted by Cello 4 MEMPACK from UCL was used to study the transmembrane topology. These results showed that ArDH protein has two trans-membrane domains and it is a cytoplasmic protein having two transmembrane regions: Region I consists of 29-44 amino acid and Region II 59-77 amino acid. This shows that ArDH has membrane binding properties and can be involved in transport of materials across the cell membrane.

Multiple Sequence Alignment Results

The identification of catalytic residues is a key to understanding the function of enzymes. With the information from other functionally similar sequences with known crystallographic structures we can identify the key catalytic residues. ClastalW2 server was used for multiple sequence alignment of ArDH with other dehydrogenases from *Mycobacterium marinum* (PDB Id: 3R1I), *Candida parapsilosis* (PDB Id: 3CTM) and *Thermotoga maritima* (PDB Id: 1VL8) shown in **Table 2 (see supplementary material).** The compared sequences varied in length but essentially conserved the key catalytic residues which have been highlighted with an asterisk (*) symbol.

Secondary structure prediction

Structure prediction was done by PsiPred and SOPMA secondary structure prediction server. The results from SOPMA are shown in **Table 3 (see supplementary material).** High percentage of helices in the structure makes the protein more flexible for folding, which might increase protein interactions. Moreover the predicted secondary structural information of ArDH was considered to improve the target-template alignment and for building 3D model of the ArDH.

Homology Modeling and Model Evaluation Results

Homology modeling was done by Modeller v9.11. The templates was selected (PDB ID: 3R1I) on CPH server. Alignment between template and model with secondary structure elements is shown in **(Figure 1A)**. Ten models were developed using Modeller v9.11 and the best model was selected on the basis of structural evaluation and stereochemical analyses by using Errat, ProSA-web Z-scores and PROCHECK Ramachandran plots (Predicted model is shown in **Figure 1B**). Superimposition of the model with the template (used as reference structure) by using UCSF Chimera v1.7.3 program **(Figure 1C)** showed a very low RMSD of 0.241 A°, proved a high similarity between them. Z-score of 7.930 indicated that both template and target protein have similar folds. Verify3D analyzes the compatibility of an atomic model (3D) with its own amino acid sequence. Verify 3D analysis revealed that 86.71 % of the residues had an average 3D-1D score of <0.2, predicting

that the model is well compatible with its sequence. Ramachandran plot was obtained from ProCheck server, showed that 91% of residues were in most favored region **(Figure 1D)**. The predicted model was submitted to Protein Model Database (PMD) and has been assigned the following PMID 0078880. Furthermore, the predicted model of (ArDH) was used to study enzyme substrate interactions by docking analysis.

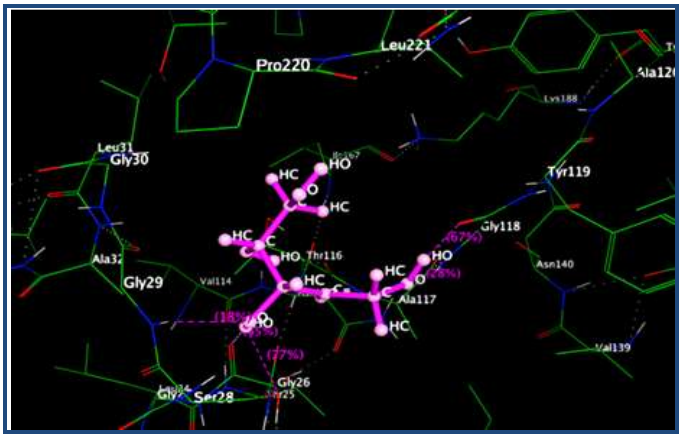

Figure 2: Molecular docking analysis by MOE software. Depicted view of interactions between a ligand (in magenta color) and ArDH atoms, dotted lines are showing the interactions.

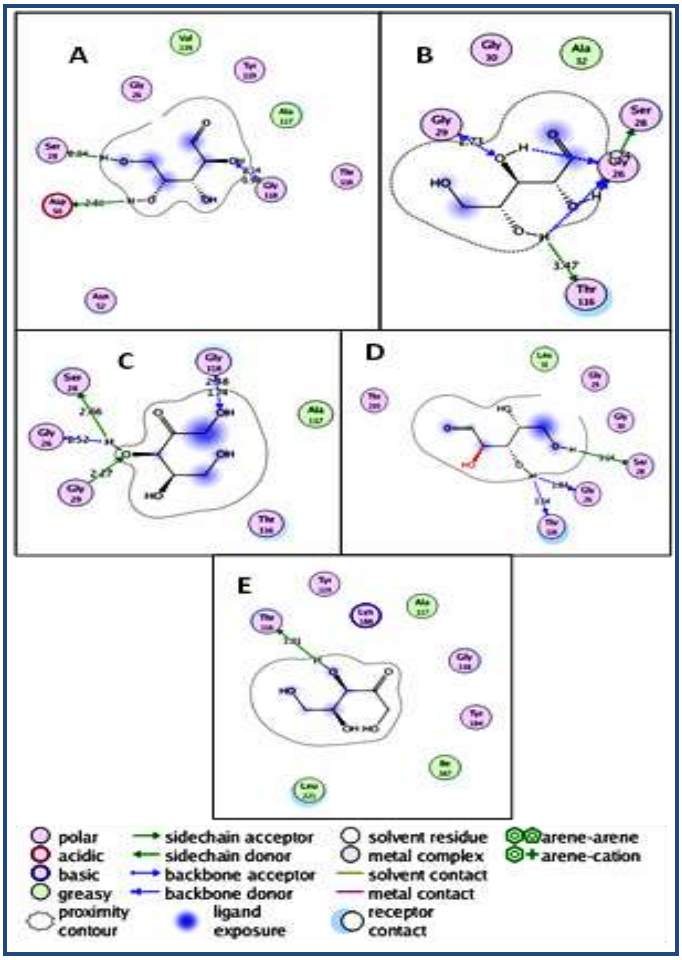

Figure 3: Ligand Interaction 2D diagram with calculated distances (Aº) showing interactions of ArDH atoms with **(A)** Arabinose **(B)** D-ribose **(C)** D-ribulose **(D)** Xylose **(E)** D-xylulose.

Docking Analysis

All substrate structures were docked with the active site of ArDH enzyme and top ranked conformations were obtained through MOE S-score. Depicted view of interaction is shown in **(Figure 2).** The values of estimated free energies of binding are as Arabinose -9.8224 Kcal/mol, d-ribose -11.3701 Kcal/mol, d-ribulose 8.9230 Kcal/mol, xylose -9.7007 Kcal/mol and d-xylulose 9.7802 Kcal/mol. These results clearly indicate that d-ribose is best substrate for the production of sugar alcohols by using ArDH. The interacting atoms and calculated distance between these, for each substrate are shown in **(Figure 3).**

Conclusion:

Present study explored the physicochemical nature, three dimensional structure and detail of interactions with substrates of ArDH enzyme by using docking analysis for very first time. Because of sweet taste and pharmaceutical applications the demand by sugar alcohols/polyols is increasing day by day. From our analysis it is revealed that d-ribose showed lower value of free energy of binding hence, it can be concluded that it can bind with ArDH more efficiently and can be a best substrate for the production of sugar alcohols. The result from this study will eventually give a better idea to experimentalist and Industrialists to design more efficient fermentation process using ArDH enzyme and d-ribose its substrate with higher yields of sugar alcohols in cost effective way. Additionally, on the basis of this study mutational analysis of ArDH can be done to enhance product yield.

Acknowledgment:

The authors would like to thank Dr. Zakir Hussain, Vice Chancellor, GC University Faisalabad, Pakistan. for providing support for completion of this project.

References:

[1] Koganti S *et al. Appl Microbiol Biotechnol.* 2011 **90:** 257 [PMID: 21127857].
[2] Suzuki S *et al. Biosci Biotechnol Biochem.* 2002 **66:** 2614 [PMID: 12596856]
[3] Link T *et al. Biochem J.* 2005 **389:** 289 [PMID: 15796718]
[4] LaFayette PR *et al. Plant Cell Rep.* 2005 **24:** 596 [PMID: 16151815]
[5] McNeil SD *et al. Plant Physiol.* 1999 **120:** 945 [PMID: 10444077]
[6] Diamant S *et al. J Biol Chem.* 2001 **276:** 39586 [PMID: 11517217]
[7] Chen X *et al. Int J Biol Sci.* 2010 **6:** 834 [PMID: 21179590]
[8] Ma T *et al. J Med Chem.* 1996 **39:** 2835 [PMID: 8709113]
[9] Granstrom TB *et al. Appl Microbiol Biotechnol.* 2007 **74:** 273 [PMID: 17216458]
[10] Gasteiger E *et al. Nucleic Acids Res.* 2003 **31:** 3784 [PMID: 12824418]
[11] Emanuelsson O *et al. Nat Protoc.* 2007 **2:** 953 [PMID: 17446895]
[12] Nugent T *et al. Bioinformatics.* 2011 **27:** 1438 [PMID: 21349872]
[13] Thompson JD *et al. Nucleic Acids Res.* 1994 **22:** 4673 [PMID: 7984417]

[14] Geourjon C & Deleage G, *Comput Appl Biosci.* 1995 **11:** 681 [PMID: 8808585]
[15] McGuffin LJ *et al. Bioinformatics.* 2000 **16:** 404 [PMID: 10869041]
[16] Batool M *et al. Bioinformation.* 2011 **7:** 384 [PMID: 22347779].
[17] Nielsen M *et al. Nucleic Acids Res.* 2010 **38:** W576 [PMID: 20542909]
[18] Fiser A & Sali A, *Methods Enzymol.* 2003 **374:** 461 [PMID: 14696385]
[19] Kaplan W & Littlejohn TG, *Brief Bioinform.* 2001 **2:** 195 [PMID: 11465736]
[20] Wiederstein M & Sippl MJ, *Nucleic Acids Res.* 2007 **35:** W407 [PMID: 17517781]
[21] Benkert P *et al. Nucleic Acids Res.* 2009 **37:** W510 [PMID: 19429685]
[22] Laskowski RA *et al. J Biomol NMR.* 1996 **8:** 477 [PMID: 9008363]
[23] Pettersen EF *et al. J Comput Chem.* 2004 **25:** 1605 [PMID: 15264254]
[24] Roy U & Luck LA, *Biochem Mol Biol Educ.* 2007 **35:** 238 [PMID: 21591100]
[25] Ikai A, *J Biochem.* 1980 **88:** 1895 [PMID: 7462208]

Edited by P Kangueane
Citation: **Sarwar *et al.*** Bioinformation 9(19): 952-957 (2013)

BIOINFORMATION

open access

Supplementary material:

Table 1: Pysicochemical properties of ArDH predicted by ProtParam program

Sr.No.	Parameters	Values
1	Molecular weight	30642.8 Dalton
2	Theoretical pI	5.59
3	Extinction coefficient*	51130 at Abs0.1% 1.669
4	Instability Index	44.16
5	Aliphatic index	85.77
6	Grand average of hydropathicity(GRAVY)	-0.118

Extinction Coefficient units $M^{-1}cm^{-1}$ at 260 nm

Table 2: CLUSTAL 2.1 Multiple sequence alignment of ArDH protein from *(Candida Albican)* with dehydrogenases: from *Mycobacterium Marinum* (PDB Id:3R1I), *Candida Parapsilosis* (PDB Id:3CTM) and *Thermotoga Maritima* (PDB Id:1VL8)

```
CLUSTAL 2.1 multiple sequence alignment

ArDH            ----------------MDSAYWSYDNIVPSFRLDGKLVILTGGSGGLAAVVSRALLAKGA
3R1I|A          ---MAHHHHHHMGTLEAQTQGPGSMSVLDLFDLSGKRALITGASTGIGKKVALAYAEAGA
3CTM|A          MGEIESYCNKELGPLPTKAP-TLSKNVLDLFSLKGKVASVTGSSGGIGWAVAEAYAQAGA
1VL8|A          --------------MGSDKIHHHHHHHMKEVFDLRGRVALVTGGSRGLGFGIAQGLAEAGC
                                .        :   * * *: . :**.* *:.  :: .    *.
ArDH            DVALVDMNLERTQQAARDVLQWGEEQMKGKYESPIGQVSAWSCNIGDAEAVDLTFKAINE
3R1I|A          QVAVAARHSDALQVVA---------DEIAGVG---GKALPIRCDVTQPDQVRGMLDQMTG
3CTM|A          DVAIWYNSHPADEKAE---------HLQKTYG---VHSKAYKCNISDPKSVEETISQQEK
1VL8|A          SVVVASRNLEEASEAA--------QKLTEKYG---VETMAFRCDVSNYEEVKKLLEAVKE
                .*.:         . .          .          .   .  *:: : . *   :.
ArDH            HHGKISSVLVNTAGYAENFPAEEYP-AKNAENLMKVNGLGSFYVS-QAFARPLIQNNMTG
3R1I|A          ELGGIDIAVCNAGIVSVQA-MLDMP-LEEFQRIQDTNVTGVFLTA-QAAARAMVDQGLGG
3CTM|A          DFGTIDVFVANAGVTWTQGPEIDVDNYDSWNKIISVDLNGVYYCS-HNIGKIFKKNGKG-
1VL8|A          KFGKLDTV-VNAAGINRRHPAEEFP-LDEFRQVIEVNLFGTYYVCREAFS--LLRESDNP
                . * :.    *:.    .    :    .. ..: ..:  * :  . .  .  :  :.
ArDH            SIILIGSMSGTIVNDPQPQCMYNMSKAGVIHLARSLACEWAKYNIRVNTLSPGYILTPLT
3R1I|A          TIITTASMSGHIINIPQQVSHYCTSKAAVVHLTKAMAVELAPHQIRVNSVSPGYIRTELV
3CTM|A          SLIITSSISGKIVNIPQLQAPYNTAKAACTHLAKSLAIEWAP-FARVNTISPGYIDTDIT
1VL8|A          SIINIGSLTVEEVTMPN-ISAYAASKGGVASLTKALAKEWGRYGIRVNVIAPGWYRTKMT
                ::*  .*::   :. *:  . *  :*..   *::::* * .    *** ::**:  * :.
ArDH            RNVISGHTEMKTEWESKIPMKRMAEPKEFVGSILYLASESASSYTTGHNLVVDGGYECW
3R1I|A          EPLA----DYHALWEPKIPLGRMGRPEELTGLYLYLASAASS-YMTGSDIVIDGGYTCP
3CTM|A          DFASK---DMKAKWWQLTPLGREGLTQELVGGYLYLASNAST-FTTGSDVVIDGGYTCP
1VL8|A          EAVFS-DPEKLDYMLKRIPLGRTGVPEDLKGVAVFLASEEAK-YVTGQIIFVDGGWTAN
                        :         *: * . .::: *  ::***  :. : **  :.:***: .
```

Table 3: Secondary structure elements as predicted by SOPMA for ArDH

Secondary Structure Element	No. of Amino acid involved	Percentage
Alpha helix	110	39.15%
Extended strand	58	20.64%
Beta turn	29	10.32%
Random coil	84	29.89%