Structure Prediction of Oxidoreductase from *Anoxybacillus sp.* SK3-4 using Computational Methods

[*1]Musa A. A Diso; [3]Umar Shittu; [2]Umar L. M; [1]Sani K. S & [1]Aisha A. H

[1]Department of Science Laboratory Technology, Kano State Polytechnic, Nigeria.
[2]Department of Pre-ND Science and Technology, S.G.S, Kano State Polytechnic, Nigeria.
[3]Department of Biology, Isa Kaita College of Education, Dutsin-ma, Katsina State, Nigeria.
**Corresponding author's e-mail: aamusadiso@yahoo.com, +2348061619400*

Abstract: Determining the structure and function of a novel protein is a cornerstone of many aspects of modern biology. For this study, the structure of a unique enzyme from *Anoxybacillus sp.* SK3-4 was predicted and identified based on its primary sequence analysis in which various computational tools were used. The structure was predicted to be identified as Oxidoreductase which belongs to alcohol dehydrogenase (ADH) family or iron-containing alcohol dehydrogenase (Fe-ADH) super-family after when the sequence is being analysed by Uniprot databases in which a maximum identity similarity of 100% was obtained with organism *Anoxybacillus sp.* SK3-4. Characteristic features of the enzyme were also predicted by computing both physical and the chemical parameters of the primary sequence using ProtParam software in Expasy. Active-sites, conserved regions or domains, secondary structure, 3D-model and transmembrane were all predicted using different software programs. Cn3D and Jalview software were also used to visualize the nature of the aligned sequences and 3D-model. Therefore, in modern science and biotechnology, structure and functions of many proteins can be predicted by analysing their sequences using various tools in bioinformatics.

Keywords: *Anoxybacillus sp.; Bioinformatics tools; Primary sequence; Oxidoreductase; Secondary structure; 3D-model*

1. Introduction

Protein structure prediction is the inference of the three-dimensional (3D) structure of a protein from its amino acid sequence—that is, the prediction of its folding and its secondary and tertiary structure from its primary structure (*Majorek et al., 2008)*. Nowadays, over six million unique protein sequences have been deposited in the public databases, and this number is growing enormously (http://www.ncbi.nlm.nih.gov/RefSeq/). However, despite the progress of high-throughput structural genomics initiatives, just over 50,000 protein structures have so far been experimentally determined (Lawrence & Michael, 2009). This rapid disparity between the number of sequences and structures has driven research toward computational methods for predicting protein structure from sequence. Computational methods grounded in simulation of the folding process using only the sequence itself as input have been pursued for decades and are showing some progress (CAPS, 2007).

The most successful general approach for predicting the structure of proteins involves the detection of homologs of known three-dimensional (3D) structure—the so-called template-based homology modelling. Given a protein sequence of interest, one may compare this sequence with the sequences of proteins with experimentally determined structures and if a homolog can be found, an alignment of the two sequences can be generated and used directly to build a 3D model of the sequence of interest (Fiser and Sali, 2001; Kryshtafovych and Fidelis, 2009*).* For this study, a unique protein (oxidoreductase) was predicted from the sequence of amino acid (386 length) after undergoing some analysis using various computational tools (**Uniprot** and **NCBI databases)**. It was also found that oxidoreductase belongs to the family butanol dehydrogenase after specific hits and super-family Iron-containing alcohol dehydrogenase (Fe-ADH). The source origin organism is ***Anoxybacillus sp.* SK3-4** with taxonomic identifier 654421 [NCBI]. The species of *Anoxybacillus* are widespread in geothermal springs, manure, and milk-processing plants. The genus is composed of 22 species and two subspecies among others (Goh *et al.,* 2013). Thus, *Bacillaceae* family remains an important microbial contributor to industrial biotechnology, mainly due to its abundance of useful proteins and enzymes (Goh *et al.,* 2014).

Oxidoreductases are generally enzymes that part take in oxidation-reduction reactions in biological systems naturally. As nature's own catalysts, enzymes possess very diverse specificity, reactivity, and other physiochemical, catalytic, and biological properties highly desirable for various industrial and medical applications. Now enzymes are being vigorously and systematically developed, as economically viable and industrial biocatalysts along with the fast advancement and expansion Modern Science and Biotechnology (Bommarius, 2004).

One of the **uniqueness of oxidoreductases** is that, they are widely distributed among microbes, plants and animals. They are found to exist in catalysing the exchange of electrons or redox equivalents between donor and acceptor molecules, in reactions involving electron transfer, proton abstraction, hydride transfer, oxygen insertion, or other key steps. Another special feature about oxidoreductases is that they employ various **redox active centres**

(Munro, 2000). The centres are shielded by the polypeptide backbone of oxidoreductases, which can modulate their selectivity, redox potency, reactivity, stability, and inhibition-resistance. Common redox centres include **amino acid residues** (e.g. cystein and tyrosine), **metal ions** (e.g. Fe, Cu, Fe-S cluster or heme), and **co-enzymes** (e.g. NADP, FAD, FMN e.t.c).

Among metal ions or complexes, Zn and Fe metals ions are special in that they are present more than other ions in functionally important locations within proteins. Some of these functions include residue binding, catalysis, structural stability and regulation (as in Fe-containing alcohol dehydrogenase).

So generally, ADH catalyses the reversible oxidation of alcohol to acetaldehyde with simultaneous reduction of $NAD(P)^+$ to NAD(P)H. Oxidoreductases can be classified according to their sequence or three-dimensional structure, which is very informative for the study of structure-function relationship, enzyme relationship and functional genomics **(**Munro *et al.,* 2000). This study was aimed to predict and identify the structure of a unique protein from *Anoxybacillus sp.* SK3-4 by analysing its primary sequence using various tools in Bioinformatics.

2. Materials and Methods

In the first place, **Bioinformatics** is an interdisciplinary field that develops and improves on methods for storing, retrieving, organizing and analysing biological data. A major activity in bioinformatics is to develop **software tools** to generate useful biological knowledge. Bioinformatics uses many areas of computer science, mathematics and engineering to process biological data. Complex machines are used to read in biological data at a much faster rate than before. Databases and information systems (Uniprot databases) are used to store and organize biological data (Fiser *et al.,* 2000).

UniprotKB and NCBI databases provide the scientific community with a comprehensive high quality and freely accessible resource of protein sequence and functional information (Majorek *et al.,* 2008). Uniprot tools have been categorised into (i) **similarity search tools** (Fasta, Blast, ScanProsite, e.t.c), (ii) **multiple sequence alignment tools** (Clustal W, Megasoftware e.t.c), (iii) **batch retrieval tools and** (iv) **proteomics tools** such as Protparam tool which allows the computation of various physical and chemical parameters for a given protein stored in SwissProt or for a user entered sequence. For this study, Blast and ScanPosite tools were used in Uniprot for the identification of query sequence (386 length) and similarity against a template chosen sequence (known protein).

```
>Query
MENFIFHNPTKLIFGRGQIEHLKKELHSYEHILIVYGGGSIKKNGVYDDVVSILRSLNKS
WSELAGVEPNPRLSTVQKGIHICREEKVDFILAVGGGSVIDCAKAIAAGALYDGEAWDFI
SRKATVERALPIGTVLTLAATGSEMNANSVITNWETKEKYGWSSPAVFPQFSILDPVYTT
TVPKDHTVYGIVDIMSHVLEQYFHHAPNTPLQDRMCEAILRTVIETAPKLIEDLQNVDHR
ETILYCGTMALNGILRMGLRGDWATHNIEHAVSAVHDIPHAGGLAILFPNWMKHVLDEHI
DRFKQLAVRVFDVYPEGKGDREIALEGIEKLRAFWNRLGAPCRLADYHIGEESLPIIVEK
AMAFGPFGNFKKLHHDDVMTILQASL
```

Figure 1 Query Sequence

2.1 Query Sequence Identification

Blasting was performed using computer system by copying, pasting and submitting the query sequence as an input in which after some seconds to minutes the result delivery will be retrieved as output. This will be explained fully in results and discussions chapter. For blastp (mainly for protein), the following links were used for the analysis: http://blast.ncbi.nlm.nih.gov/, http://www.uniprot.org/, and http: //web.expasy.org/blast/.

For multiple alignment of the identified query protein sequence, **Clustal W2 software** was used to compare the percentage similarity between the identified query sequence with the known protein sequences of the same family. P**rotParam software** was also used to compute the various physical and chemical parameters of the query protein. Such includes molecular weight (MW), amino acid composition, hydrophaticity, theoretical Ip, 5instability index, extinction coefficient, estimated half-life, e.t.c (Goh *et al.,* 2012).

2.2 Secondary Structure Prediction

Expasy tools such as CSFFP (Chou and Fasman) and GOR4 were utilised to predict the secondary structure of the protein (Edwards, 2003). SSpro8-SCRATCH is also another server for predicting the secondary structure, 3D structures and other structural features. The SCRATCH software suite includes predictors for relative solvent accessibility, disordered regions, domains, disulfide bridges, single mutation stability, residue contacts versus average, individual residue contacts and tertiary structure. The user for SSpro8 simply provides an amino acid sequence and selects the desired predictions, then submits to the server. The server is available at http://www.igb.edu/server/press.html. Different tools were used to determine the most clearly understandable feature of the structure.

2.3 3D-Structure Prediction

Researchers in biology are facing one of the frequent problems which are the functional characterization of a protein sequence that is usually related to the accurate three dimensional (3D) structure of the studied protein. In the absence of an experimentally determined

structure, comparative or homology modelling is responsible for providing a useful 3D model for a protein (target) that is related to at least one known protein structure, the template. The field of protein structure prediction is based on two classes, known as template-based homology modelling and template-free, *de novo* modelling (Fiser *et al.*, 2000). Sites are offered for calculating and displaying the 3-D structure of oligosaccharides and proteins. With two protein analysis sites, the query protein is compared with existing protein structures as revealed through homology analysis.

The selection of the template sequence is based on "template selection rule", that the template is considered as a "good" template if the sequence identity between query and template sequence is at least 30%. But, it is better to get the target-template sequence identity above 40% for accurate alignment. If the target-template sequence identity is below 40%, there is a possibility of having gaps in the alignment and needs manual interventions to minimize the number of misaligned residues (Fiser *et al.*, 2001).

Several web servers are used to generate a 3D model of protein sequences. For this study, we are restricted to only three servers: *Phyres2, Swiss-model,* and *CPH model.* ROSETTA is also available online (Eswar *et al.*, 2006). These models are used most often to predict the 3D structure of large molecules such as proteins and nucleic acids. These data typically obtained by X-ray crystallography or NMR spectroscopy (Lawrence & Michael, 2009; McGuffin and Bryson, 2000; Qian *et al.*, 2007). Swiss model results come by E-mail usually and require a viewer such as Swiss Pdb Viewer, Rasmol, Cn3D v3.0 or WebMol Java PDB Viewer to visualize (Fiser and Sali, 2001; Kelley and Sternberg, 2009).

'Normal' model modelling by Phyres2 produces a set of potential 3D models based on alignment to know protein structures. The pipeline involves detecting sequence homologues with PSI-blast, predicting secondary structure with Psi-pred and Diso-pred, construct 3D model of your protein based on the alignment between the HMM of your sequence and the HMMs of known structure (Luthy *et al.*, 1992; Kelley and Sternberg, 2009). Also can be used in binding site and trans-membrane predictions.

A lots of evaluation, validation and refinement tools are available for possible error cross-checking after the model is built such includes Verfy_3D, Errat, Procheck, Ramachandran plot, What if, e.t.c (Goh *et al.*, 2012). For this study, *Errat, Ramachandran plot,* and *Verify_3D validation software* were used to check for any possible error during the homology modelling. Thus, the validation tools confirmed us with the best predictor in determining the 3D structures built by Phyres2 model, Swiss-model, and CPH model.

3.0 Results and Discussion

After undergoing several events for query sequence analysis using various bioinformatics tools in Uniprot and NCBI, the protein name, origin, structure and its important characteristic properties have been identified as follows:

3.1 Primary Sequence identification

Based on 'blastp' result in Uniprot, the protein sequence shows 100% similarity to **oxidoreductase** from ***Anoxybacillus sp*. SK3-4** with taxonomic identifier 654421 [NCBI]. Oxidoreductase is an enzyme that catalyses the conversion of aldehyde to alcohol with the co-factor NAD(P)H being oxidized in the process. This protein family belongs to the so called iron-containing alcohol dehydrogenase super family. Since members of this family use different divalent ions preferentially iron or zinc, it has been suggested to be renamed to family III metal-dependent polyol dehydrogenases.

Detailed BLAST results Customize

› Show hits from complete proteomes only

Alignments	Entry	Entry name	Status	Protein names	Organism	Length	Identity	Score	E-value	Gene names
	T0CQH6	T0CQH6_9BACI		Oxidoreductase	Anoxybacillus sp. SK3-4	386	100.0%	2.041	0.0	C289_0009
	M5QTD9	M5QTD9_9BACI		Oxidoreductase	Anoxybacillus sp. DT3-1	386	95.0%	1.954	0.0	F510_1831
	M5JHE4	M5JHE4_9BACI		NADH-dependent butanol dehydrogenase A	Anoxybacillus flavithermus TNO-09.006	386	91.0%	1.871	0.0	bdhA AF6_2307
	M8CTZ6	M8CTZ6_9BACI		Oxidoreductase	Anoxybacillus flavithermus AK1	386	91.0%	1.862	0.0	H919_12289
	B7GEU5	B7GEU5_ANOFW		Uncharacterized oxidoreductase, Fe-dependent ...	Anoxybacillus flavithermus (strain DSM 21510 / WK1)	386	89.0%	1.830	0.0	Afw_2386
	R4G182	R4G182_9BACI		Oxidoreductase	Anoxybacillus flavithermus NBRC 109594	386	89.0%	1.830	0.0	KN10_2109
	F8D014	F8D014_GEOTC		Alcohol dehydrogenase (NADP(+))	Geobacillus thermoglucosidasius (strain C56-YS93)	387	75.0%	1.571	0.0	Geoth_0631
	E3ICS7	E3ICS7_GEOS0		Iron-containing alcohol dehydrogenase	Geobacillus sp. (strain Y4.1MC1)	387	75.0%	1.571	0.0	GY4MC1_0561

Figure 3.1.1 Results showing organism having the highest maximum identity score with the query amino acid sequence which shows **100%** similarity with ***Oxidoreductase [Anoxybacillus sp*. SK3-4]**

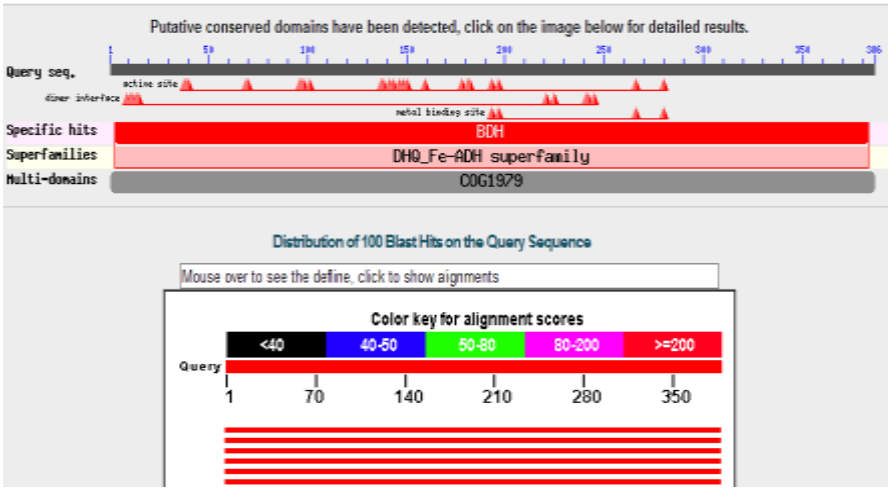

Figure 3.1.2 Results obtained from blast (http://blast.ncbi.nml.nih.gov/blast)

Putative conserved domains are detected as seen in "fig. 3.1.1" above in which active sites, dimer interfaces, and metal binding sites are located all along the sequence. The active site residues constitute 24 of 24, dimer interface 20 of 10, and metal binding site 4 of 4 of the conserved domains.

According to blast results obtained from ScanProsite program, the metal binding sites and locations were indicated to be at amino acid positions **172-200** and **263-283** of query sequence as seen in *(*fig. 3.1.3).

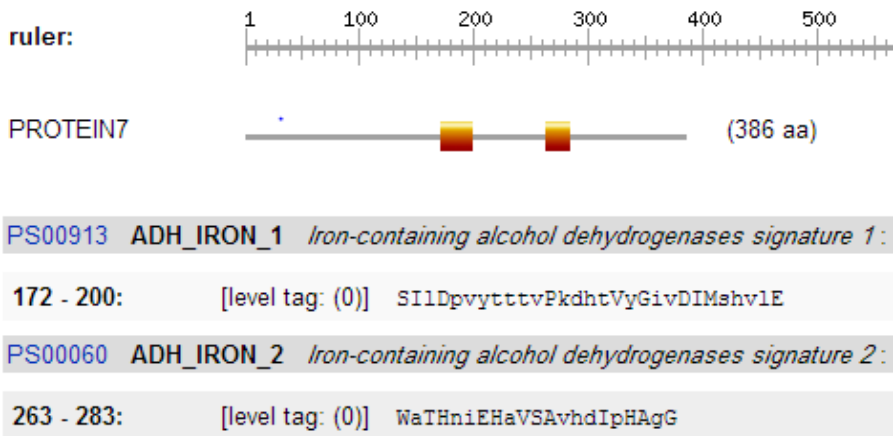

Figure 3.1.3 Results obtained from ScanProsite software showing metal binding site

ProtParam software in Uniprot has also predicted some important physical and chemical parameters of the amino acid sequence such includes, molecular weight (MW), amino acid composition, hydrophaticity, theoretical Ip, instability index, extinction coefficient, estimated half-life (table 3.1).

Table 3.1 Computed Physical and Chemical Parameters obtained from ProtParam-Expasy tool

S/N	Computed Parameters of the sequence of amino acids	Values
1	Number of amino acids	386
2	Molecular weight	43137.5
3	Theoretical pI	6.01
4	Number of negatively charged residue (Asp + Glu)	48
5	Number of positively charged residue (Arg + Lys)	38
6	Ext. Coefficient	55140m^{-1}
7	Instability index/Aliphatic index	39.4/0.074

As shown in table 3.1, the instability index computed was 39.4 and this means that the protein molecule is stable since the value is not greater than 40.0. The grand average of hydrophaticity value computed is -0.074 meaning the protein is hydrophobic (because the gravy value computed was negative) as seen in the simple diagram below (fig. 3.1.4). The molecular formular of the protein is $C_{1948}H_{3038}N_{526}O_{554}S_{14}$ with total number of atoms 6078 as computed by ProtParam software.

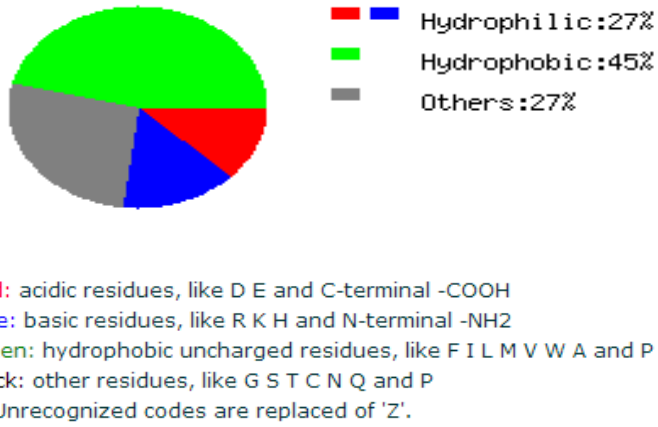

Figure 3.1.4 A sample diagram showing different parts of the protein as computed from ProteinBioedit software

3.2 Multiple Sequence Alignment

Jalview Clustal-W software provided a clear vision to analyse the matching regions between the query sequence and the template sequence chosen. It shows both conserved and non-conserved regions on the sequence.

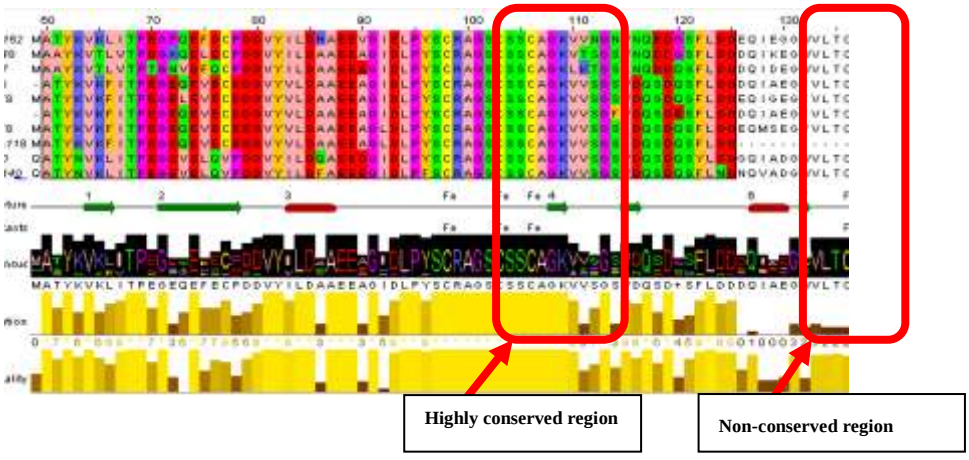

Figure 3.2 Multiple sequence alignment results from Jalview Clustal-W showing conserved and non-conserved regions

Based on the reports of Goh *et al.*, (2012) that, jalview is an important software for phylogenetic tree construction. For this unique protein, the phylogenetic tree constructed possesses high similarity towards *Anoxybacillus sp. SK3-4*. The taxonomic lineage is Bacteria>Firmicutes>Bacilli>Bacillales>Bacillaceae> Anoxybacillus. This result is similar to the result shown in BLAST search. Important amino acid residues of the enzyme which help

in metal ion-binding for stability and catalytic activity were also predicted using 3DLigandSite program such includes, **Leu, His, Asp, Glu, Asn, Val, Pro, Ser, Thr, Tyr, and Lys.** Leu, Val, Asp, Thr and Val were predicted to function as catalytic residues for substrate binding. His were found to have metal-ion binding activity (Fe or Zn) for enzyme structure stability. Heterogens such as **NAD, NAP, NDP, FE** and **ZN** were also found present in predicted binding sites.

3.3 Secondary Structure Prediction

In Chou-Fasman analysis, a helix, a beta strand, and a turn in the protein were all predicted. The method is at most about 50–60% accurate in identifying correct secondary structures, which is significantly less accurate than the modern machine learning-based techniques. . The method is based on analysis of the relative frequencies of each amino acid in alpha helices, beta sheets, and turns.

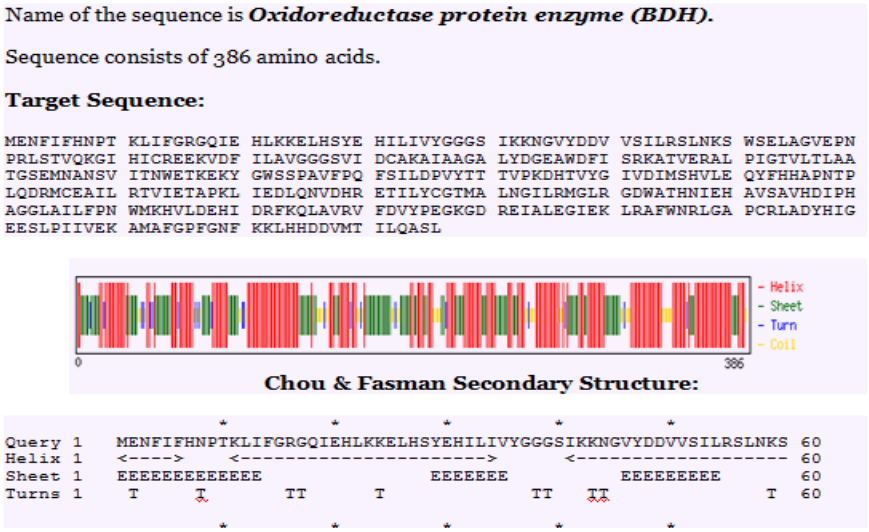

Figure 3.3.1 Results from Chou-Fasman analysis showing helix, sheet, turn and coil in different colours

GOR method takes into account not only the propensities of individual amino acids to form particular secondary structures, but also the conditional probability of the amino acid to form the secondary structure.

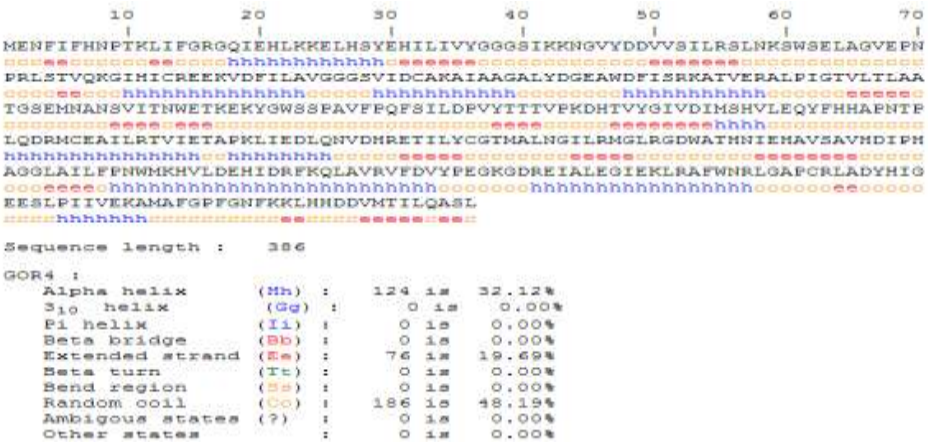

Figure 3.3.2 Secondary structure prediction from GOR4 method showing the mount of alpha helix, strand, and random coil calculated

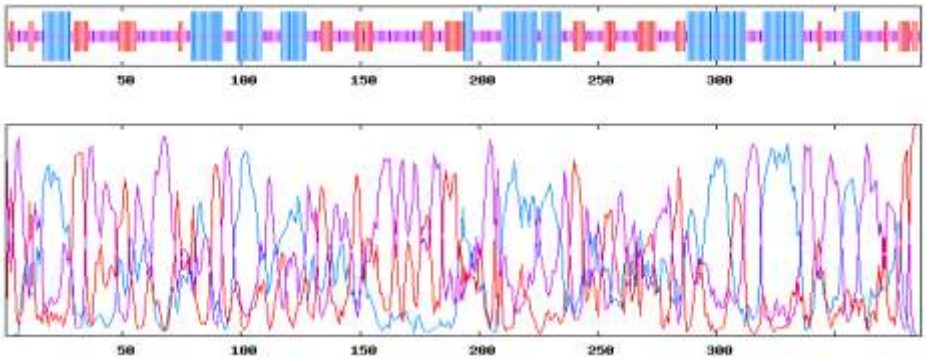

Figure 3.3.3 Graphical results from GOR4 analysis providing better comparison

For SSpro8 software, instead of using three classes which are helix, beta sheet, and the rest to assign the secondary structure of a protein, SSpro8 adopts the full DSSP 8-class output classification including H: alpha-helix, G: 3-10 helix, I: Pi-helix, (extremely rare), E: extended strand, B: beta-bridge, T: turn, S: bend, C: the rest.

One major drawback with SSpro8 method is that the result did not provide us with any visualized material. So it is hard to know if anything wrong with the prediction. And for Chou-Fasman method, the limitations are due to low accuracy, unreliable parameters, and over prediction. For GOR4 method, apart from providing useful data information about alpha helix, pi helix, beta bridge, beta turn, and random coil, also the colour and the graph of the prediction results would be a great help to compare and contrast with. Therefore, based on the above mentioned features of the methods, we best recommended GOR4 as the most reliable of the three methods used.

3.4 3D-Structures Prediction

The three models used for the study (Phyres2, CPH model, and Swiss-model), all provided better results for the prediction as the target-template sequence identity value

obtained for all the methods was above 40%. For Swiss model was 47.23%, Phyres2-model 46.0% and for CPH-model was 45.3%.

Figure 3.4.1 A model built by Swiss-model with sequence identity **47.23%** against a template 1vIj.1.A

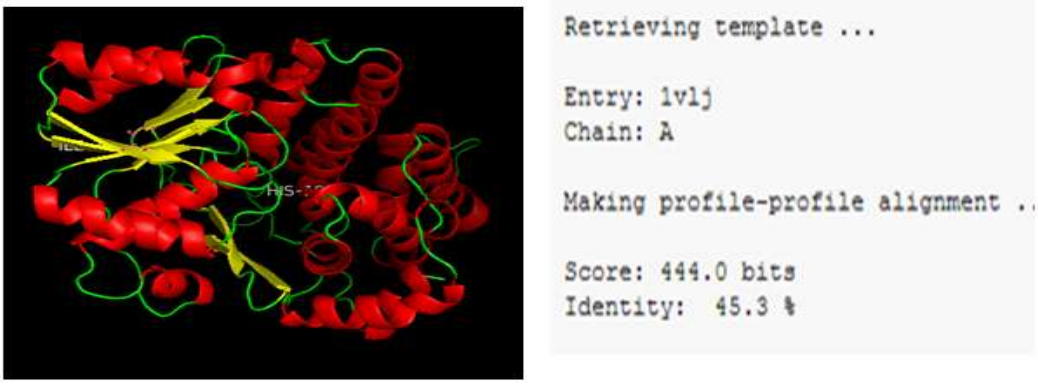

Figure 3.4.2 A model built by CPH-model with sequence identity **45.3%** against a template 1vIj

Figure 3.4.3 3D model built by Phyres2 software with sequence identity **46.0%** against a template used d1vIja_

As shown from the 3D models built with different percentage values in template-identity similarity, the Swiss-model has the highest target-template identity of **47.23%** with highest

resolution (1.78Å), followed by Phyres2-model with identity **46.0%** which leads to accurate alignment. Thus, Swiss-model and Phyres2 software happened to be the best predictors to our model which were also verified using ERRAT, Verify_3D, and Ramachandran plot software.

Table 3.2 Showing results of validation and refinement using ERRAT, Verify_3D, and Ramachandran plot software for the three models

Validation software	Swiss-model	Phyres2 model	CPH-model
ERRAT (%)	74.536	63.298	89.276
Verify_3D (%)	96.94	95.38	94.98
Ramachandran plot (%)	93.5	96.4	94.5

Ramachandran plot shows a better result by checking the stereo-chemical conformational angles of the model structure in which the residue percentage in the most favoured region for Swiss-model was **93.5%**, for Phyres2 was **96.4%** and for CPH-model was **94.5%**. Thus, leads to accuracy in model construction (model building).

However, some variations in refinement and validation results were noticed in which CPH-model was predicted to have the highest quality model of **89.276%** obtained from ERRAT software. And Phyres2 model got to have the highest quality model of **96.5%** from Ramachandran plot. Hence, only small errors were predicted in both models which lead to accuracy in the homology modelling exercise. Active sites and Trans-membrane were also predicted by ProteinBioedit software and Phyres2 model with accuracy over 90% in a cross-validation experiment for active sites and average accuracy of 89% for transmembrane on the analysed protein sequence.

For 3D models, you may use a pdb-file viewer software like Swiss-PDB viewer, PyMol, Rasmol, Cn3D e.t.c to visualize the position of important amino acids on the 3D structure.

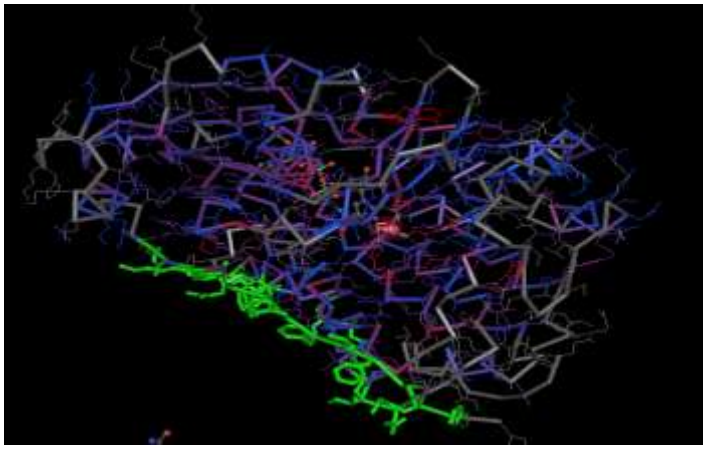

Figure 3.4.4 3D model generated by Cn3D viewer with iron-binding site at the center

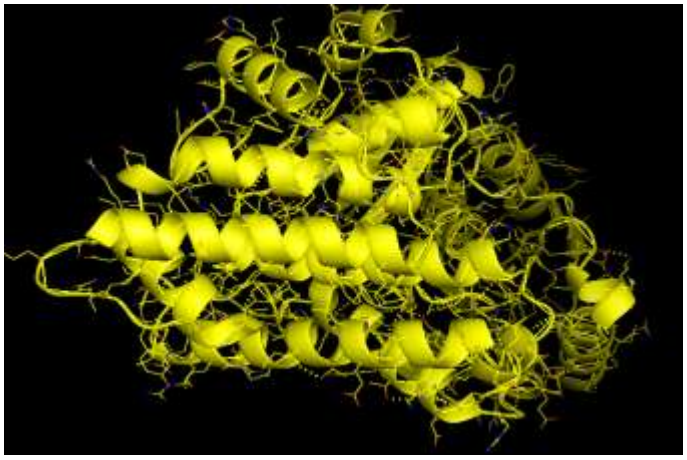

Figure 3.4.5 Swiss-3D-model (cartoon view) generated by PyMol viewer

4. Conclusion

As observed from the results being analysed by different bioinformatics tools, the structure of a unique protein from *Anoxybacillus sp.* SK3-4 was identified to be oxidoreductase achieved through "blastp" prediction method in UniprotKB and NCBI in which the sequence identity similarity shows maximum score of 100% with the organism *Anoxybacillus sp.* SK3-4 (fig. 3.1.1), together with taxonomic lineage of the organism as predicted by jalview-Clustal W software in multiple sequence alignment. Hence, jalview is very important for visualizing conserved (which contains the domains or the motifs) and non-conserved regions in a protein..

Many of the important characteristics of the protein sequence have been predicted by computing both physical and the chemical parameters of the query sequence by ProtParam software in Expasy and ProteinBioedit program (fig. 3.1.4 and table 3.1.).

Chou-Fasman, GOR4, and SSpro8 software prediction methods have been found to predict the secondary structure of the query sequence (oxidoreductase) in which GOR4 software was recommended as the best program in the current study due to the reason being that it provided more precise and most reliable results (fig 3.3.3).

The 3D-model structures built by **Swiss-model** and **Phyres2 software** were found to be the best among the three predictors due to the reason being that the target-template sequence identity value obtained for Swiss-model was 47.23% and for Phyres2 was 46.0%, and all were best verified and validated by Ramachandran plot (table 3.2).

The active sites predicted by ProteinBioedit with accuracy over 90% are usually found in a 3-D groove or pockets of the enzyme, lined with amino acid residues. Therefore, the function of this protein is attributed to the presence of these active sites. Transmembrane helix of the protein was also predicted with average accuracy of 89% using Phyres2 model, hence the protein is extracellular with Pi value of 6.01.

The unique protein identified as oxidoreductase enzyme was uncharacterized, belonging to the family alcohol dehydrogenase (ADH) and super family iron-containing alcohol dehydrogenase (Fe-ADH) which was found to catalyse the exchange of electrons between molecules. Thus, oxidoreductases employ redox active centers for reactivity and stability and hence, useful for industrial application. Therefore, in modern science and biotechnology, structures and functions of many proteins can easily be predicted by analysing their sequences using various tools in bioinformatics.

REFERENCES

Bommarius AS, and Riebel B. Biocatalysis (Wiley-VCH, Weinheim, 2004).

CASP 7 special issue. Proteins 69 (Suppl. 8), 1–207 (2007).

Edwards Y., J. (2003). Bioinformatics Methods to predict protein structure and function. A practical approach. Mol Biotechnol. Feb;23(2):139-66.

Eswar, N., Marti-Renom, M. A., Webb. B., Madhusudhan, M. S., Eramian, D., Shen, M.., Pieper, U., and Sali, A. (2006). COMPARATIVE Protein Structure Modeling with MODELLER. Current Protocols, Bioinformatics. (15:5.6.1-5.6.30.). Califonia: John Wiley & Sons, Inc.

Fiser, A., and Sali, A. (2001). Comparative protein structure modeling with Modeller: A practical approach, Comparative protein structure modeling.

Goh KM, Gan HM, Chan K-G, Chan GF, Shahar S, Chong CS, et al. (2014). Analysis of *Anoxybacillus* Genomes from the Aspects of Lifestyle Adaptations, Prophage Diversity, and Carbohydrate Metabolism. PLoS ONE 9(3): e90549. https://doi.org/10.1371/journal.pone.0090549.

Goh K.M, Kahar UM, Chai YY, Chong CS, Chai K.P, *et al.* (2013). Recent discoveries and applications of *Anoxybacillus*. Appl Microbiol Biotechnol 97: 1475–1488.

Goh Mau Goh, Goh Poh Hong, Ng Han Chyi @ Pearly Ng, Chai Kian Piaw, Raja Noor Zaliha Raja Abdul Rahman (2012). Trends and Tips in Protein Engineering; A Review. *Journal of Technology* (*sciences and Engineering*). *59 Suppl 1, 21–31*.

Goh Mau Goh, Ummirul Mukminin Kahar, Chyan Leong Ng, Kok-Gan Chan (2010). Characterization of a type I pullulanase from *Anoxybacillus* sp. SK3-4 reveals an unusual substrate hydrolysis. Volume 100, Issue 14, pp 6291–6307.

Guzzo, AV (1965). "Influence of Amino-Acid Sequence on Protein Structure". Biophys. J. **5** (6): 809–822. Bibcode:1965bpj.....5..809G.Doi:10.1016/S0006-3495(65)86753-4. PMC 1367904 . PMID 5884309.

Kelley LA and Sternberg MJE. A case study using the Phyres server. Nature Protocols 4, 363-371 (2009).

Kryshtafovych, A.; Fidelis, K. (Apr 2009). "Protein structure prediction and model quality

assessment". Drug Discov Today. **14** (7–8): 386–93. Doi:10.1016/j.drudis.2008.11.010. PMC 2808711 . PMID 19100336.

Lawrence A. Kelley & Michael J. E. Sternberg (2009). Protein structure prediction on the Web: a case study using the Phyre server. Published online 26 February 2009; doi:10.1038/nprot.2009.2.

Luthy, R., Bowie, J. U., and Eisenberg, D. (1992). Assessment of protein models with 3-dimensional profiles. Nature, 356, 84-85.

Majorek K, Kozlowski L, Jakalski M, Bujnicki, JM (2008). "Chapter 2: First Steps of Protein Structure Prediction". In Bujnicki, J. *Prediction of Protein Structures, Functions, and Interactions* (PDF).

Mcguffin, L.J., Bryson, K. & Jones, D.T (2000). The PSIPRED protein structure prediction server. Bioinformatics 16, 404–405.

Munro AW, Taylor P, and Walkinshaw MD. Curr Opin Biotechnol 11, 369-376 (2000).

Qian, B. Et al (2007). High-resolutionstructure prediction and the crystallographic phase problem. Nature 450, 259–264.