Comparative Bioinformatics Analysis Of Halophilic A-Amylases Derived From Nocardiopsis Alba And Nocardiopsis Sp.

Ashish B. Kalasava* [(1)], Kruti G. Dangar [(2)] and Dr. Geera P. Mankad [(3)]

Department of Biosciences, Saurashtra University, Rajkot, Gujarat - 360005, India [(1) (2)]
M.V.M. Science & Home Science College, Rajkot, Gujarat - 360007, India [(3)]

ABSTRACT

Halophilic actinomycetes are promising candidate for extremozymes due to their ability to survive and function potential under extreme habitats. Among these, members of the genus Nocardiopsis are recognized for producing industrially important α-amylases with multiple biotechnological potential. In the present study, a comparative bioinformatics analysis was carried out on eight α-amylase producing protein sequences derived from Nocardiopsis alba and Nocardiopsis species retrieved from the NCBI database. Physicochemical characterization using ProtParam revealed that most amylases possess low theoretical isoelectric points (pI 4.13–4.71), enrichment of acidic and aliphatic residues, and low sulphur content, reflecting supporting features for adaptations to saline environments. Homology-based three-dimensional structural models were generated using SWISS-MODEL and evaluated through GMQE, QMEAN Z-scores, MolProbity, and Ramachandran plot analysis, promote the good structural reliability and stereochemical quality to remain functional active in saline habitats. These findings highlight will be used for their potential for applications in food processing, pharmaceuticals.

Keywords : Nocardiopsis; α-amylase; Bioinformatics analysis; Protein structure modeling.

INTRODUCTION

Till date as per the literature survey, has been captured the attention about the diversity of actinomycete species along with their significance and biotechnological applications. Diverse habitats are reported with the presence of actinomycetes along with the extreme environments. These omnipresent microorganisms are found in diverse environments such as soil, plant tissues, and aquatic ecosystems, saline habitats representative exceptional potential for producing pharmaceutical products and enzymes. The amylases producing actinomycetes were reported from the saline areas, hot springs and even antarctic soils. Among these, alpha-amylases are particularly significant due to their ability to hydrolyse α-1,4-glucosidic linkages in starch, making them highly valuable products for food, textile, and pharmaceuticals (Mesbah & Wiegel, 2014).

Bioinformatics based study covers data mining of Nocardiopsis sp. and Nocardiopsis alba reported for amylase production in terms of its homologous features, uniqueness, amino acids profiles and 3D structure characterization. The selected protein sequences of Nocardiopsis sp. amylase available in NCBI were retrieved and bioinformatics analysis by ProtParam and protein 3D structure predication analysis performed for its protein characterization. A distinctive feature in primary structure, functional features, conserved motifs, and hydropathy profile of Nocardiopsis sp. amylase have been evidently illuminated to grow in presences of high salt and pH. It is used to capture the applications in biotechnological, food and pharmaceutical industries where enzymes with unique features in polyextreme conditions of salt, solvent, and high temperature are highly acceptable.

MATERIALS AND METHODS

Eight amylases producing diverse Nocardiopsis sp. were retrieved from Protein (https://www.ncbi.nlm.nih.gov/protein/) databases of NCBI in FASTA format. All the sequences were significantly different from each other. The various structural and functional analysis of sequences was carried out by using bioinformatics.

ANALYSIS OF AMINO ACID FREQUENCY

Total eight sequences were retrieved from NCBI database. The eight amylase sequences from amylase producing actinomycetes were compared to understand the amino acid composition and its role in the structural stability of the protein. The ProtParam (https://web.expasy.org/protparam/) tool was used to predict the amino acid frequency of the proteins. This tool computes physical and chemical parameters for a protein sequence, including amino acid composition.

PROTEIN 3D STRUCTURE PREDICTION AND ANALYSIS

Homology models were generated using SWISS-MODEL (https://swissmodel.expasy.org/interactive) and assessed based on GMQE, QMEAN Z-score, MolProbity, and Ramachandran plot. It makes authentic protein models and has simple access to modelling results, their visualization and elucidation (Waterhouse et al., 2018).

RESULTS

SEQUENCE RETRIEVAL AND PHYSICOCHEMICAL CHARACTERIZATION

The amino acid sequences from eight α-amylase-producing actinomycetes were analyzed using ProtParam (https://web.expasy.org/protparam/) to assess their composition and role in protein structural stability. A tool was used to predict the amino acid frequency of the proteins. The total of eight α-amylase and α-amylase–related proteins from Nocardiopsis alba and Nocardiopsis sp. were analysed using in silico approaches. The retrieved sequences ranged from 541 to 708 amino acids in length, with predicted molecular weights between 60.87 and 74.58 kDa (Tables 1 and 2).

The majority of amylases exhibited low theoretical isoelectric point (pI) values (4.13–4.71), indicating an overall acidic nature. In contrast, the catalytic domain protein from Nocardiopsis alba ATCC BAA-2165 showed a markedly higher pI (9.57), suggesting a distinct charge distribution. Instability index analysis predicted that most full-length amylases were stable, whereas several catalytic-domain-only proteins were classified as unstable. Aliphatic index values ranged from 61.24 to 81.18, indicating moderate to high thermostability across the dataset.

Protein annotation	Length (aa)	MW (kDa)	pI	Instability index	Aliphatic index	GRAVY	Stability	Functional remark
α-amylase [Nocardiopsis alba]	601	63.90	4.24	19.90	66.57	-0.391	Stable	Secreted halophilic α-amylase
α-amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2163]	555	61.75	4.64	41.99	78.76	-0.442	Unstable	GH13 catalytic core
α -amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2164]	541	60.87	4.71	41.58	79.48	-0.482	Unstable	Halophilic catalytic domain
α -amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2165]	620	69.56	9.57	52.78	65.50	-0.777	Unstable	Arginine-rich amylase
α-amylase [Nocardiopsis sp. CNT312]	708	74.58	4.38	Stable	-	-	Stable	Large salt-tolerant α-amylase
α-amylase [Nocardiopsis sp. SBT366]	606	63.80	4.13	24.68	62.48	-	Stable	Salt-adapted α-amylase

α-amylase Nocardiopsis sp. RV163]	592	63.06	4.44	23.59	64.31	-	Stable	Salt-adapted α-amylase
α-amylase [Nocardiopsis sp. JB363]	655	73.13	5.84	31.91	81.18	-	Stable	Thermostable α-amylase

Table 1. Physicochemical and Functional Characteristics

Organism / Protein annotation	No. of AA	Molecular weight (Da)	Theoretical pI	Atomic Composition					Instability index	Aliphatic index
				C	H	N	O	S		
α-amylase [Nocardiopsis alba]	601	63,905.70	4.24	2796	4200	774	921	16	19.90	66.57
α-amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2163]	555	61,748.54	4.64	2762	4168	768	837	7	41.99	78.76
α -amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2164]	541	60,875.75	4.71	2729	4116	772	809	6	41.58	79.48
α -amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2165]	620	69,562.15	9.57	3070	4740	976	866	12	52.78	65.50
α-amylase [Nocardiopsis sp. CNT312]	708	74,576.09	4.38	3277	4869	905	1070	16	28.60	61.24
α-amylase [Nocardiopsis sp. SBT366]	606	63,798.25	4.13	2789	4155	763	930	17	24.68	62.48
α-amylase Nocardiopsis sp. RV163]	592	63,059.90	4.44	2769	4137	771	895	16	23.59	64.31
α-amylase [Nocardiopsis sp. JB363]	655	73,128.35	5.84	3273	5030	934	955	12	31.91	81.18

Table 2. Comparative physicochemical characteristics of amylases from halophilic actinomycetes

ATOMIC COMPOSITION AND AMINO ACID CLASS DISTRIBUTION

Atomic composition analysis revealed high proportions of carbon, hydrogen, nitrogen, and oxygen, with consistently low sulphur content across all proteins (Table 2). Amino acid class distribution showed dominance of aliphatic residues (28.1–39.8%) and substantial acidic residue content (17.5–22.0%) (Table 3; Figure 1). Sulphur-containing residues were present in minimal proportions (1.1–2.8%), while hydroxy, aromatic, and heterocyclic residues were moderately represented.

Organism / Protein annotation	Aliphatic (%)	Sulphur Containing (%)	Hydroxy (%)	Aromatic (%)	Heterocyclic (%)	Basic (%)	Acidic (%)
α-amylase [Nocardiopsis alba]	39.2	2.6	12.2	8.8	7.0	7.7	22.0
α-amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2163]	38.1	1.3	9.7	9.9	7.7	12.3	20.9
α -amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2164]	37.1	1.1	10.0	10.6	8.5	12.4	20.3
α -amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2165]	35.5	1.9	6.5	8.1	11.1	19.4	17.5
α-amylase [Nocardiopsis sp. CNT312]	39.8	2.2	13.8	9.9	8.5	8.2	19.7
α-amylase [Nocardiopsis sp. SBT366]	39.2	2.8	13.7	9.4	8.5	6.5	21.6
α-amylase Nocardiopsis sp. RV163]	39.2	2.8	12.8	9.7	8.5	8.6	20.9
α-amylase [Nocardiopsis sp. JB363]	28.1	1.8	11.1	8.7	8.5	15.6	17.9

Table 3. Percentage distribution of amino acid classes in amylases from halophilic actinomycetes

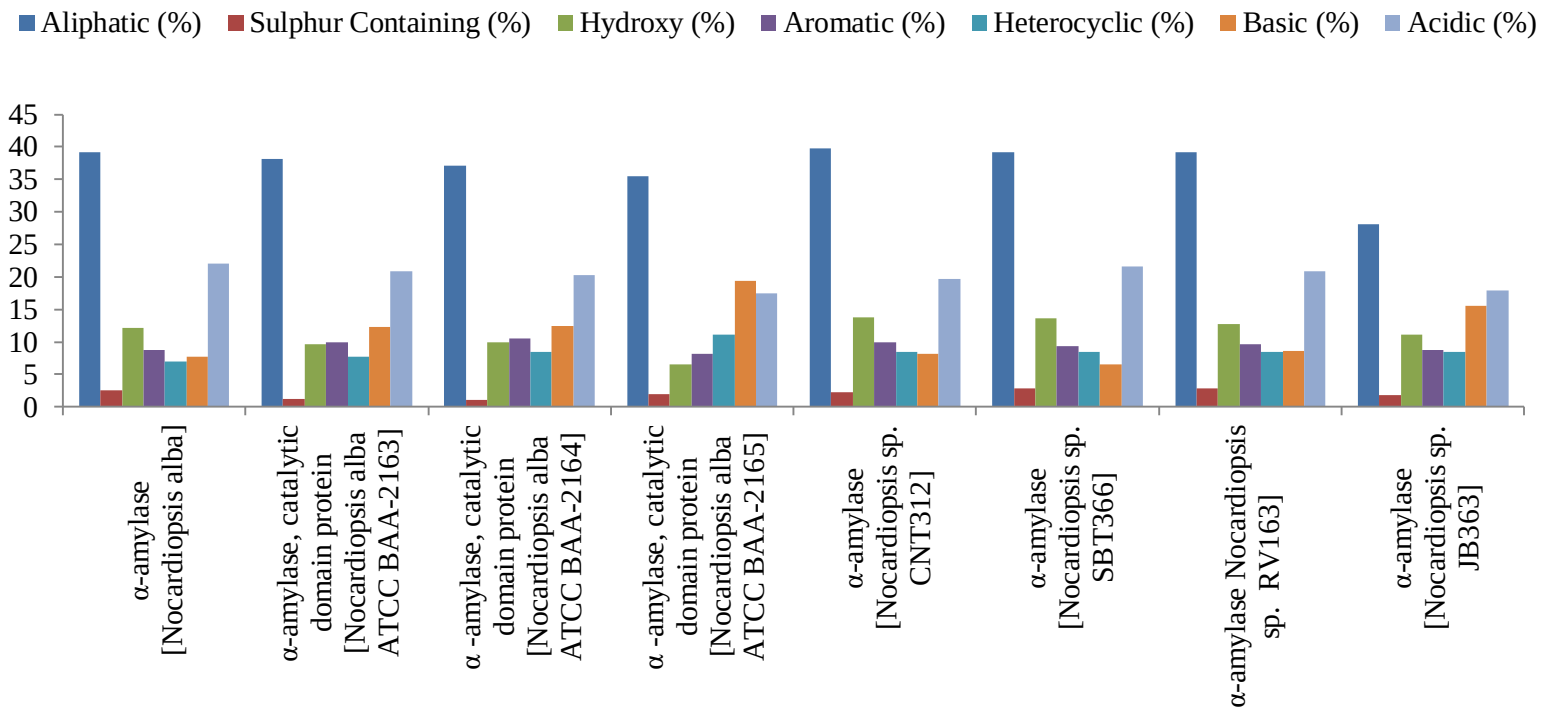


Figure 1. Amino acids frequency from halophilic actinomycetes

STRUCTURAL MODELLING AND VALIDATION

Three-dimensional structures of all amylases were predicted using SWISS-MODEL based on homologous templates with sequence identities ranging from 19.93% to 64.43% (Table 4). GMQE values (0.47–0.87) indicated moderate to high model reliability. QMEAN Z-scores ranged from −0.77 to −6.87, with models closer to zero indicating better structural quality.

Ramachandran plot analysis demonstrated high stereochemical quality, with favoured region residues ranging from 86.17% to 95.39%. Notably, Nocardiopsis sp. JB363 exhibited the highest GMQE (0.87), best QMEAN score (−0.77) and maximum Ramachandran favoured residues (95.39%), indicating superior structural stability (Tables 4 and 5).

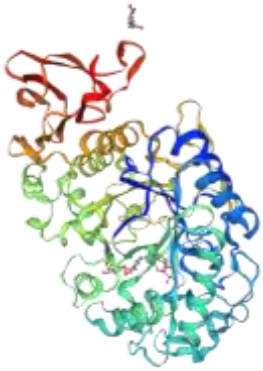
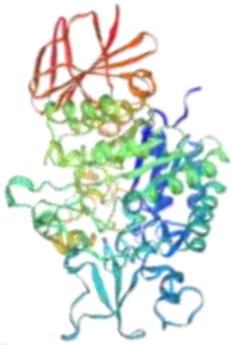
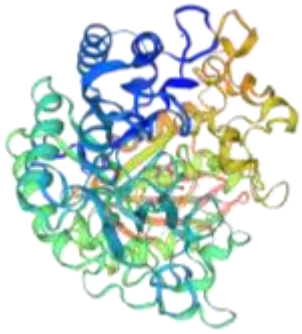
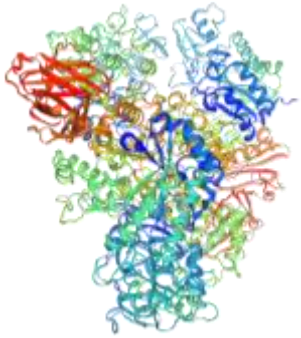
Organism	NCBI Accession No.	Length (aa)	SWISS-MODEL Template	Identity (%)	Query Coverage	GMQE	QMEAN Z-score	Mol Probity	Ram. Favoured (%)
α-amylase [Nocardiopsis alba]	WP_014910485.1	601	1xh2.1.A	39.29	40–497	0.55	-3.89	1.92	89.69
α-amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2163]	AFR09166.1	555	6aav.1.A	40.57	4–504	0.68	-1.41	1.81	93.59
α -amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2164]	AFR06474.1	541	3wy1.1.A	40.43	3–552	0.64	-4.01	1.76	90.99
α -amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2165]	AFR08628.1	620	5h2t.2.B	61.86	2–536	0.69	-2.71	1.54	91.02
α-amylase [Nocardiopsis sp. CNT312]	WP_028649691.1	708	3bmv.1.A	19.93	36–438	0.51	-3.04	2.02	86.17
α-amylase [Nocardiopsis sp. SBT366]	WP_049578795.1	606	1ciu.1.A	26.52	49–708	0.47	-5.18	2.12	90.61
α-amylase Nocardiopsis sp. RV163]	WP_047867596.1	592	1clv.1.A	42.89	26–485	0.58	-4.11	1.78	89.30
α-amylase [Nocardiopsis sp. JB363]	SIO90115.1	655	7mgy.1.A	64.43	1–653	0.87	-0.77	1.31	95.39

Table 4. Predicted three-dimensional structural models and validation parameters of amylase-producing halophilic actinomycetes

GMQE, Global Model Quality Estimation; QMEAN, Qualitative Model Energy Analysis. Higher GMQE and Ramachandran favored values indicate better model reliability, while QMEAN Z-scores closer to zero represent higher structural quality.

Protein / Enzyme	Sequence Similarity	Coverage	Range	GMQE	QMEAN	Cβ	All-atom	Solvation	Torsion	QMEAN DisCo (Global)
α -amylase [Nocardiopsis alba]	0.39	0.75	40–497	0.55	-3.89	-3.05	-2.33	-2.38	-2.66	0.68±0.05
	0.31	0.86	48–593	0.48	-4.46	-2.86	-3.22	-2.47	-2.95	0.54±0.05
	0.34	0.88	51–596	0.35	-6.10	-3.61	-3.58	-4.42	-3.78	0.35±0.05
α -amylase, catalytic domain [N. alba ATCC BAA-2163]	0.39	0.88	4–504	0.68	-1.41	-0.99	-1.73	-1.15	-0.81	0.73±0.05
	0.35	0.93	3–552	0.64	-4.01	-1.75	-2.75	-1.35	-3.05	0.64±0.05
α -amylase, catalytic domain [N. alba ATCC BAA-2164]	0.39	0.94	2–536	0.69	-2.71	-1.57	-2.30	-1.31	-1.89	0.69±0.05
α -amylase, catalytic domain [N. alba ATCC BAA-2165]	0.48	0.66	36–438	0.51	-3.04	-3.12	-2.63	-0.79	-2.30	0.73±0.05
α -amylase [Nocardiopsis sp. CNT312]	0.31	0.84	53–708	0.47	-4.52	-2.81	-2.66	-3.10	-2.91	0.52±0.05
	0.35	0.81	60–707	0.40	-5.90	-3.95	-3.51	-3.67	-3.86	0.41±0.05
	0.33	0.83	60–707	0.41	-5.92	-4.56	-3.31	-3.58	-3.83	0.41±0.05
	0.34	0.83	60–707	0.40	-5.88	-3.61	-3.19	-3.48	-3.97	0.41±0.05
	0.35	0.82	60–707	0.38	-6.26	-3.23	-3.74	-3.49	-4.38	0.39±0.05
	0.34	0.77	100–707	0.33	-5.65	-2.46	-3.27	-3.34	-3.97	0.39±0.05
	0.34	0.73	60–596	0.28	-6.73	-3.68	-4.10	-4.13	-4.50	0.38±0.05
α -amylase [Nocardiopsis sp. SBT366]	0.38	0.74	39–498	0.55	-3.39	-3.54	-2.07	-2.71	-1.93	0.67±0.05
	0.31	0.85	49-597	0.48	-4.72	-3.00	-2.92	-3.30	-2.91	0.54±0.05
α -amylase [Nocardiopsis sp. RV163]	0.41	0.75	26-485	0.58	-4.11	-0.32	-2.35	-1.93	-3.02	00.71±0.05
	0.30	0.87	42-585	0.49	-5.89	-4.42	-3.26	-3.34	-3.85	0.53±0.05
	0.34	0.89	42-588	0.37	-6.87	-4.05	-3.68	-3.88	-4.62	0.39±0.05
α -amylase [Nocardiopsis sp. JB363]	0.50	0.98	1-653	0.87	-0.77	-1.18	-0.11	-0.17	-0.52	0.85±0.05

Table 5. Model quality assessment parameters of predicted three-dimensional structures of amylase-producing halophilic actinomycetes.

α-amylase [Nocardiopsis alba]	α-amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2163]	α -amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2164]	α -amylase, catalytic domain protein [Nocardiopsis alba ATCC BAA-2165]
			

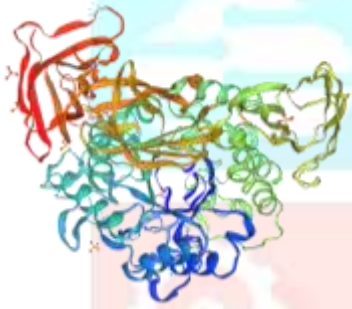
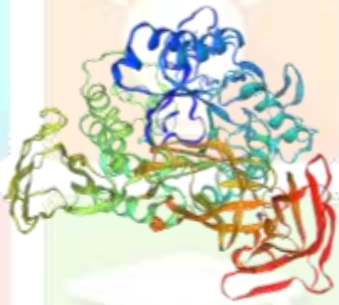
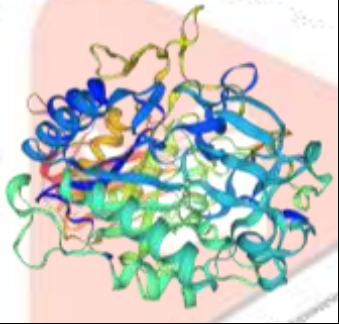
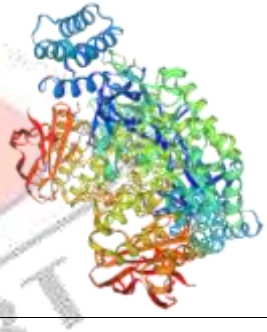
α-amylase [Nocardiopsis sp. CNT312]	α-amylase [Nocardiopsis sp. SBT366]	α-amylase [Nocardiopsis sp. RV163]	α-amylase [Nocardiopsis sp. JB363]
			

Table 6. Protein 3D Structure Prediction

DISCUSSION

The physicochemical characteristics observed in this study sharply reflect the amylase adaption in saline habitats. The predominance of acidic pI values among Nocardiopsis amylases indicates enrichment of negatively charged residues, a which strongly support for maintaining protein solubility and structural and functional stability under high-salinity conditions. The presence of acidic amino acids and lower pI values are also observed in halophilic α-amylases from Halomonasmeridiana and Kocuriavarians Natronococcus sp. strain Ah-36 (Coronado et al., 2000; Yamaguchi et al., 2011).The results are in agreement with above mentioned reports as each sequence contained higher number of acidic amino acids as compared to the basic amino acids. Acidic residues enhance the binding of hydrated ions and help to maintain water on the protein surface. While comparing all amino acids composition, each halophilic sequence showed the higher percentage of aliphatic amino acids

Looking at amino acid composition of the sequences, it was apparent that amylases contained higher number of the acidic amino acids as compared to the basic amino acids. This is in confirmation to the fact that majority of proteins obtained from halophiles and analysed for the amino acids compositions showed the higher abundance of acidic amino acids (Sumit et al., 2016). The sequences of the amylase catalytic domains from different actinomycetes were analysed through Protopram to understand the lineage of α-amylase. Marinobacteralgicola α-amylase has a low isoelectric point (pI) of 4.87. Similarly, 11 of 12 sequences exhibited low pI values, indicating adaptation to saline habitats.

High aliphatic index values across most sequencesenhanced information about the hydrophobic core packing, which contributes leads to the thermal, structural and functional stability.

Amino acids composition further strengthens halophilic adaptation. Elevated aliphatic and acidic residue content supports rigidity, salt tolerance, and catalytic efficiency, while low sulphur content support to reduce susceptibility to oxidative stress.

Lack of three-dimensional structures restricts our understanding about the structure–function association of the halophilic α-amylases. Various structural features present in halophilic proteins make them adapted to the saline conditions. In order to correlate the salt stability of halophilic α-amylase, the three-dimensional structure of twelve protein sequences were considered for producing the amylase were modeled using the online SWISS MODEL for protein 3D structure prediction. Figure No. 1 shows the model with tertiary structure topology. The Ramachandran plot visualizes energetically allowed regions of backbone dihedral angles for amino acid residues in proteins. As per SWISS MODEL guide, the score of 98% of the Ramachandran is considered as ideal. In each structure studied, 86.17% to 95.39% score was observed. Thereby, we can strongly validate our structure along with the stability. GMQE is a quality metric that combines properties from the target-template alignment and the template structure. As per the SWISS MODEL guide, the value of GMQE must be between 0 to 1. The values were in the range of 0.28 to 0.87 in our studies. It showed the reliability of the quality estimation with reference to alignment and structure.

CONCLUSION

This comprehensive bioinformatics analysis was used to understand the α-amylase properties from Nocardiopsis alba and Nocardiopsis sp. possess hallmark features of halophilic enzymeswith its structures and amino acids composition including low theoretical pI values, enrichment of acidic and aliphatic residues, high aliphatic indices, and reduced sulphur content. It showed a high inclination of the negatively charged amino acids attributing salt stability. Present investigation has led to better understanding of the structural features responsible for the adaptation of the Halophilic α-amylases in salt.

REFERENCES

- Coronado, María-José, Carmen Vargas, Jürgen Hofemeister, Antonio Ventosa, and Joaquín J. Nieto. "Production and biochemical characterization of an α-amylase from the moderate halophile Halomonas meridiana." FEMS microbiology letters 183, no. 1 (2000): 67-71.
- Mesbah, Noha M., and Juergen Wiegel. "Halophilic alkali-and thermostable amylase from a novel polyextremophilic Amphibacillus sp. NM-Ra2." International journal of biological macromolecules 70 (2014): 222-229.
- Mesbah, Noha M., and Juergen Wiegel. "Purification and biochemical characterization of halophilic, alkalithermophilic protease AbCP from Alkalibacillus sp. NM-Fa4." Journal of Molecular Catalysis B: Enzymatic 105 (2014): 74-81.
- Waterhouse, Andrew, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, Florian T. Heer et al. "SWISS-MODEL: homology modeling of protein structures and complexes." Nucleic acids research 46, no. W1 (2018): W296-W303.
- Yamaguchi, Yoshihiro, Jung-Ho Park, and Masayori Inouye. "Toxin-antitoxin systems in bacteria and archaea." Annual review of genetics 45, no. 1 (2011): 61-79.