Egypt. J. Vet. Sci. **Vol. 54,** No. 2, pp. 211-235 (2023)

Egyptian Journal of Veterinary Sciences
https://ejvs.journals.ekb.eg/

Mitochondrial Genome with Intrinsic Protein Disorder Potentiality

Maged M. Mahmoud[1,2,3], Sarwar Jamal[4] , Talal Qadah[2,5], Kawther Ali Zaher[6] , Haitham A. Yacoub[7*]

[1] *King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia 21589*
[2]*Department of Medical Laboratory Sciences, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia 21589*
[3]*Molecular Genetics and Enzymology Department, Human Genetics and Genome Research Institute, National Research Centre, Cairo, Egypt*
[4]*Integrative Biosciences Center, 6135 Woodward Ave, Wayne State Univeristy, Detroit MIchigan- 48202*
[5]*Hematology Research Unit, King Fahd Medical Research Center (KFMRC), King Abdulaziz University, Jeddah, 21589, Saudi Arabia (SA).*
[6]*Immunology unite, King Fahd Medical Research Center (KFMRC), King Abdulaziz University , Jeddah, 21589, Saudi Arabia (SA).*
[7]*Cell Biology Department, Biotechnology Research Institute, National Research Centre, Dokki, Giza, P.O. Box 12622, Egypt*

THE current Study intends to study the mitochondrial genome proteins specifically focusing on the distribution of intrinsically disordered regions within these proteins. The UniProt were employed to obtain amino acid sequences of the mitochondrial proteins. Next, the obtained sequences were studied to identify their unique features. We used the PONDR VSL2 algorithm to assess the proteins for any order or disorder. According to study results, mitochondrial genome proteins were expected to depict unexpected levels of intrinsic disorder. The results also indicated the possibility of presence of 7 intrinsic disorder regions within these proteins. The disordered regions were mostly found to be associated with protein–protein interactions. The research indicated the presence of two to nine molecular recognition features in the assessed mitochondrial proteins. These features included short structure-prone segments, location of mitochondrial proteins inside long disordered regions and the feature of showing disorder-to-order transition during binding. Moreover, such regions were attributed with a number of sites characterized with posttranslational modifications. According to the analysis, there may be many disordered regions present within mitochondrial genome proteins leading to multifunctional nature of mitochondrial genome proteins in the signal transduction pathways where these proteins regulate the cell metabolism pathways.

Keywords: Mitochondria, Genome, Proteins, IDPs, Prediction.

Introduction

It is obvious that the analysis can be better performed using mitochondrial genome in animals in comparison to using nuclear genome due to introns deficiency and limited exposure to recombination and haploid inheritance [1].

Corresponding author: Haitham A. Yacoub, E- mail.: ha.mohamed@nrc.sci.eg, Tel. +201110545634
(*Received* 28/09/2022; *accepted* 22/10/2022)
DOI. 10.21608/EJVS.2022.165833.1402
©2023 National Information and Documentation Centre (NIDOC)

The use of robust primers allows mitochondrial genome recovery [2,3]. Formerly, the researchers performed phylogenetic study mainly on mitochondrial genes encoding ribosomal (12S, 16S) DNA. However, it must be noted that the insertions and deletions (indels) in these DNA gives rise to various complexities in sequence alignment specifically when employed in broader taxonomic analyses [4].

The sequences of (partial) control region (CR; D-loop: nucleotide position 1–1232; NC_007235) was mainly used in the research performed on Chicken mtDNA. The CR does not depict high resolution mtDNA phylogeny due to its smaller size and higher mutation rate in this region in comparison to coding regions [5]. The high mutation rate means higher frequency and recurrence of mutations that leads to blurred matrilineal genealogy structure. It has become possible to obtain reconstruction of animal domestication history of cattle [6-8], horses [9] dogs [10] and pig [11] with the help of fine-grained analyses based on the employment of complete mtDNA genome. This kind of analysis allows the researchers to better comprehend animal domestication.

The study of the whole mitochondrial genome is better in comparison to study of partial genome sequencing since it allows categorization into categories like synonymous and replacement sites; genic, proteins and neutral, adaptive or deleterious mutations. This categorization allows the researcher to accurately assess genealogy, divergence events, and adaptation to gradients. Another benefit of performing the analysis of entire mitochondrial genome is that it helps to check if the evolutionary conclusions associated with subsets of the genome correspond to evolutionary dynamics associated with the entire mitochondrial genome [9-13].

The latest research performed by Native Americans on colonization pattern of the New World as well as the studies on dog domestication history indicate that in comparison to the poor resolution of maternal genealogies by the sub genomes, the whole mitochondrial genome sequences lead to higher resolution of maternal genealogies [14]. The higher stability of mitochondrial genomes makes them a reliable source of population genomic information. Such information is derived from paleontological specimens of nonexistent species as well as the associated surviving species [14-18]. Hence, it can be concluded that even the smaller size of mitochondrial genome (merely~16 kb) does not prevent it from being a heterogeneous indicator and providing evolutionary pointers[5].

One of the most important organelles of eukaryotic cells is mitochondria that are involved in a number of essential functions [19,20]. Mitochondria are involved in different metabolic pathways and are essential for energy generation, controlling stress response and also metabolism. Mitochondria perform various roles in addition to ATP generation [21-23]; mitochondria facilitate many metabolic pathways including biogenesis of iron-sulfur clusters, β-oxidation of fatty acids and metabolizing oxygen. Moreover, mitochondria are involved in different regulatory pathways like $Ca2+$ buffering and signaling [21], stress-induced cell death [23-26] and developmental cell death. The highly compartmentalized organelles in mitochondria are due to the fact that intrinsically disordered protein regions (IDPRs) present inside proteins are promiscuous binders. This property facilitates the diverse functionality of mitochondria and also indicates the presence of essential IDPRs within mitochondrial proteins.

IDPs and IDPRs stand for intrinsically disordered proteins and intrinsically disordered protein regions respectively. IDPs are disorder proteins without unique structure. These are found in the form of active structural ensembles [27-41]. Such proteins are either ordered to some extent or completely disordered. Since the IDPs have different protein amino acid sequences, these proteins depict variability in their composition, degree of disorder and flexibility. These proteins are naturally existent in large quantities [27, 40, 42-45]. Particularly, the number of IDPs within eukaryotic proteomes remarkably outshines the number of bacterial and archael content in these proteomes. Moreover, viral proteomes show extremely high levels of intrinsic disorder [46-57].

The adaptable nature of IDPs /IDPRs results in formation of complex networks due to greater tendency of these proteins to undergo interaction and binding with numerous partners[14, 30-32, 34, 35, 38, 40-43, 58-62]. IDPs/IDPRs have been found to be associated with a number of diseases in human beings like cardiovascular diseases, cancer, Alzheimer's disease and Parkinson's disease [59, 63-75]. This study aimed to detect the presence and levels of intrinsic disorder in mitochondrial genome proteins. In addition to

determine if the functionality of mitochondrial genome proteins is affected in any way by the presence of intrinsic disorder. In this regard, the study obtained mitochondrial protein sequences of different chicken species and studied the amino acid composition of these proteins. The research was an attempt to get insight into how the amino acid residues are distributed within mitochondrial proteins. The study also detected disorder-based interactions sites including molecular recognition features (MoRFs) and ANCHOR-identified binding segments (AIBSs) besides evaluating the intrinsic disorder levels in the protein sequences.

Material and Methods

Collection of the date

The UniProt database was used to collect the mitochondrial proteins from wild type chicken as an organism model (UniProt Consortium, 2015).

Composition analysis of mitochondrial non-membrane proteins

The content of amino acid residues of mitochondrial Proteins was determined by using Composition Profiler (http://www.cprofiler.org/cgi-bin/profiler.cgi) based on a physico-chemical and conformation properties [76] to reveal the enriched and depleted residues in a certain protein. In addition, the amino acid distribution at given protein was analyzed by ProtParam tool of ExPasy proteomic server (http://web.expasy.org/protparam/).

Hydrophobicity level

The hydrophobicity level of mitochondrial non-membrane proteins was determined by using ProtParam tool of ExPasy proteomic server and identified as GRAVY index which calculated by the hydrophobicity of certain amino acids/amino acid number in our target protein.

Analysis of order/disorder propensities of mitochondrial non-membrane proteins

The PONDR (Predictor of Natural Disordered region) VSL2 algorithm [77] was used to predict the tendency of mitochondrial non-membrane proteins for order/disorder. The PONDR algorithm was considered as the most accurate predictor [78, 79]. The resulted output of PONDR evaluation was represented as number range Zero (order) to 1 (disorder) and the threshold of ≥ 0.5 was used to disorder deification. The FoldIndex predictor also used to determine which region of given protein is unfolded according to net charge of given sequence and hydrophobicity mean of amino acid residues. The SLIDER tool (http://biomine-ws.ece.ualberta.ca/SLIDER/index.php) which predicts whether the mitochondrial genomic proteins have a long disordered region with at least 30 consecutive disordered residues. It calculated by logistic regression model that included sequence complexity, amino acid residues composition, and selected physicochemical properties of amino acids [78, 79].

Charge-hydropathy plot

The charge-hydropathy plot (CH-plot) of a particular sequence is a linear disorder classifier that discriminate proteins with substantial amounts of extended disorder from proteins with globular conformations [40, 52]. A CH-plot shows results from a binary disorder predictor and represents an input protein as a 2D graph, in which the mean Kyte-Doolittle hydrophobicity and the mean absolute net charge are projected onto the X- and Y-coordinates, respectively. In the corresponding CH-plot, a boundary line can separate fully structured proteins and fully disordered proteins. All proteins located above this boundary line are highly likely to be extended, while proteins located below this line are likely to be compact [40, 52].

Folding rate determination

In order to predict the folding rate of mitochondrial genomic proteins and folding type from amino acid residues, CIpred tool was used (http://ibi.hzau.edu.cn/FDserver/cipred.php). It performed based on amino acids residues weight, hydrophobicity level and the state of being degenerate [80,81].

Molecular recognition features (MoRFs) Prediction

The MoRF predictor was used to identify the formation of molecular recognition features regions in mitochondrial non-membrane proteins, which has a potentiality to bind to partner protein (http://biomine-ws.ece.ualberta.ca/MoRFpred/index.html). This tool estimated based on the propensity of amino acids residues for disordering [82]

ANCHOR algorithm

The ANCHOR algorithm (http://anchor.enzim.hu/) was used to identify the anchor predicted binding regions (APBR) that occur in disordered mitochondrial proteins. This prediction is depending on three properties as follows: localization of amino acid residues inside the long disordered region, to ensure that the residues are to be able to communicate with local sequential

adjacent, so can fold and finally, the disability of given amino acids to communicate with globular protein [83, 84].

The interactivity mitochondrial genome Proteins

The interactivity network of mitochondrial proteins was estimated by using STRING databases (Search Tool for the retrieval of interacting Genes) [83] This tool provides the predicted network interaction and association of mitochondrial proteins with a certain group of proteins. The network nodes are partner proteins, while, the edges correspond to predicted or known function associations. The edge is drawn with 7 different colored lines which related to seven types of evidences that shared in association prediction. A green line corresponds with neighborhood evidence; red line relates to the presence of fusion evidence; a purple line indicates the experimental evidence; a blue line represents co- occurrence evidence; a yellow line indicates text mining evidence; A black line represents co-expression evidence; a light blue indicates databases evidence [83].

Results

This research used the Universal Protein Resource (UniProt) database to obtain complete amino acid sequences of the mitochondrial genome proteins associated with different chicken species. The obtained sequences were examined to determine the intrinsic structural disorder in these proteins. The UniProt IDs for each of the 152 MT-genome proteins obtained from different chicken species are given in Table 1.

Analysis of Mitochondrial Genome Proteins' Contents

This research employed UniProt database to obtain mitochondrial proteins associated with wild type chicken. The obtained proteins were used for evaluation of intrinsic disorder in mitochondrial genomic proteins. Moreover, the analysis of the obtained mitochondrial proteins allowed comprehending how protein functioning is facilitated by intrinsic disordered regions. The analysis presented a contrast of the proportion of amino acid in ordered proteins and domains with the proportion of amino acid within IDPs/IDPRs by highlighting unique features of amino acids composition within the mitochondrial proteins. Despite the presence of disorder-inducing residues like Ala, Arg, Gly, Gln, Ser, Pro, Glu, and Lys in abundance, the Trp, Tyr, Phe, Ile, Leu, Val, Cys, and Asn and other so-called order-promoting amino acids show depletion of amino acid sequences in IDPs/IDPRs

The Composition Profiler was employed in this research for the purpose of identifying the amino acid residues shown in Table 2. These amino acid residues are essential for depletion as well as enrichment of mitochondrial proteins. The results indicated that Arg, Asp, Asn, Glu, Lys, Val, Tyr and Cys residues showed depletion in mitochondrial proteins. Moreover, Ile, Leu, Met, Val, Ser, Thr, Trp, Phe, Gly, His and Pro amino acids show disorder-promoting tendencies. Moreover, the analysis anticipated mitochondrial non-membrane proteins to be a form of hybrid proteins characterized with ordered and disordered compositions.

The ExPasy ProtParam tool is employed to study protein parameters particularly the amino acid compositions as shown in Figure (1). It was found that each mitochondrial protein showed same amino acid composition except for the NADH dehydrogenase subunits which showed abundant amount of Leu, Ser, Ile and Thr amino acids. Other exceptions included Phe, Pro, Met, Gly, Glu and Ala that showed reasonable number of residues. Moreover, lack of Cys, Asp, Asn, Arg, Val, Tyr and Trp amino acids was observed in the NADH subunits.

The analysis also highlighted the presence of abundant quantities of Leu residues in cytochrome oxidase subunits and cytochrome b. Moreover, moderate quantity of Lys, Thr, Trp, Ile, Gly, Glu, Gln, Ser, Phe and Ala, Arg, Pro were detected in cytochrome oxidase subunits and cytochrome b. Cys, Asp, Asn, Val amino acid residues showed insignificant interaction with cytochrome oxidase subunits and cytochrome b proteins. ATPase 8 showed deficiency of certain amino acids like Val, Tyr, His, Gly, Glu, Cys, Asp and Arg. Figure 1 shows that mostly, the proteins showed abundant amounts of Leu amino acid residues (18.11%) followed with Thr residues (8.93%) and Ser residues (7.94%). As per Figure 1, moderate quantities of Pro (6.33%), Phe (5.64%), Met (4.15%), Ile (6.91%), Gly(5.45%), and Ala residues (7.42%) while low quantities of Val (4.25%), Try(2.32%), Trp(3.09%), Lys(2.51%), His(2.58%), Glu(2.93%), Gln(2.61%), Cys(1.35%), Asp(1.85%), Asn(3.42%) and Arg residues (2.26%) were detected in mitochondrial proteins.

TABLE 1. Accession numbers of mitochondrial genome proteins found in the UniProt database

MT-proteins	Accession number
NADH dehydrogenase subunit 1	**P18936**, Q7GTV3,Q4GWR2, E5DEK7, E5DEA3, E5DEP6,E5DEW1
NADH dehydrogenase subunit 2	**P18937**, E5DF79,E5DEE3,E5DEB7,E5DEW2, Q76L17, A0A0C5ANN4, K9JTR8, A0A0B4ZV97, E5DEP7, E5DFT7, E5DEJ5, Q7IT50, E5DFJ6, E5DEA4, Q4GWR1, E5DFD1, E5DFV0, E5DFI3
NADH dehydrogenase subunit 3	**P18938,** Q4GWQ5, B8PTT7, E5DFI9, E5DFV6, E5DFL5, E5DF46, E5DEB0
NADH dehydrogenase subunit 4L	**P18942**, Q7GTU8, Q4GWQ4, E5DEB1, E5DFK3, E5DFJ0
NADH dehydrogenase subunit 4	**P18939,** Q7GTU7, W8P5E2, A0A0C5AL54, Q6ZLS4, K9JTR6 , K9JV25, A0A0B5A069, E5DEP2, E5DEJ0, E5DFN0, E5DEK3, E5DEV7, E5DFJ1 Q4GWQ3, E5DEZ6, E5DF48, E5DFL7, A0A0G3F6T9, E5DEB2, E5DFH8
NADH dehydrogenase subunit 5	**P18940**, Q4GWQ2, Q76L08, Q85A52, E5DEC6, E5DEF2, E5DED9, W8NXV0, E5DEH8, E5DFP4, E5DED9, E5DEH8, E5DEF2, E5DEG5, A5HDU0, Q4GWQ2, E5DFK5, E5DEB3, E5DFV9, E5DFJ2, Q9MPA6
NADH dehydrogenase subunit 6	**P18941**, Q7GTU4, Q195H7, E5DEN2, E5DFU8, E5DFN3, Q4GWQ0, E5DFJ4, E5DEB5, E5DEX3
Cytochrome c oxidase subunit I	**P18943,** A0A0B4ZU13
Cytochrome c oxidase subunit II	**P18944**, Q76L15, Q85B99, Q4GWQ9, A0A0D3QU53, K9JV31, E5DFI5, E5DFL1, E5DFJ8, E5DEA6
Cytochrome c oxidase subunit III	**Q7GTV0**, P18945, Q195I3, U5LT96, E5DEM6, A0A0B4VK94, E5DEQ2, E5DEE8, Q4GWQ6, E5DFC3, E5DFI8, E5DEA9, E5DFD6
Cytochrome b	**P18946**, Q85GK8, Q9G233, Q9G296, Q957U2, Q9MNW7, Q9B621, Q9B623, Q9B622, Q7GTU5, E5DEE0, E5DEU6, E5DEV9, E5DEQ7, E5DFJ3, E5DEZ8, E5DFE1, E5DEX2, E5DFK6, E5DEB4, E5DF50, A0A0B4ZVA3
ATP synthase F0 subunit 6	**P14092**, Q7GTU9, E5DEN8, E5DF83, A0A0B4ZVH0, K9JTN9, Q4GWQ7, E5DFI7, E5DEA8, E5DEI6, E5DEL2, A4QIV8
ATP synthase F0 subunit 8	**P14093**

TABLE 2. Depleted and enriched amino acids in mitochondrial proteins identified by composition profiler

Protein name	Enriched amino acids	Depleted amino acids
NADH dehydrogenase subunit 1	Leu, Pro	Arg, Asp, , Cys, Gln, Glu, Gly, His, Lys, , Val
NADH dehydrogenase subunit 2	Ile, Leu, Met, Ser, Thr, Trp	Arg, Asp, Glu, Gly, Lys, Tyr, Val
NADH dehydrogenase subunit 3	Leu, Met, Phe, Trp	Lys, Val
NADH dehydrogenase subunit 4L	His, Leu, Met, Phe, Ser	Asp, Lys
NADH dehydrogenase subunit 4	Ile, Leu, Met, Pro, Ser, Thr, Trp	Arg, Asp, Glu, Gly, Lys, Val
NADH dehydrogenase subunit 5	Ile, Leu, Met, Thr	Arg, Asp, Glu, Lys, Tyr, Val
NADH dehydrogenase subunit 6	Gly, Leu, Phe, Val	Asn, Gln, Gly, His, Ile, Lys,Thr
Cytochrome c oxidase subunit I	Ile, Leu, Met, Phe, Thr, Trp	Arg, Asp, Cys, Gln, Glu, Lys
Cytochrome c oxidase subunit II	Ile, Met	Asn, Glu, Lys
Cytochrome c oxidase subunit III	His, Leu, Phe	Arg, Asn, Asp, Glu, Lys
Cytochrome b	Ile, Leu, Phe, Trp	Arg, Asp, Gln, Glu, Lys, Val
ATP synthase F0 subunit 6	Leu, Pro, Thr	Asp, Glu, lys, Val
ATP synthase F0 subunit 8	Pro, Thr, Trp	Glu, Gly, Val

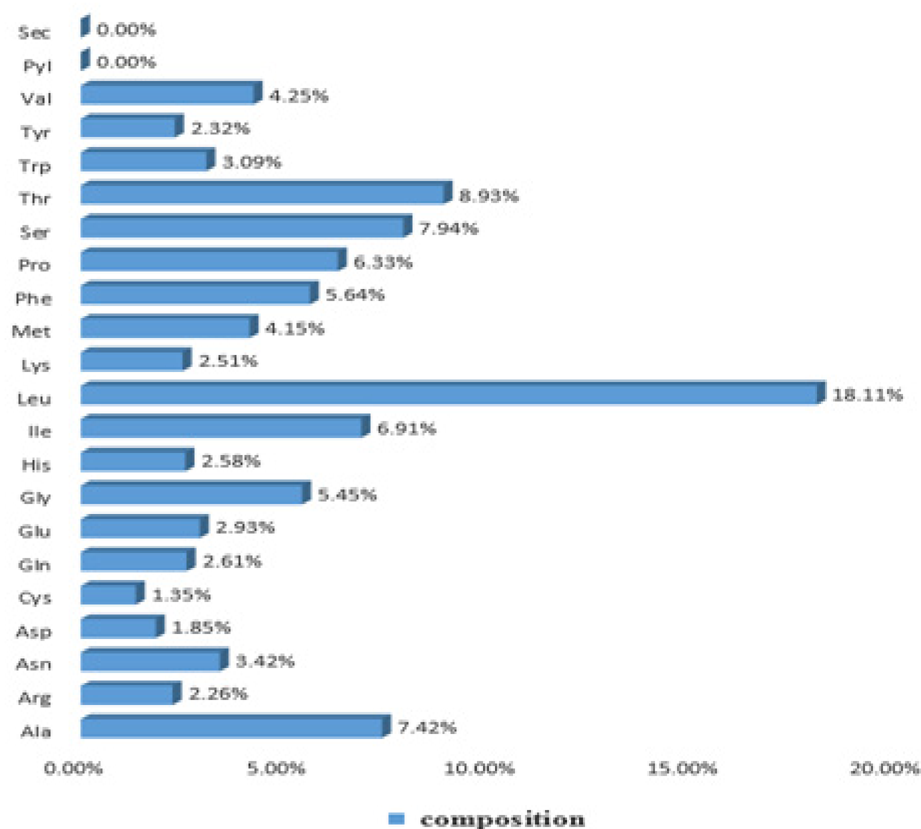


Fig. 1. The peculiarities of the amino acid compositions (mean value) of the mitochondrial genome proteins evaluated by the ProtParam tool of the ExPasy proteomic server.

Evaluation of Mitochondrial Genome Proteins' Hydrophobicity

This research also used the Grand average of hydropathicity (GRAVY) tool to evaluate the hydrophobicity of mitochondrial proteins. The mitochondrial proteins showed GRAVY scores between -0.041 and 1.42 accounting to an average score of 0.67. In this research, GRAVY scores had been evaluated on the basis of Kyte and Doolittle hydropathy scale. The results revealed Ile residues with the score of 4.5 as the most hydrophobic and Arg residues with the score of -4.5 as the most hydrophilic. The higher the score, the higher is the hydropathy and vice versa. Usually, GRAVY scores are higher than 0.5 for Membrane proteins while the mean score depicted by cytosolic proteins is ~0.2. Hence, the mitochondrial proteins being assessed in this research were found to be hydrophilic and were anticipated to depict polarity and behave like soluble proteins.

Intrinsic Disorder Propensity of Mitochondrial Proteins

The highly accurate PONDR® (Predictor of Natural Disordered region) VSL2 algorithm was employed in this research to assess and predict order/disorder propensities of mitochondrial proteins (Figure 2). The results obtained from PONDR® VSL2 were expressed between 0 (prediction of order) and 1 (prediction of disorder).

As per (Figure 2), the disorder is indicated at baseline threshold of ≥ 0.5 and scores higher than 0.5 indicates the disordered region within the assessed protein. As per the results, there is high prediction of order in the mitochondrial non-membrane proteins, which is also indicated by an aligned PONDR VSL2-disorder profile for each of the mitochondrial protein sequences. Despite the prediction of order in these proteins, there is still some possibility of disordered and flexible regions within these proteins.

Based on functionality, the Mitochondrial proteins have been classified into three main categories: NADH subunit proteins; the cytochrome oxidase subunits and cytochrome b; and lastly, ATPase (6 and 8) (Table 3). The aligned PONDR VSL2-disorder profile of the mitochondrial proteins indicates variation among proteins in terms of the numbers of disordered amino acid residues. The NADH subunit proteins from the first group depict the most disordered amino acids ranging from 18 aa (*NADH6*) to 97 aa (*NADH5*). On the other hand, the second group proteins depicted 15 aa (*COXI*) to 30 aa (*COXII*). The results indicate that ATPase 6 shows 43 disordered amino acids whereas the ATPase 8 depicts least amino acid residues (6 aa).

Table 2 clearly indicates that different sets of mitochondrial proteins depict different intrinsic disorder ratio; where, ATPase group showed highest disorder accounting to 15% while NADH subunit proteins showed 10.5% disorder and the lowest disorder of 7.5% was seen in cytochrome oxidase subunits and cytochrome b group.

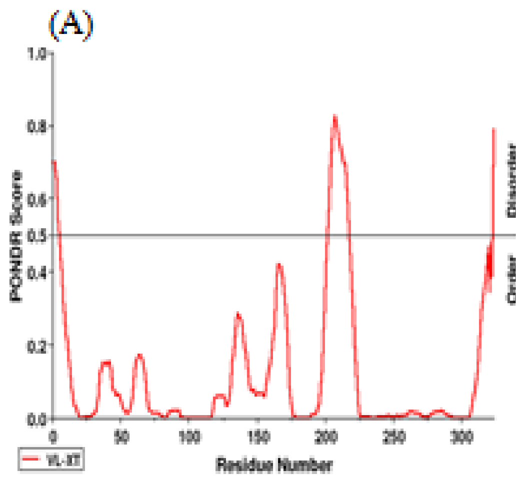
The results of the analysis indicated the presence of several intrinsic disorder protein regions in mitochondrial proteins (see Figure 2 and Table 3). The NADH dehydrogenase subunit 4 was anticipated to depict the highest number of such regions (seven) followed with NADH dehydrogenase subunit 5 which showed (six) and NADH dehydrogenase subunit 2 which showed (five) regions. On the other hand, ATP synthase subunit 8 depicted the least number of such regions (i.e., only one IDPR). Moreover, cytochrome oxidase subunits showed 2 disordered regions and cytochrome b showed 3.

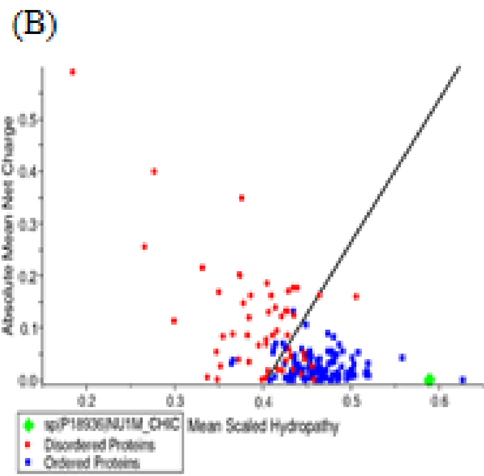
The CIpred tool was employed to determine the folding time and rate of mitochondrial proteins. The results show the protein folding rates in log (Kf)/sec. Three parameters are involved in this calculation including weight of amino acid residues, hydrophobicity level and degeneration state. The results indicated folding rates of mitochondrial proteins between (log (K_f)=-3.31) and (log(K_f)=8.11). The folding rates were found to be higher for *NADH5, COXI, NADH4, Cyt-b, NADH1* and *NADH2* in comparison to other mitochondrial proteins. ATPase subunit 8, NADH4L and NADH3 particularly showed extremely slower folding rates (see Table 4).

The mitochondrial protein sequences are checked for any features of functional intrinsic disorder like the presence of long disordered regions and occurrence of a minimum of 30 consecutive disordered residues. The researchers employed the SLIDER (http://biomine-ws.ece.ualberta.ca/SLIDER/index.php) for detecting the presence of long disordered segments in the proteins. The SLIDER had a score of 0.5380 which implied that mitochondrial protein sequence with the score higher than 0.5380 indicate the presence of long disordered region. The analysis revealed that the assessed proteins showed lower scores and hence did not contain long disordered segment.

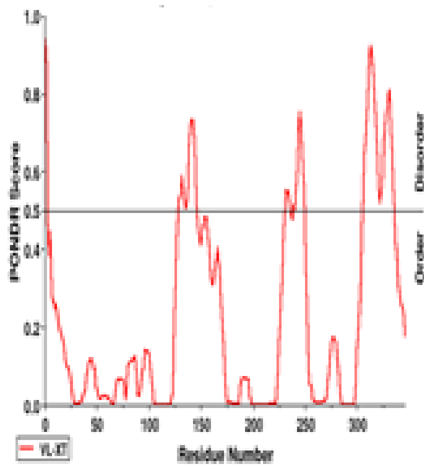
NADH1

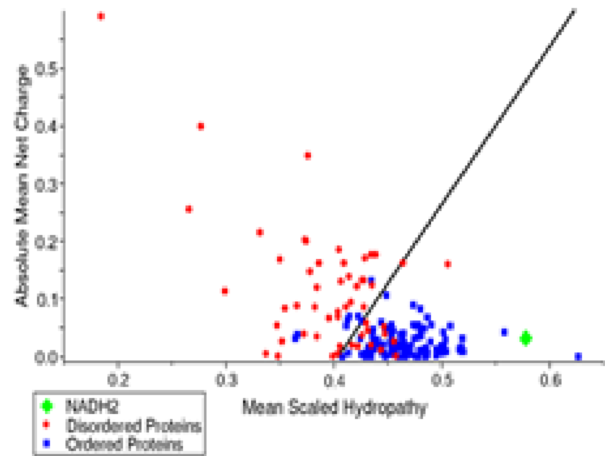
(A) (B)



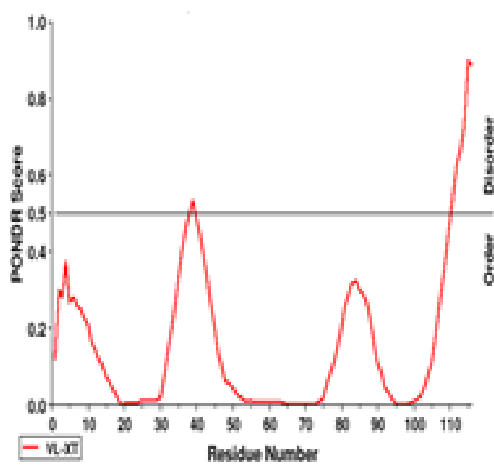


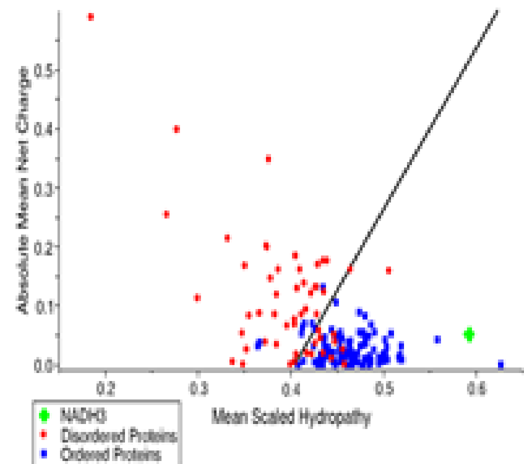
NADH2



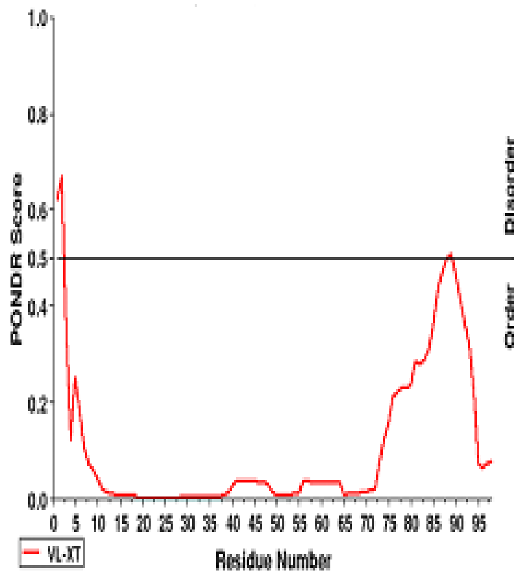


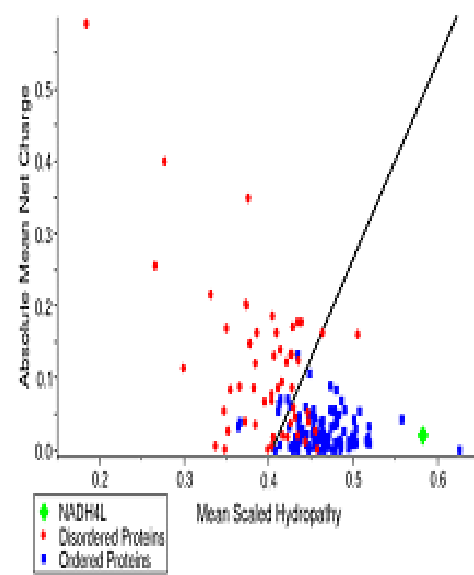
NADH3



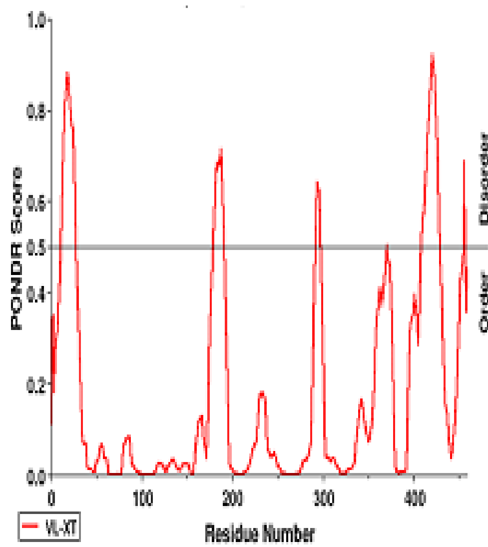


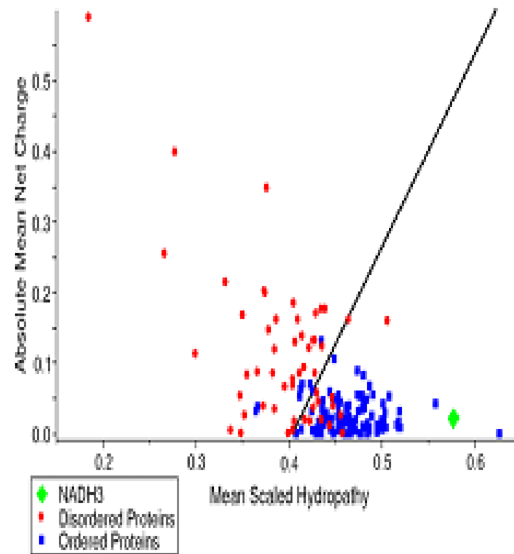
NADH4L



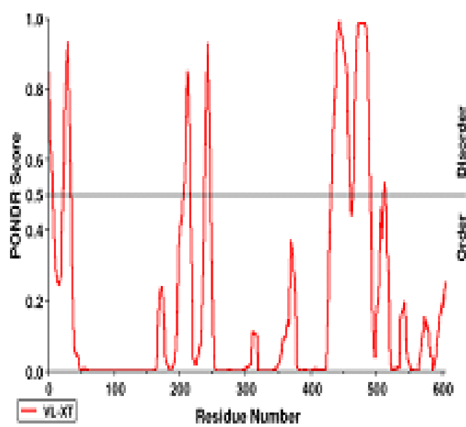


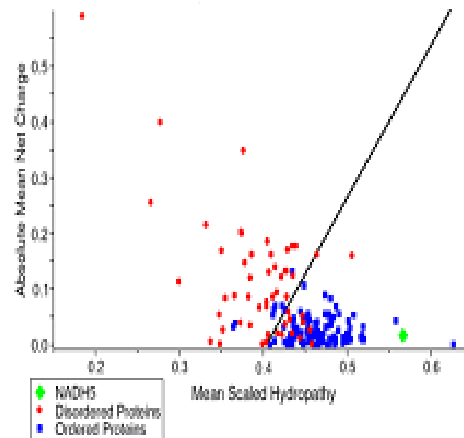
NADH4



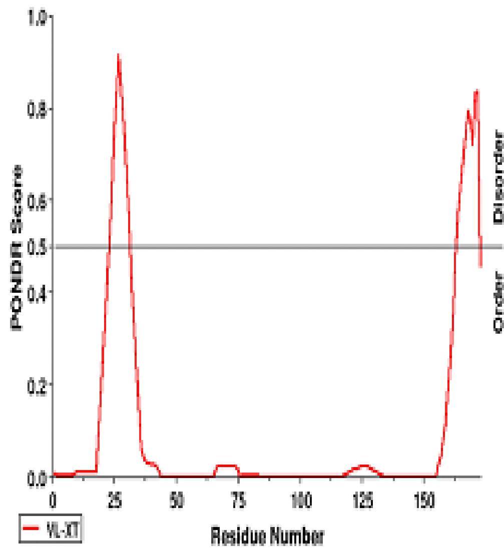


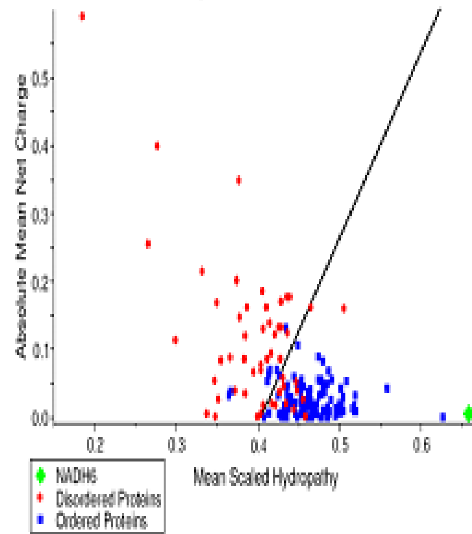
NADH5



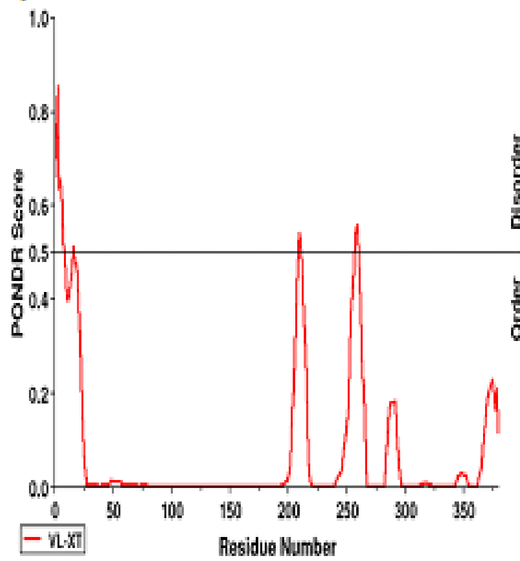


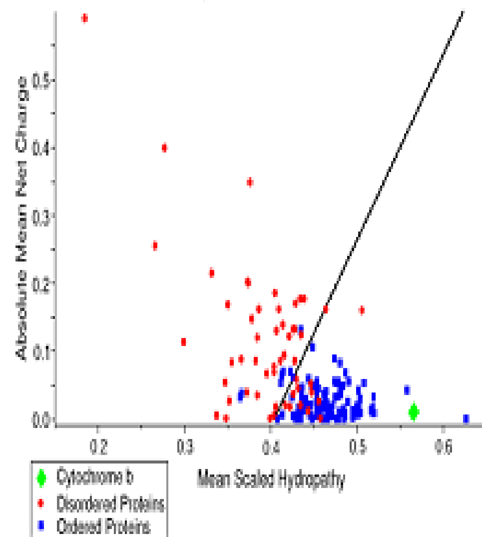
NADH6



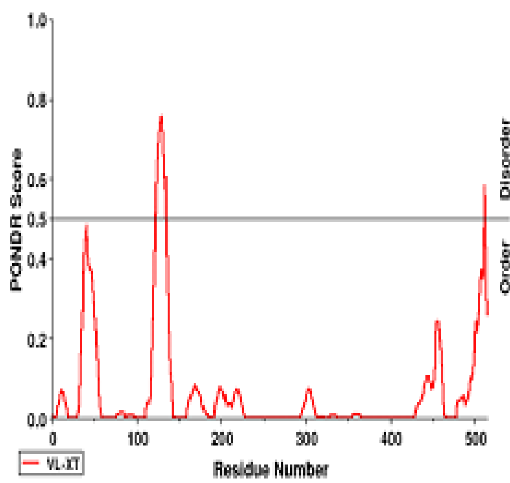


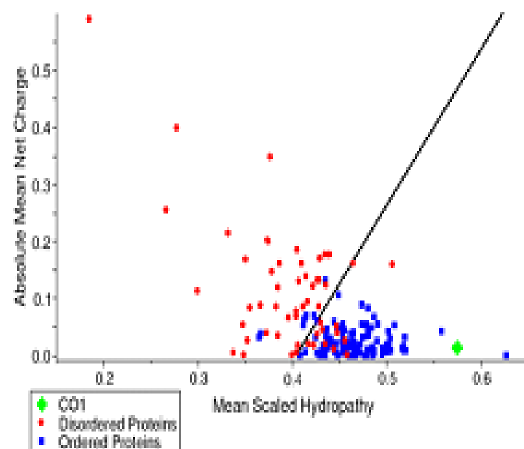
Cytochrome b



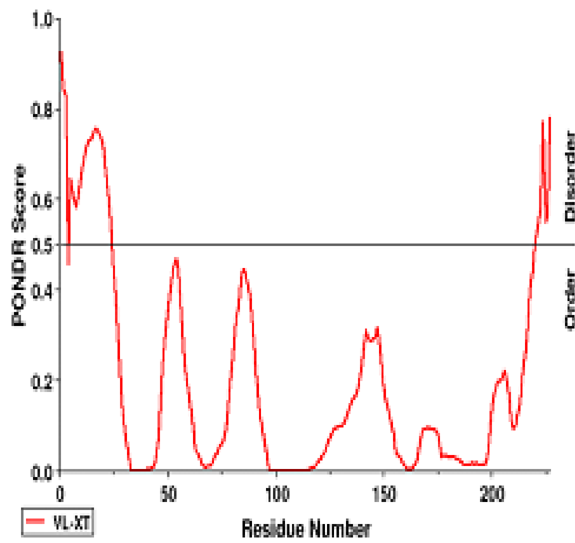


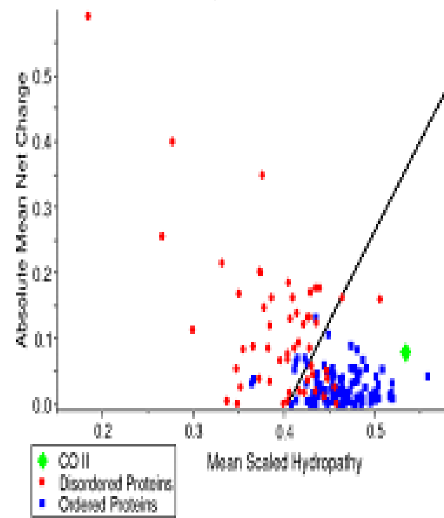
COI



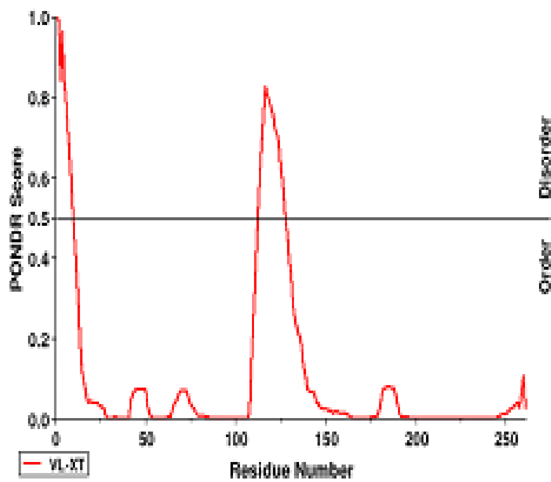


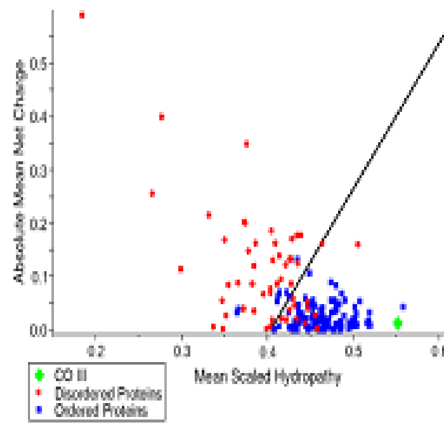
COII



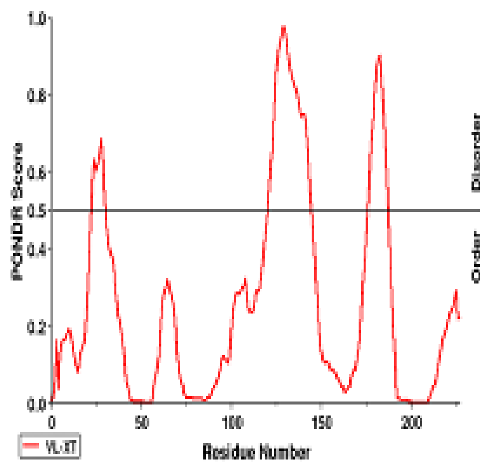


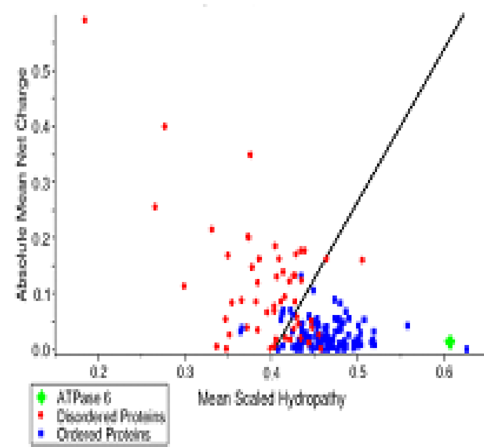
COIII





ATPase 6





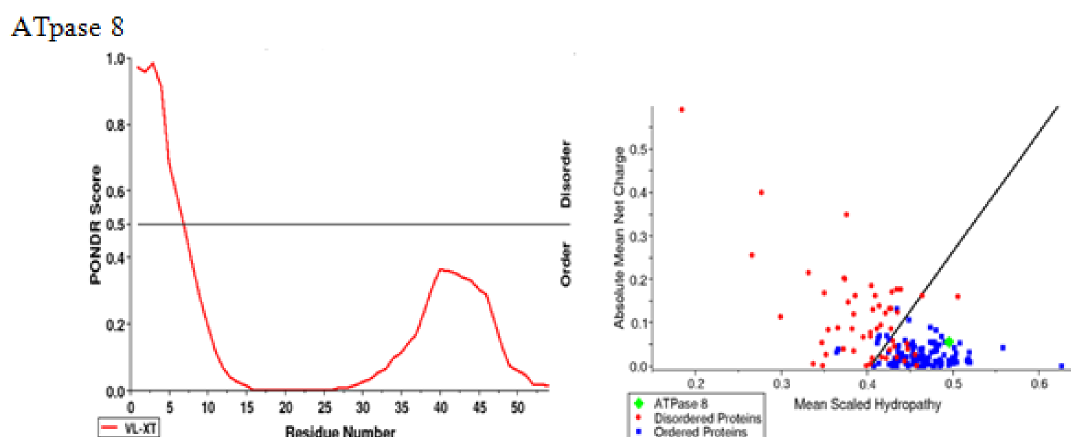


Fig. 2. (A) PONDR® VSL2 analysis of disorder distribution profiles in members of the mitochondrial genome proteins. (B) Charge-hydropathy plot (mean scaled hydropathy, <H>, against mean net charge, <R>, at neutral pH) of the mitochondrial genome proteins.

TABLE 3. Abundance of intrinsic disorder in mitochondrial proteins evaluated by PONDR® VSL2 (PONDR) algorithm.

MT-proteins	Total amino acids	Disordered amino acid (PONDR/FI)	Disorder ratio (PONDR/FI)	Disordered regions (PONDR/FI)
NADH dehydrogenase subunit 1	324	21	6.48	3
NADH dehydrogenase subunit 2	364	67	19.36	5
NADH dehydrogenase subunit 3	116	7	6.03	2
NADH dehydrogenase subunit 4L	98	3	3.06	2
NADH dehydrogenase subunit 4	459	58	12.64	6
NADH dehydrogenase subunit 5	605	97	16.03	7
NADH dehydrogenase subunit 6	173	18	10.4	2
Cytochrome c oxidase subunit I	515	15	2.91	2
Cytochrome c oxidase subunit II	227	30	13.22	3
Cytochrome c oxidase subunit III	261	25	9.58	2
Cytochrome b	380	16	4.21	4
ATP synthase F0 subunit 6	227	43	18.94	3
ATP synthase F0 subunit 8	54	6	11.11	1

TABLE 4. Folding rates of mitochondrial proteins predicted by CIpred. The folding rate was evaluated as log(K_f)/sec in the multi-state folding type.

MT-proteins	Folding rate
NADH dehydrogenase subunit 1	-2.5409
NADH dehydrogenase subunit 2	-2.5606
NADH dehydrogenase subunit 3	0.7458
NADH dehydrogenase subunit 4L	1.2965
NADH dehydrogenase subunit 4	-2.9992
NADH dehydrogenase subunit 5	-3.3112
NADH dehydrogenase subunit 6	-1.4492
Cytochrome c oxidase subunit I	-3.1748
Cytochrome c oxidase subunit II	-1.5227
Cytochrome c oxidase subunit III	-1.9486
Cytochrome b	-2.7169
ATP synthase F0 subunit 6	-1.9646
ATP synthase F0 subunit 8	8.1157
(Mean ± SD)	-0.2541±9.0

Analysis of Potential Disorder-based Functional Binding Sites

The functions of molecular recognition as well as the interactions between proteins involve the IDPs and the IDPRs. In addition, particular binding sites and regions are also present within these IDPs and IDPRs. These sites and regions have the potential to experience partial disorder-to-order transition during the binding process. The disorder-based protein-binding regions in mitochondrial proteins are predicted by experts using various computational methods including ANCHOR.

There was no anticipation of the presence of AIBSs (ANCHOR-identified binding sites) in the mitochondrial proteins as per Figure S2. The mitochondrial proteins were analyzed for detecting the presence of any (MoRFs) and the results were given in Figure 3. MoRF is short structure-prone segment of protein disorder found within long IDPR and is characterized with the potential to experience disorder-to-order transition during its binding with interaction partner. The recognition, regulation and signaling proteins usually have a large number of such binding sites as evident from the presence of a minimum of one α-helical MoRF in most of the eukaryotic proteins. It must be noted that different mitochondrial genome proteins show different number of MoRF segments and residues (length) as shown in Figure 3.

The number of MoRFs contained in the analyzed proteins ranged from two to nine and each of the MoRF was characterized with 1 to 9 residues. The proteins with the most MoRFs included NADH4L (containing 9 MoRFs), *NADH5* (containing 9 MoRFs)*, Cyt-b* (containing 8 MoRFs)*, COXIII* (containing 7 MoRFs) *and NADH2* (containing 6 MoRFs). Other proteins like ATPase 6 and COXII depicted 5 MoRFs. COI and NADH4L depicted 4 while ATPase 8, NADH6, NADH4 and NADH3 depicted 3 such regions. The lowest number of MoRFs (only 2) was seen in NADH subunit I. The MoRF regions in NADH1, NADH2, NADH3 and NADH4 were characterized with minimum amount of amino acid residues. There was abundance of arginine, cysteine, aspartic acid, leucine, serine, isoleucine, methionine, histidine, glutamine, glutamic acid and tyrosine residues in the MoRF segments of assessed mitochondrial proteins. Moreover, there was lack of lysin, glycine, glutamic acid, threonine, aspartic acid and asparagine residues

in the same mitochondrial proteins. The main location of the MoRF regions within the proteins was in N-terminus and C-terminus.

Functional Disorder and Interactomes of Mitochondrial Genome Proteins

The researcher achieved the interaction association among mitochondrial proteins and partner proteins with the help of STRING database (Search Tool for the Retrieval of Interacting Genes, http://string-db.org/). The analysis conducted on the basis of STRING allows the researchers to predict protein interaction networks that play an important role in electrons transfer from NADH to respiratory chain as well as to act as catalyst to facilitate the reduction of oxygen to water. Moreover, the oxidative phosphorylation process of production of ATP from ADP also involves these interaction networks (Figure 4).

The protein interaction network partners with high connectivity included NADH dehydrogenase subunits (1–6 subunits) that facilitated electron transfer between NADH and respiratory chain. Cytochrome oxidase subunits (COXI, COII and COXIII) are also an important segment of the protein interaction network and respiratory chain. These proteins act as catalyst to speed up reduction of oxygen to water. In addition, Cytochrome b (Cyt-b) along with cytochrome oxidase subunits and TP synthase generates electrochemical potential to facilitate reduction of oxygen to water. This protein interaction network also comprises of ATP synthase in the form of mitochondrial membrane. It helps conversion of ADP to ATP with proton gradient serving as an energy source for the ATP synthesis process. The proton gradient is formed as a result of the respiratory electron transfer chain. It is clear that the mitochondrial genome proteins are essential multifunctional proteins that play a crucial role in various biological processes. These are characterized with intrinsic disordered protein regions and consequently high binding affinity.

Discussion

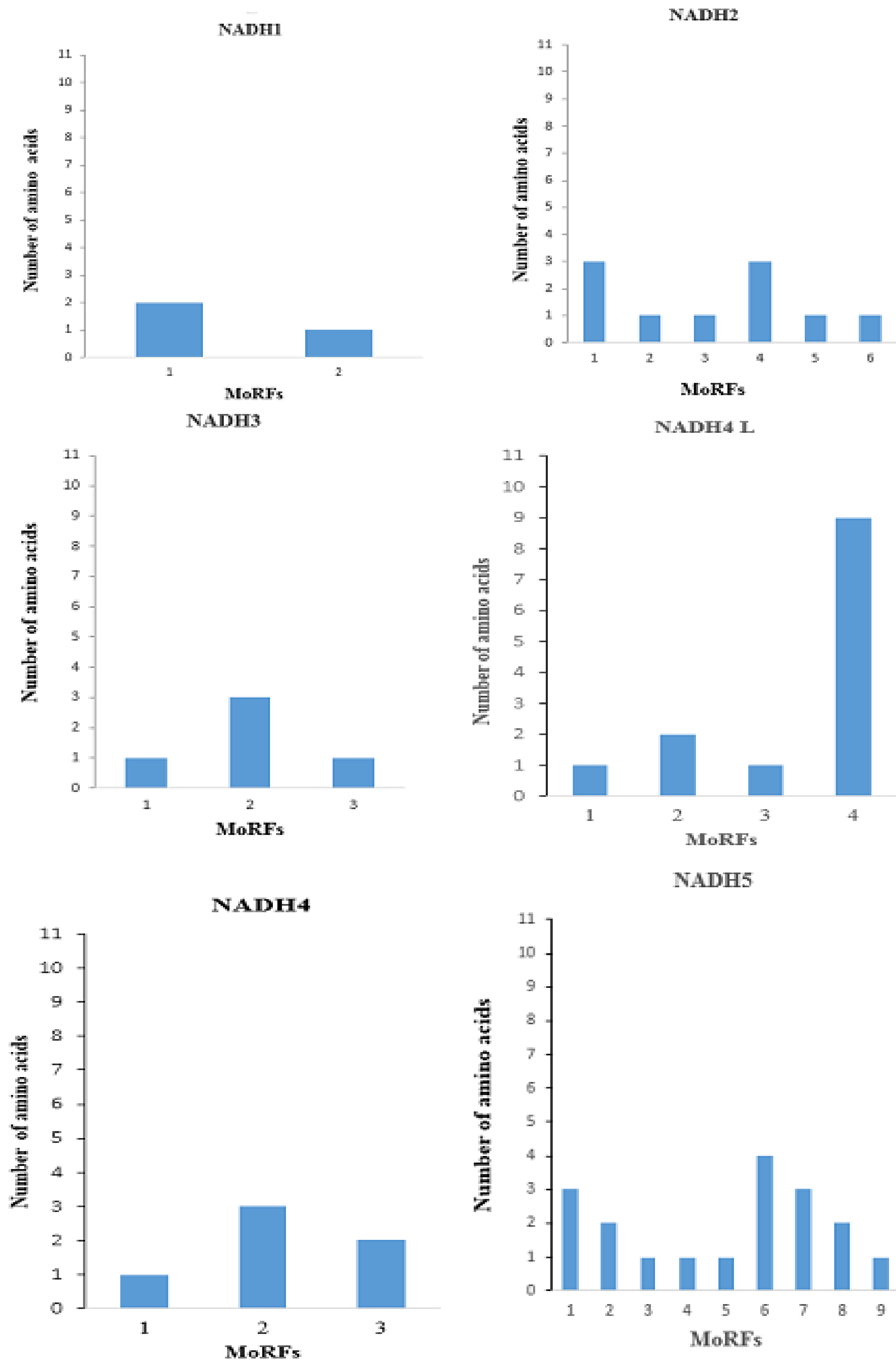
Intrinsic disordered proteins are more abundant in eukaryotes than in other domains of life [51, 55, 93, 94]. Furthermore, the highest ratio of IDPs (47%) has been found in nuclear proteins, with the average disorder level of transcription factors being as high as 63% [51, 95-98]. In contrast, mitochondrial proteins have revealed the lowest ID ratio (18% for mitochondrial membrane proteins and 13% for mitochondrial non-membrane proteins) [98].

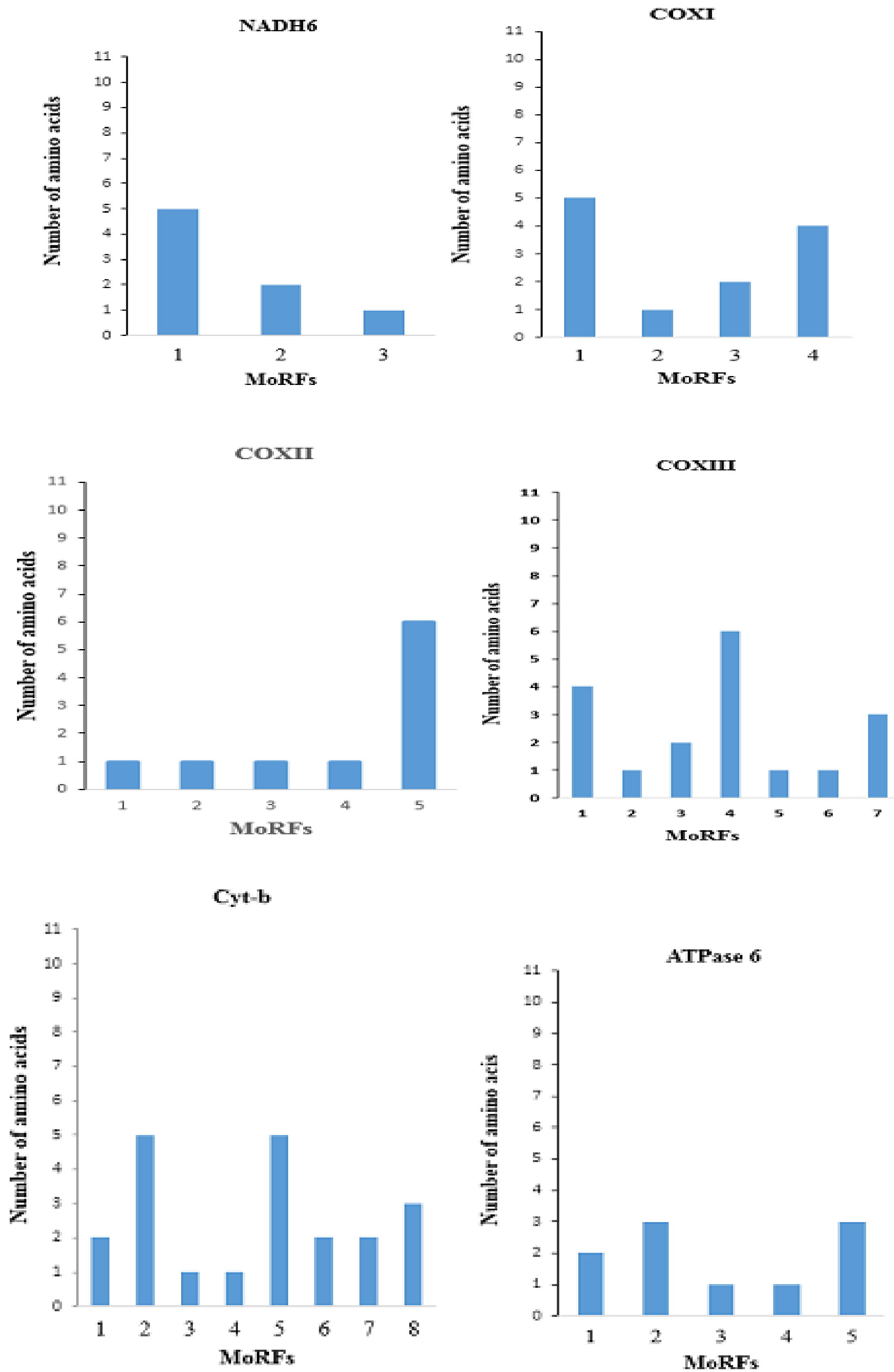
Our findings introduce the first analysis of the role of intrinsic disordered protein regions in mitochondrial genome proteins, which are important in ATP generation, oxygen metabolism, fatty acids β-oxidation, iron–sulfur bridges biogenesis, and are involved in regulation of stress-induced or programmed cell death pathways.

Our study demonstrated that the amino acid content was mostly similar for all mitochondrial genome proteins, although with many noticeable exceptions. The results of composition profile revealed that mitochondrial genome proteins are hybrid proteins constituted both of order and disorder compositions, while being closer to the compositions of disordered proteins. The intrinsic disorder of mitochondrial genome proteins sourced from the UniProt databases was evaluated using the PONDR VSL2 algorithms, regarded as among the most accurate disorder predictors [77-79]. Our finding was that these proteins are expected to be disordered and contained a considerable disordered region up to seven regions. The results showed that the intrinsic disordered ratio was different in all mitochondrial genome proteins, ranging from 7.5% to 15% as a mean value based on protein categories. This low ratio of intrinsic disorder in mitochondrial genome proteins was compatible with other studies which concluded that the ratio of IDPs was (47%) in respect of nuclear proteins, whereas it was low concerning mitochondrial proteins (13%–18%) [51, 95-98]. As such, the current study's results are consistent with other research reporting that many signaling proteins are known to capture a considerable amount of intrinsic disorder [42, 44, 99-101].

Other reports have suggested that the eukaryotic protein involved in signaling transduction network owns a larger number of intrinsic disorder protein regions than other domains of living organisms [51, 53-55, 57, 93]. Based on PONDR® VLXT analysis, eukaryotic signal transduction proteins are predicted to possess more significant disordered segments than other signal pathways; moreover, there is also considerable evidence suggesting a strong link between disorder in protein and its function [102]

The protein folding process is noted to be one of the most crucial biophysical processes, playing a fundamental role in a wide range of cellular pathways and processes. In addition, it


NADH1
Number of amino acids
MoRFs
NADH2
Number of amino acids
MoRFs
NADH3
Number of amino acids
MoRFs
NADH4 L
Number of amino acids
MoRFs
NADH4
Number of amino acids
MoRFs
NADH5
Number of amino acids
MoRFs


NADH6
Number of amino acids
MoRFs
COXI
Number of amino acids
MoRFs
COXII
Number of amino acids
MoRFs
COXIII
Number of amino acids
MoRFs
Cyt-b
Number of amino acids
MoRFs
ATPase 6
Number of amino acis
MoRFs

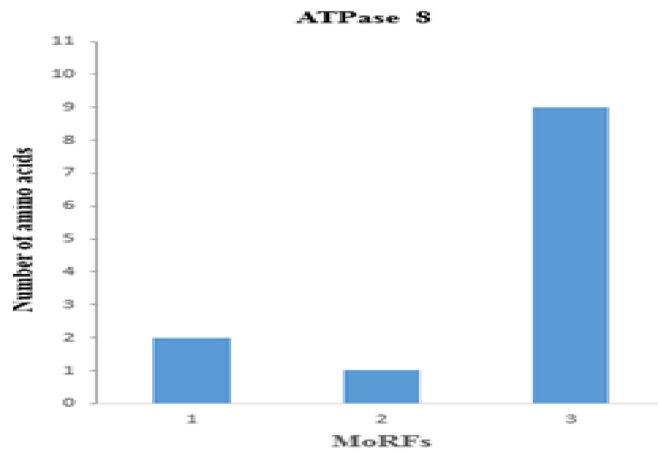


Fig. 3.The number and length of the molecular recognition features (MoRFs) in mitochondrial genome proteins based on the MoRF predictions.

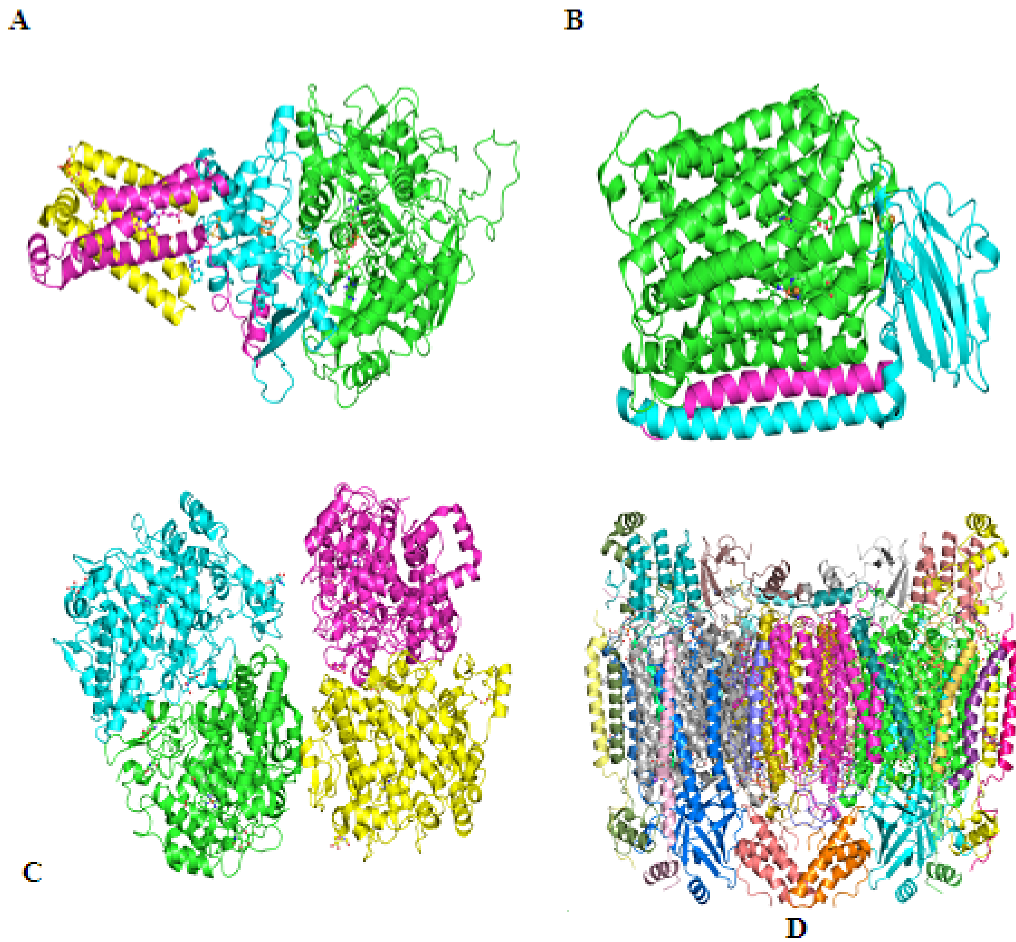


Fig. 4. Structural features of the representatives of the mitochondrial genome proteins. A. X-ray diffraction 3.35Å of the cytochrome b. B: X-ray 3.37Å crystal structure of cytochrome oxidase subunit I. C: X-ray crystal structure of cytochrome oxidase subunit II with 3.37Å resolution. D: X-ray diffraction with 3 Å resolution of cytochrome oxidase subunit III. Analysis of the interactivity of the mitochondrial genome proteins by STRING; where E: NADH subunits network interactivity; F: ATPase 6 and 8 network interactivity; G: Cytochrome oxidase subunits and cytchrome b network interactivity.

is involved in many complex events in the cell cycle and positions and transportation of protein through cell membrane and other mechanisms and pathways. Consequently, any insufficiency or defect in protein folding may lead to stoppage of many cellular events and could be associated with certain diseases [90]. So, switching of protein from disorder to ordered status is related to folding rate. This rate was determined using the log (Kf) algorithm, which is associated with folding kinetics and dynamic process, and the rate was found to be higher in respect of disorder status than in order [95]. The folding time of mitochondrial genome proteins ranged from (log (Kf)/sec = (-3.31)) to (log (Kf)/sec = (8.11)). In addition, the NADH5, COXI, NADH4, Cyt-b, NADH1 and NADH2 were all predicted to fold faster than other mitochondrial genome proteins, whereas the folding rates were very slow for ATPase subunit 8, NADH4L and NADH3.

The binding sites in disordered region in mitochondrial proteins were determined using the ANCHOR prediction. This prediction depends upon three properties as follows: localization of amino acid residues inside the long disordered region, to ensure that the residues are able to communicate with local sequential adjacent, so that the protein can fold, and finally, the inability of given amino acids to communicate with globular protein [91,92]. The results showed that the mitochondrial genome proteins were predicted to have no disorder-binding regions in any of the tested proteins.

The protein–protein interaction of mitochondrial proteins was estimated using MoRF prediction. From the results, the number and length of MoRF segments were seen to be variable among these proteins. All given protein sequences were composed of two to nine short MoRFs of 1–9 amino acid residues in length. The current study showed that the identified MoRF region in mitochondrial genome proteins contains both enriched and depleted amino acid residues. These results are consistent with other study findings that revealed many enriched amino acid residues, which promote protein disorder and depleted residues in the MoRFs [91,92]. Our analysis is consistent with other research suggesting that the IDRs are involved in proteins that are integrated in protein–protein interaction and metabolic pathways [51, 53-55, 57, 93].

Conclusions

This investigation evaluated the propensity of mitochondrial genome proteins to exhibit intrinsic disorder. The analysis revealed that intrinsic protein disorder regions are found in mitochondrial genome proteins. In addition, these disordered regions contain several short binding segments based on MoRF predictors. Moreover, this modifying structure in these kinds of proteins causes their binding to unrelated protein partners. This feature has a considerable impact on the function of protein complex networks in ATP generation and oxygen metabolism in response to stress-induced and developmental cell death programs.

Acknowledgements

This study was funded by the Deputy of Scientific Research of King Abdulaziz University (Grant Number: 1336-141-1440). The authors gratefully acknowledge the technical and financial support of the Deanship of Scientific Research.

Conflict of Interest

No to declare.

References

1. Barrera-Paez, J.D. and Moraes, C.T. Mitochondrial genome engineering coming-of-age. *Trends in Genetics,* **38** (8), 869-880(2022).

2. Klucnika, A. and Ma, H. Mapping and editing animal mitochondrial genomes: can we overcome the challenges?. *Philos Trans R Soc. Lond B Biol. Sci.,* **20 (**375),1790, 20190187 (2020).

3. Rai, P.K., Craven, L., Hoogewijs, K., Russell, O.M and Lightowlers, R.N. Advances in methods for reducing mitochondrial DNA disease by replacing or manipulating the mitochondrial genome. *Essays Biochem.,* **62**(3),455-465 (2018).

4. Doyle, J.J. and Gaut, B.S. Evolution of genes and taxa: A primer. *Plant Mol. Biol.,* **42**(1), 1-23(2000).

5. Miao, Y.W., Miao, Y-W. , Peng, M-S. , Wu, G-S , Ouyang, Y-N. , Yang, Z-Y. , Yu, N. , Liang, J-P. , Pianchou, G. , Beja-Pereira, A. , Mitra, B ., Palanichamy, M. G. , Baig, M. , Chaudhuri, T. K. , Shen, Y-Y. , Kong, Q-P. , Murphy, R. W. , Yao, Y-G. and Zhang, Y-P. .Chicken domestication: An updated perspective based on mitochondrial genomes. *Heredity (Edinb),* **110**(3), 277-282(2013).

6. Achilli, A., Olivieri, A., Pellecchia, M., Uboldi, C., Colli, L. and Al-Zahery, N . Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Curr. Biol.,* **18**(4), R157-8 (2008).

7. Bonfiglio, S., Achilli, A., Olivieri, A., Negrini, R., Colli, L. and Liotta, L.The enigmatic origin of bovine mtDNA haplogroup R: Sporadic interbreeding or an independent event of Bos primigenius domestication in Italy? *PLoS One,* **5**(12), e15760(2010).

8. Achilli, A., Bonfiglio, S., Olivieri, A., Malusa, A., Pala, M. and Kashani, B.H. The multifaceted origin of taurine cattle reflected by the mitochondrial genome. *PLoS One.,* **4**(6), e5753(2009).

9. Pang, J.F., Kluetsch, C., Zou, X.J., Zhang, A.B., Luo, L.Y. and Angleby, H. mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Mol. Biol. Evol.,* **26**(12), 2849-2864(2009).

10. Achilli, A., Olivieri, A., Soares, P., Lancioni, H., Hooshiar, K.B. and Perego, U.A. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc. Natl. Acad. Sci. U S A.,* **109**(7), 2449-2454(2012).

11. Wu, G.S., Yao, Y.G., Qu, K.X., Ding, Z.L., Li, H. and Palanichamy, M.G. Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East. *Asia. Genome Biol.,* **8**(11), R245(2007).

12. Tamm, E., Dudley, J., Nei, M. and Kumar, S. Beringian standstill and spread of Native American founders. *PLoS One.,* **2**(9), e829(2007).

13. Smith, I. N., Thacker, S., Seyfi, M., Cheng, F. and Eng, C. Conformational Dynamics and Allosteric Regulation Landscapes of Germline Pten Mutations Associated with Autism Compared to Those Associated with Cancer. *Am. J. Hum. Genet.,* **104**, 861–878(2019).

14. Smith, I. N., Thacker, S., Jaini, R. and Eng, C. Dynamics and Structural Stability Effects of Germline Pten Mutations Associated with Cancer Versus Autism Phenotypes. *J. Biomol. Struct. Dyn.,* **37**, 1766–1782(2019).

15. Campitelli, P.,Modi, T., Kumar, S. and Ozkan, S. B. The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annu. Rev. Biophys.,* **49**, 267–288(2020).

16. Ganguly, P., Madonsela, L., Chao, J. T., Loewen, C. J. R., O'Connor, T. P., Verheyen, E. M. and Allan, D. W. A Scalable Drosophila Assay for Clinical Interpretation of Human Pten Variants in Suppression of Pi3k/Akt Induced Cellular Proliferation. *PLoS Genet.,* **17**, e100977410(2021).

17. Boldon, L., Laliberte, F. and Liu, L. Review of the Fundamental Theories Behind Small Angle X-Ray Scattering, Molecular Dynamics Simulations, and Relevant Integrated Application. *Nano Rev.,* **6**, 25661(2015).

18. Georgescu, M. M., Kirsch, K. H., Kaloudis, P., Yang, H., Pavletich, N. P. and Hanafusa, H. Stabilization and Productive Positioning Roles of the C2 Domain of Pten Tumor Suppressor. *Cancer Res.,* **60**, 7033–7038(2000).

19. McBride, H.M., M. Neuspiel, and S. Wasiak, Mitochondria: More than just a powerhouse. *Curr. Biol.,* **16**(14), 551-560(2006).

20. Nunnari, J. and Suomalainen, A. Mitochondria: In sickness and in health. *Cell,* **148**(6), 1145-1159 (2012).

21. Dawson, J. E., Smith, I. N., Martin, W., Khan, K., Cheng, F. and Eng, C. Shape shifting: The multiple conformational substates of the PTEN N-terminal PIP2 -binding domain. *Protein Sci.,* **31**, e430810(2022).

22. Wallace, D.C., A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: A dawn for evolutionary medicine. *Annu. Rev. Genet.,* **39**,359-407(2005).

23. Benard, G. and Karbowski, M. Mitochondrial fusion and division: Regulation and role in cell viability. *Semin. Cell Dev. Biol.,* **20**(3), 365-74(2009).

24. Leitner, M. G., Hobiger, K., Mavrantoni, A., Feuer, A., Oberwinkler, J., Oliver, D. and Halaszovich C. R. A126 in the Active Site and Ti167/168 in the Ti Loop Are Essential Determinants of the Substrate Specificity of Pten. *Cell. Mol. Life Sci.,* **75**, 4235–4250 (20180).

25. Suen, D.F., Norris, K.L. and Youle, R.J. Mitochondrial dynamics and apoptosis. *Genes Dev.,* **22**(12), 1577-1590(2008).

26. Modi, T., Risso, V. A., Martinez-Rodriguez, S., Gavira, J. A., Mebrat, M. D., Van Horn, W. D., Sanchez-Ruiz, J. M. and Banu Ozkan, S. Hinge-shift mechanism as a protein design principle

for the evolution of β-lactamases from substrate promiscuity to specificity. *Nat. Commun.,* **12**, 1852(2021).

27. Dunker, A.K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C. and Villafranca, J.E. Intrinsically disordered protein. *J. Mol. Graph. Model.,* **19**(1), 26-59(2001).

28. Dunker, A.K., Brown,C.J. and Obradovic, Z. Identification and functions of usefully disordered proteins. *Adv. Protein Chem.,* **62**, 25-49(2002).

29. Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L. and Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry,* **41**(21), P. 6573-6582(2002).

30. Dunker, A.K. and Uversky, V.N. Signal. transduction via unstructured protein conduits. *Nat. Chem. Biol.,* **4**(4),229-230(2008).

31. Dunker, A.K. and Obradovic, Z. The protein trinity--linking function and disorder. *Nat. Biotechnol.,* **19**(9), 805-806(2001).

32. Radivojac, P., Iakoucheva, L.M., Oldfield, C.J., Obradovic, Z., Uversky, V.N. and Dunker, A.K. Intrinsic disorder and functional proteomics. *Biophys J.,* **92**(5), 1439-1456(2007).

33. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.,* **27**(10), 527-533(2002).

34. Tompa, P. Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays,* **25**(9), 847-855(2003).

35. Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.,* **579**(15), 3346-3354(2005).

36. Tompa, P., Schad, E., Tantos, A. and Kalmar, L. Close encounters of the third kind: Disordered domains and the interactions of proteins. *Bioessays,* **31**(3), 328-335(2009).

37. Wright, P.E. and Dyson, H.J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.,* **293**(2), 321-331(1999).

38. Uversky, V.N. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.,* **11**(4), 739-756(2002).

39. Uversky, V.N. What does it mean to be natively unfolded? *Eur. J. Biochem.,* **269**(1), 2-12 (2002).

40. Uversky, V.N., Gillespie, J.R. and Fink, A.L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins,* **41**(3), 415-427(2000).

41. Uversky, V.N., Oldfield, C.J. and Dunker, A.K. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.,* **18**(5), 343-384(2005).

42. Uversky, V.N. and Dunker, A.K. Understanding protein non-folding. *Biochim. Biophys. Acta.,* **1804**(6), 1231-1264(2010).

43. Uversky, V.N. Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: Which way to go? *Cell Mol. Life Sci.,* **60**(9), 1852-1871(2003).

44. Uversky, V.N. Intrinsic disorder-based protein interactions and their modulators. *Curr. Pharm. Des.,* **19**(23), 4191-4213(2013).

45. Habchi, J., Tompa, P., Longhi, S. and Uversky, V.N. Introducing protein intrinsic disorder. Introducing protein intrinsic disorder. *Chem. Rev.,* **114**(13), 6561-6588(2014).

46. Mohan, A., Sullivan, W.J., Radivojac, P., Dunker, A.K. and Uversky, V.N. Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.,* **362**(5), 1043-1059(2006).

47. Goh, G.K., Dunker,A.K. and Uversky, V.N. Protein intrinsic disorder toolbox for comparative analysis of viral proteins. *BMC Genomics,* **9** (Suppl 2), S4(2008).

48. Goh, G.K., Dunker, A.K. and Uversky,V.N. A comparative analysis of viral matrix proteins using disorder predictors. *Virol. J.,* **5**, 126(2008).

49. Uversky, V.N., Roman, A., Oldfield, C.J. and Dunker, A.K. Protein intrinsic disorder and human papillomaviruses: Increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. *J. Proteome Res.,* **5**(8), 1829-1842(2006).

50. Uversky, V.N. Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: Another illustration of the D(2) concept. *Expert. Rev. Proteomics,* **7**(4), 543-564(2010).

51. Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. and Jones, D.T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.,* **337**(3), 635-645(2004).

52. Oldfield, C.J., Cheng, Y., Cortese, M.S., Brown, C.J., Uversky, V.N. and Dunker, A.K. Comparing and combining predictors of mostly disordered proteins. *Biochemistry,* **44**(6), 1989- 2000 (2005).

53. Schad, E., Tompa, P. and Hegyi, H. The relationship between proteome size, structural disorder and organism complexity. *Genome Biol.,* **12**(12), R120(2011).

54. Pancsa, R. and Tompa, P. Structural disorder in eukaryotes. *PLoS One,* **7**(4), e34687(2012).

55. Xue, B., Dunker, A.K. and Uversky, U.V. Orderly order in protein intrinsic disorder distribution: Disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.,* **30**(2), 137-149(2012).

56. Xue, B., Williams, R.W., Oldfield, C.J., Goh, G.K., Dunker, A.K. and Uversky, U.V.Viral disorder or disordered viruses: Do viral proteins possess unique features? *Protein Pept. Lett.,* **17**(8), 932-951(2010).

57. Xue, B. and Uversky, V.N. Intrinsic disorder in proteins involved in the innate antiviral immunity: Another flexible side of a molecular arms race. *J. Mol. Biol.,* **426**(6), 1322-1350 (2014).

58. Uversky, V.N., Li, J. and Fink, A.L. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J. Biol. Chem.,* **276**(14), 10737-10744(2001).

59. Uversky, V.N., Oldfield, C.J. and Dunker, A.K. Intrinsically disordered proteins in human diseases: Introducing the D2 concept. *Annu. Rev. Biophys.,* **37**, 215-246 (2008).

60. Uversky, V.N. Multitude of binding modes attainable by intrinsically disordered proteins: A portrait gallery of disorder-based complexes. *Chem. Soc. Rev.,* **40**(3), 1623-1634(2011).

61. Tompa, P., Schad, E., Tantos, A. and Kalmar, L. Intrinsically disordered proteins: Emerging interaction specialists. *Curr. Opin. Struct. Biol.,* **35**, 49-59(2015).

62. Oldfield, C.J., Meng, J., Yang, J.Y., Yang, M.Q., Uversky, V.N. and Dunker, A.K. Flexible nets: Disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics,* **9** (Suppl 1), S1(2009).

63. Chang, C.I., Zhang, Y.A., Zou, J., Nie, P. and Secombes, C.J. Two cathelicidin genes are present in both rainbow trout (Oncorhynchus mykiss) and atlantic salmon (Salmo salar). *Antimicrob. Agents Chemother.,* **50** (185), e195(2006).

64. Ning, S., Liu, J., Liu, N. and Yan, D. The Accuracy of Force Fields on the Simulation of Intrinsically Disordered Proteins: A Benchmark Test on the Human P53 Tumor Suppressor. *J. Theor. Comput. Chem.,* **19,** 2050011(2020).

65. Kaynak, B. T., Bahar, I. and Doruker, P. Essential Site Scanning Analysis: A New Approach for Detecting Sites That Modulate the Dispersion of Protein Global Motions. *Comput. Struct. Biotechnol. J.,* **18,** 1577–1586(2020).

66. Adkins, J.N. and Lumb, K.J. Intrinsic structural disorder and sequence features of the cell cycle inhibitor p57Kip2. *Proteins,* **46**(1), 1-7(2002).

67. Milona, P., Townes, C.L., Bevan, R.M. and Hall, J. The chicken host peptides, gallinacins 4, 7, and 9 have antimicrobial activity against Salmonella serovars. *Biochem. Biophys. Res. Commun.,* **356**, 169-174(2007).

68. Hegyi, H., Buday, L. and Tompa, P. Intrinsic structural disorder confers cellular viability on oncogenic fusion proteins. *PLoS Comput. Biol.,* **5**(10), e1000552(2009).

69. Erkizan, H.V., Uversky,V.N. and Toretsky, J.A. Oncogenic partnerships: EWS-FLI1 protein interactions initiate key pathways of Ewing's sarcoma. *Clin. Cancer Res.,* **16**(16), 4077-4083 (2010).

70. Goitsuka, R., Chen, C.L., Benyon, L., Asano, Y., Kitamura, D. and Cooper, M.D. Chicken cathelicidin-B1, an antimicrobial guardian at the mucosal M cell gateway. *Proc. Natl. Acad. Sci. U. S. A.,* **104,** 15063-15068(2000).

71. Habchi, J., Tompa, P., Longhi, S. and Uversky, V.N. Introducing protein intrinsic disorder. *Chem. Rev.,* **114**, 6561-6588(2014).

72. Nan, Y.H., Bang, J.K., Jacob, B., Park, I.S. and Shin, S.Y. Prokaryotic selectivity and LPS-neutralizing activity of short antimicrobial peptides designed from the human antimicrobial peptide LL-37. *Peptides*, **35,** 239-247(2012).

73. Achanta, M., Sunkara, L.T., Dai, G., Bommineni, Y.R., Jiang, W. and Zhang, G. Tissue expression and developmental regulation of chicken cathelicidin antimicrobial peptides. *J. Anim. Sci. Biotechnol.***, 3**, 15(2012).

74. Yuan J. Zooarchaeological study on the domestic animals in ancient China. *Quaternary Sci.,* **30**, 298–306(2010).

75. Malaney, P., Pathak R. R., Xue B., Uversky V. N. and Davé, V. Intrinsic disorder in PTEN and its interactome confers structural plasticity and functional versatility. *Sci. Rep.,* **3**, 2035(2013).

76. van Oven, M. and Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.,* **30**, E386–E394(2009).

77. Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., Dunker, A.K. and Obradovic, Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J. Bioinform. Comput. Biol.,* **3**(1), 35-60(2005).

78. Peng, Z.L. and Kurgan, L. Comprehensive comparative assessment of in-silico predictors of disordered regions. *Curr. Protein Pept. Sci.,* **13**(1), 6-18(2012).

79. Fan, X. and Kurgan, L. Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. *J. Biomol. Struct. Dyn.,* **32**(3), 448-464 (2014).

80. Ma, B.G., Guo, J.X. and Zhang, H.Y. Direct correlation between proteins' folding rates and their amino acid compositions: An ab initio folding rate prediction. *Proteins,* **65**(2), 362-372 (2006).

81. Ma, B.G., Chen, L.L. and Zhang, H.Y. What determines protein folding type? An. investigation of intrinsic structural properties and its implications for understanding folding mechanisms. *J. Mol. Biol.,* **370**(3), 439-448(2007).

82. Diamond, J.M. Taiwan's gift to the world. *Nature,* **403,**709–710(2000).

83. Schad, E., Tompa, P. and Hegyi, H. The relationship between proteome size, structural disorder and organism complexity. *Genome Biol.,* **12**, R120(2011).

84. Scocchi, M., Pallavicini, A., Salgaro, R., Bociek, K. and Gennaro, R.The salmonid cathelicidins: a gene family with highly varied C-terminal antimicrobial domains. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.,* **152,** 376-381(2009).

85. Smith, J.G. and Nemerow, G.R. Mechanism of adenovirus neutralization by Human alpha-defensins. *Cell Host Microbe,* **3**, 11-19(2008).

86. Sugiarto, H. and Yu, P.L. Avian antimicrobial peptides: the defense role of beta-defensins. *Biochem. Biophys. Res. Commun.,* **323**, 721-727(2004).

87. Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M. and Stromstedt, L. Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genet.,* **4**, e1000010(2008).

88. Uversky, V.N. Disordered competitive recruiter: Fast and foldable. *J. Mol. Biol.,* **418**(5), 267-268(2012).

89. Uversky, V.N. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta.,* **1834**(5), 932-951(2013).

90. Dyson, H.J. and Wright, P.E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.,* **12**(1), 54-60(2002).

91. Meszaros, B., Simon, I. and Dosztanyi, Z. Prediction of protein binding regions in disordered proteins. *PLoS Comput. Biol.,* **5**(5), e1000376(2009).

92. Dosztanyi, Z., Meszaros, B. and Simon, I. ANCHOR: Web server for predicting protein binding regions in disordered proteins. *Bioinformatics,* **25**(20), 2745-2746(2009).

93. Dunker, A.K. and Obradovic, Z. Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform.,* **11**, 161-171(2000).

94. Peng, Z., Yan, J., Fan, X., Mizianty, M.J., Xue, B., Wang, K., Hu, G., Uversky, V.N. and Kurgan, L. Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol. Life Sci.,* **72**(1), 137-51(2015).

95. Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N. and Dunker, A.K. Intrinsic disorder in transcription factors. *Biochemistry,* **45**(22), 6873-6888(2006).

96. Minezaki, Y., Homma, K., Kinjo, A.R. and Nishikawa, K. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.,* **359**(4), 1137-1149 (2006).

97. Fukuchi, S., Homma, K., Minezaki, Y., Gojobori, T. and Nishikawa, K. Development of an accurate classification system of proteins into structured and unstructured regions that uncovers novel structural domains: Its application to human transcription factors. *BMC Struct. Biol.,* **9**, 26(2009).

98. Ito, M., Tohsato, Y., Sugisawa, H., Kohara, S., Fukuchi, S., Nishikawa, I. and Nishikawa, K. Intrinsically disordered proteins in human mitochondria. *Genes Cells,* **17**(10), 817-825 (2012).

99. 99. Dunker, A.K., Babu, M., Barbar, E., Blackledge, M., Bondos, S.E., Dosztanyi, Z., Dyson, H.J., Forman-Kay, J., Fuxreiter, M., Gsponer, J., Han, K.-H., Jones, D.T., Longhi, S., Metallo, S.J., Nishikawa, K., Nussinov, R., Obradovic, Z., Pappu, R., Rost, B., Selenko, P., Subramaniam, V., Sussman, J.L., Tompa, P. and Uversky, V.N. What's in a name? Why these proteins are intrinsically disordered. Intrinsically Disord. *Proteins,* **1**, e241579(2013).

100. Fan, X. and Kurgan, L. Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. *J. Biomol. Struct. Dyn.,* **32**, 448-464 (2014).

101. Tantos, A., Han, K.H. and Tompa, P. Intrinsic disorder in cell signaling and gene transcription. *Mol. Cell Endocrinol.,* **348**(2), 457-465(2012).

102. Vacic, V., Markwick, P.R., Oldfield, C.J., Zhao, X., Haynes, C., Uversky, V.N. and Iakoucheva, L.M. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. *PLoS Comput. Biol.,* **8,** e1002709(2012).

جينوم الميتوكوندريا مع احتمال الاضطراب الجوهري لبروتينها

ماجد مصطفى محمود[١،٢،٣] ، **سروار جمال**[٤]، **طلال قداح**[٢،٥]، **كوثر على الزاهر**[٦]، **هيثم احمد يعقوب**[٧*]
[١]مركز الملك فهد للبحوث الطبية ، جامعة الملك عبد العزيز ، جدة ، المملكة العربية السعودية ٢١٥٨٩
[٢]قسم علوم المختبرات الطبية، كلية العلوم الطبية التطبيقية ، جامعة الملك عبد العزيز ، جدة ، المملكة العربية السعودية ٢١٥٨٩
[٣]قسم الوراثة الجزيئية والإنزيمات، معهد الوراثة البشرية وأبحاث الجينوم ، المركز القومي للبحوث ، القاهرة ، مصر
[٤]مركز العلوم البيولوجية التكاملي ، ٦١٣٥ وودوارد افي ، جامعة واين ستيت ، ديترويت ميتشيغان- ٤٨٢٠٢
[٥]وحدة أبحاث أمراض الدم ، مركز الملك فهد للبحوث الطبية ، جامعة الملك عبد العزيز ، جدة ، ٢١٥٨٩ ، المملكة العربية السعودية
[٦]وحدة علم المناعة ، مركز الملك فهد للبحوث الطبية ، جامعة الملك عبد العزيز ، جدة ، ٢١٥٨٩ ، المملكة العربية السعودية
[٧]قسم بيولوجيا الخلية ، معهد بحوث التكنولوجيا الحيوية ، المركز القومي للبحوث ، الدقي ، الجيزة ، ١٢٦٢٢ مصر

تهدف الدراسة الحالية إلى دراسة بروتينات جينوم الميتوكوندريا مع التركيز بشكل خاص على توزيع المناطق المضطربة جوهريًا داخل هذه البروتينات. تم استخدام UniProt للحصول على تسلسل الأحماض الأمينية لبروتينات الميتوكوندريا. بعد ذلك ، تمت دراسة التسلسلات التي تم الحصول عليها لتحديد ميزاتها الفريدة. استخدمنا خوارزمية PONDR VSL2 لتقييم البروتينات لأي ترتيب أو اضطراب. وفقًا لنتائج الدراسة ، كان من المتوقع أن تصور بروتينات جينوم الميتوكوندريا مستويات غير متوقعة من الاضطراب الجوهري. أشارت النتائج أيضًا إلى إمكانية وجود ٧ مناطق اضطراب جوهري داخل هذه البروتينات. تم العثور على المناطق المضطربة في الغالب مرتبطة بتفاعلات البروتين والبروتين. أشار البحث إلى وجود ميزتين إلى تسع سمات تمييز جزيئي في بروتينات الميتوكوندريا التي تم تقييمها. تضمنت هذه الميزات مقاطع قصيرة معرضة للهيكل ، وموقع بروتينات الميتوكوندريا داخل مناطق مضطربة طويلة وميزة إظهار انتقال الاضطراب إلى الترتيب أثناء الارتباط. علاوة على ذلك ، تم تخصيص هذه المناطق بعدد من المواقع التي تتميز بتعديلات ما بعد عملية الترجمة. وفقًا للتحليل ، قد يكون هناك العديد من المناطق المضطربة الموجودة داخل بروتينات جينوم الميتوكوندريا مما يؤدي إلى طبيعة متعددة الوظائف لبروتينات جينوم الميتوكوندريا في مسارات نقل الإشارة حيث تنظم هذه البروتينات مسارات استقلاب الخلية.

الكلمات الدالة: الميتوكوندريا، جينوم، برويتن، المناطق المضطربة جوهريا، التنبؤ.