*Research Article*

# Functional and Structural Analysis of Predicted Proteins Obtained from *Homo sapiens'* Minisatellite 33.15-Tagged Transcript pAKT-45 Variants

**Mohd Shahbaaz,**[1,2] **Awad Saeed Al-Samghan,**[3] **Arshi Malik** (iD)**,**[4] **Sarah Afaq,**[4] **Afaf S. Alwabli,**[5] **Irfan Ahmad,**[6,7] **Mostafa A. Hussien,**[8] **Mohammad Zubair,**[9] **Farha Fatima,**[10] **Shamina Begum,**[9] **Othman Alzahrani,**[11,12] **and Mohammed Tarique** (iD)[13]

[1]*South African Medical Research Council Bioinformatics Unit, South African National, Bioinformatics Institute, University of the Western Cape, South Africa*

[2]*Laboratory of Computational Modeling of Drugs, South Ural State University, 76 Lenin Prospekt, Chelyabinsk, Russia 454080*

[3]*Department of Family and Community Medicine, College of Medicine, King Khalid University, Abha 61421, Saudi Arabia*

[4]*Department of Clinical Biochemistry, College of Medicine, King Khalid University, Abha, Saudi Arabia*

[5]*Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia*

[6]*Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, King Khalid University, Abha, Saudi Arabia*

[7]*Research Center for Advanced Materials Science, King Khalid University, Abha, Saudi Arabia*

[8]*Department of Chemistry, Faculty of Science, King Abdulaziz University, P.O. Box 80203 Jeddah 21589, Saudi Arabia*

[9]*Department of Medical Microbiology, Faculty of Medicine, University of Tabuk, Tabuk 71491, Saudi Arabia*

[10]*Department of Zoology, Faculty of Life Science, Aligarh Muslim University, Aligarh 202002, India*

[11]*Department of Biology, Faculty of Science, University of Tabuk, Tabuk, Saudi Arabia*

[12]*Genome and Biotechnology Unit, Faculty of Sciences, University of Tabuk, Tabuk, Saudi Arabia*

[13]*Center for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, Jamia Nagar, New Delhi 110025, India*

Correspondence should be addressed to Arshi Malik; arshimalik@gmail.com and Mohammed Tarique; tariqueaiims@gmail.com

The spermatozoa are transcriptionally dormant entities which have been recognized to be an archive of mRNA, coding for a variety of functionally crucial cellular proteins. This significant repository of mRNA is predicted to be associated with early embryogenesis and postfertilization. The mRNA transcripts which are tagged with minisatellites have been involved in the regulation of the gene functions as well as their organization. However, very little information is available regarding the expression of the transcripts tagged with minisatellites in spermatozoa. Therefore, in order to understand the functions and the conformational behavior of the proteins expressed from these minisatellite-tagged transcripts, we have performed a detailed *in silico* analysis using the sequences of the transcripts. The protein predicted from KF274549 showed the functionalities similar to uncharacterized C4orf26 proteins, while that obtained from KF274557 predicted to be a metallophosphoesterase. Furthermore, the structural folds in the structure of these predicted proteins were analyzed by using the homology modeling and their conformational behaviors in the explicit water conditions were analyzed by using the techniques of Molecular Dynamics (MD) simulations. This detailed analysis will facilitate the understanding of these proteins in the spermatozoon region and can be used for uncovering other attributes of the metabolic network.

## 1. Introduction

The ejaculated spermatozoa correspond to fixed terminally differentiated cells, which are devoid of the transcription as well as translation of the nuclear-encoded mRNAs. Consequently, the spermatozoa carry only the paternal genome to the ooplasm. The knowledge obtained from the discovery of a variety of soluble signaling molecules, transcription factors, and the molecules carried by the spermatozoa into the zygotic cytoplasm during fertilization has revolutionized this assessment (Saunders, Larman et al. [1–3]). Although the spermatozoa generally retain the transcriptionally dormant state, still the constituent mRNA transcripts present in its framework expressed into a variety of the transcription factors and these proteins may be involved in cell proliferation, signal transduction, regulation of sperm motility, acrosome reaction chromatin condensation, and capacitation [1, 2, 4–6]. Furthermore, the release of spermatozoal transcripts into the ooplasm is predicted to have a significant role during the process of fertilization and later stages. Generally, the spermatozoa contain around 3000–5000 mRNA transcripts, which may be involved in the expression and regulation of various constituent molecules [7–9].

The repeats present in the DNA sequences are dynamic elements of the genome and form the major portion of the satellites' regions and transposable components [10, 11]. These repeats are usually observed in the noncoding fragments of the genomes whereas a minor portion is preserved inside the transcriptome [12–14] and involved in gene regulation during gene silencing, transcription, and translation [9, 15, 16]. The mechanisms related to the expression and organization of these conserved repeats in the mammalian transcriptomes, predominantly in the spermatozoa, still remain undiscovered. Therefore, an *in silico* methodology was used for understanding the functions and conformational behavior of proteins expressed from such satellite DNA sections. The different functionalities were observed for the two proteins, KF274549 and KF274557, that were uncovered from *Homo sapiens'* minisatellite 33.15-tagged transcript sequence. The KF274549 showed similarities to the family of C4orf26 proteins, which are the group of hypothetical proteins. The hypothetical or uncharacterized proteins are the predicted polypeptides which do not have the experimental evidences at the biochemical levels [17–21]. The information regarding the functionalities of such proteins can be helpful in understanding the hidden mechanism behind the pathogenesis of a variety of microbial organisms [17–20]. Furthermore, the protein obtained from KF274557 showed the presence of metallophosphoesterase activities. These analyses can be useful in understanding the expression behavior of the satellite regions and how they are modulating the functionalities of other biomolecules present in the metabolic networks.

## 2. Material and Methods

The *in silico* methodology used for the analyses of the predicted proteins was obtained from KF274549 and KF274557. The primary stages involve the establishment of the phylogenetic relationships among the close homologs of the predicted proteins. Then, sequences of these proteins were utilized as the inputs in order to predict the conserved motifs and domains along with the functions they may perform in the metabolic network. We have also observed the availability of the possible interaction partners in the biological databases that may modulate their functionalities. Furthermore, the three dimensional (3-D) structures of these predicted proteins were modeled by using the X-ray crystal structures present in the publically available databases. We have also analyzed the conformational changes of these proteins in the explicit water environment by using the available methods of the MD simulations. The phases of the adopted methodology are explained here in details:

*2.1. Protein Translation.* The protein sequences for KF274549 and KF274557 were predicted by using the "Translate Tool" present in the Expasy server on the basis of standard genetic code and validated by using the "Translation" Module of the Discovery Studio 4.0 [22]. The Expasy's translate tools predicted six different protein sequences for each query DNA sequence. The BLASTx [23] was used for the selection of the most suitable translated protein sequence for further study. For KF274549, the output based on $3'$-$5'$ Frame 2 was selected while for KF274557, the $5'$-$3'$ Frame 3 was selected.

*2.2. Sequence Analyses.* A diverse range of bioinformatics was utilized for the functional analyses of the obtained proteins. In order to obtain the close homologs for the respective proteins, the sequence similarity tools such as BLASTp [23], HMMER [24], and HHpred [25] were utilized. For each sequence search, the proteins with low sequence identity (<20%) and query coverage (<50%) were excluded. Similarly, homologs with high sequence identities (>40%) were considered as the close homologs. Furthermore, the sequences of the close homologs were compared by using the multiple sequence alignment methods such as PRALINE [26]. The ClustalW method [27] was also utilized for inferring the information obtained from the alignment of the multiple sequences. On the basis of these alignments, the phylogenetic relationships were established by using the PHYLIP software package (http://evolution.genetics.washington.edu/phylip .html). The functional domains were annotated by using a variety of curated databases such as Pfam [28], SUPERFAMILY [29], PANTHER [30], SVMProt [31], CDART [32], SMART [33], InterPro [34], and ProtoNet [35]. Similarly, the conserved motifs in the sequences of KF274549 and KF274557 proteins were identified by using the MEME suite [36]. Moreover, the interaction partners in the metabolic network were predicted by using the STRING [37] database.

*2.3. Structural Analyses.* The understanding of the protein functionalities was further explored by analyzing their structural elements. The structures of the proteins were predicted by using the MODELLER [38] module of DS by satisfying the spatial restraints. The accuracy of the predicted models was evaluated using the Ramachandran plot. The topology of
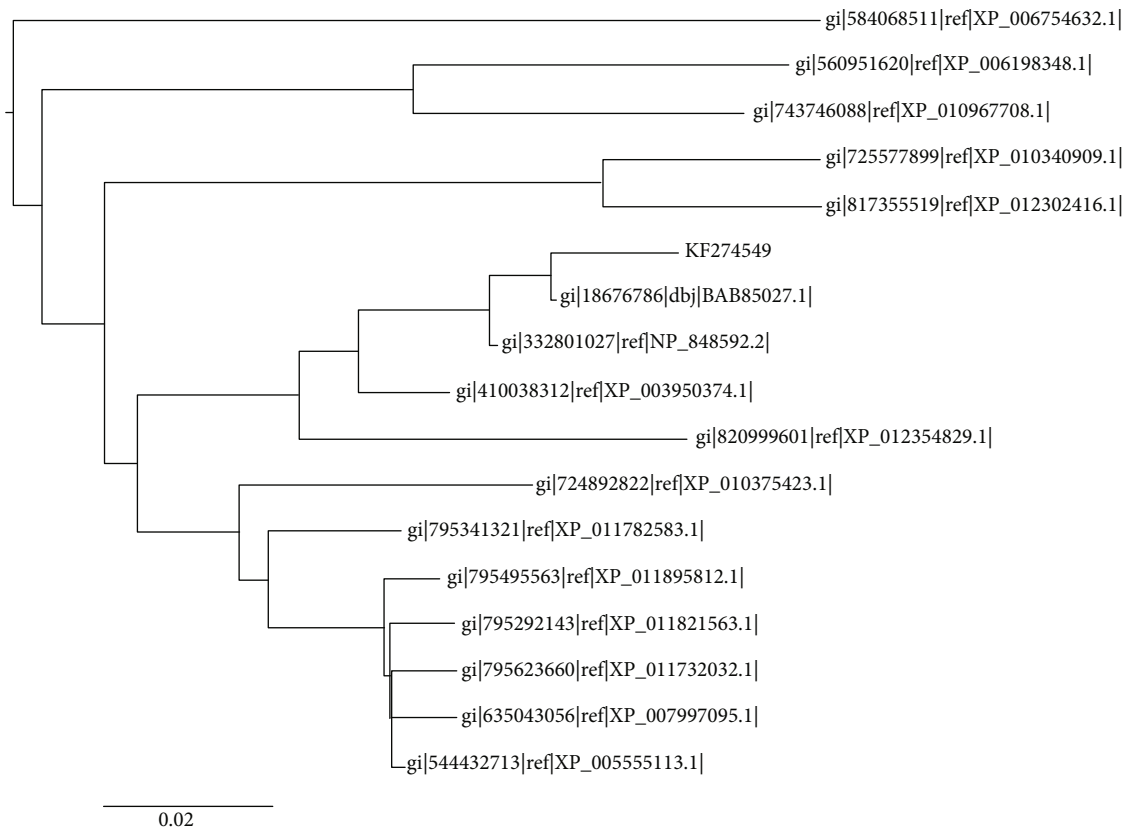
FIGURE 1: The phylogenetic analyses of T1 showing closeness to C4orf26 proteins.



(a)                                                                                          (b)
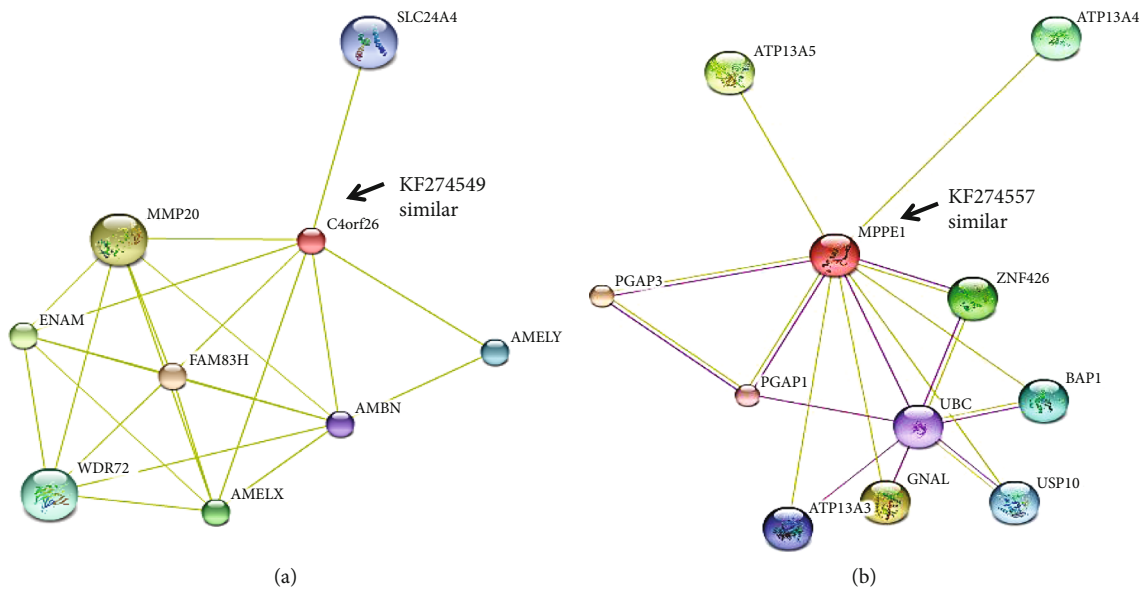
FIGURE 2: Predicted interaction partners of T1 and T2 in the metabolic network according to STRING database.

the generated models was analyzed by using the PDBsum [39], and the structural homologs were searched by using the Dali server [40]. Furthermore, the conformational behavior of the predicted protein was understood by performing the Molecular Dynamics (MD) simulations using the GRO-

MACS 4.6.5 [41] software package. The proteins were solvated by using the SPC/E water model [42], and through the steepest descent algorithm, the energy minimization was performed with a convergence criterion of 0.005 kcal - $mol^{-1}$. The equilibration phase was carried out under NVT
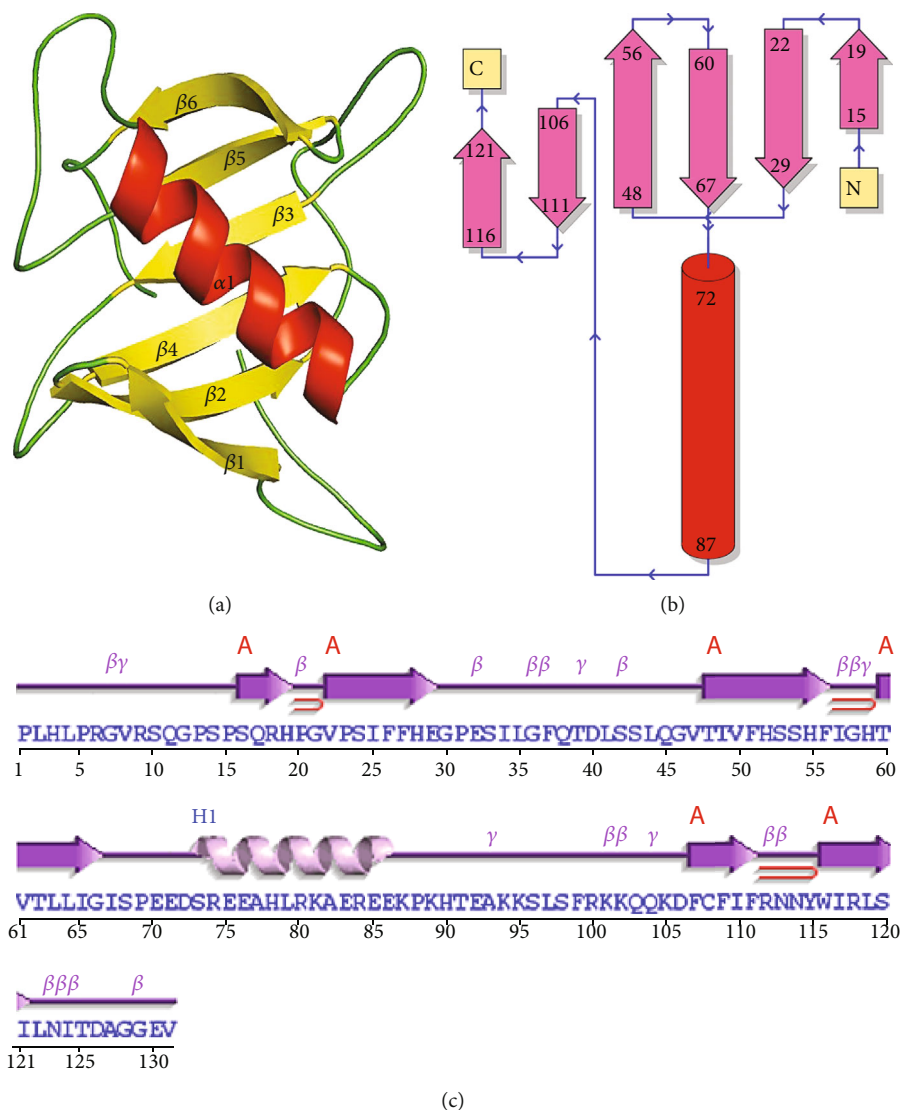
(a)

(b)

(c)

Figure 3: (a) The predicted model of T1 protein. (b, c) The topology generated by the PDBsum server on the basis of the modeled T1 protein.

(constant volume) and NPT (constant pressure) ensemble conditions. The temperature of the system was maintained at 300 K by using the Berendsen weak coupling method in both ensemble conditions along with pressure which was maintained at 1 bar by utilizing the Parrinello-Rahman barostat in constant pressure ensemble. The final MD was produced by using the LINCS algorithm for the time scale of 200 ns. The knowledge extracted from the trajectory files was utilized for understanding the behavior of protein in the explicit solvent environment. The Root Mean Square Deviations (RMSD), radius of gyration (Rg), and Root Mean Square Fluctuations (RMSF) were analyzed.

## 3. Result and Discussion

The minisatellites are found to be involved in the regulation of gene expression, the unstable regions of the chromosomes, and the genome imprinting [7–9]. Yet, expression profiles and biological implication of their correlation with the cod-

ing regions still remain principally unresolved. In this study, we have performed an *in silico* analyses that enable us to understand the characteristic of translated protein obtained from the KF274549 and KF274557. The purpose of choosing these transcripts is explained in a previously published work [7–9]. The detailed analysis of each protein is discussed here separately.

*3.1. KF274549 (T1).* T1 showed high sequence similarities to uncharacterized C4orf26 proteins (Figure S1), with the highest similarity found with unnamed protein GI:18676786 (Figure 1), whereas the outcomes of Pfam and InterPro showed that the T1 belongs to the DUF4721, a family of protein with a domain of unknown functions. Similarly, PANTHER classifies the T1 into the family of uncharacterized proteins PTHR40376. The SMART server identifies that the T1 may belong to the superfamily of ubiquitin-like proteins. The proteins belonging to the ubiquitin-proteasome (ubiquitin-proteasome) system have an established role in the
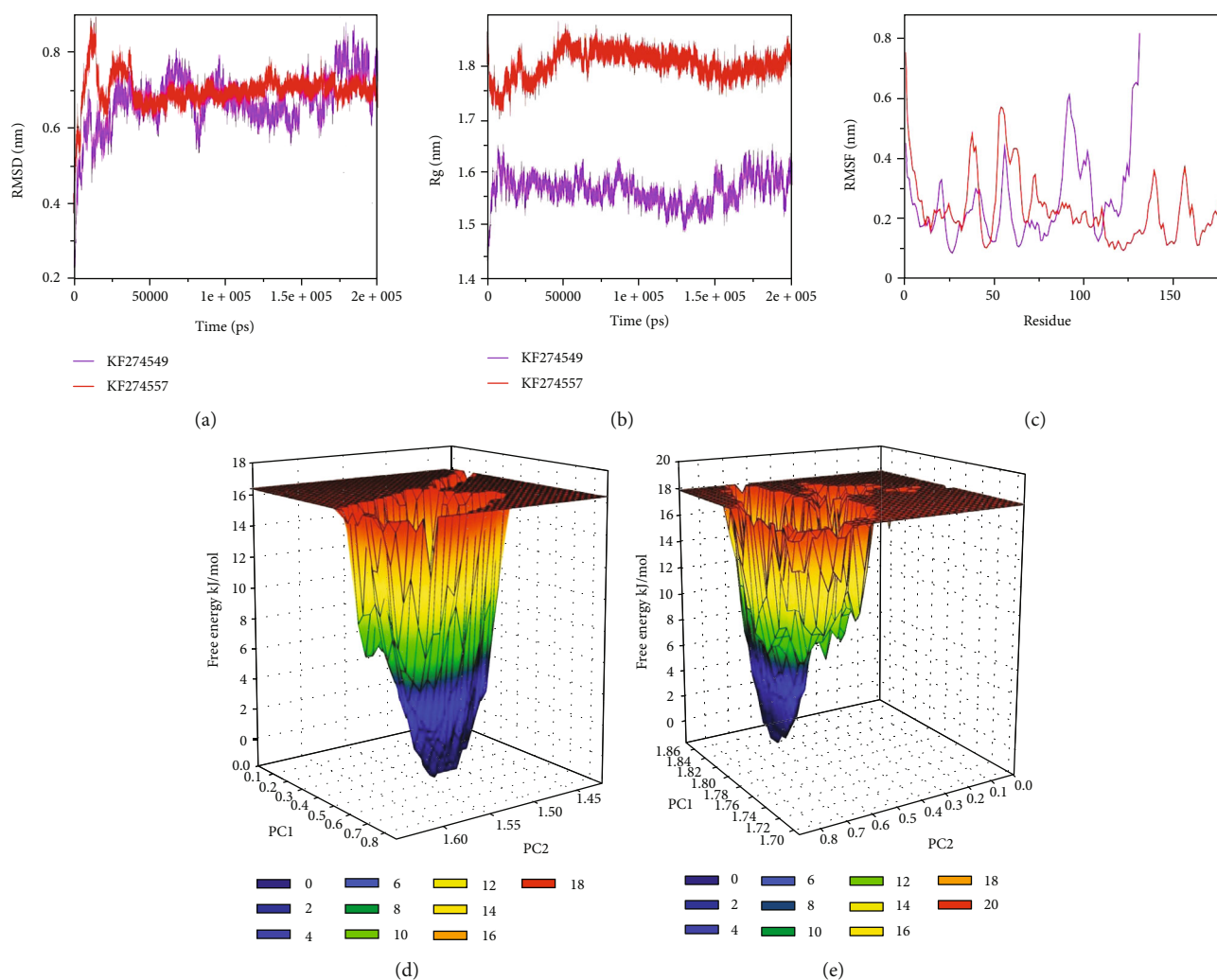
(a)

(b)

(c)



(d)

(e)

FIGURE 4: The outcomes obtained from the MD simulations for 200 ns time scale for both T1 and T2. (a) The plot showing the changes in the RMSD values of both the protein during the course of MD simulations. (b) The changes in the radius of gyrations (Rg) showing the variations observed in the compactness of the proteins. (c) The plot showing the fluctuations observed in the constituent residues. (d) The generated free energy landscape of T1. (e) The projected free energy landscape of T2 showing relative structure stability.

process of protein degradation in the human cell [43]. During spermatogenesis, the ubiquitination enzymes play a major role in the formation of the normal sperm by replacing histone with protamine [43]. Lately, the histone ubiquitin ligases were discovered to play crucial roles in several stages of spermatogenesis, such as DNA damage response, meiotic sex chromosome inactivation, and spermiogenesis [43]. Furthermore, the ProtoNet classifies the T1 into cluster 3965083, which is the collection of mammalian secreted proteins. The STRING database predicted that the T1 interact with protein member H, matrix metallopeptidase 20, enamelin, amelogenin, WD repeat domain 72, amelogenin, solute carrier family 24 (sodium/potassium/calcium exchanger), and ameloblastin (Figure 2(a)). In addition, we were able to observe three motifs in the sequence of KF274549 by using the MEME suite, namely, 47′-CNHRFPFQ, 119′-IKYPKHHLGRW, and 82′–SEGRET.

The structure of T1 was predicted by using the crystal structure of the hypothetical protein ORF126 (PDB ID-2X5R), which is involved in the molecular process of metal ion binding. The predicted structure contains one alpha and six beta strand secondary structural elements (Figure 3(a)). The stereochemical analyses on the basis of Ramachandran plots showed that 97.7% of the residues were present in the allowed region of the plot, which is indicative of the reliability of the predicted models. Similarly, PDBsum analyzed the presence of 1 sheet, 3 beta hairpins, 1 psi loop, 3 beta bulges, 6 strands, 1 helix, 16 beta turns, and 5 gamma turns in the topological description of the modeled structure (Figures 3(b) and 3(c)). The Dali server identified the structural homologs with functionalities as carbamoyl-phosphate synthetase large chain. Therefore, T1 may be involved in the pyrimidine biosynthesis in mammalian spermatozoa [44]. The topology of the predicted model was generated by using the OPLS all-atom force field. Then, it was solvated by using the SPCE water model and minimized for 1000 steps of steepest descent. The NVT and NPT equilibration each for 100 ps time scale was carried out. Finally, the protein was simulated for a 200 ns time scale. The trajectories of the simulations were analyzed by using GROMACS
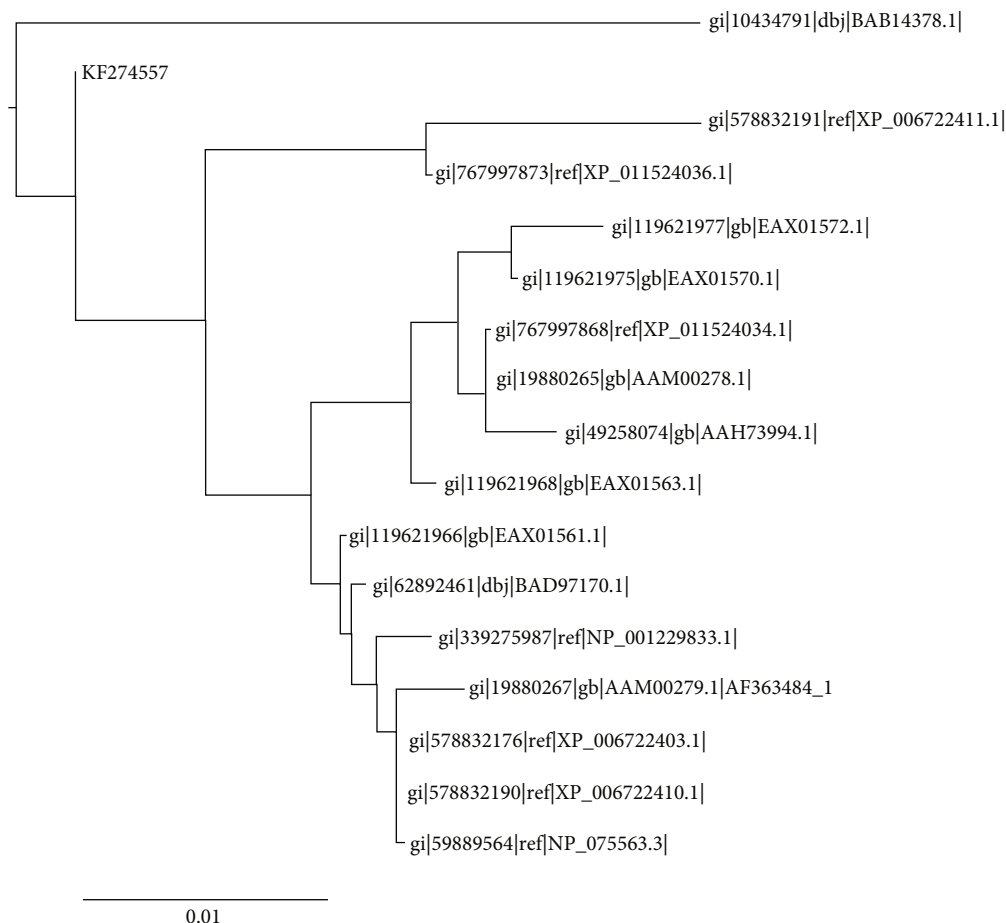
FIGURE 5: The outcomes of molecular phylogenetic analyses for T2 which showed higher homology towards metallophosphoesterases.

utilities. The RMSD values which provide the assessment regarding the system stability showed a continuous increase till 25 ns and become stable after 40 ns with fluctuations which were observed between 0.6 nm and 0.8 nm (Figure 4(a)). Similarly, the radius of gyration (Rg) is the measure of the compactness of the protein structures that were observed to be around 1.6 nm (Figure 4(b)). Furthermore, constituent residues showed relatively higher fluctuations around the residue count of 75-100 which correspond to the alpha-helix region of the predicted structure (Figure 4(c)). The free energy landscapes were projected and showed the presence of relatively unstable conformational behavior in T1 (Figure 4(d)).

### 3.2. KF274557 (T2).

The sequence similarities for T2 showed high closeness to metallophosphoesterase (Figure S2) as well as showed similarities with a diversity of proteins (Figure 5). The outputs generated by CDART, SMART, and ProtoNet also validated the presence of metallophosphatase activities in T2. The mammalian phosphodiesterases in the spermatozoa were found to be significant for the capacitation and fertilization [45]. This category of the protein is also involved in the modulation of cyclic nucleotide cellular levels by catalyzing their degradation [46]. The cyclic nucleotides are involved in the regulation of sperm motility as well as

acrosome reaction. The predicted interaction partners of T2 are post-GPI attachment to protein 3, guanine nucleotide-binding protein (G protein), zinc finger protein 426, ubiquitin-specific peptidase 10, ATPase-type 13A3, ubiquitin C, and post-GPI attachment to protein 1 (Figure 2(b)). Moreover, the MEME suite identified three motifs, namely, $18'$-MNSDFGEQ, $119'$-IYYLAIFQCNW, and $137'$-SDGEQT.

The three dimensional (3-D) structure of T2 was predicted by using the restriction endonuclease MspI (PDB ID-1SA3). The predicted structure of T2 contains the characteristic six alpha and four beta strands (Figure 6(a)). The modeled structure showed 98.3% of the constituent residues in the allowed region of the Ramachandran plot. In addition, the PDBsum identified 1 sheet, 1 beta alpha beta unit, 2 beta hairpins, 2 beta bulges, 4 strands, 7 helices, 5 helix-helix interacts, 19 beta turns, and 2 gamma turns in the predicted structure of T2 (Figures 6(b) and 6(c)). The structural homologs identified by the Dali server showed functionalities of type 2 restriction endonucleases and can be considered as significant for the sperm acrosome reaction [47]. Furthermore, the solvated model of T2 was minimized for 1200 steps of the steepest descent algorithm. After equilibration, the final MD simulations were executed for a 200 ns time scale. The dynamical analyses showed that RMSD values showed a continuous perturbation up to 40 ns, while after that, the values become
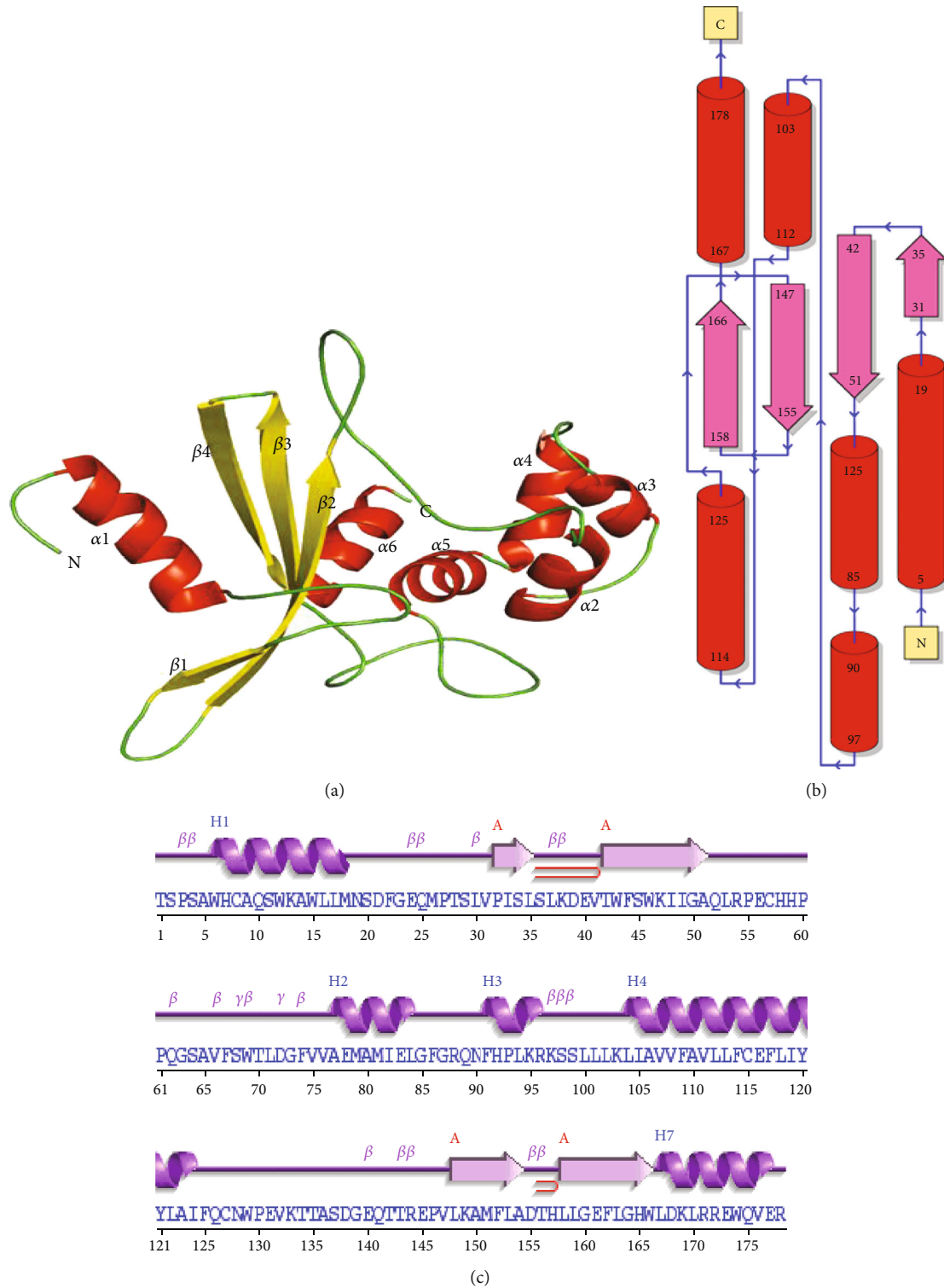
(a)

(b)



(c)

FIGURE 6: (a) The obtained 3-D structure of T2 protein by using homology modeling. (b, c) The PDBsum-based observed topology of the T2.

constant at 0.7 nm (Figure 4(a)). Moreover, Rg values showed continuous fluctuations around 1.75–1.85 nm (Figure 4(b)), while constituent residues showed higher RMSF values in the residue region of 25-75 corresponding to the beta-sheet of the predicted structure (Figure 4(c)). Furthermore, the free energy landscapes showed a comparatively stable structural conformation of T2 (Figure 4(e)).

## 4. Conclusions

The current study identified the possible functions and conformational behavior associated with the proteins translated from *Homo sapiens*' minisatellite 33.15-tagged transcript pAKT-45 variants. The T1 was not showing any functional identities at the primary structure level as high similarities

were observed with hypothetical proteins. Therefore, structure-based methods were utilized which predicted the T1 homology to carbamoyl-phosphate synthetase large chain, while the sequence and structure-based function of T2 was predicted with high confidence. Furthermore, the T2 showed a higher stability profile as compared to the T1 during the course of 200 ns MD simulations. This is attributed to the difference in the structural compactness of both the proteins. This study provides a better understanding of the behavior exhibited by the minisatellite regions at the protein level and their possible functionalities in the metabolic networks.

## Data Availability

The data used to support the findings of this study are included in the article.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgments

## Supplementary Materials

Figure S1: the multiple sequence alignment showing high similarities of T1 with C4orf26 proteins. Figure S2: the multiple sequence alignment showing high similarities of T2 with metallophosphoesterase proteins. (Supplementary Materials)

## References

[1] S. A. Krawetz, "Paternal contribution: new insights and future challenges," *Nature Reviews Genetics*, vol. 6, no. 8, pp. 633–642, 2005.

[2] D. Miller, G. C. Ostermeier, and S. A. Krawetz, "The controversy, potential and roles of spermatozoal RNA," *Trends in Molecular Medicine*, vol. 11, no. 4, pp. 156–163, 2005.

[3] C. M. Saunders, M. G. Larman, J. Parrington et al., "PLC zeta: a sperm-specific trigger of Ca(2+) oscillations in eggs and embryo development," *Development*, vol. 129, no. 15, pp. 3533–3544, 2002.

[4] S. Lambard, I. Galeraud-Denis, G. Martin, R. Levy, A. Chocat, and S. Carreau, "Analysis and significance of mRNA in human ejaculated sperm from normozoospermic donors: relationship to sperm motility and capacitation," *Molecular Human Reproduction*, vol. 10, no. 7, pp. 535–541, 2004.

[5] D. Miller, "Analysis and significance of messenger RNA in human ejaculated spermatozoa," *Molecular Reproduction and Development*, vol. 56, 2 Suppl, pp. 259–264, 2000.

[6] S. M. Wykes, D. W. Visscher, and S. A. Krawetz, "Haploid transcripts persist in mature human spermatozoa," *Molecular Human Reproduction*, vol. 3, no. 1, pp. 15–19, 1997.

[7] J. Srivastava, S. Premi, D. Pathak et al., "Transcriptional status of known and novel genes tagged with consensus of 33.15 repeat loci employing minisatellite-associated sequence amplification (MASA) and real-time PCR in water buffalo, Bubalus bubalis," *DNA and Cell Biology*, vol. 25, no. 1, pp. 31–48, 2006.

[8] G. Toth, Z. Gaspari, and J. Jurka, "Microsatellites in different eukaryotic genomes: survey and analysis," *Genome Research*, vol. 10, no. 7, pp. 967–981, 2000.

[9] G. Vergnaud and F. Denoeud, "Minisatellites: mutability and genome architecture," *Genome Research*, vol. 10, no. 7, pp. 899–907, 2000.

[10] B. Charlesworth, P. Sniegowski, and W. Stephan, "The evolutionary dynamics of repetitive DNA in eukaryotes," *Nature*, vol. 371, no. 6494, pp. 215–220, 1994.

[11] A. J. Jeffreys, N. J. Royle, V. Wilson, and Z. Wong, "Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA," *Nature*, vol. 332, no. 6161, pp. 278–281, 1988.

[12] P. Bennett, "Demystified ... microsatellites," *Molecular Pathology*, vol. 53, no. 4, pp. 177–183, 2000.

[13] B. Borstnik and D. Pumpernik, "Tandem repeats in protein coding regions of primate genes," *Genome Research*, vol. 12, no. 6, pp. 909–915, 2002.

[14] A. Jasinska and W. J. Krzyzosiak, "Repetitive sequences that shape the human transcriptome," *FEBS Letters*, vol. 567, no. 1, pp. 136–141, 2004.

[15] Y. C. Li, A. B. Korol, T. Fahima, and E. Nevo, "Microsatellites within genes: structure, function, and evolution," *Molecular Biology and Evolution*, vol. 21, no. 6, pp. 991–1007, 2004.

[16] E. P. Rocha, I. Matic, and F. Taddei, "Over-representation of repeats in stress response genes: a strategy to increase versatility under stressful conditions?," *Nucleic Acids Research*, vol. 30, no. 9, pp. 1886–1894, 2002.

[17] M. Shahbaaz, F. Ahmad, and M. I. Hassan, "Structure-based function analysis of putative conserved proteins with isomerase activity from Haemophilus influenzae," *3 Biotech*, vol. 5, no. 5, pp. 741–763, 2015.

[18] M. Shahbaaz, K. Bisetty, F. Ahmad, and M. I. Hassan, "Current advances in the identification and characterization of putative drug and vaccine targets in the bacterial genomes," *Current Topics in Medicinal Chemistry*, vol. 16, no. 9, pp. 1040–1069, 2015.

[19] M. Shahbaaz, K. Bisetty, F. Ahmad, and M. I. Hassan, "Functional insight into putative conserved proteins of Rickettsia rickettsii and their virulence characterization," *Current Proteomics*, vol. 12, no. 2, pp. 101–116, 2015.

[20] M. Shahbaaz, K. Bisetty, F. Ahmad, and M. I. Hassan, "Towards new drug targets? Function prediction of putative proteins of Neisseria meningitidis MC58 and their virulence characterization," *OMICS: A Journal of Integrative Biology*, vol. 19, no. 7, pp. 416–434, 2015.

[21] M. Shahbaaz, M. I. Hassan, and F. Ahmad, "Functional annotation of conserved hypothetical proteins from Haemophilus influenzae Rd KW20," *PLoS One*, vol. 8, no. 12, article e84263, 2013.

[22] D. S. Biovia, *Discovery Studio Modeling Environment*, Dassault Systèmes, San Diego, 2013.

[23] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.

[24] R. D. Finn, J. Clements, and S. R. Eddy, "HMMER web server: interactive sequence similarity searching," *Nucleic Acids Research*, vol. 39, supplement, pp. W29–W37, 2011.

[25] J. Soding, A. Biegert, and A. N. Lupas, "The HHpred interactive server for protein homology detection and structure prediction," *Nucleic Acids Research*, vol. 33, no. Web Server, pp. W244–W248, 2005.

[26] V. A. Simossis and J. Heringa, "PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information," *Nucleic Acids Research*, vol. 33, no. Web Server, pp. W289–W294, 2005.

[27] J. D. Thompson, T. J. Gibson, and D. G. Higgins, "Multiple sequence alignment using ClustalW and ClustalX," *Current Protocols in Bioinformatics*, no. 1, pp. 2.3.1–2.3.22, 2002.

[28] A. Bateman, E. Birney, L. Cerruti et al., "The Pfam protein families database," *Nucleic Acids Research*, vol. 30, no. 1, pp. 276–280, 2002.

[29] J. Gough, K. Karplus, R. Hughey, and C. Chothia, "Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure," *Journal of Molecular Biology*, vol. 313, no. 4, pp. 903–919, 2001.

[30] H. Mi, A. Muruganujan, J. T. Casagrande, and P. D. Thomas, "Large-scale gene function analysis with the PANTHER classification system," *Nature Protocols*, vol. 8, no. 8, pp. 1551–1566, 2013.

[31] C. Z. Cai, L. Y. Han, Z. L. Ji, X. Chen, and Y. Z. Chen, "SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3692–3697, 2003.

[32] L. Y. Geer, M. Domrachev, D. J. Lipman, and S. H. Bryant, "CDART: protein homology by domain architecture," *Genome Research*, vol. 12, no. 10, pp. 1619–1623, 2002.

[33] I. Letunic, T. Doerks, and P. Bork, "SMART 7: recent updates to the protein domain annotation resource," *Nucleic Acids Research*, vol. 40, no. D1, pp. D302–D305, 2012.

[34] S. Hunter, R. Apweiler, T. K. Attwood et al., "InterPro: the integrative protein signature database," *Nucleic Acids Research*, vol. 37, no. Database issue, pp. D211–D215, 2009.

[35] N. Rappoport, S. Karsenty, A. Stern, N. Linial, and M. Linial, "ProtoNet 6.0: organizing 10 million protein sequences in a compact hierarchical family tree," *Nucleic Acids Research*, vol. 40, no. D1, pp. D313–D320, 2012.

[36] T. L. Bailey, M. Boden, F. A. Buske et al., "MEME SUITE: tools for motif discovery and searching," *Nucleic Acids Research*, vol. 37, no. Web Server, pp. W202–W208, 2009.

[37] D. Szklarczyk, A. Franceschini, M. Kuhn et al., "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored," *Nucleic Acids Research*, vol. 39, no. Database issue, pp. D561–D568, 2010.

[38] N. Eswar, B. Webb, M. A. Marti-Renom et al., "Comparative protein structure modeling using Modeller," *Current Protocols in Bioinformatics*, vol. 15, no. 1, pp. 5.6.1–5.6.30, 2006.

[39] R. A. Laskowski, E. G. Hutchinson, A. D. Michie, A. C. Wallace, M. L. Jones, and J. M. Thornton, "PDBsum: a web-based database of summaries and analyses of all PDB structures," *Trends in Biochemical Sciences*, vol. 22, no. 12, pp. 488–490, 1997.

[40] L. Holm and P. Rosenstrom, "Dali server: conservation mapping in 3D," *Nucleic Acids Research*, vol. 38, supplement_2, pp. W545–W549, 2010.

[41] S. Pronk, S. Pall, R. Schulz et al., "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, vol. 29, no. 7, pp. 845–854, 2013.

[42] J. Zielkiewicz, "Structural properties of water: comparison of the SPC, SPCE, TIP4P, and TIP5P models of water," *The Journal of Chemical Physics*, vol. 123, no. 10, p. 104501, 2005.

[43] K. Sheng, X. Liang, S. Huang, and W. Xu, "The role of histone ubiquitination during spermatogenesis," *BioMed Research International*, vol. 2014, Article ID 870695, 8 pages, 2014.

[44] E. A. Carrey, C. Dietz, D. M. Glubb, M. Loffler, J. M. Lucocq, and P. F. Watson, "Detection and location of the enzymes of de novo pyrimidine biosynthesis in mammalian spermatozoa," *Reproduction*, vol. 123, no. 6, pp. 757–768, 2002.

[45] R. W. Baxendale and L. R. Fraser, "Mammalian sperm phosphodiesterases and their involvement in receptor-mediated cell signaling important for capacitation," *Molecular Reproduction and Development*, vol. 71, no. 4, pp. 495–508, 2005.

[46] M. Bajpai, S. E. Fiedler, Z. Huang et al., "AKAP3 selectively binds PDE4A isoforms in bovine spermatozoa," *Biology of Reproduction*, vol. 74, no. 1, pp. 109–118, 2006.

[47] S. Meizel, "The importance of hydrolytic enzymes to an Exocytotic event, the mammalian sperm acrosome reaction," *Biological Reviews*, vol. 59, no. 1, pp. 125–157, 1984.