DOI: https://doi.org/10.36094/sc.v91.2025.Identification_and_Characterization_of_Putative.Das_Sarkar.293

ARTICLE

Sci. and Cult. 91 (5–6) : 293-304 (2025)

IDENTIFICATION AND CHARACTERIZATION OF PUTATIVE DRUG TARGET IN DRUG RESISTANT TUBERCULOSIS USING SUBTRACTIVE GENOMICS APPROACH

MONISH MUKUL DAS[1] AND KEKA SARKAR[2]

Recently, drug-resistant tuberculosis is of great concern owing to lack of new drugs. Rapid identification of unique drug targets against Mycobacterium tuberculosis DKC2 strain is crucial. In silico subtractive genomics methods identified 230 essential non-homologous proteins. Druggability analysis revealed 27 potential targets, including WP_003899075.1 (Catalase-peroxidase), with high virulence and antibiotic resistance and physico-chemical analysis supported its stability. Structure-based modelling and validation confirmed WP_003899075.1 as a promising drug target which could be useful for new antitubercular compounds.

Introduction

M tuberculosis, the most dreadful scourge that causes morbility and mortality to humankind, has remained uncontrolled though a lot of advances has been achieved on the study of this deadly pathogen since its discovery by Robert Koch in 1882.

In spite of availability of DOT, BCG vaccine and many antitubercular compounds[1,2], upward trend in drug resistance coupled with inadequacy of new drugs causes disease and deaths. Besides, treatment by antitubercular drugs is very often associated with severe side effects and non-adherence of patients. In addition, it is reported that the nature and extent of virulence is not sufficiently understood[3]. Again, the advent of MDR-MTB strain and XDR-TB strains have made the scenario worse. The complexities in texture of bacterial cell wall, efficient drug efflux mechanism, enzymes coding genes and super families of enzymes have contributed unbounded success of this bacteria to spread tuberculosis amongst millions of human beings all over the world[3,4]. However, inactivation of important biomolecules and genes produce significant decrease in virulence of pathogen[5]. Moreover, proteins of cell surface secreted by various types of transporter systems, are involved in antibiotic resistance[6,7]. In recent years, drug resistance is observed as an usual phenomenon in respect of antitubercular drugs and resistance to combination of drugs or to INH alone, is the most popular type of resistance which is visible amongst tuberculosis infected patients[8].

Therefore, it is still barrier to combat INH-resistance in M. tuberculosis pathogens. The continuous genesis of MDR-strains of M. tuberculosis and want of authentic drug having no cross reactions, have necessitated the need to identify potential non-homologous drug targets perceived as core point in the process of drug discovery.

However, the development of genomics has facilitated its application in the field of research on infectious disease: research on contagious diseases[9,10]. The bioinformatic methods that incorporate systematic use of algorithms and biological databases, are set up based on wet lab-oriented results and reports of available literature indicating its application with great success[9,10]. During last two decades,

[1] Department of Computer Science and Engineering, University of Kalyani, Kalyani, West Bengal, India, e-mail: edwardcullen11295@gmail.com

[2] Department of Microbiology, University of Kalyani, Kalyani, West Bengal, e-mail: keka@klyuniv.ac.in

combination of computational methods with biological data, has defeated the stumbling block with a view to realize the genomics of pathogens if its genetic information is available. The foremost application of subtractive genomics which is one of the such bioinformatics approaches, is regarded as a way to find out drug or vaccine targets in pathogens and the approach is perceived as the key point of drug development.

The basics of this approach is to entail comparing host proteins with that of pathogen with a view to identify essential proteins which are critical for survival of pathogen[9]. These pathogen specific proteins, are considered as putative drug or vaccine targets. This genomic method was applied in various research for identification of vaccine candidates and therapeutic targets against various pathogens[11-15]. In recent years, same approach was employed in various pathogens[16,17] to predict potential drug targets. Moreover, the subcellular localization provides required information with regard to functions and locations of an essential protein. Therefore, identification of location of a particular protein in cell is vital for therapeutic target[18]. In addition, the secreted and membrane localized essential protein has natural tendency to evoke immune response and it could be acted as vaccine target while cytoplasmic essential protein is more favourable as drug target[19,20]. Virulence factors which cause tuberculosis and have key role for survival of M. tuberculosis[21], were considered as therapeutic targets instead of focussing on conventional drugs. It has been reported that emergence of M. tb_DKC2 strain having remarkable virulence, has been increasing alarmingly[22] and the strain has developed resistance to conventional antitubercular drugs which results in tremendous threat as there is big gap between demand and supply for new potential drugs. Therefore, these demands have necessitated the need to explore potential drug targets using computational approaches. So fat, no published record is available on therapeutic targets against this pathogen whole only complete genome sequence is recorded[23].

The present investigation has therefore, been oriented to in silico prediction of most virulent and antibiotic resistant pathogen specific essential protein as therapeutic target which could be of further help to discover efficient drug compound against strain of M. tb_DKC2 using various computational approaches. The proposed computational approaches are designed based on scientific reports and experimental results denoting its applicability with great success[9,24]. In this study, amino acid sequence was retrieved and CodonW, was executed to calculate genomic features of amino acid usages. Finally critical features obtained using GA, were employed as inputs of trained DNN[25] to identify essential genes. Systems of CD-HIT, BLASTp, DEG, NCBI, DRUGBANK, VFDB, CARD and Prot Param, were used employing a technique to identify, characterize and analyse the essential proteins. Homology models of proposed essential protein were built employing SWISS-MODEL server, refined and validated using PROCHECK, PSIPRED, ERRAT and PROSA computational tools. STRING database was used to evaluate functional as well as physical interactions of this predicted drug target with other partner proteins. The technique includes data extraction, subtractive genomic approach, structure-based study, Druggability of proteins and its localization, and physico-chemical properties of drug target protein. This investigation is an attempt to uncover the importance of essential proteins in virulence and to identify the potent therapeutic targets.

Material and Methods: In developing effective drug against M. tb_DKC2 strain, various advanced computational tools and subtractive genomic methods were used to identify potential drug target.

Phase I: In phase I, genomic data was obtained by predicting the essential genes of M. tb_DKC2 strain by using trained and newly designed deep neural network based on key genomic sequence features of 20 strains of M. tuberculosis obtained through application of CodonW

Table.1: Various steps involved in subtractive genomics analysis with corresponding shortlisted protein sequence number from Mycobacterium tuberculosis DKC2 strain.

SL. No.	Description	No. of Protein sequence
1.	*Predicted Proteins*	3456
2.	*Predicted Essential Proteins*	1192
3.	*Removal of Paralogs with CD-HIT (60% Threshold)*	1100
4.	*BLASTp of non-paralogous proteomes against human proteome (e value 10^{-5})*	596
5.	*BLASTp of non-homologous essential proteins against DEG (e value 10^{-100})*	230
6.	*BLASTp of non-homologous essential proteins against DRUGBANK (e value 10^{-5})*	27
7.	*BLASTp of non-homologous essential proteins against VFDB (e value 10^{-5})*	6
8.	*BLASTp of non-homologous essential proteins against CARD (e value 10^{-5})*	2

and genetic algorithm[26] and proteomic sequence data corresponding to essential genes were retrieved from NCBI GenBank. Simultaneously, human proteome was also obtained from databank of NCBI.

Phase II: In the second phase of study, an effective subtractive genomics protocol was executed for essentiality, specificity, selectivity and eradication of paralogs. This method was used to select essential non-homologous protein (Table 1). The proteomes of pathogen was subjected at 60% threshold by using CD-HIT server[27,28] for down-stream analysis and removal of paralogous sequences[29]. After excluding paralogs, proteomes were treated with BLASTp against human proteome keeping e-value cutoff of 10^{-5} and bit score of 100[30] and the output data was non-homologous essential proteins. The shortlisted essential non-homologous proteins was treated with BLASTp against DEG for further validation[31,32] of proteome essentiality keeping e-value cutoff score of 1e-100 and bit score cutoff of 100[29]. Any target is considered as druggable only when it possesses strength of binding interaction with drug compounds. The target proteins were treated with BLASTp against DrugBank v 5.0 with e-value of 1e-5[33] for alignment study and prediction of potent target[29] which was selected for next stage of subtractive genomics method.

Accordingly, target proteins were treated with BLASTp against PSORTb v 3.0.2[34] in order to obtain information about target proteins and its predicted locations in cells. All proteins were subjected to next phase of analysis. The probable drug targets were treated with BLASTp against VFDB[35] for obtaining target proteins with higher virulence properties. The virulent proteins were subjected to further phase of analysis. CARD v 3.2.5 was employed for identification of shortlisted antibiotic resistant[36] proteins by using E-value cutoff of 10^{-5}. Eventually, subtractive genomics analysis[37], produced a protein (WP_003899075.1) as drug target having maximum virulence scores and antibiotic resistance against the pathogen i.e. M. tb_DKC2 (NCBI Acc. No. ID = NZ_LR027516.1).

Phase III: In the third phase, the shortlisted protein sequence (VFG001396) was taken into consideration of subsequent structure-based studies i.e. protein structure modelling, modelled structure validation, analysis of physico-chemical properties and protein-protein

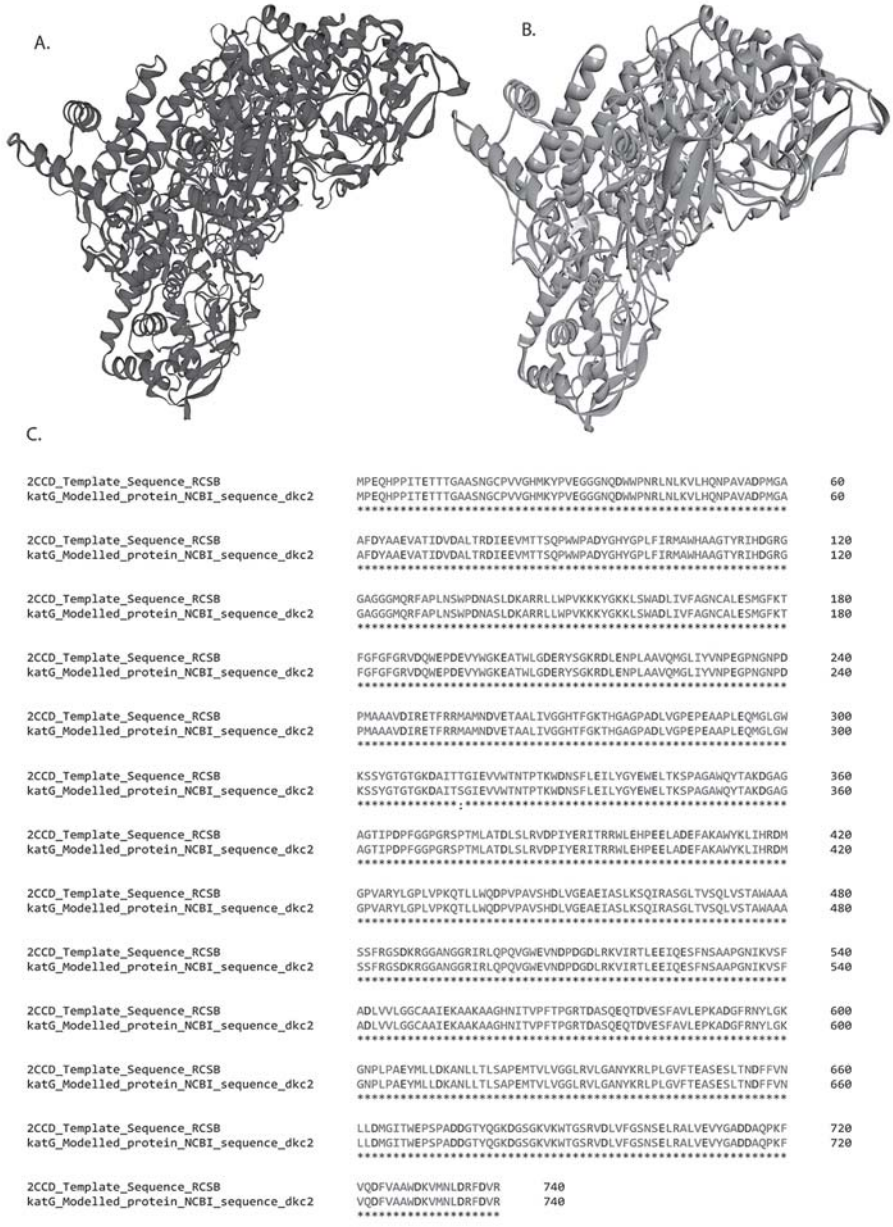

Fig.1: (A) Modelled protein structure using 2CCD as template. (B) Superimposing of template protein (green) with modelled protein (Catalase-peroxidase) in yellow colour. (C) Protein sequence alignment of Catalase-peroxidase & template protein (2CCD).

interactions. The drug target protein (WP_003899075.1) was screened by using BLAST[38] and HHblits[39] against the SWISS-MODEL protein template library[40] linked with RCSB-PDB for suitable templates to predict 3D protein structure. 2CCD template having higher QSQE value (0.80) was considered for final model of target protein[40]. The modelled protein was validated employing PROCHECK, PSIPRED, PROSA and ERRAT computational tools. PROCHECK[41] was used to generate a Ramachandran plot with a view to determine stereochemical properties of modelled protein by analysing its overall structural geometry along with residue-by-residue geometry. The PSIPRED v 4.0[42] was used to estimate secondary structures of target protein viz. 'helix', 'strand' and 'coil'[43]. The ERRAT v 2.0 program was executed to confirm the quality i.e. percentage (%) of protein present in 3D protein structure and identify erroneous regions which are constructed incorrectly in modelled protein[44] Finally, the modelled protein was subjected to ProSA-web server[45] to determine the quality of modelled protein against experimentally derived protein structure form PDB database on the basis of Z-score value.

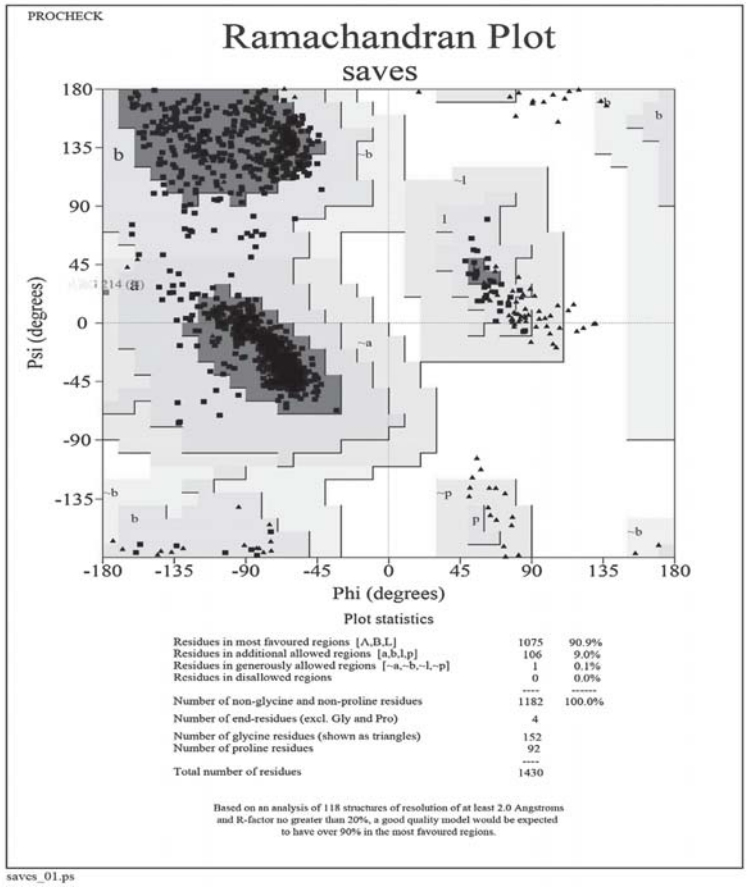

Fig.2: Evaluation of predicted modelled protein (WP_003899075.1) by using a Ramachandran plot produced by PROCHECK.

Prot Param[46] was employed to calculate various physical and chemical properties and percentage (%) composition of each amino acid of drug target. The properties like molecular weight, iso-electric point, composition of amino acid, composition of atoms, instability index value, aliphatic index value and GRAVY. The PPI of drug target was generated employing STRING database in 2021[47] to elucidate evolutionary, physical and functional information on target protein and its least possible interactions with other surrounding partners. PPI analysis supported that drug target is a 'hub protein' with 858 edges, 95 nodes and 18.1 average nodes degree and as such aiming this protein might lead to disruption of functional activities of interacting partner proteins.

Results and Discussion: In the present study, essential genes of M. tb_DKC2 strain were predicted using newly designed and trained DNN[25] and thereafter, proteomic data corresponding to essential genes were also obtained from NCBI GenBank. Eventually, complete proteomic data of Homo sapiens was also retrieved from NCBI databank. The subtractive genomics approaches were explored to predict and identify potent drug targets against M. tb_DKC2. Finally, a cytoplasmic protein i.e. Catalase-peroxidase (KatG) having no cross-reactions with the host, was observed as antibiotic resistant drug target with highest virulence properties (Table 1, Table S1-S4). This KatG protein having catalytic and peroxidic activities, could be of use as effective drug target.

Finally, structure based method was explored to build 3D model of KatG protein (Fig.1) using '2CCD' template having higher QSQE score value of 0.80[40] and higher GMQE score of 0.96. Model protein structure attained QMEAN global score of 0.87 ± 0.005 (Table S5 and S6) that stood between '0' and '1'[48]. The quality of modelled protein was verified employing computational tools like PSIPRED, PROSA, PROCHECK, ERRAT. Input modelled protein structure displayed a validation result of 94.468% around 95% as an average overall quality factor for ERRAT, z-score of -11.77 for PROSA (S3 Fig.) and high confidence score ranging from '0' to '9' for PSIPRED (S1 Fig.). The obtained results i.e. 90% residues in the most favoured regions of Ramachandran plots supported the quality and reliability of 3D protein model (Fig 2)[49]. ERRAT findings (S2

Table S1: Essential Non-homologous 27 Identified Proteins as Drug Targets with its Proposed Drug Bank ID(s).

No.	NCBI protein ID	Protein Name	Protein Function	Drug bank target Id
1.	*WP_003402118.1*	Glycine oxidase	flavin adenine dinucleotide binding / glycine oxidase activity	*DB02713, DB03147*
2.	*WP_003402354.1*	UDP-N-acetylenolpyruvoylglucosamine reductase	flavin adenine dinucleotide binding / UDP-N-acetylmuramate dehydrogenase activity	*DB03147, DB07296*
3.	*WP_003900222.1*	Indole-3-pyruvate decarboxylase	indolepyruvate decarboxylase activity / magnesium ion binding / thiamine pyrophosphate binding	*DB01987*
4.	*WP_003900287.1*	Bifunctional protein PutA	1-pyrroline-5-carboxylate dehydrogenase activity / aldehyde dehydrogenase (NAD) activity etc.	*DB03051, DB03147, DB04398*
5.	*WP_003406845.1*	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	UDP-N-acetylglucosamine 1-carboxyvinyltransferase activity	*DB00828*
6.	*WP_003407318.1*	Riboflavin synthase	oxidoreductase activity / riboflavin synthase activity	*DB00140*
7.	*WP_003408024.1*	Cytochrome D Ubiquinol oxidase subunit II	oxidoreductase activity	*DB05154*
8.	*WP_003408027.1*	Cytochrome D ubiquinol oxidase subunit I	electron transfer activity / metal ion binding	*DB05154*
9.	*WP_003899049.1*	Urease subunit alpha	nickel cation binding / urease activity	*DB00551, DB05265*
10.	*WP_003899075.1*	Catalase-peroxidase	catalase activity / heme binding / metal ion binding / NADH binding etc.	*DB00609, DB00951*
11.	*WP_003899116.1*	Ribonucleoside-diphosphate reductase subunit beta	metal ion binding / ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor	*DB04272*
12.	*WP_003411214.1*	Cell division protein	penicillin binding / peptidoglycan glycosyltransferase activity	*DB01413, DB09050, DB06211, DB14879*
13.	*WP_003411387.1*	Anthranilate phosphoribosyltransferase	anthranilate phosphoribosyltransferase activity / magnesium ion binding	*DB01632*
14.	*WP_256382343.1*	Nicotinate-nucleotide adenylyltransferase	ATP binding / nicotinate-nucleotide adenylyltransferase activity	*DB04272*
15.	*WP_003412624.1*	Probable pyruvate-flavodoxin oxidoreductase	4 iron, 4 sulfur cluster binding / iron ion binding / pyruvate-flavodoxin oxidoreductase activity / thiamine pyrophosphate binding	*DB00698*
16.	*WP_003413027.1*	Chorismate synthase	chorismate synthase activity	*DB03247, DB03350*
17.	*WP_003413468.1*	Pyridoxal 5'-phosphate synthase subunit Pdx1	catalytic activity / glutaminase activity / identical protein binding etc.	*DB11638*
18.	*WP_003899465.1*	Thymidylate synthase ThyX	flavin adenine dinucleotide binding / thymidylate synthase (FAD) activity	*DB01903, DB03147, DB03761, DB03800*
19.	*WP_003415934.1*	Streptogramin A acetyltransferase	transferase activity, transferring acyl groups	*DB01669, DB01992, DB01764*
20.	*WP_003416922.1*	Flavohemoprotein	FAD binding / fatty acid binding / heme binding etc.	*DB03147*
21.	*WP_023637501.1*	Alanine racemase	alanine racemase activity / pyridoxal phosphate binding	*DB00260*
22.	*WP_003418351.1*	DNA-directed RNA polymerase subunit alpha	DNA binding / DNA-directed RNA polymerase activity / zinc ion binding	*DB00615, DB11753*
23.	*WP_003419253.1*	Fumarate reductase flavoprotein subunit	metal ion binding / succinate dehydrogenase activity	*DB03147*
24.	*WP_003899596.1*	Dihydropteroate synthase	dihydropteroate synthase activity / metal ion binding	*DB00250*
25.	*WP_003899610.1*	D-alanyl-D-alanine carboxypeptidase DacB	carboxypeptidase activity / endopeptidase activity / penicillin binding / serine-type D-Ala-D-Ala carboxypeptidase activity	*DB00760, DB01329, DB01328, DB00303, DB04570, DB00417, DB14879, DB00578, DB09319*
26.	*WP_003420798.1*	UDP-galactopyranose mutase	UDP-galactopyranose mutase activity	*DB03147, DB03709*
27.	*WP_003899733.1*	Ferredoxin-dependent glutamate synthase 2	3 iron, 4 sulfur cluster binding / glutamate synthase (ferredoxin) activity / glutamate synthase activity / metal ion binding	*DB03247*

Table S2: Sub-cellular Localization of Essential Non-homologous Drug Target Proteins with its NCBI Protein IDs in Various Area of Cell.

No.	NCBI Protein IDs	NCBI Locus Tags	Protein's Name	PSORTb results
1.	*WP_003402118.1*	DKC2_0445	Glycine oxidase	*Cytoplasmic*
2.	*WP_003402354.1*	DKC2_0514	UDP-N-acetylenolpyruvoylglucosamine reductase	*Cytoplasmic*
3.	*WP_003900222.1*	DKC2_0912	Indole-3-pyruvate decarboxylase	*Cytoplasmic Membrane*
4.	*WP_003900287.1*	DKC2_1269	Bifunctional protein PutA	*Cytoplasmic*
5.	*WP_003406845.1*	DKC2_1400	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	*Cytoplasmic*
6.	*WP_003407318.1*	DKC2_1503	Riboflavin synthase	*Cytoplasmic*
7.	*WP_003408024.1*	DKC2_1707	Cytochrome D Ubiquinol oxidase subunit II	*Cytoplasmic Membrane*
8.	*WP_003408027.1*	DKC2_1708	Cytochrome D ubiquinol oxidase subunit I	*Cytoplasmic Membrane*
9.	*WP_003899049.1*	DKC2_1956	Urease subunit alpha	*Cytoplasmic*
10.	*WP_003899075.1*	DKC2_2019	Catalase-peroxidase	*Cytoplasmic*
11.	*WP_003899116.1*	DKC2_2095	Ribonucleoside-diphosphate reductase subunit beta	*Unknown*
12.	*WP_003411214.1*	DKC2_2287	Cell division protein	*Cytoplasmic Membrane*
13.	*WP_003411387.1*	DKC2_2316	Anthranilate phosphoribosyltransferase	*Cytoplasmic Membrane*
14.	*WP_256382343.1*	DKC2_2556	Nicotinate-nucleotide adenylyltransferase	*Cytoplasmic*
15.	*WP_003412624.1*	DKC2_2588	Probable pyruvate-flavodoxin oxidoreductase	*Cytoplasmic*
16.	*WP_003413027.1*	DKC2_2676	Chorismate synthase	*Cytoplasmic*
17.	*WP_003413468.1*	DKC2_2744	Pyridoxal 5'-phosphate synthase subunit Pdx1	*Cytoplasmic*
18.	*WP_003899465.1*	DKC2_2892	Thymidylate synthase ThyX	*Cytoplasmic*
19.	*WP_003415934.1*	DKC2_3224	Streptogramin A acetyltransferase	*Cytoplasmic*
20.	*WP_003416922.1*	DKC2_3434	Flavohemoprotein	*Cytoplasmic*
21.	*WP_023637501.1*	DKC2_3645	Alanine racemase	*Cytoplasmic*
22.	*WP_003418351.1*	DKC2_3687	DNA-directed RNA polymerase subunit alpha	*Cytoplasmic*
23.	*WP_003419253.1*	DKC2_3767	Fumarate reductase flavoprotein subunit	*Cytoplasmic Membrane*
24.	*WP_003899596.1*	DKC2_3838	Dihydropteroate synthase	*Unknown*
25.	*WP_003899610.1*	DKC2_3855	D-alanyl-D-alanine carboxypeptidase DacB	*Unknown*
26.	*WP_003420798.1*	DKC2_4054	UDP-galactopyranose mutase	*Cytoplasmic*
27.	*WP_003899733.1*	DKC2_4111	Ferredoxin-dependent glutamate synthase 2	*Cytoplasmic Membrane*

Fig.) also justified input modelled protein to be a good protein with high resolution[26]. The outcome of PROSA i.e. negative z-score value (-11.77) which was very near to that of template protein (-11.87) again justified the reliability and quality tertiary structure of 3D modelled protein (S4 Fig.)[50]. A high confidence score ranging from '0' to '9' for PSIPRED further confirmed 2D secondary structure of modelled protein[51]. It is revealed from Table 2 that the isoelectric point (5.09 < 7), instability index value (3.90 < 40) and negative GRAVY index value (-0.325) for ProtParam tool further justified that target protein was negatively charged, stable, hydrophilic and soluble in aqueous medium. The high aliphatic index (76.24) of target protein also indicated its rising thermostability. Table S7 also displayed various amino acid composition of KatG protein.

Table 2: Estimation of Physicochemical Properties of Shortlisted Drug Target Protein (WP_003899075.1) having High Virulence Factors (score) by Using Prot Param

SL. No.	Physical and Chemical properties of target proteins	Values
1.	Amino acid residue number	740
2.	Molecular weight (Dalton)	80604.87
3.	Isoelectric point (Value)	5.09
4.	positive charged residues (Number)	71
5.	negative charged residues (Number)	94
6.	Atoms (Number)	11225
7.	Index value of instability	31.90
8.	Aliphatic index of protein	76.24
9.	Index value of GRAVY	-0.325

Table S3: Essential Non-homologous 6 Identified Proteins as Drug Targets with its Proposed Virulence Factor Database (VFDB) IDs.

No.	NCBI Protein ID	NCBI Locus Tag	Protein Name	VFDB ID	Score
1.	*WP_003899049.1*	DKC2_1956	Urease subunit alpha	VFG006476	538
2.	*WP_003899075.1*	DKC2_2019	Catalase-peroxidase	VFG001396	1442
3.	*WP_003411387.1*	DKC2_2316	Anthranilate phosphoribosyltransferase	VFG001398	530
4.	*WP_003415934.1*	DKC2_3224	Streptogramin A acetyltransferase	VFG005938	64
5.	*WP_003416922.1*	DKC2_3434	Flavohemoprotein	VFG003391	61
6.	*WP_003420798.1*	DKC2_4054	UDP-galactopyranose mutase	VFG049062	283

Table S4: Antibiotic Resistant Non-homologous Essential Proteins Using Comprehensive Antibiotic Resistance Database [CARD]

No.	NCBI protein ID	Protein Name	Protein Function	Drug Bank Target ID	VFDB ID
1.	*WP_003415934.1*	Streptogramin A acetyltransferase	transferase activity, transferring acyl groups	*DB01669, DB01992, DB01764*	*VFG005938*
2.	*WP_003899075.1*	Catalase-peroxidase	catalase activity / heme binding / metal ion binding / NADH binding etc.	*DB00609, DB00951*	*VFG001396*

Table S5: Modelled Protein Structure Showing GMQE and QMEAN DisCo Global Score.

Protein ID	Details of the Protein	Model Identity with Template (%)	PDB ID of the Template	GMQE of Modelled Protein	QMEAN DisCo Global of Modelled Protein
WP_003899075.1	Catalase-Peroxidase	99.86	2CCD.1	0.91	0.87±0.05

Table S6: Templates of Protein Showing QSQE and GMQE Score.

PDB ID of the Template	Model Identity with Template (%)	QSQE of the Template	GMQE of Template
2CCD.1	99.86	0.80	0.96
2CCA.1	100.00	0.79	0.96

In this study, KatG protein was also evaluated based on Physical as well as functional interactions with other proteins employing STRING database in 2021[47]. The results justified that KatG protein acts as 'hub' protein (S5 Fig.) having interactions with other proteins. The nodes are closely connected in 'PPI' network owing to high clustering co-efficient value of 0.638[52]. Therefore, KatG protein having higher virulence factor was considered as a potent target. The subtractive genomic methods employed here, correlates with previous study on another pathogen[53]. The computational applications has made this study for identification of potential drug target, a straight forward in the drug development process. The application of genomes for potential drug targets, depends on those non-homologous essential proteomes[54].

Conclusion

Persistent emergence of multidrug resistant M.tb_DKC2 strains and demands for proper drug compounds, have necessitated urgent requirement for identification of potent therapeutic target which might be used for development and designing drug molecules against drug resistant MTB-pathogens. It is noteworthy that in silico subtractive genomics and structure-based methods employed in this investigation, have facilitated identification and characterization of target proteins against this pathogen. In silico strategies employed here, might be applied to identify and characterize targets in respect of all other strains of MTB-pathogens subject to in vitro and preclinical trials for validation.

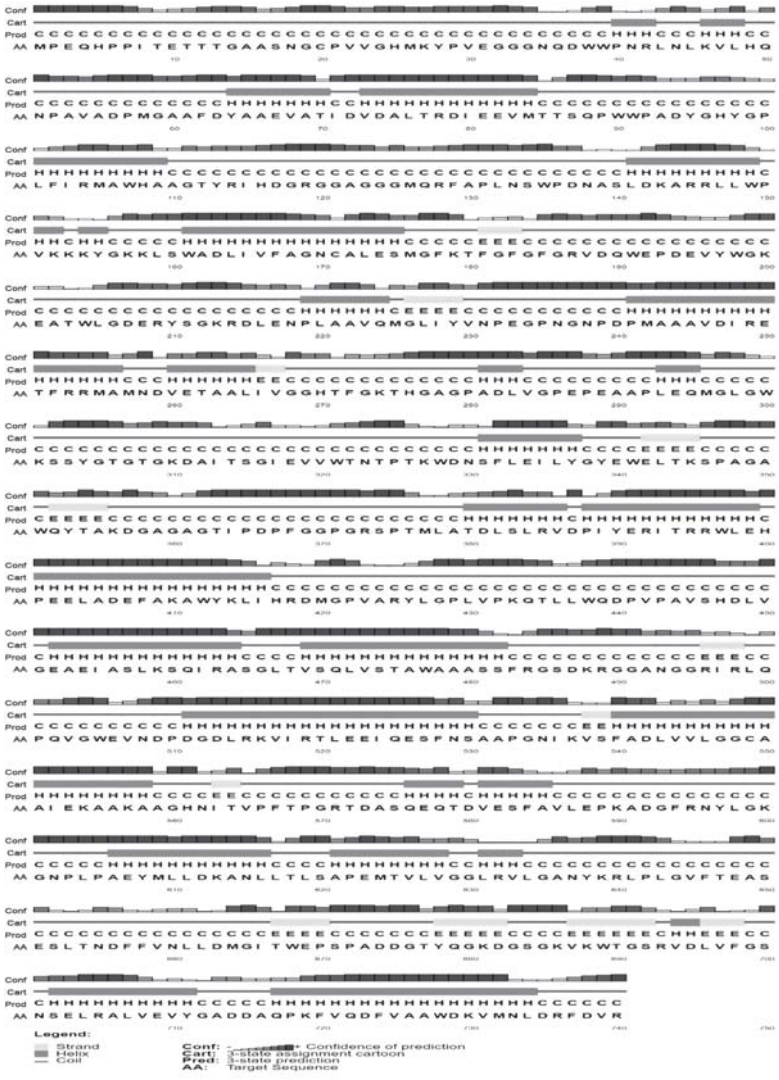

Fig. S1: Two-dimensional Graphical Output from the Prediction of Protein Secondary Structures Produced by PSIPRED.

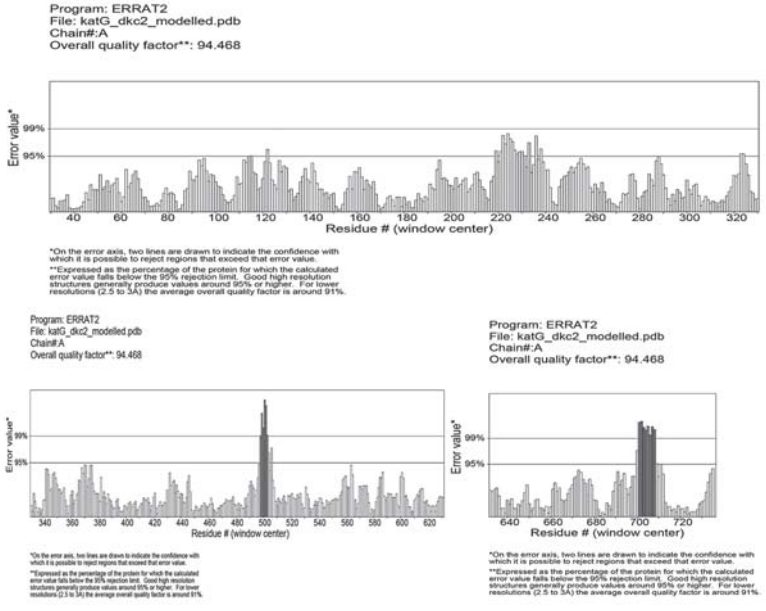

Fig. 2: ERRAT Plot (version 2.0) Displaying the Target Protein Structure Validation of WP_003899075.1.

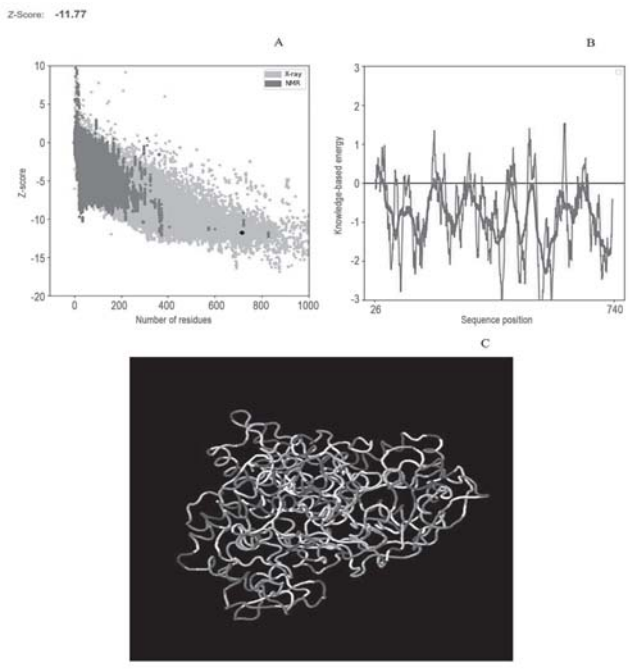

Fig. S3: (A) ProSA-web Z-score of Catalase-Peroxidase (represented in dot) Present in the Range of all Protein Chains in PDB Database. (B) ProSA-web Energy Plot of Catalase-peroxidase Model. (C) ProSA-web Molecule Visualization with Jmol.

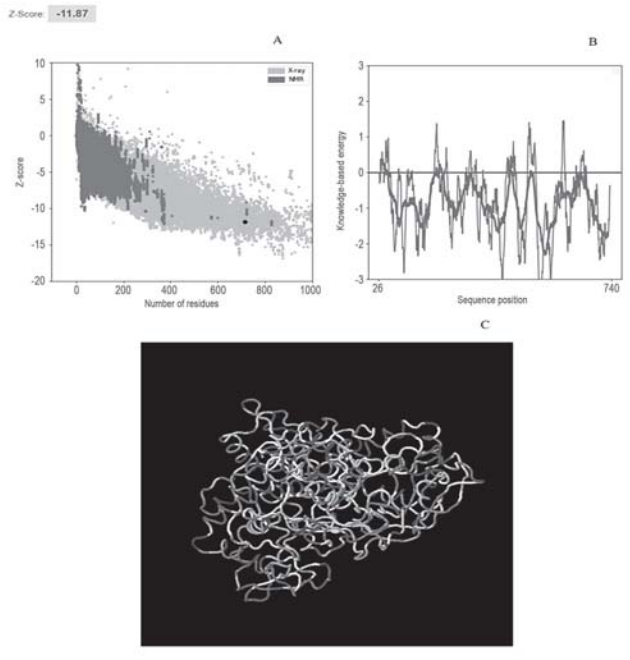

Fig. S4: ProSA-web Z-score of Template (2CCD), (B) ProSA-web Energy Plot of Template (2CCD), (C) ProSA-web Molecule Visualization with Jmol.

Table S7: Percentage Composition of Amino Acids in Target Protein using Prot Param.

Amino acid	Composition in Number	Composition in Percentage(%)
Ala(A)	77	10.04
Arg(R)	36	4.9
Asn(N)	28	3.8
Asp(D)	50	6.8
Cys(C)	3	0.4
Gln(Q)	21	2.8
Glu(E)	44	5.9
Gly(G)	79	10.7
His(H)	12	1.6
Ile(I)	25	3.4
Leu(L)	62	8.4
Lys(K)	35	4.7
Met(M)	18	2.4
Phe(F)	26	3.5
Pro(P)	50	6.8
Ser(S)	36	4.9
Thr(T)	42	5.7
Trp(W)	24	3.2
Tyr(Y)	21	2.8
Val(V)	51	6.9
Pyl(O)	0	0.0
Sec(U)	0	0.0

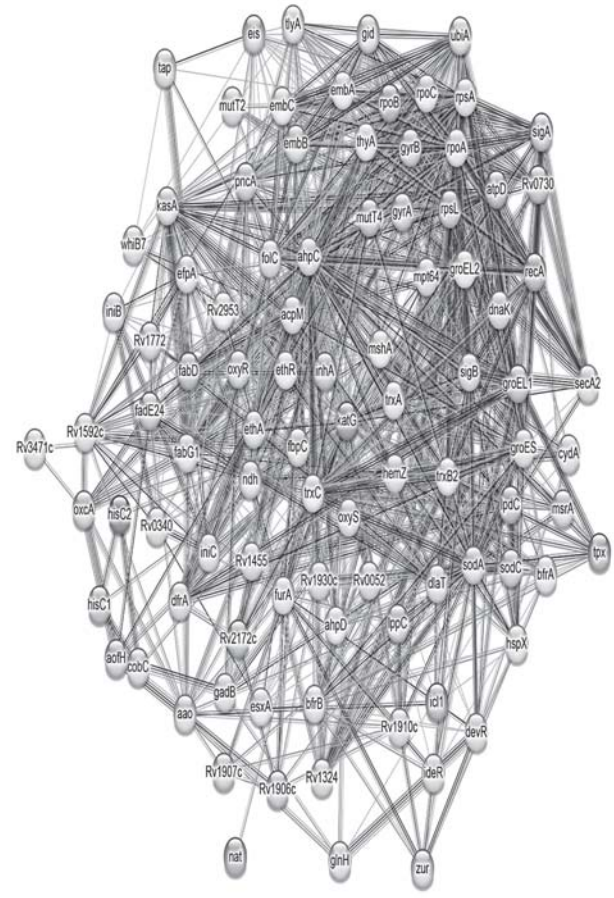

Fig. S5: Representation of PPI Network (schematic) Generated by STRING Database for Catalase-peroxidase.

Acknowledgements

The authors acknowledge University of Kalyani, Department of Computer Science & Engineering and Department of Microbiology for providing the resources and platform for conducting this study. ❐

References

1. R.E. Dalbey, M.O. Lively, S. Bron, and J.M.V. Dijl, Protein Science **6**, 1129 (1997).
2. G. Von Heijne, Journal of molecular biology **184**, 99 (1985).
3. S. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. Gordon, K. Eiglmeier, S. Gas, and C. Barry Iii, Nature, **396**, 190 (1998).
4. S.E. Hasnain, N. Z. Ehtesham, and S. Grover, Mycobacterium tuberculosis: molecular infection biology, pathogenesis, diagnostics and new interventions (Springer, 2019).
5. G. Echeverria-Valencia, S. Flores-Villalva, and C. I. Espitia, Virulence factors and pathogenicity of Mycobacterium (InTech Rijeka, 2018), Vol. 4.
6. J.E. Posey, T.M. Shinnick, and F.D. Quinn, *Journal of bacteriology* **188**, 1332 (2006).
7. B. Saint-Joanis, C. Demangel, M. Jackson, P. Brodin, L. Marsollier, H. Boshoff, and S.T. Cole, *Journal of bacteriology* **188**, 6669 (2006).
8. A. Wright and M. Zignol, Anti-tuberculosis drug resistance in the world: fourth global report: the world health organization/international union against tuberculosis and lung disease (who/union) global project on anti-tuberculosis drug resistance surveillance, 2002-2007 (World Health Organization, 2008).
9. D. Barh, S. Tiwari, N. Jain, A. Ali, A.R. Santos, A.N. Misra, V. Azevedo and A. Kumar, *Drug Development Research* **72**, 162 (2011).
10. R. Uddin and K. Saeed, *Computational Biology and Chemistry* **48**, 55 (2014).
11. K. Hema, V.I. Priyadarshini, D. Pradhan, M. Munikumar, S. Sandeep, N. Pradeep, M. Suchitra, and A. Umamaheswari, *Biochemistry and Analytical Biochemistry* **4**, 1 (2015).
12. M.A. Rahman, M.S. Noore, M.A. Hasan, M.R. Ullah, M.H. Rahman, M.A. Hossain, Y. Ali, and M.S. Islam, *Computational Biology and Chemistry* **52**, 66 (2014).
13. M. Hasan, K.F. Azim, M.A.S. Imran, I.M. Chowdhury, S.R.A. Urme, M.S.A. Parvez, M.B. Uddin, and S.S.U. Ahmed, *PloS one* **15**, e0237181 (2020).
14. M.T. Khan, A. Mahmud, A. Iqbal, S.F. Hoque, and M. Hasan, *Informatics in Medicine Unlocked* **20**, 100385 (2020).
15. V. Solanki and V. Tiwari, *Scientific Reports* **8**, 9044 (2018).
16. H. Kaur, M. Kalia, and N. Taneja, *Microbial Pathogenesis* **152**, 104608 (2021).
17. A.J. Fatoba, M. Okpeku, and M.A. Adeleke, *Pathogens* **10**, 921 (2021).
18. M.I. Hosen, A.M. Tanmoy, D.-A. Mahbuba, U. Salma, M. Nazim, M.T. Islam, and S. Akhteruzzaman, Interdisciplinary Sciences: Computational Life Sciences **6**, 48 (2014).
19. A. Kumar Jaiswal, S. Tiwari, S.B. Jamal, D. Barh, V. Azevedo, and S.C. Soares, *International journal of Molecular Sciences* **18**, 402 (2017).
20. M. Duffield, I. Cooper, E. McAlister, M. Bayliss, D. Ford, and P. Oyston, Molecular BioSystems **6**, 2482 (2010).
21. A.E. Clatworthy, E. Pierson and D.T. Hung, *Nature Chemical Biology* **3**, 541 (2007).
22. D.B. Folkvardsen, A. Norman, Å.B. Andersen, E.M. Rasmussen, T. Lillebaek, and L. Jelsbak, *Scientific Reports* **8**, 11869 (2018).
23. A. Norman, D.B. Folkvardsen, S. Overballe-Petersen, and T. Lillebaek, *Microbiology Resource Announcements* **8**, 10.1128/mra. 01554 (2019).
24. R. Uddin, Q.N. Siddiqui, M. Sufian, S.S. Azam and A. Wadood, *Genes & Genomics* **41**, 1281 (2019).
25. M.M. Das and K. Sarkar, *Bioinformation* **18**, 1126 (2022).
26. J.L. Gardy, M.R. Laird, F. Chen, S. Rey, C. Walsh, M. Ester, and F.S. Brinkman, *Bioinformatics* **21**, 617 (2005).
27. W. Li and A. Godzik, *Bioinformatics* **22**, 1658 (2006).
28. L. Fu, B. Niu, Z. Zhu, S. Wu, and W. Li, *Bioinformatics* **28**, 3150 (2012).
29. K. Palaniappan and S. Mukherjee, in 2011 10th International.Conference on Machine Learning and Applications and Workshops (IEEE, 2011), pp. 189.
30. M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, and T. L. Madden, *Nucleic acids Research* **36**, W5 (2008).
31. R. Zhang, H.Y. Ou, and C.T. Zhang, *Nucleic Acids Research* **32**, D271 (2004).
32. H. Luo, Y. Lin, F. Gao, C.-T. Zhang, and R. Zhang, *Nucleic Acids Research* **42**, D574 (2014).
33. D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, and Z. Sayeeda, *Nucleic Acids Research* **46**, D1074 (2018).
34. N.Y. Yu, J.R. Wagner, M.R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S.C. Sahinalp, M. Ester, and L.J. Foster, *Bioinformatics* **26**, 1608 (2010).
35. L. Chen, D. Zheng, B. Liu, J. Yang, and Q. Jin, *Nucleic Acids Research* **44**, D694 (2016).
36. B.P. Alcock, A.R. Raphenya, T.T. Lau, K.K. Tsang, M. Bouchard, A. Edalatmand, W. Huynh, A.-L.V. Nguyen, A.A. Cheng, and S. Liu, *Nucleic Acids Research* **48**, D517 (2020).
37. M. Goyal, C. Citu, and N. Singh, *Asian J. Pharm. Clin. Res* **11**, 230 (2018).
38. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T.L. Madden, *BMC Bioinformatics* **10**, 1 (2009).
39. M. Steinegger, M. Meier, M. Mirdita, H. Vöhringer, S.J. Haunsberger, and J. Söding, *BMC bioinformatics* **20**, 1 (2019).
40. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. de Beer, C. Rempfer, and L. Bordoli, *Nucleic Acids Research* **46**, W296 (2018).
41. R.A. Laskowski, M.W. MacArthur, D.S. Moss, and J.M. Thornton, *Journal of Applied Crystallography* **26**, 283 (1993).

42. D.W. Buchan and D.T. Jones, *Nucleic Acids Research* **47**, W402 (2019).

43. D.T. Jones, *Journal of Molecular Biology* **292**, 195 (1999).

44. S. Agnihotry, R.K. Pathak, D.B. Singh, A. Tiwari, and I. Hussain, in *Bioinformatics* (Elsevier, 2022), pp. 177.

45. M. Wiederstein and M.J. Sippl, *Nucleic Acids Research* **35**, W407 (2007).

46. E. Gasteiger, C. Hoogland, A. Gattiker, S.E. Duvaud, M.R. Wilkins, R.D. Appel, and A. Bairoch, Protein identification and analysis tools on the ExPASy server (Springer, 2005).

47. D. Szklarczyk, A.L. Gable, K.C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N.T. Doncheva, M. Legeay, T. Fang, and P. Bork, *Nucleic Acids Research* **49**, D605 (2021).

48. G. Studer, G. Tauriello, S. Bienert, M. Biasini, N. Johner, and T. Schwede, *PLoS Computational Biology* **17**, e1008667 (2021).

49. R. Laskowski, M. MacArthur, and J. Thornton, (2006).

50. S.K. Kwofie, B. Dankwa, E.A. Odame, F.E. Agamah, L.P. Doe, J. Teye, O. Agyapong, W.A. Miller III, L. Mosi, and M. D. Wilson, *Molecules* **23**, 1550 (2018).

51. V.V. Khrustalev, T.A. Khrustaleva, and V.V. Poboinev, *Biosystems* **168**, 45 (2018).

52. J. Bultinck, S. Lievens, and J. Tavernier, *Current Pharmaceutical Design* **18**, 4619 (2012).

53. T. Hossain, M. Kamruzzaman, T.Z. Choudhury, H.N. Mahmood, A. Nabi, and M.I. Hosen, *BioMed Research International* **2017** (2017).

54. A.M. Butt, I. Nasrullah, S. Tahir, and Y. Tong, (2012).