Why Hypothetical Protein KPN00728 of *Klebsiella pneumoniae* Should Be Classified as Chain C of Succinate Dehydrogenase?

Sy Bing Choi · Yahaya M. Normi · Habibah A. Wahab

Abstract Twenty percent of genes that encode for hypothetical proteins from *Klebsiella pneumoniae* MGH 78578 were identified, leading to KPN00728 and KPN 00729 after bioinformatics analysis. Both open reading frames showed high sequence homology to Succinate dehydrogenase Chain C (SdhC) and D (SdhD) from *Escherichia coli*. Recently, KPN00729 was assigned as SdhD. KPN00728 thus remains of particular interest as no annotated genes from the complete genome sequence encode for SdhC. We discovered KPN00728 has a missing region with conserved residues important for ubiquinone (UQ) and heme group binding. Structure and function prediction of KPN00728 coupled with secondary structure analysis and transmembrane topology showed KPN00728 adopts SDH-(subunit C)-like structure. To further probe its functionality, UQ was docked on the built model (consisting KPN00728 and KPN00729) and formation of hydrogen bonds between UQ and Ser27, Arg31 (KPN00728) and Tyr84 (KPN00729) further reinforces and supports that KPN00728 is indeed SDH. This is the first report on the structural and function prediction of KPN00728 of *K. pneumoniae* MGH78578 as SdhC.

Abbreviations

HP	Hypothetical protein
ORF	Open reading frame
SDH	Succinate dehydrogenase
SdhA	Succinate dehydrogenase chain A
SdhB	Succinate dehydrogenase chain B
SdhC	Succinate dehydrogenase chain C
SdhD	Succinate dehydrogenase chain D
UQ	Ubiquinone
TM	Transmembrane
PDB	World wide web Protein Data Bank
DOPE	Discrete optimized potential energy
FAD	Flavin adenine dinucleotide
RMSD	Root mean square deviation
KIE	Kinetic isotopes effect
TST	Transition states theory
NQM	Nuclear quantum mechanics

S. B. Choi · H. A. Wahab
Pharmaceutical Design and Simulation (PhDS) Laboratory,
School of Pharmaceutical Sciences, Universiti Sains Malaysia,
11800 Minden, Pulau Pinang, Malaysia

Y. M. Normi
School of Biological Sciences, Universiti Sains Malaysia, 11800
Minden, Pulau Pinang, Malaysia

H. A. Wahab (✉)
Centre for Advanced Drug Delivery, Malaysian Institute
of Pharmaceuticals and Nutraceuticals, Ministry of Science,
Technology and Innovation, SAINS@USM, No 10,
11900 Persiaran Bukit Jambul, Pulau Pinang, Malaysia
e-mail: habibah@ipharm.gov.my; habibahw@usm.my

1 Introduction

*Klebsiella pneumonia*e is a Gram negative, non-motile and rod-shaped bacterium. The genus *Klebsiella* is a member of the family Enterobacteriaceae which causes a wide range of infections. To date, 7 known species of *Klebsiella* which had shown DNA homology had been identified. These are

Klebsiella pneumoniae, Klebsiella ozaenae, Klebsiella rhinoscleromatis, Klebsiella oxytoca, Klebsiella planticola, Klebsiella terrigena, and *Klebsiella ornithinolytica*. *Klebsiella pneumoniae* is one of the most medically important species of the group [15]. *K pneumoniae* is known as an opportunistic pathogen found in the environment and in mammalian mucosal surfaces. They appeared as normal flora of the intestinal tract but usually low in number compared to *Escherichia coli* [38]. Generally, *K. pneumoniae* infections tend to occur in patient with a weakened immune system and people with underlying diseases [14]. The principal pathogenic reservoirs of infection are the gastrointestinal tract of patients and the hands of hospital personnel. These infections can spread rapidly, often leading to nosocomial outbreaks which can be fatal. Studies conducted in Asia estimate that the incidence rate in elderly persons to be 15–40% [18, 27, 35], which is equal to or greater than that of *Haemophilus influenzae* and the occurrences are far more common in Asia than elsewhere [15]. Although the incidents of community acquired *K. pneumoniae* has apparently decreased, the mortality rate remains significant as a result of other underlying disease(s) that tend(s) to be aggressively present in the affected patient [21] including alcoholics [13] despite optimal medical treatment [6]. These rapid incidences deserved to be investigated, understood and delineated.

Recently, genome sequencing determination for the complete genome of *K. pneumoniae* MGH 78578 was completed in 2007 by Genome Research Center of Washington University of St. Louise [26]. It consists of about 5 million of nucleotides and a total of 4,894 genes and out of that, 4,776 genes encode proteins. Further analysis showed that from the 4,776 protein coding genes, there are about 20% of the genes which are annotated poorly and are classified as hypothetical genes. In theory, these hypothetical genes (nucleic acid sequence) are eventually translated into proteins known as hypothetical proteins. In addition, these hypothetical proteins have not shown to exist by experimental protein evidence. They also normally have low sequence identity to known annotated proteins and majority of the functional aspect of these proteins are not known [19]. It is therefore worth to predict their structures which give clues to the functions of these proteins in the view of the fact that they are coded by 20% of genes in the genome of *K. pneumoniae*. In this study we aimed to analyze and interrogate these proteins via computational approach to give us insight of their possible function and mechanisms.

There are a total of 1,003 hypothetical proteins in *K. pneumoniae* MGH 78578, of which one that is the focus of our discussion has been assigned as KPN00728 (gi: 152969292) (Table 1). Recently, a revision of the genome map of this organism assigned the function of KPN00729 as 'provisionally Chain D of Succinate dehydrogenase'. When we started this work, (from December 2007 to July 2008) this protein along with KPN00728 were classified as hypothetical proteins. To date, although the function of KPN00729 is provisionally known, the structure of this protein is yet to be determined. KPN00728 and KPN00729 have 91 and 115 amino acids, respectively. BLAST result showed that both of them have more than 90% sequence identity with Succinate dehydrogenase of Enterobacteriaceae family. Since it is believed that the function of an unknown protein can be inferred from other known homologous proteins based on their sequence and structure similarity, therefore, we postulated that these hypothetical are subunits of Succinate dehydrogenase enzyme.

Succinate dehydrogenase plays an important role in the aerobic respiratory chain and Krebs cycle in both eukaryotic and prokaryotic organisms (Fig. 1). In general, it is encoded by four different genes namely *SdhA*, *SdhB*, *SdhC* and *SdhD*, respectively. It is believed that the mutation of human genes encoding Succinate dehydrogenase subunits leads to cancer and aging although this rarely happen [32]. However, no details of this mechanism have been reported so far. Inhibition of Succinate dehydrogenase by carboxin and thenoyltrifluoroacetone in Krebs cycle results in total termination of respiration in the pathway [20]. This is known as metabolic poisoning which can be fatal for both eukaryotic and prokaryotic organisms.

Succinate dehydrogenase comprising of four chains structurally contribute to a heterotetramer complex [16]. It is divided into three domains: Chain A-SdhA (catalytic domain), Chain B-SdhB (electron transfer subunit) and Chain C-SdhC and Chain D-SdhD (heme binding domain). The first two domains or chains are located in the matrix of the mitochondria. The third domain [composed of Chain C (SdhC) and D (SdhD)] forms dimeric membrane unit anchored together with a heme group at the transmembrane of the mitochondria. *SdhA* and *SdhB* have shown hydrophilic characteristic where they are attached to the inner cytoplasmic surface of the membrane [39]. Both SdhA and SdhB were found to interact with the hydrophobic subunit of SdhC and SdhD [39]. It is observed that SdhA and SdhB are more structurally conserved and have higher sequence similarity [39] but SdhC and SdhD have higher sequence variation amongst organisms in the same family of Succinate dehydrogenase [39]. It is interesting to note here, that the genome map of *K. pneumonia* MGH78578 failed to reveal the sequence of SdhC and only recently assigned KPN00729 as SdhD which led us to believe the protein is coded as hypothetical protein.

In this work, we present results from computational approaches to determine the structure of KPN00729 and hypothetical protein KPN00728 from *K. pneumoniae* MGH 78578 in order to elucidate the function of

Table 1 Sequence identity and similarity between KPN00728 and KPN00729 and available templates obtained from BLAST search against PDB

ID	Significant hit	*E*-value	SI (%)[a]	SS (%)[b]	Organism	Protein function	Ref
KPN00728	**Complex II (Succinate dehydrogenase) from *E. coli* with Ubiquinone bound: 1NEK chain C**	**8e-42**	**90**	**95**	***E. coli***	**SDH**	[43]
	Complex II (Succinate dehydrogenase from *E. coli* with Dinitrophenol-17 inhibitor co-crystallized at the ubiquinone binding site: 1NEN chain C	8e-42	90	95	*E. coli*	SDH	[43]
	Complex II (Succinate dehydrogenase from *E. coli* with Atpenin A5 inhibitor co-crystallized at the ubiquinone binding site: 2ACZ Chain C	8e-42	90	95	*E. coli*	SDH	[10]
KPN00729	**Complex II (Succinate dehydrogenase) from *E. coli* with Ubiquinone bound: 1NEK chain D**	**5e-48**	**90**	**95**	***E. coli***	**SDH**	[43]
	Complex II (Succinate dehydrogenase from *E. coli* with Dinitrophenol-17 inhibitor co-crystallized at the ubiquinone binding site 1NEN chain D	5e-48	90	95	*E. coli*	SDH	[43]
	Complex II (Succinate dehydrogenase from *E. coli* with Atpenin A5 inhibitor co-crystallized at the ubiquinone binding site 2ACZ Chain D	5e-48	90	95	*E. coli*	SDH	[10]

The significant hits which are in bold for both KPN00728 and KPN00729 were selected as the template or the structure modeling

[a] Sequence identity

[b] Sequence similarity

Fig. 1 An overview of Krebs cycles. The section (highlighted in *red*) shows the catalytic scheme of Succinate dehydrogenase. Succinate is oxidized into fumarate, converting FAD to $FADH_2$ which facilitates electron transfer to ubiquinone and form QH_2
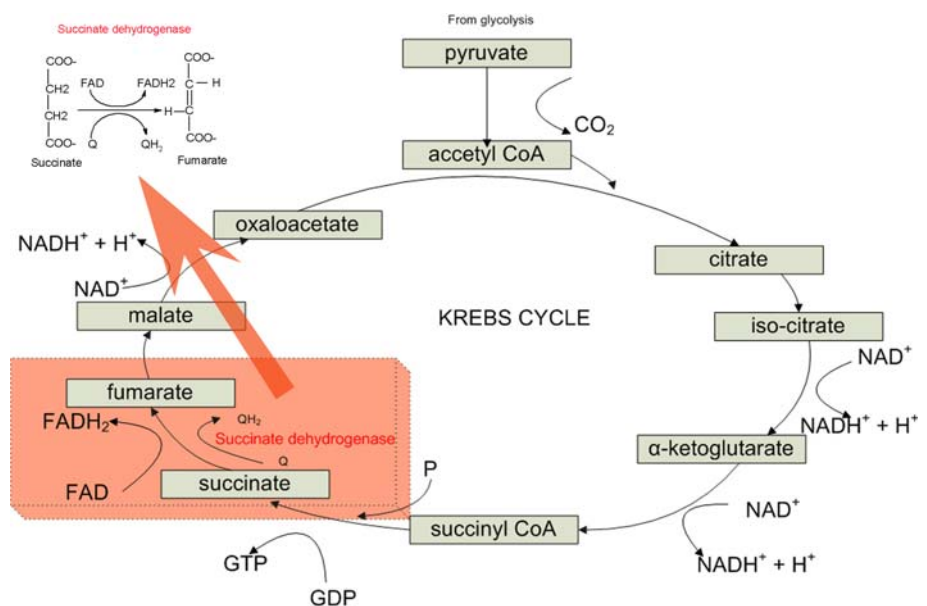
KPN00728. This is intriguing from the fact that this protein actually shared $\sim 90\%$ sequence identity with Sdhs from other microorganisms. Sequence analysis of the genome revealed that there might be a missing region representing 38 translated amino acid residues in KPN00728 which are important for the protein to function as Succinate dehydrogenase. 1NEK, crystal structure of Succinate dehydrogenase from *E. coli* was selected as the template for homology modeling. From the predicted structure of both proteins, we found that the built model showed similar structural features with the template used (that from *E. coli*) in terms of its transmembrane (TM) topology and their secondary structural arrangement. Binding of ubiquinone at the active site was also observed from docking simulations performed on the built model. This feature helped to distinguish Succinate dehydrogenase Chain C and D from other peptide function. In addition, we observed that the active site was active during docking simulation. Possible hydrogen bond is postulated to exist between O1 of ubiquinone and Tyr83

from KPN00729 similar to what observed with the binding of ubiquinone in the crystal structure of Succinate dehydrogenase from *E. coli* [10]. This allowed us to make a hypothesis on the structure–function relationship for both of the selected proteins (KPN00729 and the HP, KPN00728) from *K. pneumoniae* MGH78578,

2 Computational Methods

Common bioinformatics computational approach that combines database search, comparative homology modeling and docking simulation were employed in our quest to predict the structure and function of KPN00728 and KPN00729. The complete genome of *K. pneumoniae* subsp. *pneumoniae* MGH 78578 was obtained from NCBI database (http://www.ncbi.nlm.nih.gov/sites). Primary sequence of these proteins was used to search through the non-redundant database BLAST local alignment tool [1]. KPN00728 and KPN00729 were further searched against Protein Data Bank (PDB) [3] with BLAST. Multiple sequence alignment within members of Enterobacteriaceae was performed using CLUSTAL-W program [37]. Based on the sequence identity obtained form BLAST and ClustalW results for both proteins, Succinate dehydrogenase Chain C and D from *E. coli* (PDB id: 1NEK Chain C and D) were then selected as the template for structure prediction of KPN00728 and KPN00729.

Next, three dimensional (3D) models for KPN00728 and KPN00729 were built using MODELLER 9 version 2 [33]. 20 models were generated randomly. 1NEK Chain C was used as the template for KPN00728 and 1NEK Chain D was used as the template for KPN00729. Subsequently, the best model with the highest Discrete Optimized Potential Energy (DOPE) score was chosen. To further eliminate unfavorable contacts and steric clashes, the built model underwent 2,000 cycles of energy minimization using Sander module in Amber 8 program package [7, 31]. Verification of the best model was done using PROCHECK Ramachandran plot [17]. MGenthreader secondary prediction tool by Jones and co-workers [12] and STRIDE [9] were used for secondary structure prediction [23]. Comparison between 1NEK Chain C and D with built model on the transmembrane segment were performed using Toppred web server [40].

Docking of ubiquinone to the putative Succinate dehydrogenase Chain C and D was performed using AutoDock 3.0.5 software [25]. The polar hydrogen atoms, Kollman-amber united atom partial charges and solvation parameters were added on the built model with the aid of AutoDock tools. Partial charges of ubiquinone were assigned with Gasteiger charges. Non-polar hydrogen atoms of ubiquinone were merged and 7 rotatable bonds were assigned.

Grid map of $40 \times 40 \times 40$ grid points and 0.375 Å spacing were generated using Autogrid3 [25] program and centered around the potential binding site. Molecular docking simulation was carried out using Lamarckian genetic algorithm and the Solis and Wets local search method [34] with Autodock 3.0.5 [25]. A total of 300 runs with 250 population size, root mean square tolerance 1.0 Å were set for the docking simulation. The lowest docked energy of each conformation in the most populated cluster was selected.

3 Results and Discussions

3.1 Selection of Template

For selection of an appropriate template, KPN00728 and KPN00729 underwent a local alignment search against the non-redundant database using BLAST tool. The result yielded remarkable similarity with Succinate dehydrogenase subunit C and D for other microorganisms with indication of *E*-value above the threshold (Table 1). From the result, sequence identity for KPN00728 and KPN00729 with *E. coli* are ranked second and fifth, respectively, from the top 10 hits showed in Table 2. Subsequently, both proteins were further searched against PDB using BLAST. Results showed sequences of KPN00728 (gi: 152969292) and KPN00729 (gi: 152969293) recorded 90.5% sequence identity with that of Succinate dehydrogenase group of *E. coli* (PDB id: 1NEK). In addition, the *E*-values are above the threshold values with those of *E. coli* Succinate dehydrogenase (Table 1). Complex II (Succinate dehydrogenase) from *E. coli* with Ubiquinone bound (PDB id: 1NEK), Complex II (Succinate dehydrogenase) from *E. coli* with Dinitrophenol-17 inhibitor co-crystallized at the ubiquinone binding site (PDB id: 1NEN) and Complex II (Succinate dehydrogenase) from *E. coli* with Atpenin A5 inhibitor co-crystallized at the ubiquinone binding site (PDB id: 2ACZ) have the same sequence but the structures were solved crystallographically with different interacting ligand (Table 1). Based on both BLAST results and the fact that Succinate dehydrogenase from *E. coli* is the only current available crystal structures, 1NEK was selected as the template for subsequent modeling for KPN00728 and KPN00729. In addition, it has the best crystallographic resolution amongst those Succinate dehydrogenase solved for *E. coli*. [43].

3.2 Sequence and Structural Analysis

In the *K. pneumoniae* MGH78578 complete genome map, hypothetical proteins KPN00728 and KPN00729 were

Table 2 BLAST search result with different organism

Accession number	Title/function	Organism	SI (%)[a]	SS (%)[b]	E-value
KPN00728					
Gb:AAO39685.1	Succinate dehydrogenase cytochrome b-556 subunit; SdhC	*Enterobacter cloacae*	87	96	2e-39
Ref:NP_286437	Succinate dehydrogenase cytochrome b556 large membrane subunit	*Escherichia coli* O157:H7 EDL933	90	95	2e-39
Ref:YP_001453991	Succinate dehydrogenase cytochrome b556 large membrane subunit	*Citrobacter koseri* ATCC BAA-895	90	95	4e-39
Ref:YP_688171.1	Succinate dehydrogenase cytochrome b556 large membrane subunit	*Shigella flexneri* 5 str. 8401	89	95	9e-39
Ref:NP_455288.1	Succinate dehydrogenase cytochrome b556 large membrane subunit	*Salmonella enterica* subsp. *enterica serovar Typhi* str. CT18	86	93	2e-38
Ref:ZP_0203463.1	Succinate dehydrogenase, cytochrome b556 subunit	*Escherichia albertii* TW07627	90	95	3e-38
Ref:YP_049463.1	Succinate dehydrogenase cytochrome b556 large membrane subunit	*Pectobacterium atrosepticum* SCRI1043	80	90	3e-34
Ref:ZP_00828037.1	Succinate dehydrogenase/fumarate reductase, cytochrome B subunit	*Yersinia frederiksenii* ATCC 33641	72	90	8e-32
Ref:ZP_02959665.1	Hypothetical protein PROSTU_01551	*Providencia stuartii* ATCC 25827	73	86	1e-31
Ref:YP_454552.1	Succinate dehydrogenase cytochrome b556 large membrane subunit	*Sodalis glossinidius* str. 'morsitans'	72	88	6e-31
KPN00729					
Ref:YP_001175955	Succinate dehydrogenase cytochrome b556 small membrane subunit [*Enterobacter* sp. 638]	*Enterobacter* sp. 638]	88	95	9e-52
Ref:YP_001453990.1	Succinate dehydrogenase cytochrome b556 small membrane subunit	*Citrobacter koseri* ATCC BAA-895	88	96	4e-51
Gb:AAO39686.1	Succinate dehydrogenase hydrophobic membrane anchor protein:SdhD	*Enterobacter cloacae*	87	94	5e-51
Ref:ZP_02903524.1	Succinate dehydrogenase, hydrophobic membrane anchor protein	*Escherichia albertii* TW07627	87	93	1e-49
Ref:YP_001438701.1	Succinate dehydrogenase cytochrome b556 small membrane subunit	*Enterobacter sakazakii* ATCC BAA-894	83	93	2e-48
Ref:NP_415250.1	Succinate dehydrogenase, membrane subunit, binds cytochrome b556	*Escherichia coli* str. K-12 substr. MG1655	90	95	2e-45
Ref:NP455289.1	Succinate dehydrogenase cytochrome b556 small membrane subunit	*Salmonella enterica* subsp. *enterica serovar Typhi* str. CT18	88	95	3e-45
Ref:YP_001879377.1	Succinate dehydrogenase, hydrophobic membrane anchor protein	*Shigella boydii* CDC 3083-94	89	94	4e-45
Ref:YP049464.1	Succinate dehydrogenase cytochrome b556 small membrane subunit	*Acterium atrosepticum* SCRI1043	73	89	2e-43
Ref:NP_928725.1	Succinate dehydrogenase cytochrome b556 small membrane subunit	*Photorhabdus luminescens* subsp. laumondii TTO1	71	89	8e-42
Ref:ZP_00825352.1	Succinate dehydrogenase, hydrophobic anchor subunit	*Yersinia mollaretii* ATCC 43969	70	89	6e-41

Top 10 hits with different organisms from BLAST search against Non-Redundant database with KPN00728 and KPN00729

[a] Sequence identity

[b] Sequence similarity

Table 3 Location of the protein coding gene for KPN00728 and KPN00729 in the whole genome of *Klebsiella pneumoniae* MGH78578

Location	Length	PID	Code	Product
818319–818594	91	152969292	KPN00728	Hypothetical protein
818588–818935	115	152969293	KPN00729	SdhD
818954–820701	588	152969294	KPN00730	SdhA
820717–821433	238	152969295	KPN00731	SdhB

coded by two protein coding genes which are located from 818319 to 818594 and from 818588 to 818935, respectively (Table 3). We found that the location of protein coding genes *sdhA* and *sdhB* encoding Succinate dehydrogenase catalytic subunit Chain A and Chain B are located after both protein coding genes that coded for KPN00728 and KPN00729 (Table 3). Since both KPN00728 and KPN00729 shared 90% sequence identity with Succinate dehydrogenase of *E. coli* (1NEK) as well as the location of the genes, we believe that KPN00728 and KPN00729 might be Chain C and Chain D of Succinate dehydrogenase.

Nevertheless, the length of KPN00728 is 38 residues shorter than the chosen template (1NEK Chain C) (Fig. 2a). Iwata and co-workers [10] suggested that Ser27 and Arg31 from Chain C of Succinate dehydrogenase of *E. coli* might have some interactions with ubiquinone at the binding site where ubiquinone is bound. Based on similar argument, we hypothesized that if those 38 residues are missing or do not exist, KPN00728 might not be able to interact with ubiquinone, as it requires the corresponding Ser27 (in *E. coli*) which is essential for the protein to play its role as a Succinate dehydrogenase. Thus, an effort was made to search for this region in the genome map of *K. pneumoniae* MGH78578.

Referring to Fig. 3a and b, there are a total of 770 nucleotides before KPN00728 gene (red coloured region in Fig. 3a) in which the feature is not being identified yet (white region in Fig. 3b). Translations were done from nucleotide to amino acids for 114 nucleotides at the beginning of KPN00728 gene in a reverse direction (Fig. 3b). From there, these translated 38 residues of amino acids were taken to perform a manual local alignment between the *E. coli* Succinate dehydrogenase Chain C from residues 1 to 38 (Fig. 3c). Among these 38 residues, only 3 residues are different from each other and the sequence identity is 92% within these 38 residues (Fig. 2b). Residues which are involved in the interaction with the ubiquinone were shown to be conserved including the position of Ser27 and Arg31 in KPN00728. Based on this result, it strengthens the possibility further that KPN00728 and along with KPN00729 are indeed Succinate dehydrogenase Chain C and D, respectively.

3.3 Multiple Sequence Alignment

Multiple sequence alignment among 7 other Enterobacteriaceae was done for both KPN00728 (inclusive of the missing 38 residues) and KPN00729 (Fig. 4). The length of KPN00728 and KPN00729 are consistent with 7 other Enterobacter's Succinate dehydrogenase Chain C and D. Ser27 and Arg31 from KPN00728, Tyr83 from KPN00729 are found to be highly conserved among 7 other Succinate dehydrogenases from different Enterobacteriaceae. These three residues are deemed important for ubiquinone binding [39]. Two His residues which are known to be centering around the heme group from Chain C and D of Succinate dehydrogenase [36] have also been identified in both KPN00728 and KPN00729.

A

```
KPN_00728    1   MAVGILLWLLGTSLSSPEGFLTAASIMDSFFVKFIMWGILTALAYHVVVGIRHLMMDFGY   60
                 +AVGILLWLLGTSLSSPEGF  A++IM SFFVKFIMWGILTALAYHVVVGIRH+MMDFGY
1NEK_C      39   VAVGILLWLLGTSLSSPEGFEQASAIMGSFFVKFIMWGILTALAYHVVVGIRHMMMDFGY   98

KPN_00728   61   LDETLEAGKRSAKISFVITVVLSLLAGVLVW   91
                 L+ET EAGKRSAKISFVITVVLSLLAGVLVW
1NEK_C      99   LEETFEAGKRSAKISFVITVVLSLLAGVLVW   129
```

B   Local alignment

E.coli 1NEK:C (1-38res) MIRNVKKQRPVNLDLQTIRFPITAIASILHRVSGVITF

KPN00728   MVRNMKKQRPVNLDLQTIRFPITAIASILHRVFGVITF

Fig. 2 a Local alignment of KPN00728 with selected template 1NEK Chain C. KPN00728 aligned with 1NEK chain C starting at Residue 39. The sequence identity is 90.0%. **b** Sequence alignment of the newly found 1–38 residues of KPN00728. Local alignment of the newly found 1–38 residues of KPN00728 were done between the 1 and 38 residues of Succinate dehydrogenase from *E. coli*. There are three different residues (represented in *red*) between both of the proteins. The sequence identity of the newly found 1–38 residues as compared to *E. coli* is 92.3%
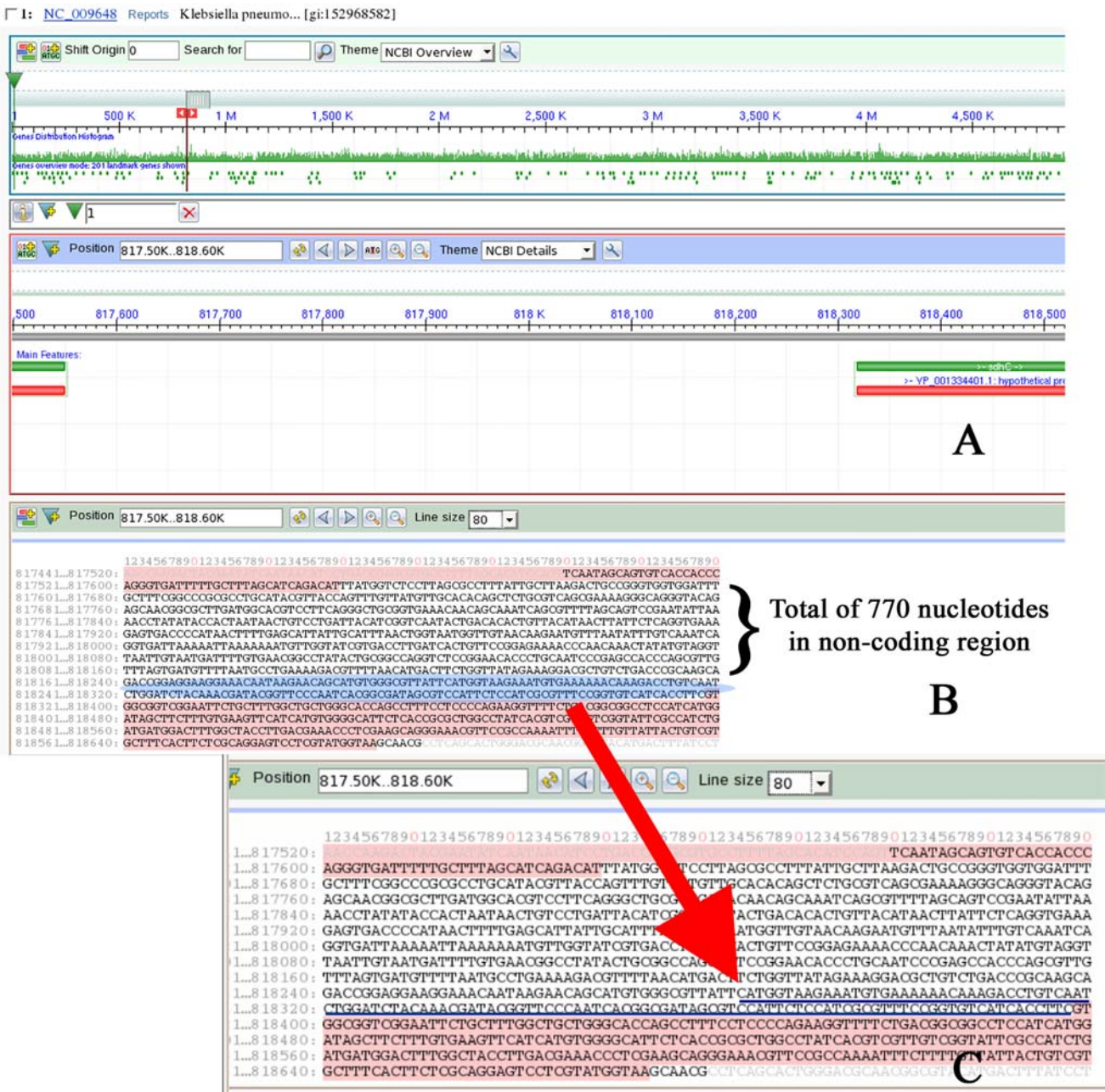
Fig. 3 a–c Snapshots of *Klebsiella pneumoniae* MGH78578 complete genome map around the 817500–818500 nucleotides. Nucleotides that are not shaded (*white region*) namely none coding region in **b** are yet to be classified. The *pink shaded* nucleotides in **b** are classified as protein coding gene. Total of 117 nucleotides at the non coding region in the genome which *underlined* in *blue* at the *bottom* in 3C is suspected to belong to KPN00728 (a total 276 nucleotides = 91 residues of amino acids)

3.4 Model Building and Validation

Comparison of Succinate dehydrogenase (PDB: 1NEK) and both KPN00728 and KPN00729 showed some consistency in the built model. Root mean square deviation (RMSD) calculated between them gave the value of 3.91 Å (Fig. 5). There are three helices from each Chain C and D of 1NEK and these were also observed in the built model.

Furthermore, topology and the packing of six helices of both built model and 1NEK were similar (Table 4). This showed that 1NEK Chain C and D are indeed appropriate templates for both proteins, respectively. The similarities of the helices length and transmembrane topology gave a deeper conviction that KPN00728 and KPN00729 are in fact, the suspected Succinate dehydrogenase Chain C and D, respectively.

Fig. 4 Multiple sequence alignment of KPN00728 and KPN00729 with seven other members of the Enterobacterociae family. Ser27 and Arg31 in KPN00728 and Tyr83 in KPN00729 are highly conserved among the other seven enterobacters. Possible interactions with ubiquinone are found among these residues. His30, His84 and His91 from KPN00728 and His71 from KPN00729 are also highly conserved. These His residues were located closely with the heme group of Succinate dehydrogenase. These residues are postulated to be either directly or indirectly involved in the binding of heme group to KPN00728 and KPN00729.
* indicated as conserved regions

```
SdhC[Enterobacter  MTRSYRKGTLLDPHRSPEEGNNKNSMWALFMIRNVKKQRPVNLDLKTIRFPVTAIASIIH  60
SdhC[Shigella      ------------------------------MIRNVKKQRPVNLDLQTIRFPITAIASIIH  30
SdhC[Salmonella    ------------------------------MIRNVKKQRPVNLDLQTIRFPITAIASIIH  30
SdhC[Erwinia       ------------------------------MGKSVVKKQRPVNLELQTIQFPVTAIASIIH  30
SdhC[Yersinia      ---------------------MENNKSCVGKNVKKQRPVNLDLQTIRFPVVNLVALASIIH  37
SdhC[Sodalis       ------------------------------MGKTVKKQRPVNLDLQTIRFPVTAIVSIIH  30
1NEK_C[E.coli]     ------------------------------MIRNVKKQRPVNLDLQTIRFPITAIASIIH  30
KPN_00728          ------------------------------MVRNMKKQRPVNLDLQTIRFPITAIASIIH  30
                                                 : :.:*********:*:**:**:***.****

SdhC[Enterobacter  RVSGVITFVAVGILLWLLGTSLSSPEGFLQASAIMNSFFVKFIMWGILTALAYHVVVGVR  120
SdhC[Shigella      RVSGVITFVAVGILQWLLGTSLSSPEGFEQASAIMGSFFVKFIMWGILTALAYHVVVGIR  90
SdhC[Salmonella    RVSGVITFIAVGILLWLLGTSLSSPEGFQQAADIMDGFIVKFIMWGILTALAYHVIVGIR  90
SdhC[Erwinia       RVSGVITFVAVGILLWLLGTSLSSEEGFLRAAEIMDSFIVKFIVWGILVALAYHIVGGIR  90
SdhC[Yersinia      RVSGVITFVAVGILLWLLGLSVVSSQEGFMQAAAIMNSFFVKFIFWGILTALAYHICGGIR  97
SdhC[Sodalis       RVSGVIMFVAIGILLWVLGLSLSSPEGFIQAAAMMDSFIVKFIFWGILTALAYHTLCGIR  90
1NEK_C[E.coli]     RVSGVITFVAVGILLWLLGTSLSSPEGFEQASAIMGSFFVKFIMWGILTALAYHVVVGIR  90
KPN_00728          RVFGVITFMAVGILLWLLGTSLSSPEGFLTAASIMDSFFVKFIMWGILTALAYHVVVGIR  90
                   ** *** *:*:*** *:** *:** *** *: :*..*.****.****.*****   *:*

SdhC[Enterobacter  HMLMDFGYLEETFEAGKRSANISFVITVVLSLLAGVLVW  159
SdhC[Shigella      HMMMDFGYLEETFEAGKRSAKISFVITVVLSLLAGVLVW  129
SdhC[Salmonella    HMLMDFGYLEETFEAGQRSAKISFVITVVLSLLAGVLVW  129
SdhC[Erwinia       HLLMDFGYIEEDLAAGKRSANLSFIITVVLSILAGVLVW  129
SdhC[Yersinia      HLLMDFGYIEESLAAGTRSAQVAMVLTLVLSVLAGVLVW  136
SdhC[Sodalis       HMLMDFGYLEETLAMGKTTANVTFLLTVVLSILAGVLV-  128
1NEK_C[E.coli]     HMMMDFGYLEETFEAGKRSAKISFVITVVLSLLAGVLVW  129
KPN_00728          HLMMDFGYLDETLEAGKRSAKISFVITVVLSLLAGVLVW  129
                   *::*****::*  :  *  :*::::::*:***:******

SdhD[Enterobacter  MVSNASALGRNGVHDFILVRATAIVLTLYIIYMIGFFATSGTLTWEIWSGFFGSAFTKVF  60
SdhD[Shigella      MVSNASALGRNGVHDFILVRATAIVLTLYIIYMVGFFATSGELTYEVWIGFFASAFTKVF  60
SdhD[Salmonella    MVSNASALGRNGVHDFILVRATAIVLTLYIIYMVGFFATSGELTFEAWTGFFSSAFTKVF  60
SdhD[Erwinia       MVSNASALGRNGVHDWLLIRASAIVIVLYVIYLIGFIATAGDITYEIWRGFFAMALTKVF  60
SdhD[Yersinia      MVSNASALGRNGVHDWLLLRASAIVITLYVLYILGFVVIVPDITYEIWRGFFASHITKVF  60
SdhD[Sodalis       MISNASALGRNGVHEWLLVRASAIVICLYIIYLLGFILLADTLTYDVWRGFFATPMTKVF  60
1NEK_D[E.          MVSNASALGRNGVHDFILVRATAIVLTLYIIYMVGFFATSGELTYEVWIGFFASAFTKVF  60
KPN_00729[Klebsiella  MVSNASALGRNGVHDFILVRATAIVLTLYIIYMVGFFATTGEISWEVWTGFFASGFTKVF  60
                   *:************:::*:**:***: **::*::**.    :::: * ***. :****

SdhD[Enterobacter  TLLALFSILIHAWIGMWQVLTDYVKPLAIRLPLQLAIVVALVVYVIYGFVVVWGV  115
SdhD[Shigella      TLLALFSILIHAWIGMWQVLTDYVKPLALRLMLQLVIVVALVVYVIYGFVVVWGV  115
SdhD[Salmonella    TLLALFSILIHAWIGMWQVLTDYVKPLAVRLILQLVIVVALVVYVIYGFVVVWGV  115
SdhD[Erwinia       TLLTLFSILVHAWIGMWQVLTDYIKPLALRLTLQLAIVIALLVYVIYGTIVVWGA  115
SdhD[Yersinia      TLLTLLSILAHAWIGLWQVLTDYIKPLALRLVLQLAVVITLLVYLLYGTIVVWGA  115
SdhD[Sodalis       TLLTLFSILVHTWIGMWQVLTDYVKPLALRLILQLTIVIVLMVYLLYGTVVVWGA  115
1NEK_D[E.          TLLALFSILIHAWIGMWQVLTDYVKPLALRLMLQLVIVVALVVYVIYGFVVVWGV  115
KPN_00729[Klebsiella  TLLALVSILIHAWIGMWQVLTDYVKPLALRLVLQLAIVVALVAYLYGFVVVWGV  115
                   ***:*.*** *:***:*******:****:**** ***.:*:.*:.*::** :****.
```

PROCHECK Ramachandran plot [17] was used to check the stereochemical quality of the built model. PROCHECK result indicated that more than 97% of the residues have phi and psi angles falling in the most favored regions (Table 5). The overall *G*-factor quality was 0.2, indicating a good quality model (*G*-factor values between 0 to 0.5 are acceptable). The validity of the built model was further confirmed by using both PROCHECK and DOPE. DOPE energy score was comparable to that of the template (−32556.3 and −32339.1 for model and template, respectively).

3.5 Docking of Ubiquinone

In general, Succinate dehydrogenase Chain A (SdhA) catalyzes oxidation of succinate to fumarate. The catalytic power of the enzyme gives rise to the proposals of ideas generating from transition state theory (TST), nuclear quantum mechanical effects (NQM) as discussed by Olsson et al. [28]. These quantum studies have led to the

understanding of kinetic isotope effect (KIE) using quantum mechanical methods as showed in Mavri et. al. and Meyer et. al. [22, 24], where their studies demonstrated interesting findings on the hydrogen transfer process in soybean lipoxygenase-1 (SLO-1). Although the catalytic activity with its isotope effect may apply to SDH, this and its rate constant are not studied here as it is out of the scope of the study.

Succinate dehydrogenase chain A contains a flavin adenine dinucleotide (FAD) cofactor that is covalently linked to a conserved His. Subsequently, FAD is reduced to $FADH_2$ by losing two electrons in a process. Electrons from SdhA are transferred to SdhB via the iron sulfur cluster. These electrons are then transferred to ubiquinone which is bound to SdhC and SdhD, reducing it to ubiquinol (QH_2) (Fig. 1).There is a heme group in place between His residue from SdhC and Cys residue from SdhD each for *Saccharomyces cerevesiae* [29]. Mutation of His46 and His113 residues in SdhC demonstrated reduction of
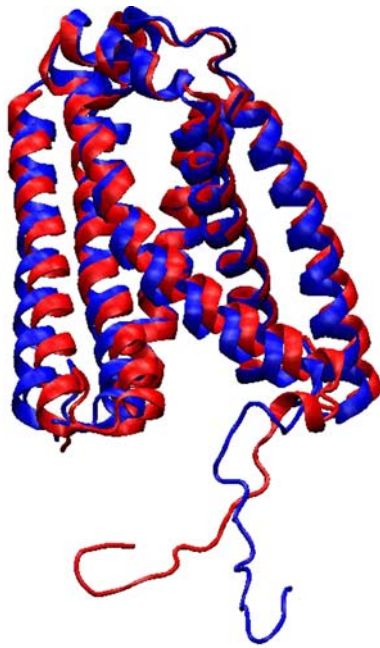
Fig. 5 The major deviation of both template and built model structures (1 NEK chain C and D in *red*, built model in *blue*) is at the N-terminal region (1–21 residues) of the postulated chain C which is located at the cytoplasm. The deviation at the helix bundle is not significant as shown above

ubiquinol formation but the mechanism is yet to be resolved [30]. The present study showed that the SdhC and SdhD of Succinate dehydrogenase bind with a heme group and provide a binding site for ubiquinone. In *E. coli*, ubiquinone binding site in Succinate dehydrogenase namely Q-site is known to be mediated solely by hydrogen bonding between O1 carbonyl group of quinine (Fig. 6) and the side chain of conserved tyrosine residue at the Chain D [10]. It is also suggested by Iwata and co-workers [10] that this

tyrosine residue forms an additional hydrogen bond with Arg31 residue in Chain C. In addition, Ser27 in Chain C of Succinate dehydrogenase from *E. coli* is located at a position where interaction with O3 of ubiquinone might occur. This is also consistent with the conservation of Ser27 residues in Succinate dehydrogenase in all other organisms [10] as shown in the multiple sequence alignment (Fig. 4).

To date, all Succinate dehydrogenases identified contain at least one heme group and ubiquinone reduction site [30]. There are also two histidine residues, His84 and His71 in the Chain C and D of the enzyme involved in heme binding [30, 42]. As shown in the result of multiple sequence alignment, a total of three His residues in KPN00728 and 1 in KPN00729 were found to be highly conserved among other species of Enterobacteriaceae (Fig. 4). In this study, the heme group that was docked onto the built model was found to have the same conformation arrangement as the one observed in the experimental data [8, 39]. Based on these observations, it was found that the His84 residue in Chain C and His71 residue in Chain D indeed played a role in heme axial ligand binding similar to that observed with the previous experiments (Fig. 7).

It is known that Succinate dehydrogenase in *E. coli* carries a ubiquinone by forming a direct hydrogen bond with O_H Tyr83 (Chain D). Previous reports showed that mutation of Ser27, Arg31 from Chain C and Tyr83 from Chain D of Succinate dehydrogenase of *E. coli* had shown a drastic defect in the conversion of ubiquinone to ubiquinol and a reduction in Succinate dehydrogenase physiological activities [10, 39, 42, 43]. Based on these observations, molecular docking simulation of ubiquinone at sites covering these neighbouring residues using different grid centres was performed to further ascertain that the built model has its function as a Succinate dehydrogenase.

Table 4 Topology and secondary structure comparison of 1NEK chain C and chain D with KPN00728 and KPN00729

Total of 4 helix regions had been identified among them. Similar arrangement of secondary structure topology is also shown between the 1 NEK (template) and KPN00728 and KPN00729

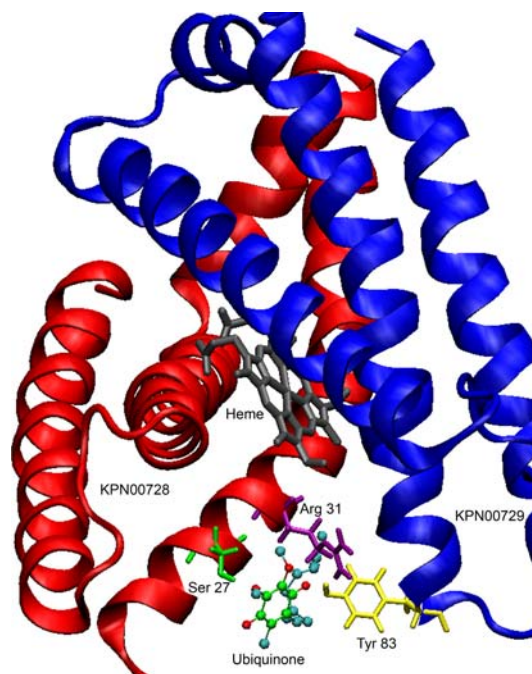| Secondary structure	Residue number Start–End			
	1NEK chain C	KPN00728	1NEK chain D	KPN00729
Helix				
1	22–53	22–53	12–38	10–38
2	55–65	55–66	45–53	43–51
3	68–96	68–96	55–83	53–81
4	103–128	103–128	88–114	85–112
Topology				
Cytoplasmic loop	1–31	1–31	1–20	1–20
Transmembrane	32–52	32–52	21–41	21–41
External loop	53–61	53–68	42–58	42–58
Transmembrane	62–82	69–89	59–79	59–79
Cytoplasmic loop	83–108	90–108	80–94	80–94
Transmembrane	109–129	109–129	95–115	95–115

Table 5 Comparison of Ramachandran scores for built model and template (1NEK chain C and from *E. coli*)

Ramachandran plot (%)	Structure	
	Built model (*K. pneumoniae* MGH78578)	1NEK chain C and D (*E. coli*)
Core	97.7	93.9
Allowed	2.3	5.6
Generous	0.0	0.5
Disallowed	0.0	0.0

A total of 97.7 and 2.3% residues from the built model fall within the core and allowed regions. Validation of built model using Ramachandran plot had demonstrated better quality as compared to the 1NEK from *E. coli*
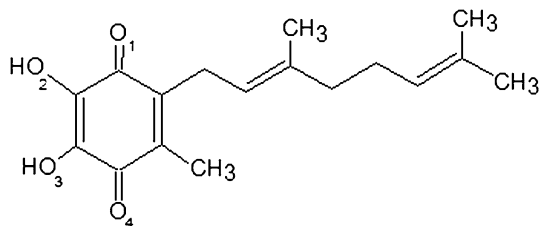
Fig. 6 Structure of ubiquinone. Ubiquinone is labeled with the *numbering* of the *oxygen atom*
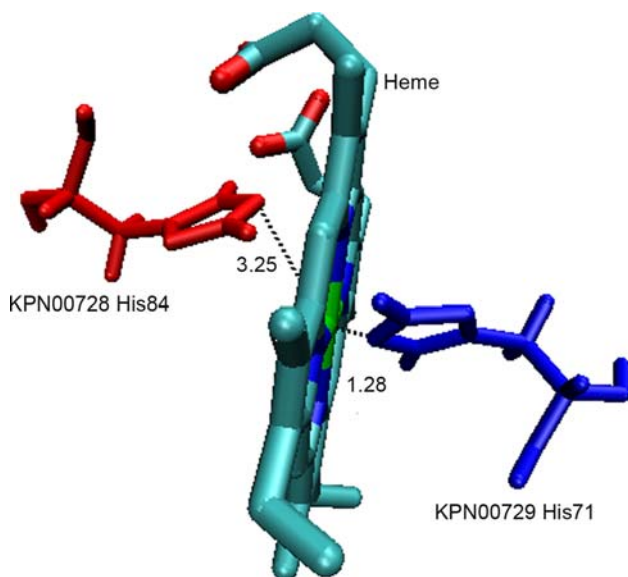
Fig. 7 Heme group was sandwiched between His84 (KPN00728) and His71 (KPN00729). Similar orientation of heme group is observed in the built model as the distance between His84 and His71 is 3.25 and 1.28 Å, respectively. The schematics licorice representation of the model is generated using VMD 1.8.5 [11]

Fig. 8 Ubiquinone docking simulation with built model. Built model: KPN00728 (*red*) and KPN00729 (*blue*) in secondary structure presentation with heme group (*grey*). Ser27 (*green*) and Arg31 (*purple*) from KPN00728 and Tyr83 (*yellow*) from KPN00729 which are located at the ubiquinone (*ball and stick* representation) binding site is shown above

Docking simulation showed that the most possible ubiquinone binding site was located at O_H of Tyr83 in KPN00729 (Fig. 8). Ubiquinone binds at the location where the distance of O1 ubiquinone is 2.58 Å away from the O_H of Tyr83 in KPN00729 (Fig. 9). This resulted in a bond angle of 124.5° between O_H of Tyr83 and O1 of ubiquinone (Fig. 9) which are in agreement with previous experimental data [10, 39, 43]. On the basis of these distances and angles, a hydrogen bond exists between O1 of ubiquinone and O_H of Tyr83 in which case the latter acts as a hydroxyl group donor while the former acts as the acceptor. This result strongly suggests that KPN00729 might potentially interact with ubiquinone by forming a possible hydrogen bond with the side chain of Tyr83 residue that acted as one of the interacting residues to facilitate ubiquinone binding, which correlated well with ubiquinone binding of Succinate dehydrogenase from *E. coli* [39]. The docking result demonstrated that KPN00729 had preserved the functionality of ubiquinone binding, thus confirming it to be Chain D of Succinate dehydrogenase.
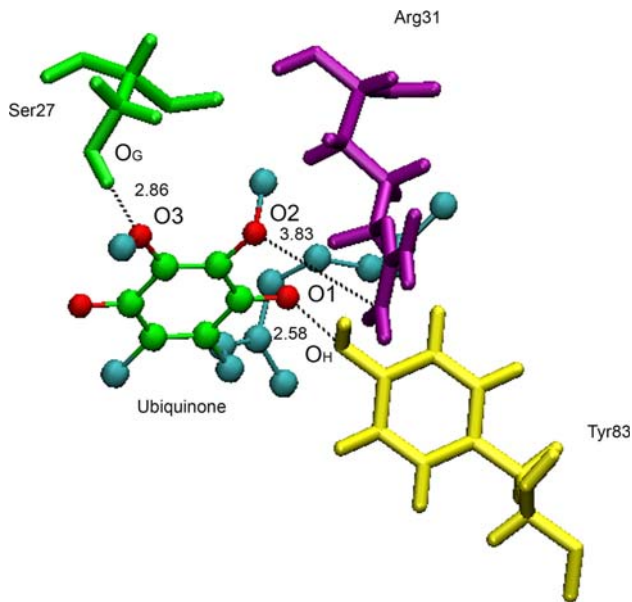
Fig. 9 Close up snapshot of Ser27 and Arg31 from KPN00728 and Tyr 83 from KPN00729 with ubiquinone. The distance between UQO1 (*ball and stick* representation) and OH of Tyr83 (*yellow*) is 2.58 Å which is in agreement of hydrogen bond formation. Both Ser27 (*green*) and Arg31 (*purple*) lie closely with ubiquinone with less than 4 Å which might also contribute as a major structural component of the binding site of ubiquinone

Apart from Tyr83, Ser27 of Chain C was also previously suggested to play an important role in ubiquinone binding and reduction process [42]. Mutation of this residue inflicts the cell growth in succinate and Succinate dehydrogenase prepared from these mutants cell showed low Succinate dehydrogenase activity and no sign of incorporation of ubiquinone at the mutated residue. Their result indicated that both hydroxyl group of Ser side chain are critical in ubiquinone binding [42]. This is supported by [39] that mutation of Ser27 residues in *E. coli* had diminished the reduction activity towards ubiquinone. Our results showed that O3 of ubiquinone was positioned at 2.86 Å from O_G of Ser27 KPN00728. This distance is adequate for a potential hydrogen bond to be formed. It had been reported by [29] that ligation of Ser27 with O3 of ubiquinone increase the stability of semiubiquinone intermediate generated during catalytic cycle based on the theoretical model generated from 1NEK Succinate dehydrogenase X-ray structure. The position of O3 ubiquinone with O_G of Ser27 KPN00728 had demonstrated the potential as the hydrogen bonding partner and it might adopt similar characteristic as mentioned by Oyedotun and Lemire [29]. Moreover, the multiple sequence alignment result had shown that Ser27 residue in KPN00728 is strictly conserved throughout all species of Enterobacteriaceae (Fig. 4). Based on these results, we postulated that Ser27 from KPN00728 in our built model is indeed an important residue that might serve

in forming hydrogen bond with ubiquinone similar to the Ser27 residue of Chain C of *E. coli* Succinate dehydrogenase [29, 39, 42].

In addition to the above two residues, the distance of O2 ubiquinone with N_{H1} of Arg31 from KPN00728 is 3.83 Å. This value is in proximity with the previous 3.1 Å value reported by Horsefield et al. [10]. According to [39] Arg31 from Chain C of *E. coli* Succinate dehydrogenase is a major structural component of ubiquinone binding site as it lies equidistant between the heme group and ubiquinone. In our built structure, similar arrangement of Arg31 of KPN00728 was observed where it was sandwiched between the heme group and ubiquinone (Fig. 9).

4 Further Discussion

Prior to July 2008, KPN00729 was still classified as a hypothetical protein along with 1,043 other proteins in *K. pneumoniae*. Interestingly, the current revised genome map of this organism has provisionally identified this protein as SdhD and the number of hypothetical proteins reduced from 1,044 (in December 2007) to 1,003 (Revision date 1 May 2009). Therefore, the genome map has now SdhA, SdhB and SdhD. It is known that the protein Succinate dehydrogenase is composed of four catalytic chains namely A, B, C and D. Albeit, all the four chains are needed to function as Succinate dehydrogenase. This poses a question as to where the Chain C of the enzyme is. Initially when the sequence of KPN00728 and KPN00729 were analyzed using BLAST search, potential templates (1NEK, 2ACZ and 1NEN) with ∼90% sequence identity were obtained. This leads to another question as to why sequences with more than 90% sequence identity were classified as hypothetical proteins in the complete genome map of *Klebsiella* sp. while it should be functionally classified. Based on this, we revisited the genome map and we discovered that the complete genome of *Klebsiella* sp. already consists of three genes encoding Succinate dehydrogenase Chain A, B and D. KPN00728 and KPN00729 (postulated to be Chain C in this study and D as provisionally assigned function) are located prior to the genes encoded for Chain A and B in the genome map. This again, led to our postulation that these two proteins might actually be Chain C and D of Succinate dehydrogenase.

During BLAST search for KPN00728, there were 38 residues of amino acids missing in the beginning of the sequence when aligned to the templates: 1NEK, 2ACZ and 1NEN (which are Succinate dehydrogenase enzymes present in *E. coli*). Previous studies [39, 42, 43] showed that this missing region (38 residues) contributed to the functionality of Succinate dehydrogenase. For this reason, we reanalyzed KPN00728 to look for the missing regions

in the genome map. Reverse translation on KPN00728 nucleotide sequences with a total of 114 nucleotides at the start of the gene which can translate into 38 residues of amino acid was carried out. The translated 38 residues were found to be unsurprisingly almost identical to the residues 1-38 of 1NEK with 92% sequence identity. Together with the missing region and the original sequence of KPN00728, BLAST search was performed again and the sequence identity is ∼90%. Although there is no improvement in terms of sequence identity, from the multiple sequence alignment result it showed that the missing region is highly conserved among other microorganisms. Furthermore, residues that are essential for the functionality as Succinate dehydrogenase such as Ser27 and Arg31 are found within this region (Fig. 3b). Thus, this further convinces us that KPN00728 might be the missing Chain C of the enzyme in question.

From our understanding, Chain C and D of Succinate dehydrogenase in general is anchored into the inner membrane of mitochondria as transmembrane region of this protein. In addition to this, in order for the Chains to be in the transmembrane region, it must require a polypeptide chain which can traverse into the membrane bilayer. This portion of the protein that is embedded in the bilayer must therefore have residues that are hydrophobic or not polar [41]. Commonly, these residues form a coil, or helix, that is hydrophobic and therefore be stable within the bilayer [2, 5]. By analyzing our built homology model, besides the transmembrane topology and secondary structure which is consistent to the structure of 1NEK, we also found that a total of 80% of the polypeptide sequences of KPN00728 and KPN00729 formed helices. A bundle of eight helices made up from four helices in KPN00728 and KPN00729, respectively are found (Table 3). The length of the secondary structure (helices bundles) is approximately 40 Å. This allow the structure to integrate into the membrane bilayer, which in general is within a thickness of 30 Å [41].

In addition to this, we observed significant presence of amino acid residues such as Val and Leu in the model, situated very close to the transmembrane region similar to the observation reported elsewhere [2]. In terms of hydrophobicity, there is more than 50 and 40% of amino acid residues in both KPN00728 and KPN00729, respectively that are hydrophobic. This is in agreement to the general rules of the transmembrane protein structure, where multiple helices with hydrophobic characteristic on the outer side are essential for the chain to anchor on the membrane as well as to maintain its stability [4].

Moreover, sequence analysis showed the presence of conserved residues such as Ser and Arg from Chain C and Tyr from Chain D of Succinate dehydrogenase are involved in the binding of ubiquinone from other microorganisms (E. coli and S. cerevisae). They are also found to

be located close to each other in our model. Both His residues from KPN00728 and KPN00729 were found to arrange themselves in almost axial position enabling the Heme group to sit comfortably between them. Furthermore from our molecular docking result, the formation of hydrogen bonds between ubiquinone with both proteins support our postulation of KPN00728 as the chain C and further proved that KPN00729 is in fact Chain D of Succinate dehydrogenase in Klebsiella pneumoniae MGH 78578.

In addition, they have high sequence identity with Succinate dehydrogenase from other organisms. From the genome analysis, we managed to find the conserved residues within the missing region which is critical for ubiquinone binding. The transmembrane analysis of the developed homology model showed an agreement with the secondary structure profile of the Chains C and D of the enzyme certainly convince us that both proteins are indeed part of Succinate dehydrogenase. All in all, the missing genomic region of KPN00728 is possibly the most important reason why this protein is still classified as hypothetical protein (KPN00729 is reclassified provisionally as SdhD on July 2008). Inclusion of this region in the protein, supported by all the sequence analysis and molecular modeling results, has yielded conclusive evidence that it is in effect Chain C of Succinate dehydrogenase.

5 Conclusions

In this work, a combination of genome analysis, protein sequence analysis, structural modeling and molecular docking simulation approaches were employed to provide an understanding of the possible functions and characteristics of hypothetical proteins with unknown structure and biochemical function. In this present study, we have discovered that both KPN00728 shares similarity in terms of functions and characteristic to Succinate dehydrogenase of E. coli. Ser27 and Arg31 from KPN00728 which are highly conserved within this region had demonstrated an important role in binding of ubiquinone in Succinate dehydrogenase. Formations of hydrogen bonds between ubiquinone with Ser27, Arg31 and Tyr84 from KPN00728 and KPN00729 further implied that these two proteins had the functionality of ubiquinone binding, thus increasing the possibility of them being Chain C and D of Succinate dehydrogenase. The work presented above thus answer the question as to where the missing Chain C of Succinate dehydrogenase is and the analysis have provided an answer beyond doubt that KPN00728 is the missing Chain C of Sdh. Succinate dehydrogenase is very important in all living things and in prokaryote they consist of four chains

or subunits to function in the Krebs cycle. It is hoped that this work will serve as a stimulant for further structure to function characterization of hypothetical proteins.

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) J Mol Biol 215:403–410
2. Arce J, Sturgis JN, Duneau JP (2009) Biopolymers (in press)
3. Berman H, Henrick K, Nakamura H, Markley JL (2007) Nucleic Acids Res 35:D301–D303
4. Booth PJ, Templer RH, Meijberg W, Allen SJ, Curran AR, Lorch M (2001) Crit Rev Biochem Mol Biol 36:501–603
5. Brändén C-I, Tooze J (1991) Introduction to protein structure. Garland Publishing, New York, p 302
6. Carpenter JL (1990) Rev Infect Dis 12:672–682
7. Case DA, Darden TA, Cheatham TEI, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA (2004) Amber 8
8. De Baere I, Perutz MF, Kiger L, Marden MC, Poyart C (1994) Proc Natl Acad Sci 91:1594–1597
9. Heinig M, Frishman D (2004) Nucleic Acids Res 32:W500–W502
10. Horsefield R, Yankovskaya V, Sexton G, Whittingham W, Shiomi K, Omura S, Byrne B, Cecchini G, Iwata S (2006) J Biol Chem 281:7309–7316
11. Humphrey W, Dalke A, Schulten K (1996) J Mol Graphics Modell 14: 33–38, 27–38
12. Jones DT (1999) J Mol Biol 287:797–815
13. Jong GM, Hsiue TR, Chen CR, Chang HY, Chen CW (1995) Chest 107:214–217
14. Kawai T (2006) Clin Infect Dis 42:1359–1361
15. Ko WC, Paterson DL, Sagnimeni AJ, Hansen DS, Von Gottberg A, Mohapatra S, Casellas JM, Goossens H, Mulazimoglu L, Trenholme G, Klugman KP, McCormack JG, Yu VL (2002) Emerg Infect Dis 8:160–166
16. Lancaster CRD (2002) Biochim Biophys Acta 1553:1–6
17. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) J Appl Crystallogr 26:283–291
18. Lee KH, Hui KP, Tan WC, Lim TK (1994) J Hosp Infect 27:299–305
19. Lubec G, Afjehi-Sadat L, Yang JW, John JPP (2005) Prog Neurobiol 77:90–127
20. Maklashina E, Cecchini G (1999) Arch Biochem Biophys 369:223–232
21. Marra AR, Wey SB, Castelo A, Gales AC, Cal RG, Filho JR, Edmond MB, Pereira CA (2006) BMC Infect Dis 6:24
22. Mavri J, Liu HB, Olsson MHM, Warshel A (2008) J Phys Chem B 112:5950–5954
23. McGuffin LJ, Jones DT (2003) Bioinformatics 19:874–881
24. Meyer MP, Tomchick DR, Klinman JP (2008) Proc Natl Acad Sci 105:1146–1151
25. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) J Comput Chem 19:1639–1662
26. NCBI (2007) NCBI Entrez genome—Klebsiella pneumoniae subsp. pneumoniae MGH 78578, complete genome, http://www.ncbi/nlm.nih.gov/sites/entrez?Db=genome&Cmd=ShowDetailView&TermToSearch=21205. Accessed Dec 2007
27. Obiamiwe U, Leonard B (2008) Klebsiella infections, http://www.emedicine.com/med/TOPIC1237.HTM. Accessed 12 Jun 2008
28. Olsson MHM, Mavri J, Warshel A (2006) Philos Trans R Soc Lond B 361:1417–1432
29. Oyedotun KS, Lemire BD (2004) J Biol Chem 279:9424–9431
30. Oyedotun KS, Yau PF, Lemire BD (2004) J Biol Chem 279:9432–9439
31. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, Debolt S, Ferguson D, Seibel G, Kollman P (1995) Comput Phys Commun 91:1–41
32. Rustin P, Munnich A, Rotig A (2002) Eur J Hum Genet 10:289–291
33. Sali A, Blundell TL (1993) J Mol Biol 234:779–815
34. Solis FJ, Wets RJB (1981) Math Oper Res 6:19–30
35. Tang LM, Chen ST, Hsu WC, Chen CM (1997) Epidemiol Infect 119:135–142
36. Tang Q, Kalsbeck WA, Olson JS, Bocian DF (1998) Biochemistry 37:7047–7056
37. Thompson JD, Higgins DG, Gibson TJ (1994) Nucleic Acids Res 22:4673–4680
38. Thomson ABR, Shaffer EA (1994) First principles of gastroenterology: the basis of disease and an approach to management, Canadian Association of Gastroenterology/Astra Pharma, pp 587
39. Tran QM, Rothery RA, Maklashina E, Cecchini G, Weiner JH (2006) J Biol Chem 281:32310–32317
40. Vonheijne G (1992) J Mol Biol 225:487–494
41. White SH, Wimley WC (1999) Annu Rev Biophys Biomol Struct 28:319–365
42. Yang XD, Yu L, He DY, Yu CA (1998) J Biol Chem 273:31916–31923
43. Yankovskaya V, Horsefield R, Tornroth S, Luna-Chavez C, Miyoshi H, Leger C, Byrne B, Cecchini G, Iwata S (2003) Science 299:700–704