Electronic Journal of Biotechnology ISSN: 0717-3458
Vol.6 No.1, Issue of April 15, 2003

Received June 26, 2002 / Accepted March 25, 2003
RESEARCH ARTICLE

A predicted structure of the cytochrome *c* oxidase from *Burkholderia pseudomallei*

Mohd. Firdaus Mohd. Raih
Centre for Gene Analysis and Technology
School of BioSciences and Biotechnology
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Malaysia
Tel: +6 03 89267446
Fax: +6 03 89267972
E-mail: mfirr@cgat.ukm.my

Ahmad Tarmidi Sailan
Centre for Gene Analysis and Technology
School of BioSciences and Biotechnology
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Malaysia
Tel: +6 03 89267446
Fax: +6 03 89267972
E-mail: tarmidi@medic.ukm.my

Zulkeflie Zamrod
Centre for Gene Analysis and Technology
School of BioSciences and Biotechnology
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Malaysia
Tel: +6 03 89267446
Fax: +6 03 89267972
E-mail: zza@pkrisc.cc.ukm.my

Mohd. Noor Embi
Centre for Gene Analysis and Technology
School of BioSciences and Biotechnology
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Malaysia
Tel: +6 03 89267446
Fax: +6 03 89267972
E-mail: noormb@pkrisc.cc.ukm.my

Rahmah Mohamed*
Centre for Gene Analysis and Technology
School of BioSciences and Biotechnology
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Malaysia
Tel: +6 03 89267446
Fax: +6 03 89267972
E-mail: ram@cgat.ukm.my

http://cgat.ukm.my

Financial support: This work was funded by the Intensification of Research in Priority Areas (IRPA) grants IRPA 01-02-02-001 and IRPA-TOPDOWN 09-02-02-T001 provided by the Ministry of Science, Technology and the Environment, Malaysia.

Keywords: *Burkholderia pseudomallei*, cytochrome *c* oxidase, protein structure prediction, sequence alignments, structure-function extrapolation.

Abbreviations: Cox1: *B. pseudomallei* predicted cytochrome *c* oxidase subunit 1 protein; *Bp cox1*: *B. pseudomallei* predicted cytocrome *c* oxidase subunit 1 gene; SU: subunit; ORF: open reading frame; BLAST: Basic Local Alignment Search Tool; TM: transmembrane; TMH: transmembrane helices; CVFF: consistent valence force field.

*Corresponding author

Cytochrome *c* oxidase, the terminal enzyme of the respiratory chains of mitochondria and aerobic bacteria, catalyzes electron transfer from cytochrome *c* to molecular oxygen. The enzyme belongs to the haem-copper-containing oxidases superfamily. A recombinant plasmid carrying a 2.0 kb insert from a *Burkholderia pseudomallei* genomic library was subjected to automated DNA sequencing utilizing a primer walking strategy. Analysis of the 2002 bp insert revealed a 1536 bp open reading frame predicted to encode a putative cytochrome *c* oxidase. Further analysis using sequence alignments and tertiary structure analysis tools demonstrated that the hypothetical *B. pseudomallei* cytochrome *c* oxidase is similar to cytochrome *c* oxidases from other organisms such as *Thermus thermophilus* (36% protein sequence identity), *Paracoccus denitrificans* and bovine heart mitochondrial, the latter two which crystal structures available. The deduced 512 residue protein sequence includes the six canonical histidine residues involved in binding the low spin heme B and the binuclear center Cu_B/hemeA. The predicted tertiary structure of the hypothetical protein is consistent with previous models of electron transfer for cytochrome *c* oxidase.

Cytochrome *c* oxidase, a protein complex located in the inner membrane of mitochondria and many bacteria, is the terminal enzyme of most respiratory chains. It is a respiratory enzyme catalyzing the energy-conserving reduction of molecular oxygen to water. The system catalyzes the final electron transfer steps from cytochrome *c* to molecular oxygen, and is a member of the superfamily of heme-copper containing terminal oxidases. Some bacterial terminal oxidases use quinol as substrates (quinol oxidases). Due to the homology of their sequences, both types of enzymes belong to the superfamily of haem copper oxidases (Ferguson-Miller and Babcock, 1996; Michel et al. 1998). The cytochrome *c* oxidase from the soil bacterium *Paracoccus denitrificans* for instance, consists of the three core subunits - I, II, and III and a small non-conserved subunit IV of unknown function (Witt and Ludwig, 1997). Subunit (SU) I is the largest and best conserved subunit of cytochrome *c* oxidase. Protein SU I of most cytochrome *c* oxidase contains two heme A molecules (heme *a* and heme a_3) and copper B (Cu_B). These enzymes receive electrons either from quinols or from cytochrome *c* to heme a_3. Cu_B forms a binuclear active centre where dioxygen is reduced to water. Four protons are consumed in water formation per oxygen molecule. Electron transfer is coupled to proton translocation (up to four) across the membrane electrogenetically ('pump'), resulting in a proton and charge gradient that is then employed by the F_0F_1-ATPase to synthesize ATP. Members of this family contain one low spin cytochrome and a bimetallic structure consisting of a high spin heme in close proximity to a copper ion. These metals reside in the large subunit of the complex, subunit I, where they are coordinated to at least six conserved histidine residues. The cytochrome *c* oxidase bind two additional copper ions in the Cu_A site to conserved histidine and cysteine residues of subunit II.

Burkholderia pseudomallei is the causative agent of melioidosis, a serious disease of humans and animals that occurs primarily in South East Asia, Northern Australia and other tropical areas (Dance, 2002). This pathogenic bacterium survives in diverse environmental conditions and secretes various extracellular products that have been implicated as factors involved in pathogenesis of this disease. In this paper, we summarize the identification of a cytochrome *c* oxidase gene in *B. pseudomallei* and its tertiary structure determination utilizing available protein structure prediction (comparative modeling) methods based on the predicted protein sequence. The postulated amino acid residues that play the major roles in oxygen reduction, proton and electron transfer pathways were used as reference points to gauge functional probability. The sequence we have identified and its predicted corresponding structure is the first functional annotation for this family of proteins in *B. pseudomallei*.

Materials and Methods

Genomic library preparation. A *Burkholderia pseudomallei* genomic DNA library was prepared in pSV-SPORT1 vector (Gibco, BRL) and transformed in *Escherichia coli* strain DH5a. After screening with a heterologous oligonucleotide probe, a recombinant plasmid containing a 20 kb *Eco*R1 insert was isolated. Plasmid preparations were carried out by standard procedures (Birnboim and Doly, 1979) and then subjected to automated DNA sequencing.

DNA sequencing. Automated DNA sequencing was carried out utilizing *Taq* DyeDeoxy™ Terminator Cycle Sequencing Kit, Amplitaq® DNA Polimerase (Perkin Elmer, USA), FS enzyme and electrophoresed via the ABI PRISM 377 Automated DNA sequencer. The insert was then subjected to a *primer walking* strategy initially utilizing universal primers USP6, UT7 (Gibco, BRL) and followed by four synthetic primer pairs.

The complete sequence of the ORF2 sequence has been deposited in the GenBank Database (http://www.ncbi.nlm.nih.gov/Genbank/) and was assigned the GenBank nucleotide accession number AF087002 (AAF13732 protein accession number).

Gene prediction. Open reading frames were identified with the aid of DNAsis (Hitachi Software Engineering America Ltd.) and GeneMark ver. 2.0 programs (Lukashin and Borodovsky, 1998). Sequence database alignments and comparisons were done with the BLAST family of programs (blastx, blastp) against database specifications of non-redundant protein, SWISS-PROT and PDB which were available from the BLAST website at the National Center for Biotechnology Information webserver,

(http://www.ncbi.nlm.nih.gov/blast/) (Altschul et al. 1997). Multiple sequence alignments were done using ClustalW 1.8 (Thompson et al. 1994). PROSIS was used to calculate the amino acid composition encoded by the ORFs and some predicted properties of the individual proteins.

Protein structure prediction. The *B. pseudomallei* Cox subunit I (AF087002) predicted protein sequence was submitted to several transmembrane prediction programs accessed via their web interfaces. The programs used were DAS (Cserzo et al. 1997; http://www.sbc.su.se/~miklos/DAS/), TOPPRED (Claros and von Heijne, 1994; http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html), TMHMM (Krogh et al. 2001; http://www.cbs.dtu.dk/services/TMHMM-2.0/), Split (Juretic et al. 1999), MEMSAT (Jones et al. 1994; http://www.psipred.net) and SOSUI (Hirokawa et al. 1998; http://sosui.proteome.bio.tuat.ac.jp/cgi-bin/sosui.cgi?/sosui_submit.html). A survey of evaluated TM regions prediction program (Moller et al. 2001) was used as a reference to gauge accuracy and reliability of the TM prediction results.

The 512 residue amino acid sequence was submitted for further automated prediction of secondary structures and protein fold. The programs chosen were based on the results of CAFASP 2 (Fischer et al. 2001; http://www.cs.bgu.ac.il/~dfischer/CAFASP2/) for fully automated protein structure prediction. The following programs were used via their web browser interfaces: GenThreader (Jones, 1999; http://www.psipred.net), PHD (Rost et al. 1994; http://cubic.bioc.columbia.edu/predictprotein/), bioinbgu (Fischer, 2000; http://www.cs.bgu.ac.il/~bioinbgu/) and 3D-PSSM (Kelley et al. 2000; http://www.sbg.bio.ic.ac.uk/~3dpssm/). These servers were accessed via a web directory interface of online protein structure prediction resources – SPORes: Structure Prediction with Online Resources website (http://cgat.ukm.my/spores/).

Structural topology was built from the secondary structure prediction data while template selection was done using the fold prediction data. The target sequence was then aligned to the template sequence using the Homology module of InsightII (MSI, ver. 98). Manual editing for optimal alignments were done where deemed necessary. The alignment was translated into tertiary structure using Modeller Release 6 (Sali and Blundell, 1993). The initial strain of the predicted structure was relieved by carrying out energy minimizations using Discover (MSI ver. 98) utilizing the CVFF force field. Short contacts were removed by manually rotating the side chains. Model evaluation was done using Procheck (Laskowski et al. 1993), Errat (Colovos and Yeates, 1993), What If (Vriend, 1990) and Verify3D (Luthy et al. 1992). Refinements of the structure and geometry optimizations, when necessary, were carried out using the Insight interface (MSI, ver. 98).

Results and Discussion

Sequence analysis. Complete sequencing of the 2.0 kb DNA insert from the recombinant clone showed that it contained a 1536 bp ORF designated *Bp cox1* potentially encoding a subunit 1 cytochrome *c* oxidase termed Cox.

A thorough search of available sequence databases showed that the putative *B. pseudomallei* Cox (AF087002) sequence is homologous to other known cytochrome *c* oxidases with varying levels of sequence identity and appears to have structural features similar to the largest subunit of the heme/copper-requiring cytochrome *c* and quinol oxidases (Figure 1 and Figure 2). BLAST queries of the predicted *B. pseudomallei* Cox protein sequence yielded eight other sequences with E values lower than 10^{-50}. All of these sequences were identified as being cytochrome *c* oxidases (6 sequences) or hypothetical cytochrome *c* oxidases (2 sequences) (Figure 1). Alignments of the predicted protein sequence against the PDB database also showed significant homology with the sequence of a solved crystal structure for cytochrome *c* oxidase (PDB identification = 1EHK) from *Thermus thermophilus* (Soulimane et al. 2000). Multiple sequence alignments on these sequences (*P. denitrificans*, *T. thermophilus* and bovine heart mitochondria aligned to the *B. pseudomallei*) showed several regions of high conservation despite some sequences being phylogenetically distant (Figure 1 and Figure 2). Several of these highly conserved regions were identified as crucial residues in other proposed models for electron transfer in cytochrome *c* oxidases.

The *Thermus thermophilus* sequence exhibited 36% identity to the *B. pseudomallei* AF087002 sequence. Figure 1 illustrates further the observed sequence conservation by showing the alignments of the amino acid sequences of the *B. pseudomallei* cytochrome *c* oxidase towards the other sequences chosen from the GenBank (non-redundant protein databases, SWISSPROT, PDB) BLAST searches. The *B. pseudomallei* subunit is shorter at both termini than the subunit from these cytochrome *c* oxidases, the degree of conservation however remains clear, as shown by the number residues that are identical across the sequences aligned. For clarity, the sequence alignment in Figure 2 shows only segments that include the six histidines present in every representative of these enzymes. The six histidine residues and other conserved amino acids are placed in a similar pattern along the putative membrane spanning hydrophobic segments.

Protein structure prediction. Multiple methods of transmembrane prediction were used to enable a consensus confirmation of predicted transmembrane helices via differing approaches (data not shown). The TM region search for *B. pseudomallei* Cox subunit I revealed 12 possible TM helices (Figure 2). The objective of this TM search step was to identify and map out regions of transmembrane helices to confirm our sequence database based hypothesis of the putative sequence being identified

as a cytochrome c oxidase subunit I. Cythochrome *c* oxidases are known to have these transmembrane regions and the identification of these regions served as a confirmatory step in gauging the validity of the predicted structure. The observation from the transmembrane prediction step is consistent with the subunit I of most other haem copper oxidases and acts as confirmatory data for correctness of the protein fold generated by the Modeller program. The fold prediction methods used, proposed the crystal structure of a ba3-cytochrome *c* oxidase (PDB identification 1ehk) from *Thermus thermophilus* as a suitable template. The crystal structure for 1ehk (Soulimane et al. 2000) was solved a relatively low resolution of 2.4Å, was however still selected as the template structure. The target sequence showed a sequence identity of 36% towards the template sequence (Figure 3). The 1ehk structure, has an unusual property for proteins in the oxidase superfamily as its subunit I contained a 13th TM helix instead of just the usual 12 TM segments.

The predicted structure was found to have an RMS fit of 3.1 Å to the template structure. The Ramachandran plot from the Procheck validation revealed five residues in the disallowed regions of the plot while 80.7% of the residues were in the most favoured regions of the plot and the remainder residues in the additional and generously allowed regions of the plot (Figure 3a). A check of the 1ehk template structure revealed two residues which violated the Ramachandran region. Validation by the Errat and Verify 3D programs showed that the predicted structure had a generally acceptable three dimensional profile (Figure 3a; Figure 3b and Table 2). Evaluation methods used generally agree on the correct threading of the backbone. Comparisons of the initial Errat and Verify-3D results with those conducted after refinement and subsequent geometry optimisations showed marked improvements (results not shown). Regions of bad geometry were found to be located mainly in regions with unaligned target-template sequences and structurally variable regions (Figure 1 and Figure 2).

Despite the phylogenetically distant relationship of the target-template sequences, the sequence structure alignment yielded sufficient information of structurally conserved regions to enable a functionally probable model to be generated. The predicted protein structure for *B. pseudomallei* Cox consists of 12 discernible TM helices (Figure 4). An initial assessment for functional plausibility of the predicted protein fold was gauged from comparisons of the primary protein structure to the predicted tertiary structure. The folding of the functionally crucial residues, such as the haem and Cu ligands, electron transfer pathway residues and proton pathway residues were found to fold closely together in 3D space even though some of these residues were distant to each other in the primary structure. Furthermore, the structural placement of these residues, were found to be similar when compared to the template crystal structure. The overall structure shows a clear hydrophobic core with pores A, B and C, described in Iwata et al. 1995, visible when viewed from the periplasmic side.

Electron transfer. The structure-based sequence alignment of subunit I (Figure 2) between other cytochrome *c* oxidase sequences, the *P. denitrificans* and bovine heart oxidases shows that functionally vital residues, such as heme and Cu ligands or the residues proposed for the electron transfer from Cu_A to the hemes, are conserved. In addition to these residues, a highly conserved motif (VLYTFYPP, located between Val84 and Pro91) can be discerned from the alignment especially to *T. thermophilus* (Figure 3 and Figure 4). His236 in *B. pseudomallei cox1*, was postulated to be one ligand for Cu_B comparable to His326 and His 291 in *P. denitrificans* and bovine heart mitochondrion, respectively. This residue might form an electron transfer pathway from Cu_A directly to Cu_B. The current understanding of the oxygen reduction mechanism at the binuclear centre requires the input of at least one of the electrons via Cu_B (Hill, 1994; Michel et al. 1998). The electron transfer from Cu_A via haem *a*/*b* to heme a_3 at the binuclear centre is well established (Hill, 1994), and the corresponding residues (Arg401, Arg400 and Phe338) are conserved in the *B. pseudomallei cox1* (Table 1 and Figure 5). We propose the above-described pathway via Tyr89, Trp183 and His236 as an additional electron transfer pathway that could be used for electrons that are provided from Cu_B to the catalytic oxygen intermediates.

Proton pathways. Two possible proton transfer pathways have been suggested based on the crystal structure of the *P. denitrificans* enzyme and in agreement with the results of site directed mutagenesis (Garcia-Horsman et al. 1995) *i.e.* K-pathway and D-pathway. The shorter K-pathway, leads to the binuclear centre via the highly conserved residues SU I-Thr 351 and SU I-Tyr280 located in the TMH VI and VIII and the hydroxyl group or the heme a_3 hydroxyethylfarnesyl chain (Michel et al. 1998). Two residues were postulated to have similar functions in *Bp CoxI i.e.* Su I-Thr262 and Thy190 (Table 1). Nevertheless, none of the residues in *Bp CoxI* were found to have counterparts in the longer D-pathway as for *Paracoccus sp*.

Acknowledgments

Computational facilities and resources used were based at the Bioinformatics Laboratory of the National Biotechnology and Bioinformatics Network (NBBnet), National Biotechnology Directorate (BIOTEK), Ministry of Science, Technology and the Environment, Malaysia.

References

ALTSCHUL, S.F.; THOMAS, L.M.; ALEJANDRO, A.S.; JINGHUI Z.; ZHENG Z.; WEBB, M. and DAVID, J.L. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, September 1997, vol. 25, no. 17, p. 3389-3402.

BIRNBOIM, H.C. and DOLY, J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, November 1979, vol. 7, no. 6, p. 1513-1523.

CLAROS, M.G. and VON HEIJNE, G. TopPred II: An improved software for membrane protein structure predictions. *CABIOS*, December 1994, vol. 10, no. 6, p. 685-686.

COLOVOS, C. and YEATES, T.O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science*, September 1993, vol. 2, p. 1511-1519.

CSERZO, M.; WALLIN E.; SIMON, I.; VON HEIJNE, G. and ELOFSSON, A. prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Engineering*, June 1997, vol. 10, no. 6, p. 673-676.

DANCE, D.A.B. Melioidosis. *Current Opinion in Infectious Diseases*, April 2002, vol. 15, no. 2, p. 127-132.

FERGUSON-MILLER, S. and BABCOCK, G.T. Heme/copper terminal oxidases. *Chemical Reviews*, November 1996, vol. 96, no. 7, p. 2889–2907.

FISCHER, D.; ELOFSSON, A.; RYCHLEWSKI, L.; PAZOS, F.; VALENCIA, A.; ROST, B.; ORTIZ, A.R. and DUNBRACK, R.L. Jr. CAFASP2: The second critical assessment of fully automated structure prediction methods. *Proteins*, 2001, vol. 45, Suppl. 5, p. 171-183.

FISCHER, D. Hybrid fold recognition: Combining sequence derived properties with evolutionary information. In: *Pacific Symposium on Biocomputing*, (January 4-9 2000 Hawaii), World Scientific, p. 119-130.

GARCIA-HORSMAN, J.A.; PUUSTINEN, A.; GENNIS, R.B. and WIKSTRÖM, M. Proton transfer in cytochrome ba_3 ubiquinol oxidase of *Escherichia coli*: second-site mutations in subunit I that restore proton pumping in the mutant Asp-135 Asn. *Biochemistry*, April 1995, vol. 34, no. 13, p. 4428-4433.

HILL, B.C. Modelling the sequence of electron transfer reactions in the single turnover of reduced, mammalian cytochrome *c* oxidase with oxygen. *Journal of Biological Chemistry*, January 1994, vol. 28, no. 269, p. 2419-2425.

HIROKAWA, T.; BOON-CHIENG, S. and MITAKU, S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, 1998, vol. 14, no. 4, p. 378-379.

IWATA, S.; OSTERMEIER, C.; LUDWIG, B. and MICHEL, H. Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature*, August 1995, vol. 24, no. 376, p. 660–669.

JONES, D.T. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *Journal of Molecular Biology*, April 1999, vol. 287, no. 4, p. 797-815.

JONES, D.T.; TAYLOR, W.R. and THORNTON, J.M. A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry*, March 1994, vol. 33, no. 10, p. 3038-3049.

JURETIC, D.; JERONCIC, A. and ZUCIC, D. Sequence analysis of membrane proteins with the web server SPLIT. *Croatica Chemica Acta*, 1999, vol. 72, no. 4, p. 975-997.

KELLEY, L.A.; MACCALLUM, R.M. and STERNBERG, M.J.E. Enhanced genome annotation using structural profiles in the Program 3D-PSSM. *Journal of Molecular Biology*, June 2000, vol. 299, no. 2, p. 501-522.

KROGH, A.; LARSSON, B.; VON HEIJNE, G. and SONNHAMMER, E.L.L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, January 2001, vol. 305, no. 3, p. 567-580.

LASKOWSKI, R.A.; MACARTHUR, M.W.; MOSS, D.S. and THORNTON J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 1993, vol. 26, p. 283-291.

LUKASHIN, A. and BORODOVSKY, M. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Research*, February 1998, vol. 26, no. 4, p. 1107-1115.

LUTHY, R.; BOWIE, J.U. and EISENBERG, D. Assessment of protein models with three-dimensional profiles. *Nature*, March 1992, vol. 356, p. 83-85.

MICHEL, H.; BEHR, J.; HARRENGA, A. and KANNT, A. Cytochrome *c* oxidase: structure and spectroscopy. *Annual Reviews Biophysics and Biomolecular Structure*, 1998, vol. 27, p. 329–356.

MOLLER, S.; CRONING, M.D. and APWEILER, R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, July 2001, vol. 17, no. 7, p. 646-653.

ROST, B.; SANDER, C. and SCHNEIDER, R. PHD-an automatic server for protein secondary structure prediction. *CABIOS*, February 1994, vol. 10, no. 1, p. 53-60.

SALI, A. and BLUNDELL, T.L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, December 1993, vol. 234, p. 779-815.

SOULIMANE, T.; BUSE, G.; BOURENKOV, G.P.; BARTUNIK, H.D.; HUBER, R. and THAN, M.E.

and mechanism of the aberrant ba_3-cytochrome *c* oxidase from *Thermus thermophilus*. *European Molecular Biology Organization Journal*, April 2000, vol. 19, no. 8, p. 1766-1776.

THOMPSON, J.D.; HIGGINS, D.G. and GIBSON, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, November 1994, vol. 22, p. 4673-4680.

VRIEND, G. What if: a molecular modelling and drug design program. *Journal of Molecular Graphics*, March 1990, vol. 8, p. 52-56.

WITT, H. and LUDWIG, B. Isolation, analysis, and deletion of the gene coding for subunit IV of cytochrome *c* oxidase in *Paracoccus denitrificans*. *Journal of Biological Chemistry*, February 1997, vol. 272, no. 9, p. 5514-5517.

APPENDIX

Tables

Table 1. Functions and residue sequence position comparisons table for functionally important amino acid residues in subunits of the cytochrome *c* oxidase from *P. denitrificans*, bovine heart mitochondria and predicted protein sequence from *B. pseudomallei.*

			Residue position			
		Residue	*P. denitrificans*	Bovine heart mitochondria	*T. thermophilus*	*B. pseudomallei*
Heme ligands	a	His	94	61	72	24
		His	413	378		339
	a_3	His	411	376		337
Cu_B ligands		His*	276	240	233	186
		His	325	290	282	235
		His	326	291	283	236
Proton pathways	K-pathway	Thr	351	316	309 (Ser)	262
		Tyr*	280	244	237	190
Electron transfer	Cu_A heme a	Arg	473	438	450	400
		Arg	474	439	451	401
	heme heme a_3	Phe	412	377	385	338

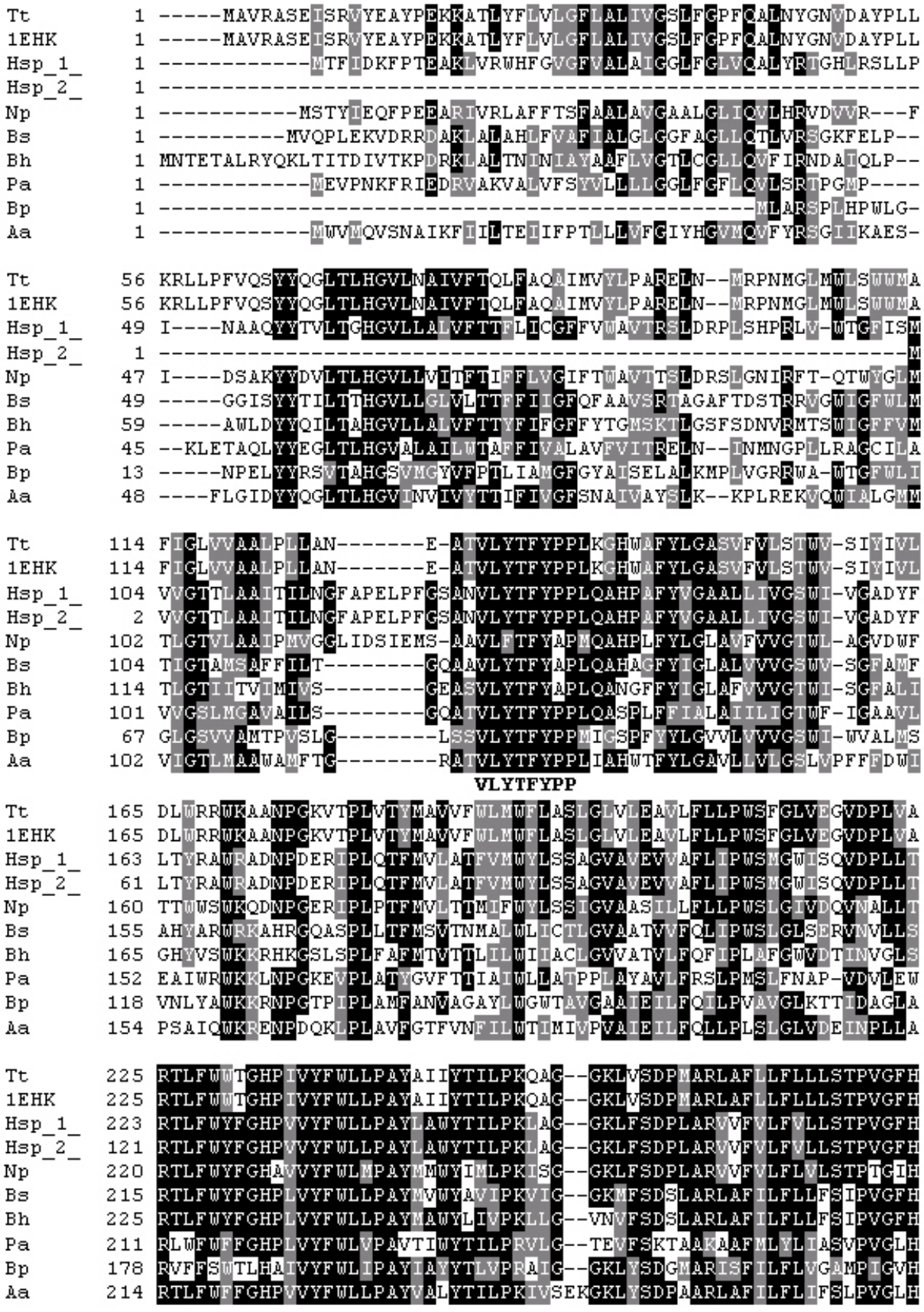
*Tyrosine is covalently bound to a Cu_B histidine ligand.

Table 2. Overall quality of the predicted model as assessed by various methods.

Method	Template structure (1ehk)	Predicted structure	Recommended value for good structures
Procheck[a]	-0.37	-0.45	>-0.5
Errat[b]	96.7%	93.5%	>95%
Whatif[c]	-0.624	-1.199	>-1.0

[a] G factors average score as listed in the "Procheck Summary" output.
[b] Confidence limit.
[c] Quality control value.

Figures


Tt 1 -----MAVRASEISRVYEAYPEKKATLYFLVLGFLALIVGSLFGPFQALNYGNVDAYPLL
1EHK 1 -----MAVRASEISRVYEAYPEKKATLYFLVLGFLALIVGSLFGPFQALNYGNVDAYPLL
Hsp_1_ 1 ----------MTFIDKFPTEAKLVRWHFGVGFVALAIGGLFGLVQALYRIGHLRSLLP
Hsp_2_ 1 ----------
Np 1 ----------MSTYIEQFPEEARIVRLAFFTSFAALAVGAALGLIQVLHRVDVVR---F
Bs 1 ----------MVQPLEKVDRRDAKLALAHLFVAFIALGLGGFAGLLQTLVRSGKFELP--
Bh 1 MNTETALRYQKLTITDIVTKPDRKLALTNINIAYAAFLVGTLCGLLQVFIRNDAIQLP--
Pa 1 ----------MEVPNKFRIEDRVAKVALVFSYVLLLLGGLFGFLQVLSRIPGMP----
Bp 1 ----------MLARSPLHPWLG-
Aa 1 ----------MWVMQVSNAIKFIILTEIIFPTLLLVFGIYHGVMQVFYRSGIIKAES-
Tt 56 KRLLPFVQSYYQGLTLHGVLNAIVFTQLFAQAIMVYLPARELN--MRPNMGLMWLSWWMA
1EHK 56 KRLLPFVQSYYQGLTLHGVLNAIVFTQLFAQAIMVYLPARELN--MRPNMGLMWLSWWMA
Hsp_1_ 49 I----NAAQYYTVLTGHGVLLALVFTTFLICGFFVWAVTRSLDRPLSHPRLV-WTGFISM
Hsp_2_ 1 ----------M
Np 47 I----DSAKYYDVLTLHGVLLVITFTIFFLVGIFTWAVTTSLDRSLGNIRFT-QTWYGLM
Bs 49 -----GGISYYTILTTHGVLLGLVLTTFFIIGFQFAAVSRTAGAFTDSTRRVGWIGFWLM
Bh 59 -----AWLDYYQILTAHGVLLALVFTTYFIFGFFYTGMSKTLGSFSDNVRMTSWIGFFVM
Pa 45 --KLETAQLYYEGLTLHGVALAILWTAFFIVALAVFVITRELN--INMNGPLLRAGCILA
Bp 13 -----NPELYYRSVTAHGSVMGYVFPTLIAMGFGYAISELALKMPLVGRRWA-WTGFWLI
Aa 48 ----FLGIDYYQGLTLHGVINVIVYTTIFIVGFSNAIVAYSLK--KPLREKVQWIALGMM
Tt 114 FIGLVVAALPLLAN-------E-ATVLYTFYPPLKGHWAFYLGASVFVLSTWV-SIYIVL
1EHK 114 FIGLVVAALPLLAN-------E-ATVLYTFYPPLKGHWAFYLGASVFVLSTWV-SIYIVL
Hsp_1_ 104 VVGTTLAAITILNGFAPELPFGSANVLYTFYPPLQAHPAFYVGAALLIVGSWI-VGADYF
Hsp_2_ 2 VVGTTLAAITILNGFAPELPFGSANVLYTFYPPLQAHPAFYVGAALLIVGSWI-VGADYF
Np 102 TLGTVLAAIPMVGGLIDSIEMS-AAVLFTFYAPMQAHPLFYLGLAVFVVGTWL-AGVDWF
Bs 104 TIGTAMSAFFILT--------GQAAVLYTFYAPLQAHAGFYIGLALVVVGSWV-SGFAMF
Bh 114 TLGTIITVIMIVS--------GEASVLYTFYAPLQANGFFYIGLAFVVVGTWI-SGFALI
Pa 101 VVGSLMGAVAILS--------GQATVLYTFYPPLQASPLFFIALAIILIGTWF-IGAAVL
Bp 67 GLGSVVAMTPVSLG--------LSSVLYTFYPPMICSPFYYLGVVLVVVGSWI-WVALMS
Aa 102 VIGTLMAAWAMFTG--------RATVLYTFYPPLIAHWTFYLGAVLLVLGSLVPFFFDWI
VLYTFYPP
Tt 165 DLWRRWKAANPGKVTPLVTYMAVVFWLMWFLASLGLVLEAVLFLLPWSFGLVEGVDPLVA
1EHK 165 DLWRRWKAANPGKVTPLVTYMAVVFWLMWFLASLGLVLEAVLFLLPWSFGLVEGVDPLVA
Hsp_1_ 163 LTYRAWRADNPDERIPLQTFMVLATFVMWYLSSAGVAVEVVAFLIPWSMGWISQVDPLLI
Hsp_2_ 61 LTYRAWRADNPDERIPLQTFMVLATFVMWYLSSAGVAVEVVAFLIPWSMGWISQVDPLLI
Np 160 TTWWSWKQDNPGERIPLPTFMVLTTMIFWYLSSIGVAASILLFLLPWSLGIVDQVNALLT
Bs 155 AHYARWRKAHRGQASPLLTFMSVTNMALWLICTLGVAATVVFQLIPWSLGLSERVNVLLS
Bh 165 GHYVSWKKRHKGSLSPLFAFMTVTTLILWIIACLGVVATVLFQFIPLAFGWVDTINVGLS
Pa 152 EAIWRWKKLNPGKEVPLATYGVFTTIAIWLLAIPPLAYAVLFRSLPMSLFNAP-VDVLEW
Bp 118 VNLYAWKKRNPGTPIPLAMFANVAGAYLWGWTAVGAAIEILFQILPVAVGLKTTIDAGLA
Aa 154 PSAIQWKRENPDQKLPLAVFGTFVNFILWTIMIVPVAIEILFQLLPLSLGLVDEINPLLA
Tt 225 RTLFWWTGHPIVYFWLLPAYAIIYTILPKQAG--GKLVSDPMARLAFLLFLLLSTPVGFH
1EHK 225 RTLFWWTGHPIVYFWLLPAYAIIYTILPKQAG--GKLVSDPMARLAFLLFLLLSTPVGFH
Hsp_1_ 223 RTLFWYFGHPVVYFWLLPAYLAWYTILPKLAG--GKLFSDPLARVVFVLFVLLSTPVGFH
Hsp_2_ 121 RTLFWYFGHPVVYFWLLPAYLAWYTILPKLAG--GKLFSDPLARVVFVLFVLLSTPVGFH
Np 220 RTLFWYFGHAVVYFWLMPAYMMWYIMLPKISG--GKLFSDPLARVVFVLFLVLSTPTGIH
Bs 215 RTLFWYFGHPLVYFWLLPAYMVWYAVIPKVIG--GKMFSDSLARLAFILFLLFSIPVGFH
Bh 225 RTLFWYFGHPLVYFWLLPAYMAWYLIVPKLLG--VNVFSDSLARLAFILFLLFSIPVGFH
Pa 211 RLWFWFFGHPLVYFWLVPAVTIWYTILPRVLG--TEVFSKTAAKAAFMLYLIASVPVGLH
Bp 178 RVFFSWTLHAIVYFWLIPAYIAYYTLVPRAIG--GKLYSDGMARISFILFLVGAMPIGVH
Aa 214 RTLFWFFGHPVVYFWLLPAYVALYTILPKIVSEKGKLYSDPAARLAFILFLIFSLPVGLH

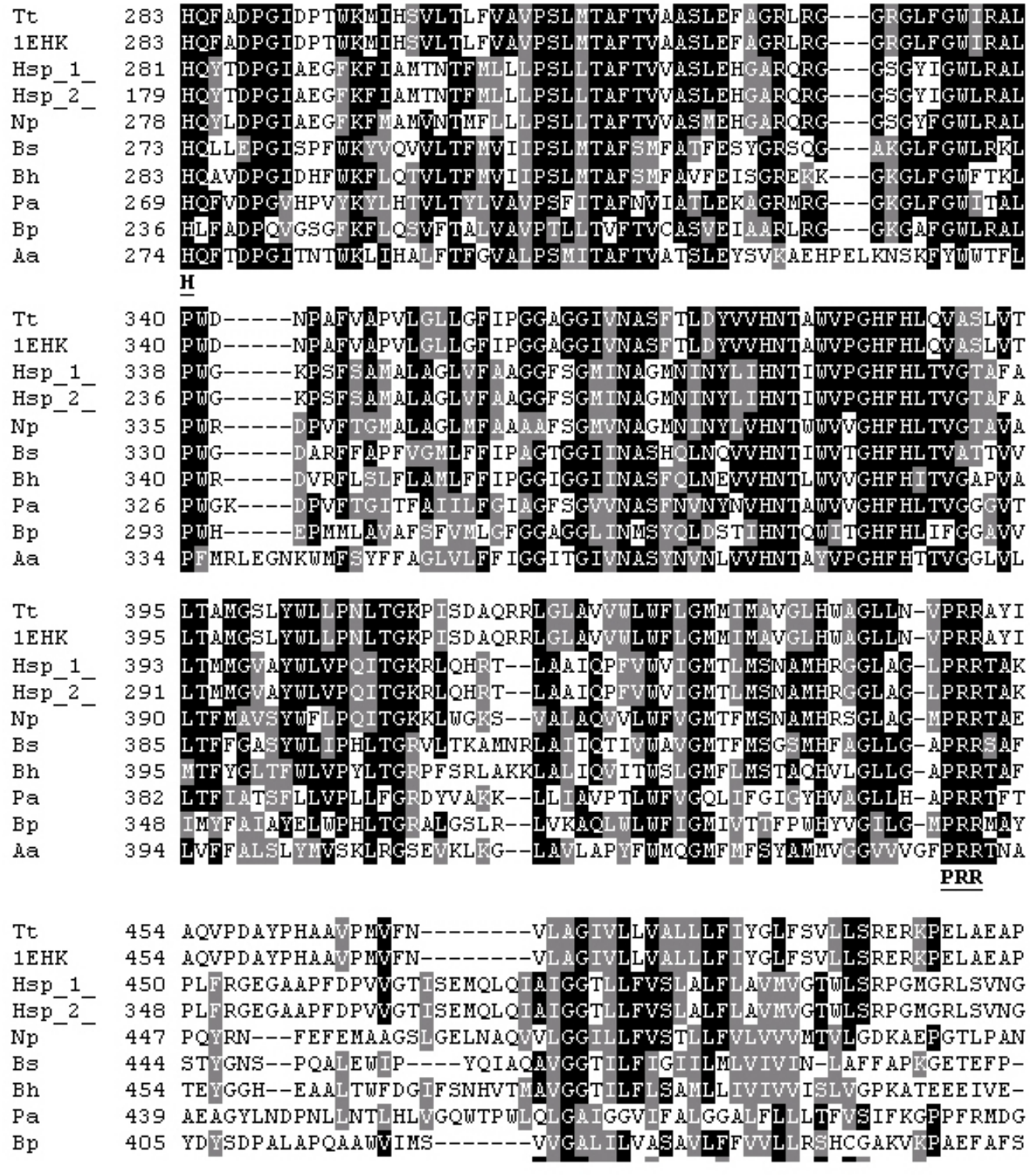


Figure 1. ClustalW alignment of BLAST (blastp) results for 8 hits with E values lower than 10^{-50} and the 1EHK sequence used as the comparative modelling template, GenBank accession numbers are bracketed: Tt, *Thermus thermophilus* (Q56408); 1EHK, *Thermus thermophilus* (13399682); Hsp_1, *Halobacterium sp*. (NP44237); Hsp_2, *Halobacterium sp*. (AAG20326); Np, *Natronobacterium pharaonis* (T44942); Bs, *Bacillus stearothermophilus* (T43835); Bh, *Bacillus halodurans* (NP241065); Pa, *Pyrobaculum aerophilum* (NP559246); Bp, *Burkholderia pseudomallei* (AF087002); Aa, *Aquifex aeolicus* (NP214506). Functionally crucial conserved residues are underlined.

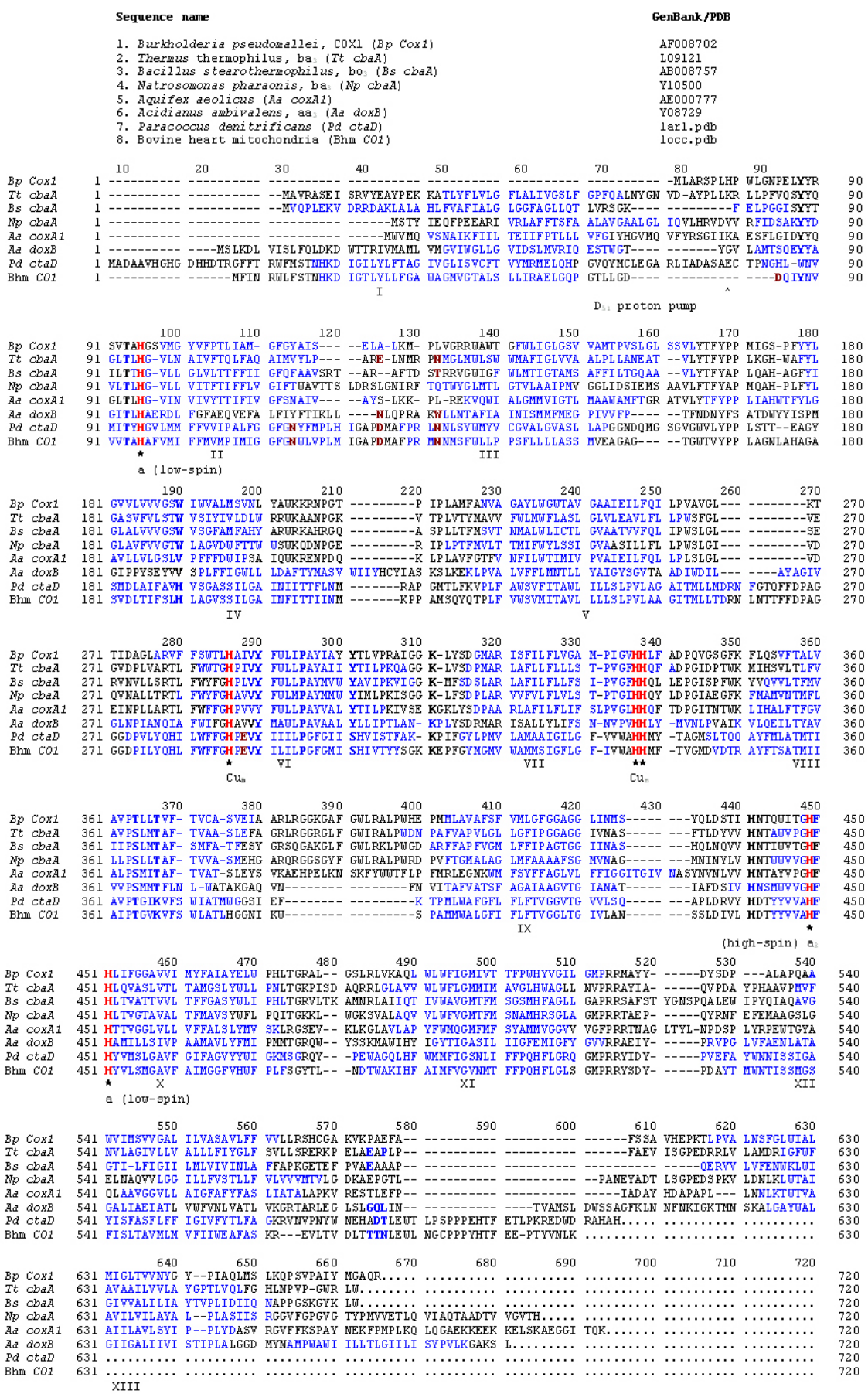

Sequence name
GenBank/PDB
1. Burkholderia pseudomallei, COX1 (Bp Cox1) AF008702
2. Thermus thermophilus, ba3 (Tt cbaA) L09121
3. Bacillus stearothermophilus, bo3 (Bs cbaA) AB008757
4. Natrosomonas pharaonis, ba3 (Np cbaA) Y10500
5. Aquifex aeolicus (Aa coxA1) AE000777
6. Acidianus ambivalens, aa3 (Aa doxB) Y08729
7. Paracoccus denitrificans (Pd ctaD) 1ar1.pdb
8. Bovine heart mitochondria (Bhm CO1) 1occ.pdb
Bp Cox1
Tt cbaA
Bs cbaA
Np cbaA
Aa coxA1
Aa doxB
Pd ctaD
Bhm CO1
I
D51 proton pump
II
III
a (low-spin)
IV
V
Cu
VI
VII
VIII
IX
X
XI
XII
(high-spin) a3
XIII

Figure 2. **Multiple sequence alignment of a diverse range of COX sequences and sequences from PDB structures.** TMH: Transmembrane helices are denoted by Roman numerals I to XIII and colored in blue type. Positions of the six histidine residues (H), proton pump and residues crucial to electron transport are depicted and highlighted in red.

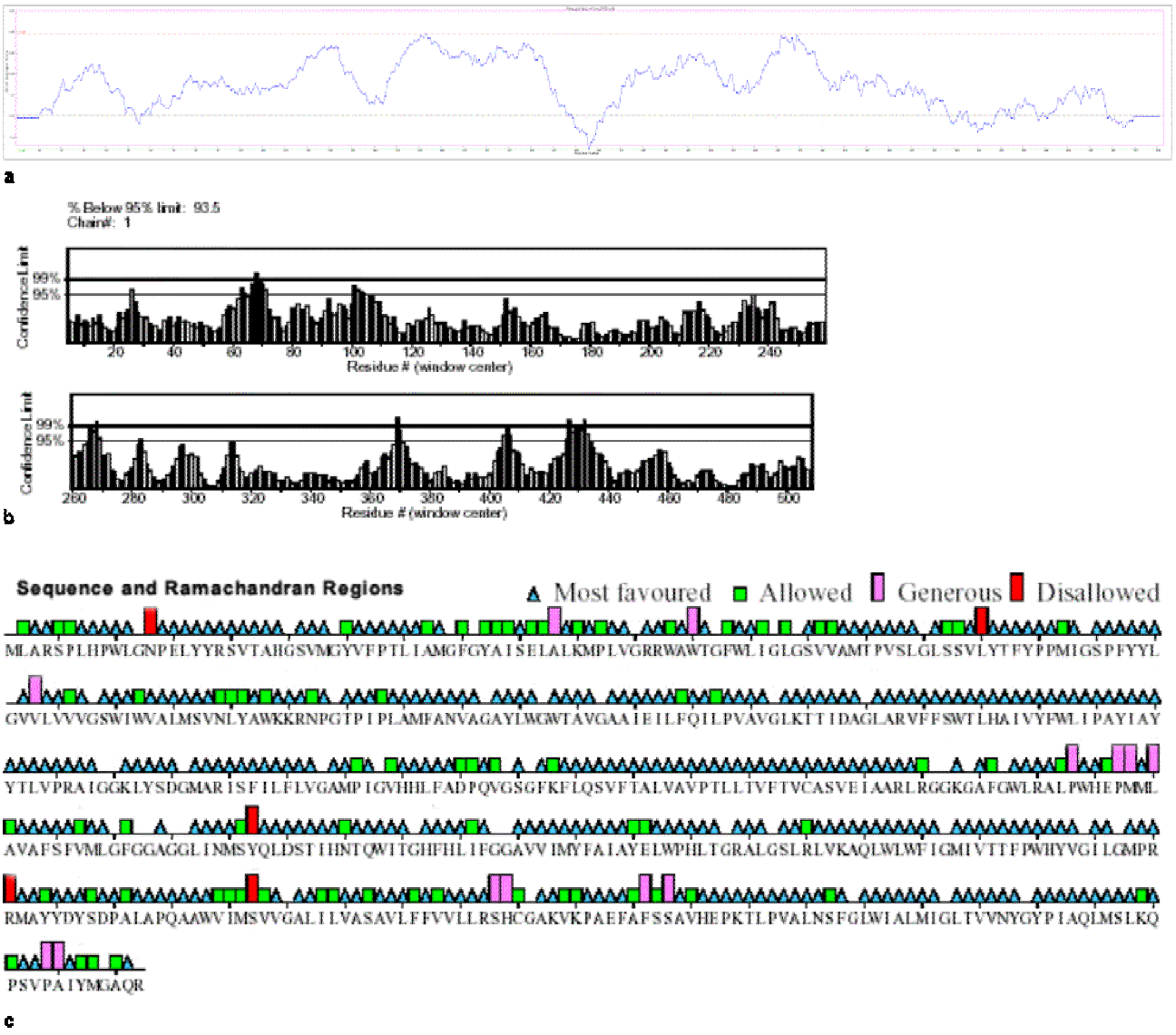
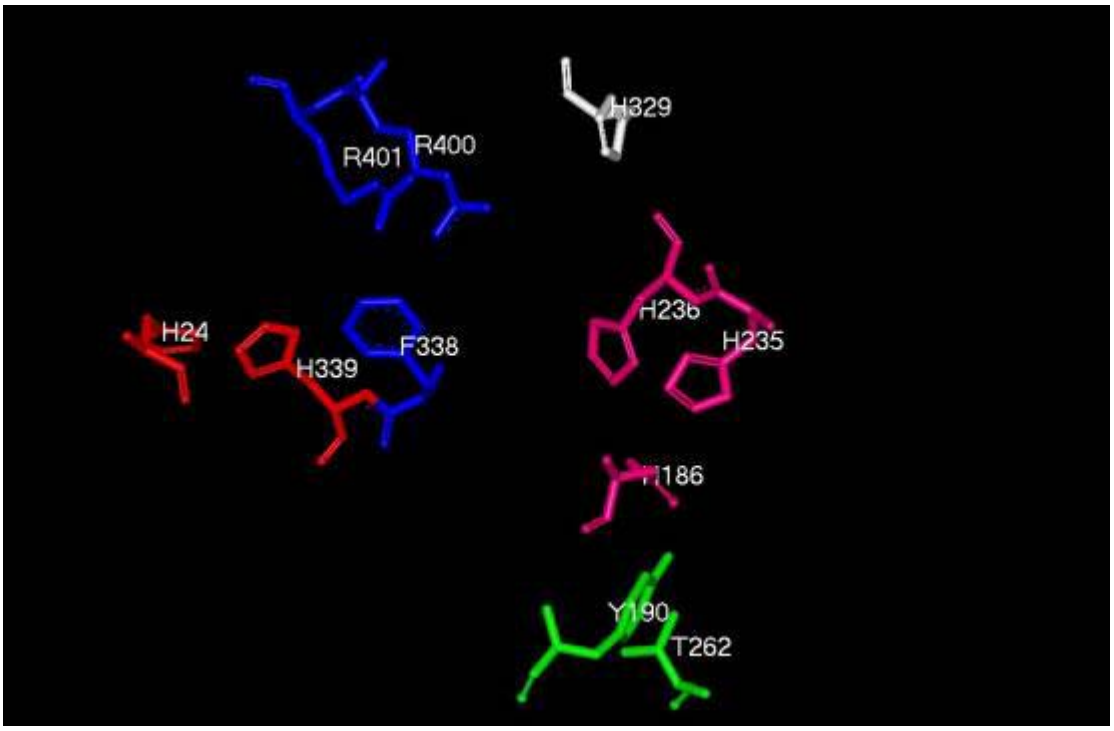


Figure 3. Validation of the overall structure using Verify-3D (a), Errat (b) and Ramachandran plot (Procheck) (c).

a

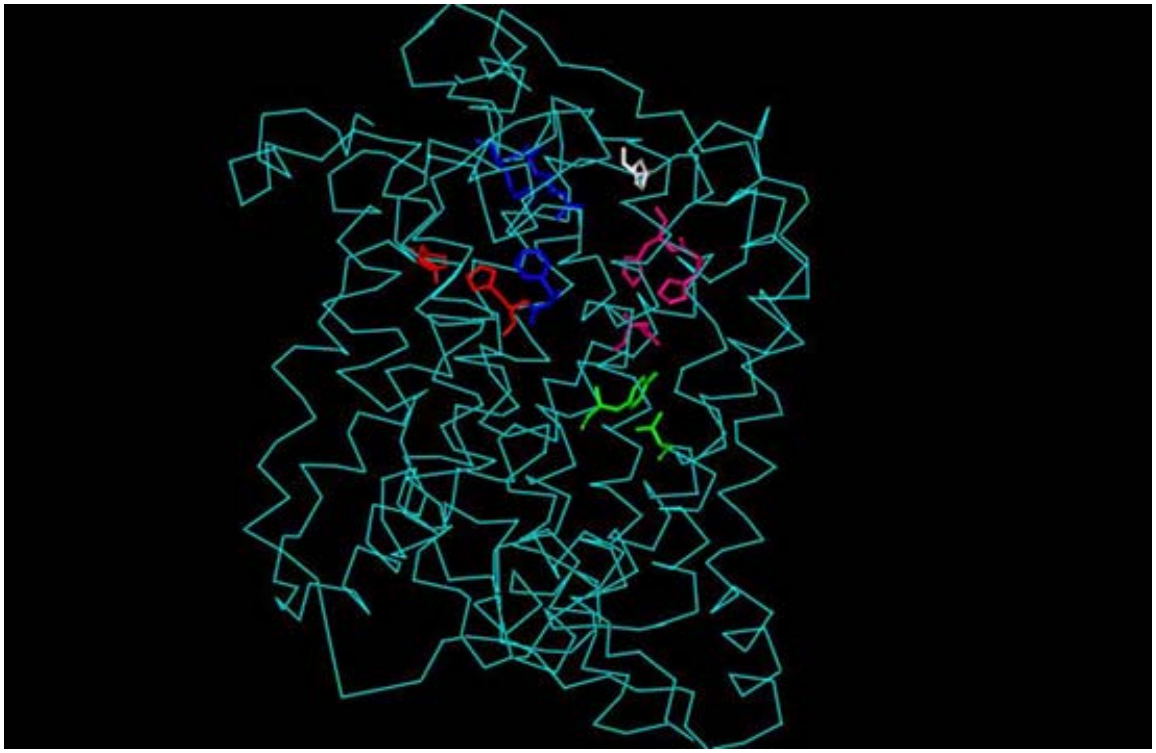


b

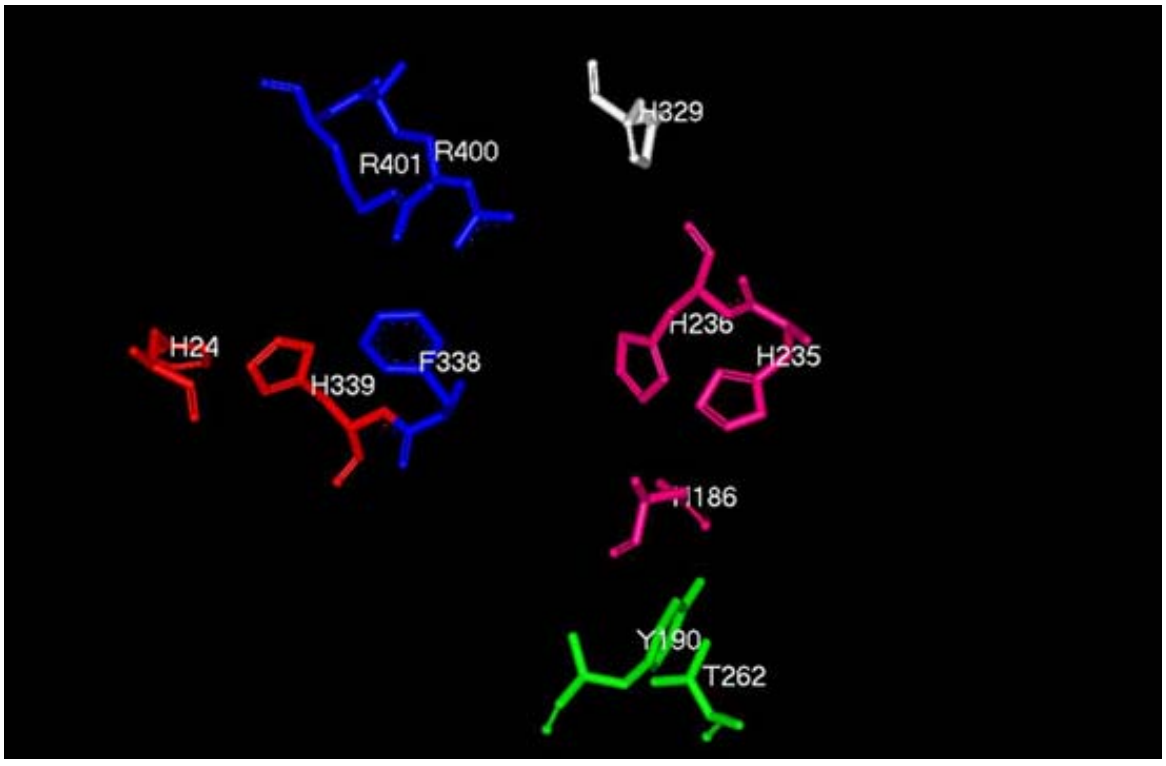
Figure 4. Functional overview of the predicted cytochrome *c* oxidase structure for *B. pseudomallei*.

a. The figure shows the 12 discernible predicted transmembrane helices with the electron transport, proton pathway and heme ligand residues rendered as balls and sticks.

b. Magnification of the space with the functionally crucial residues mentioned and the backbone removed for clarity.



a



b

Figure 5. The location of residues in tertiary structure space for heme ligands (H24, H339), electron transfer pathway residues (F338, R400, R401), Mg ligands (H329), CuB ligands (H186, H235, H236), proton pathway residues (Y190, T262). Figure 5a shows an overall view of the structure (line trace back-bone) with the above residue positions highlighted as sticks. Figure 5b is a magnification of the residue positions in structure space.