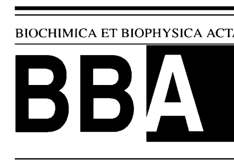
Biochimica et Biophysica Acta 1410 (1999) 32–50



Sequence analysis of cytochrome *bd* oxidase suggests a revised topology for subunit I

Jeffrey P. Osborne, Robert B. Gennis *

School of Chemical Sciences, University of Illinois, Urbana, IL 61801, USA

Received 4 November 1998; accepted 19 November 1998

Abstract

Numerous sequences of the cytochrome *bd* quinol oxidase (cytochrome *bd*) have recently become available for analysis. The analysis has revealed a small number of conserved residues, a new topology for subunit I and a phylogenetic tree involving extensive horizontal gene transfer. There are 20 conserved residues in subunit I and two in subunit II. Algorithms utilizing multiple sequence alignments predicted a revised topology for cytochrome *bd*, adding two transmembrane helices to subunit I to the seven that were previously indicated by the analysis of the sequence of the oxidase from *E. coli*. This revised topology has the effect of relocating the N-terminus and C-terminus to the periplasmic and cytoplasmic sides of the membrane, respectively. The new topology repositions I-H19, the putative ligand for heme b_{595}, close to the periplasmic edge of the membrane, which suggests that the heme b_{595}/heme *d* active site of the oxidase is located near the outer (periplasmic) surface of the membrane. The most highly conserved region of the sequence of subunit I contains the sequence GRQPW and is located in a predicted periplasmic loop connecting the eighth and ninth transmembrane helices. The potential importance of this region of the protein was previously unsuspected, and it may participate in the binding of either quinol or heme *d*. There are two very highly conserved glutamates in subunit I, E99 and E107, within the third transmembrane helix (*E. coli* cytochrome *bd*-I numbering). It is speculated that these glutamates may be part of a proton channel leading from the cytoplasmic side of the membrane to the heme *d* oxygen-reactive site, now placed near the periplasmic surface. The revised topology and newly revealed conserved residues provide a clear basis for further experimental tests of these hypotheses. Phylogenetic analysis of the new sequences of cytochrome *bd* reveals considerable deviation from the 16sRNA tree, suggesting that a large amount of horizontal gene transfer has occurred in the evolution of cytochrome *bd*. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Cytochrome bd oxidase; Sequence analysis; Topology

1. Introduction

In *Escherichia coli*, cytochrome *bd* is a terminal oxidase in the branched electron transport chain. It is expressed during both aerobic and anaerobic growth conditions, but the expression level is highest when the bacterium is growing in aerobic stationary phase or in low oxygen environments [1,2] during which times it is the primary respiratory oxidase. Cytochrome *bd* has a remarkably high affinity for oxygen [3,4] and catalyzes the four-electron reduction of oxygen to water. When the reductant, quinol, is oxidized it releases protons on the periplasmic side of the membrane. Since protons are taken up from the

* Corresponding author. Fax: +1-217-244-3186;
E-mail: gennis@scs.uiuc.edu

0005-2728/99/$ – see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0005-2728(98)00171-6

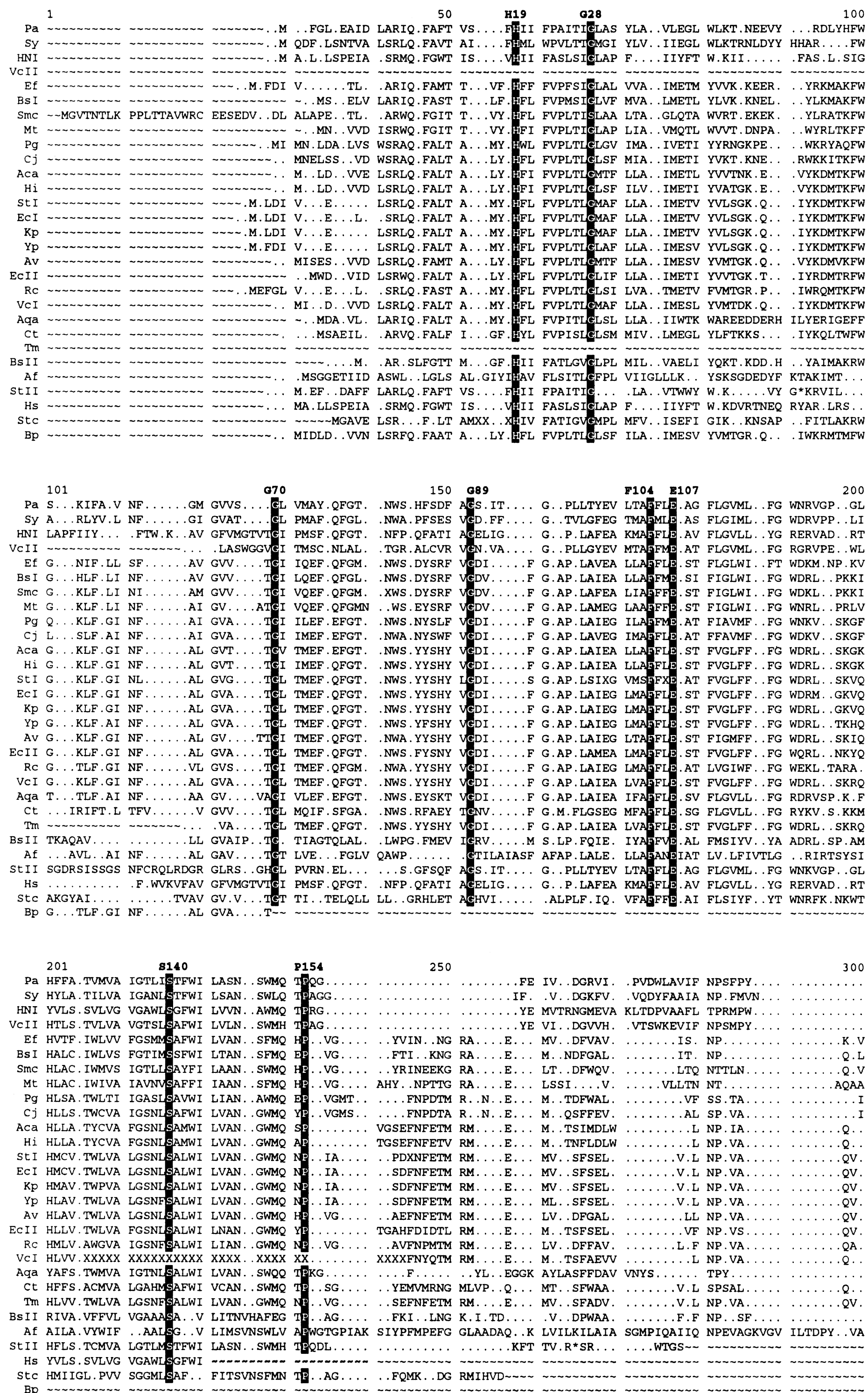


Fig. 1. Alignment. These were created using PILEUP in the GCG software package with gap creation and extension penalties of 3.0 and 1.0, respectively, followed by manual adjustment. VC23 is a fragment that must come from either VcII or VcIII, but which one is not known.

```
       301       H186                                          350                                   ↓ Q-loop start  400
    Pa ........R. LLHMSVAAFL A..TAFFV.G ..ASA..AWH LLRGRDNP.. .......... ...A.IR..K MLSMAM.WMA LIVAPVQALI GDAHGLNTLE
    Sy .........S FLHMFFAT.L ET..SMFVIG ..GIS..AWC ILTGR.QPQ. .......... ....FFA..R SLQVIL.VVV MAVAPLQIFI G..H.LSA.E
   HNI .........M YVHMMNASVI S..VALLVAG VS..AYLVW. .....KKPDA D......... ......AWNS ALKLAV.VLL LVSAPFQAVH GDAYGRH.VE
  VcII ........R. LAHTLLASAL T..ASFLIAG IS..AYQ... VLRNAKH... Q......... ......AAKM GLKVAVSVAA LAI.PVQILV GDLHGLNTLE
    Ef WYE.F..S.. ..HVIMGAFT L.GGMA.VAG MA..A...FQ LLKKR..... DIS....FHK ASM.RI.... GL.WVTLFGS IGVL.L...A GDLQMKALIE
   BsI WVE.F..... .PHVIFGALA T.GAF.FIAG VS..A...FK LLKKK..... EV....PFFK QSF.KLAMIV GL.CA...G. LGV.GL...S GHMQAEHLME
   Smc F......... ..HSFSAAFL TGGAF.MV.G IA..A...FH LMRKK..... HIP....VMR TSL.RL.... GL..VTL..A VGGL.LTAVS GDTLGKVMYE
    Mt F......... .THTVSGALL TAGTF..V.A .AVSA...WW LVRSSTT... HADSDTQAM. ..Y.RPATIL GC.WVAL.AA TAGLLFT... GDHQGKLMFQ
    Pg ..NKF..... .WHTISSCWT L.GSVF.ALG VC..G...IY LLR.K..... D.DKHKDFAL KNIKII.APF GL.AA....S LIT....AFT GDTSAYNVAQ
    Cj ..SKF..... .LHTIGSGYV I.SALF.VMG IS..A...WF ML..K..... ..GRHIIEAK KSL.VVGASF GL.VC....S IFL....FFS GDESAYRVTQ
   Aca ..SKF..... .LHTLSAGYV S.AAFF.VLA IS..S...YY LLKGR..... DI....GFAK RSFS.VASTF GI.ISVV..S LLIM...... GDESGYEIGY
    Hi ..SKF..... .LHTLSAGYV T.GAFF.VLA IS..S...YF LLKGR..... DF....EFAK RSFS.VAATF GF.IASI..S VLIL...... GDESGYDIGK
   StI ...KF..V.. ..HTVASGYV T.GAMF.ILG IS..A...YY MLKGR..... DFA....FAK RSFA.IAASF GM.AAVL..S V.IV.L.... GDESGYEMGD
   EcI ...KF..V.. ..HTVASGYV T.GAMF.ILG IS..A...WY MLKGR..... DFA....FAK RSFA.IAASF GM.AAVL..S V.IV.L.... GDESGYEMGD
    Kp ...KF..V.. ..HTVASGYV T.GAMF.ILA IS..S...WY MLKGR..... DFA....FAK RSFA.IAASF GM.AAIL..S V.IV.L.... GDESGYEMGD
    Yp ...KF..V.. ..HTVAAGYV T.GAMF.ILG IS..S...YY LLKGR..... DIP....FAK RSFA.IAASF GL.ASVL..S V.IV.L.... GDESGYEMGD
    Av ...KF..V.. ..HTVASGYV T.GAVF.VLA IS..S...YY LLKKR..... DL....GFAR RSFA.IASAF GM.ASIL..S V.IV.L.... GDESGYEVGE
  EcII ...KF..... .VHTVMAGYV T.GAMF.IMA IS..A...WY LLRGR..... E....RNVAL RSFA.IGSVF GT.LAIIGT. ...LQL.... GDSSAYEVAQ
    Rc ...KF..V.. ..HTVSAGYV T.ASIF.VIG VS..A...WY LLRGR..... HID....LAR RSIA.VASAF GL.AASL..S V.IV.L.... GDESGYEATH
   VcI ..VKF..... .LHTVASAYT C.GAMF.ILG VS..S...YY LLQGR..... DI....EFAR RSFA.IAATF GM.AAII..S VIVL...... GDESGYKLGD
   Aqa ...RF..... .VHMFLAGVI T.GAIF.VAG IS..A...YY ILK.R..... ...KHLDVAK ESF.RIAVTI GM.IASL... LQII.....A GDLHGYQVAR
    Ct ...RF..T.. ..HVVLGAWL S.G.IFLVLS VS..A...HY LRKERHK... D.....FAN Q......... GL.KISMFCA FLVLALQLWS ADVTARGVAK
    Tm ...KF..V.. ..HTVAAGYT T.GAMF.ILG IS..A...YY LLKGR..... DVA....FAR RSFA.IAASF GM.ASVL..S V.IV.L.... GDESGYELGE
  BsII ....F..ITA .GHVVLSAFM T.GA.FIVAS VA..A...YK MIRTRKK... E..RVYRFHR KAL.LLAA.. .L...TI.GG IFSL.LTALN GHESAQMLYE
    Af IFNPFAAISA .LHALFAAF. SVG.....VS IALLA...FS L...R...Y. .YTGGE...K RNL.KAAKVA SLVILVLFL. IQPTILGHFM GD..G..VVE
  StII ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    Hs ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   Stc ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    Bp ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

       401    K251  E257                                   450                              P289                 500
    Pa ....HQPAKI AAMEGHWDNS SGE.PT.... ......PLIL FGW.....PD MQREET.... ....RFKVEI P.VLGSLI.L KHSLTEPIPA LKDFPPEDRP
    Sy QVSVHQPAKL AAMEALWETV PAKTP..... A......... .AWSVVALPN DKAEKN.... ....DWELSI PGALSYILEL KPQLDKPIQG LKDWAPVDRP
   HNI DT...QPQKF AAMEAHYETG SADLHLLAF. ......PK.. ...SLDALTD PRTE.N.... ....LFTVSL PG.IGSFLAS GGDFDAEVIG LNEY..EENP
  VcII ....HQPAKI AAMEGVWQTE RG...A.... ......PLLL F.....ALPN EETRSN.... ....DFALGI PN.LASLI.L THDMNGEIVG LDAFAP.DHP
    Ef G....QPMKF AAMEGAYEDS .DD....... ......PA.. .AWTLIAW.. .A...D.E.R EHKQVFGIDI PYMLS.ILS. YNKLSGSVDG M.D.......
   BsI S....QPMKM AASEGLWEDS G.D....... ......PA.. .AWT..AF.. .A...TIDTK NEKSSNEIKV PYALS.YLA. YQKFSGSVKG MK........
   Smc Q....QPMKM AAAEALW.DG .EK....... ......PA.. .PFSVFAY.. .G...DVD.K GHNKV.ALEV PGLLS.FLA. HSDFESYVPG IND.......
    Mt Q....QPMKM ASAESL.CDT QTD....... ......P... .NFSVLTV.. .GRQNNCDSL ..TRV..IEV PYVLP.FLA. ....EGRISG V.........
    Pg K....QPMKL AAMEALYDSG QTDKD..GLT ADGKGLPL.. .S..LFGILN PAKETPQDDK EA.FLFNVSV PRVLS.VLG. TRNPSGYVPG INNILEGGYV
    Cj T....QPMKL AAMEGVY.QG E.HRA..GL. .....VP... .....FGILN P.KKT.IDNN ESVFLFDITI PYALS.ILG. NRDPNSFVPG IEDLIYGN..
   Aca A....QPTKL AAMEGEW... HTQ.P..A.. ......PA.. .AWNMIIV.. P....NQAEQ KND..FSLQI PYIAG.VIV. TRSFDKQFSG LIDL......
    Hi A....QPVKL AAMEAEF... ETH.P..A.. ......PA.. .PFLPVAI.. P....NTAEM KND..FAIEI PYLGG.VIA. TRSIDKEIIG LKDL......
   StI V....QKTKL AAIEAEW... ETQ.P..A.. ......PA.. .SFTLFGI.. P....DQDKQ ENH..LAIQI PYALG.IIA. TRSVDTPVIG LKDL......
   EcI V....QKTKL AAIEAEW... ETQ.P..A.. ......PA.. .AFTLFGI.. P....DQE.E ETNK.FAIQI PYALG.IIA. TRSVDTPVIG LKEL......
    Kp V....QKTKL AAIEAEW... ETQ.P..A.. ......PA.. .AFTLFGI.. P....DQ.RR RRT..FAIQI PFALG.IIA. TRSVDKQVTG LKAL......
    Yp V....QKTKL AAIEAEW... ETQ.P..P.. ......PA.. .AFTLFAI.. P....NQETM ENR..FAIQI PYALG.LIA. TRSLDTPVIG LRDL......
    Av V....QKAKL AAIEAEW... ET.HP..A.. ......PA.. .SFTLIGF.. P....NEEEQ RTD..FAVKI PWVLG.IIA. TRSLDEQVIG IKDL......
  EcII V....QPVKL AAMEGEW... QTE.P..A.. ......PA.. .PFHVVAW.. P....EQDQE RN..AFALKI PALLG.ILA. THSLDKPVPG LKNL......
    Rc N....QRMKL AAIEGMW... DTE.E..A.. ......PA.. .SFNLFGI.. P....DQEAR ET.K.YAIKI PYVMG.LIG. TRSLTTEMPG IKDL......
   VcI V....QKVKL AAIEAEW... HTE.P..A.. ......PA.. .AFTLFGL.. P....NQKEG KTD..YAIKI PYLMG.LIA. TRSFDEQVTG LHDL......
   Aqa T....QPLKL AMMEDKW... ETESG..A.. .GLEIVP... ......G... .......... .......IEI PKLLS.ILA. YHDPDAKVYG IKDLVQ...T
    Ct H....QPAKL AAFEGVF... KTQ....... ...GHTPI.. ..Y.LLGIV. .....DMK.K E..RVIGIPI PSGLSLLV.. HRNAKTPVTG L.D.......
    Tm V....QRVKL AAIESEW... HTE.P..A.. ......PA.. .AFTLFGI.. P....NQEEM RTD..YAIKI PFALG.DIA. TRSLD~~~~~ ~~~~~~~~~~
  BsII Y....QPEKL AGAEGLFET. .......... ..RSHAPL.. .A..IGGFTD PN.E...... EKVK.WAIEI PWALSFLAA. .NRFDTVVKG LN........
    Af ....MNPTKF AMME...NAK ET........ .FYN..PM.I ...ALVAYGD PSR....... .......... P.....IVGF D.EFERQCNS LGDAELGDLA
  StII ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    Hs ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   Stc ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    Bp ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

       501                                                 550                                                   600
    Pa NST.V..... .......... .......... .......... .......... .......... .......... .......... .......... ........VF
    Sy RMVGL..... .......... .......... .......... .......... .......... .......... .......... .......... ........IY
   HNI .PVAL..... .......... .......... .......... .......... .......... .......... .......... .......... ........VF
  VcII .PVAP..... .......... .......... .......... .......... .......... .......... .......... .......... ........LF
    Ef .....T.V.. .........N .....KRLVE K...Y..... G.DDK..... .......... ...NY..... .......... .......... ..YP..P.VN
   BsI .....T.L.. .QAEY..... .........E KI..YG...K G..D...... .......... ...Y.I.... .......... .......... ...P..P.VK
   Smc .....T.... .........N .....KALQE Q...F..... GPGD...... .......... ...Y...... .......... .......... ..KPIVP.V.
    Mt .....T.... .......LQG IRDLQQEYQQ R...F..... GPND...... .......... ...Y.R.... .......... .......... ......PNLF
    Pg KADGTTAI.. ..PVDSMMQR GRRAIMALND YSKAKQ...A GDMEAALQHK SVIDEN..FP YFGY.S.... .Y........ ......IQH. .KNDIVPPVG
    Cj ESKG...I.. .EPMQNRIDR GKIAIQTLKD YKLAKE...N NDTIAMANHK SILETH..FK DFGY.G.... .Y........ ......LEK. .PSDTIPPVA
   Aca ......QERN LE....RVRS GIQAY.ALLQ QLREEK...K ANGQASEETK AKFDKV..QK DLGF.GLLLK RYTDNVVD.. .ATEEQIKQA A.ADTIPKVA
    Hi ......QALN .ET...RVRS GIRAY.ELFT QLRAEK...K ANGQVNEETK AQFNEV..KK DLGF.GLLLK RYTNNVVD.. .ATEEQIKQA A.RDTIPNVG
   StI .......MVQ HEE...RFRN GMKDY.ELLE QXRA...... GSTDQA..VR DQFNRM..KK DLGY.GLLLK RYTPNVTD.. .ATEAQIQQA T.KDSIPRVA
   EcI .......MVQ HEE...RIRN GMKAY.SLLE QLRS...... GSTDQA..VR DQFNSM..KK DLGY.GLLLK RYTPNVAD.. .ATEAQIQQA T.KDSIPRVA
    Kp .......MVQ HEE...RIRN GMKAY.ALLE QLRA...... GSTDQA..VR DRFNDV..KK DLGY.GLLLK RYTDNVLD.. .ATEEQIALA A.KDSIPAVA
    Yp .......MSQ HEV...RIRN GIQAY.SLLE QLRG...... GNTDPA..VR DAFNKA..KQ DLGY.GMLLK RYTENVAD.. .ATEEQIQLA A.QDSIPRVL
    Av .......IAD HEA...RIRN GMVRYG.LLE ELRA...... GNKSPEKIA. .AFNEV..KD DLGY.GLLLK KYTPNVVD.. .ASEEQIKQA A.KDTIPSVA
  EcII .......MAE TYP...RLQR GRMAWL.LMQ EI...S...Q GNREP..HVL QAFRGL..EG DLGY.GMLLS RYAP...DMN HVTAAQY.QA AMRGAIPQVA
    Rc .......VAQ AEG...RIRS GVIAYDALL. QIRA...... AKGDVAPEVK DTFEAH..SA DLGY.AMLL. ..TTVVEDPK LATDAQISEA AWK.TVPQVA
   VcI .........R HEHL.LRIRN GMVAY.QLLE RLQA...... ..GDASVETQ RKFEQT..NH DLGY.GLLLK RYTDTITD.. .ANEAQIQQA A.DDSMPTVW
   Aqa FQ........ ......KIAK G........E EV........ GE........ DELPEL..WL D.......PR QFAGR..QSL YITEEEIP.P TW...IP...
    Ct .......... ........Q. .......... .......F.. .PKDEWPNVA FVFQ...... .......... .......... .......... ..........
    Tm ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  BsII .......... .......... .......... AF........ .PRDEWPPL. ..F....... .......... .......... .......... ........IH
    Af GQLGIT.MDA AIATAKQV.. GVKVGSE..E ......IQNA WNTE...... ..LRQI.... .......CL. ......ADLK KAESR..... .....IAVVH
  StII ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    Hs ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   Stc ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    Bp ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
```

Fig. 1 (*continued*).

cytoplasmic side to make water [5–8], an electrochemical gradient results that is available for various processes such as the production of ATP, membrane transport and flagellar motion [9–11]. Sequence alignments clearly categorize the cytochrome *bd* oxidases as unique and not homologous to the superfamily of heme–copper terminal oxidases that includes cytochrome *c* oxidase [12,13]. In contrast to

```
         601                                        Q-loop end ↓ M393 650                                                  700
    Pa ...WSF.... .......... .......... .......... .RIMAGLGML M.ILVGVWSL ....WLRWR. ........GQ DK..LFN.SK ..........
    Sy ...YSF.... .......... .......... .......... .RIMVAIGLF FAALMAVTVL Q...WL..R. ........GK LKPEIISQQK ..........
   HNI ...WSF.... .......... .......... .......... .RFMVGLGFL FIGL.ALW.G G...YLTYR. ........G. .R...LTEST ..........
  VcII ...FGF.... .......... .......... .......... .RVMVGVGVL M.LLVS.WFG A...WRVWR. ........N. .KP....LPK ..........
    Ef TLFWSF.... .......... .......... .......... .RIMAGFG.A LMLLVAA..L GL.FFTRKK. ..KPSLY... .E.......K K........W
   BsI TTFWSF.... .......... .......... .......... .RIMVGAG.V VMIL.AA..L GGLWLNRRK. ..K..L.... .E......NS K........W
   Smc .AYWGF.... .......... .......... .......... .RWMIGFGMA ...SFSLGLL GL.WLTRKRF LLPPALRTGE DEVPHLVLLK KPLGARLTRM
    Mt VTYWSF.... .......... .......... .......... .RMMIGL.MA IPVLFA..LI AL.WLTR... ........G. GQIPN..... .......QRW
    Pg LTYYSF.... .......... .......... .......... .RIMVGLGML FILLF...LM ..A.WL.LSF KPEKFSKM.. .......... ..........
    Cj LTFYSF.... .......... .......... .......... .HIMVALGSF FFLLF...IV ..TLYLTMAN DIEKFRKV.. .......... ..........
   Aca PTFWSF.... .......... .......... .......... .RVMMATGG. .AILL...LM LLAFVQNVR. .....KTVGS .......... ..........
    Hi PNFWAF.... .......... .......... .......... .RAMIAAGG. .LIAL...LT FGAFVQNLR. .....NKVTQ .......... ..........
   StI PLYXAY.... .......... .......... .......... .RIXVVCG.. FLLLA...II ALSFWSVIR. .....NRIGE .........K K.........
   EcI PLYFAF.... .......... .......... .......... .RIMVACG.. FLLLA...II ALSFWSVIR. ....N.RIGE .........K K.........
    Kp PLYFAF.... .......... .......... .......... .RIMVACG.. VLMLL...II ALSFWSVIR. .....NRIGE .........K K.........
    Yp PLYFAF.... .......... .......... .......... .RIMVACG.. FLMLL...II GLAFYNVVR. .....GRIGE .........K K.........
    Av SMFWSF.... .......... .......... .......... .RAMV..GAG FAMLI...LF VCAFWASARK ........NE ........ES K.........
  EcII PVFWSF.... .......... .......... .......... .RIMVGCGS. .LLLL...VM LIALVQTLR. .....GKIDQ HR........ ..........
    Rc PLFWSF.... .......... .......... .......... .RIMVGAG.. FFMLA...LI GYMFFT..S. ....NFRGGR ........FP T.........
   VcI PLFWSF.... .......... .......... .......... .RIMVGCGF. .IMLF...VF GAALLQTCR. .....QRITH .......... ..........
   Aqa ..FWSL.... .......... .......... .......... .RGMVFLG.. FFFAF...IT SLGFIFTVQD TIEQ...... .......... ..........
    Ct ....TY.... .......... .......... .......... .HLMVMLWGV MVLLA...LI A..F...... .......... ......AVYK KKSWSCK.KG
    Tm ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  BsII TL...F.... .......... .......... .......... .NAMVGVG.. .MLLILYSII GVVW..RK.. .......... .......VLK KDRFPT....
    Af AAYYT.KIAF GVIGFVSAIA LFAHFRKLPL LSSLVDRILG RRAMLILPIG VFL....... .......... .......... .......... ..........
  StII ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    Hs ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   Stc ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    Bp ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

         701                        G440     G447 W451       750                                                   800
    Pa ..AFLRLTLL MG...PSGLI AILAGWFTTE MGRQPWVVY. ...GL.MRTA DASSA.QSVT ..QM...SLT LLTFV.VVYF LSF.GVG.I. ...GWMMRLV
    Sy ..WLLRAWVF AA...PLGYL AVETGWIVRC VGRQPWIVY. ...GE.MRTA E..SASN.LP PGEILF.SLT GLSVM.YIVF L....IATLY .F.G..SRII
   HNI ..RYLKSMIA AS...PFGYA ALLTGWYVTE IGRQPWVIQ. ...GE.LKTS EAVSS..TLT GTE..A.TLT LVGFV.VL.Y IAL.ILTALY VL.KW...LI
  VcII ..PYMYALIG MT...FSGWV ATIAGWYVTE IGRQPWLVS. ...GV.LRTA EAVT..P.VA SSSVGI.SLT L..YL.IT.Y VVL.LVAYVH TL.FYHAR..
    Ef MLWIV..ALC .TFA.PF..L ANTTGWLVTE LGRYPWTVY. ...G.LFTIE QSVSPNVS.. VASLIT.SNV I....YFLLF AGLGSV.MVY LVI....LEL
   BsI YLRIM.IAL. ISF..PF..L ANSAGWIMTE IGRQPWTVM. ...G.LMTTA QSVSPNVT.. AGSLL.FS.I IAFGVMYMI. ..LGAL.LVF LFIR....EI
   Smc Y.WLL..ALW .TMAFPL..I ANSWGWIFTE MGRQPWVVY. ...G.VMQTR DAVSPGVS.. TAEVII.SMS V....FTLLY AVLAVI.EVK LLA....KYV
    Mt FSWL...AL. LTM..PAPFL ANSAGWVFTE MGRQPWVVVP NPTGDQL.VR LTVKAGVSDH SATVVATSL. ...LMFTLVY AVLAVI.WCW LL....KRYI
    Pg .RWFHMIAI. VCMPLAW..V ASQSGWIVAE VGRQPWTIQ. ....DLLPVQ AAVS.KLE.. AGSV....I. ITFFVFLVLF .SALLVAELN IMRKAIKK..
    Cj .LW...VCL. LSIPLGY..I AAEAGWIVAE VGRQPWAIQ. ....DLLPVH IAAT.QL... .GKV...NVQ ISFWIFAVLF .TALLIAEVK IMLTQIKK..
   Aca RPLLLKALLG .GLPLPW..I GIECGWFLAE YGRQPWAIY. ....EVLPVG VSAS.KLA.. PGDLWF.SIG L.IC...ALY .TLFLVVKMY LTF....KYG
    Hi IPLLLKVLLW .GLPLPW..I AIECGWFLAE YGRQPWATY. ....EILPVG VSAS.NLS.. TGDLWF.SIG L.IC...ALY .LAFIIVEMY LMF....KYA
   StI ..WLLRAALY .GIPLPW..I AVEAGWFVAE YGRQPWAI.. ...GEVLPTA VVNS.SLT.. VGDLLF.SMS L.ICG...LY .TLFLVAELF LMF....KFA
   EcI ..WLLRAALY .GIPLPW..I AVEAGWFVAE YGRQPWAI.. ...GEVLPTA VANS.SLT.. AGDLIF.SMV L.ICG...MY .TLFLVAELF LMF....KFA
    Kp ..WLLAAAFY .GLPLPW..I AVESGWFVAE YGRQPWAI.. ...GEVLPTA VANS.SLT.. AGDLIF.SMV L.ICG...LY .TLFLVAELF LMF....KFA
    Yp ..WLLRAALY .GIPLPW..I AVEAGWFVAE YGRQPWAI.. ...GEVLPTA VANS.SVT.. AGEILF.SMA L.ICG...LY .TLFLIAEMY LMF....KFA
    Av .PWLLKFALY .SLPLPW..I ATQTGWFVAE HGRQPWTIG. ...G.VLPTH LSAS.SLS.. TGDL.WGSLI ALI....AFY .TLLLVVEMY LM....IRFA
  EcII ..WVLKMALW .SLPLPW..I AIEAGWFMTE FGRQPWAIQ. ....DILPTY SAHS.ALT.. TGQLAF.SLI MIV....GLY .TLFLIAEVY LM....QKYA
    Rc ..WALRVAVF .AIPVPW..I AAETGWFVAE FGRQPWTVD. ...G.LLPTA LSVS.NLS.. ....IW.DVV LTLAGFVTLY .TVLLVIEIR LML....AAI
   VcI KPWVLKAALW .S~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   Aqa RRWLLKLFFY .SIPLP..IV ANILGWIVAE VGRQPWVVY. ...Y.ILKTK DAASP.LP.. AGQIL.TSII L....FSSIY FIVFLIF.LG LMVQKVIKGP
    Ct ILWILS...F .SVLCPE..L CNEIGWISTE VGRQPWVVY. ...G.LLKTK DATSPIVN.. AGQ.IWQSLI L....FSIIF ICLLSVF.VS LLL....KKI
    Tm ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  BsII ..WLL..IIF .MTAGPFSLI GIEFGWIFAC TGRQPWVIY. H....LLKTS DVVT...T.. TGSI....GV LFLF.FTFVY AVLGAAV.VY VLL....YYF
    Af .........G AAVP.....S VL..GWYVRE VGRKPWTVY. ...GILYP.E ELVTV.VGY. .GRSF..EFA VFMGVIIAAI AAFGMV.SMY FVATRGYRYL
  StII ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    Hs ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   Stc ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    Bp ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

         801                                             850        863
    Pa RKG..P..V. T...HEG.RE TNPGGAGQK. RTPARPLSAA GEGFDEEHDG HAQAADRDQD KRN
    Sy RNG..PD.L. TLPAPRG.SE ASLWGTVTKH ETNRRPAEIA .E.VKE~~~~ ~~~~~~~~~~ ~~~
   HNI R.G...E.LR SL....GVQE SSA.G...RW RGPL.PWVTS DD~~~~~~~~ ~~~~~~~~~~ ~~~
  VcII KTGHAPQ.V. S...PSSTKA AL~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
    Ef RKGP...... .....DYEAK KL.AKENEP. ALDPFDK.GV F.GE~~~~~~ ~~~~~~~~~~ ~~~
   BsI KKG.....AE HDNHHDV... ........PV STDPFSQE.V YHGISS~~~~ ~~~~~~~~~~ ~~~
   Smc KAGP...... .....P.E.. .L....TE.A DLNPPTKIG. ..GDLRDADK ..PMAFSY~~ ~~~
    Mt VEGP...... .......... .L.....E.H DAEP....AA .HGAPRD.DE VAPLSFAY~~ ~~~
    Pg ..GP..ET.E ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
    Cj ..GF..DA.H AGHTPLMGKG EK~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
   Aca RLGP..SALK TGRYY.FEQS SK~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
    Hi RLGP..SALK TGKYY.FEQS AK~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
   StI RLGP..SSLK TGRYH.FE.. .....QSTVT SQPAR~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
   EcI RLGP..SSLK TGRYH.FE.. .....QSSTT TQPAR~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
    Kp RKGP..SSLK TGRYH.FE.. .....QSSAA IQSAR~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
    Yp RLGP..SSLK TGRYH.FE.. .....QPTAP VQEAR~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
    Av RLGP..SSLH TGRYH.FEQL EQHAVKHASP SQ.ADPQQPV NA~~~~~~~~ ~~~~~~~~~~ ~~~
  EcII RLGP..SAMQ S......EQP TQQQG~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
    Rc RKGP...... T...HDIEVV EAWAAKRRPL TQPAE~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
   VcI ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
   Aqa KR.P..EEVA ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
    Ct GEGP..DEQD LIEV.DL~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
    Tm ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
  BsII RKHPVDEDLN TAES~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
    Af FRGDENE~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
  StII ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
    Hs ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
   Stc ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
    Bp ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
```

Fig. 1 (*continued*).

the heme–copper oxidases, cytochrome *bd* oxidases do not pump protons [14]. The fact that the enzyme does not pump protons may contribute energetically to its high affinity for oxygen and high catalytic efficiency [15]. In general, relatively little is known about what advantages using a cytochrome *bd* terminal oxidase instead of a heme–copper oxidase confers on an organism. In *E. coli*, cytochrome *bd* may be im-

Table 1
The reported sequences for cytochrome *bd*

Organism	Abbreviation	Domain	Classification	Q	Ref.
E. coli	EcI	Bacteria	D. Proteobacteria (Sc. gamma)	Yes	a
	EcII			Yes	b
A. vinelandii	Av	Bacteria	D. Proteobacteria (Sc. gamma)	Yes	c
H. influenzae	Hi	Bacteria	D. Proteobacteria (Sc. gamma)	Yes	d
A. actinomycetemcomitans	Aca	Bacteria	D. Proteobacteria (Sc. gamma)	Yes	e
K. pneumoniae	Kp	Bacteria	D. Proteobacteria (Sc. gamma)	Yes	f
S. typhimurium	StI	Bacteria	D. Proteobacteria (Sc. gamma)	Yes	g
	StII			–	h
Y. pestis	Yp	Bacteria	D. Proteobacteria (Sc. gamma)	Yes	i
V. cholerae	VcI	Bacteria	D. Proteobacteria (Sc. gamma)	Yes	j
	VcII			–	j
	VcIII			–	j
P. aeruginosa	Pa	Bacteria	D. Proteobacteria (Sc. gamma)	No	k
R. prowazekii	Rp	Bacteria	D. Proteobacteria (Sc. alpha)	–	l
R. capsulatus	Rc	Bacteria	D. Proteobacteria (Sc. alpha)	Yes	m
B. pertussis	Bp	Bacteria	D. Proteobacteria (Sc. beta)	–	n
C. jejuni	Cj	Bacteria	D. Proteobacteria (Sc. epsilon)	Yes	o
M. tuberculosis	Mt	Bacteria	D. Gram-positive (Sd. High GC)	No	j
S. coelicolor	Smc	Bacteria	D. Gram-positive (Sd. High GC)	No	p
B. subtilis	BsI	Bacteria	D. Gram-positive (Sd. Low GC)	No	q
	BsII			No	r
S. carnosus	Stc	Bacteria	D. Gram-positive (Sd. Low GC)	–	s
E. faecalis	Ef	Bacteria	D. Gram-positive (Sd. Low GC)	No	j
C. trachomatis	Ct	Bacteria	D. Chlamydia	No	t
P. gingivalis	Pg	Bacteria	D. Bacteroides and Cytophagales	Yes	j
Synechocystis sp. strain PCC6803	Sy	Bacteria	D. Cyanobacteria	No	u
T. maritima	Tm	Bacteria	D. Thermotogales	–	j
A. aeolicus	Aqa	Bacteria	D. Aquificaceae	Yes	v
H. salinarium	Hs	Archaea	F. Halobacteriaceae	No	w
Halobacterium sp. NRC-1	HNI	Archaea	F. Halobacteriaceae	No	x
	HNII			No	x
A. fulgidus	Af		F. Archaeoglobaceae	Yes	y

The abbreviations denote the different cytochromes *bd*, and are used in the text and in succeeding figures. Yp1 and Yp2 refer to two non-overlapping fragments. The classification follows that of Woese in [104] and [105]. The abbreviations used in the classification stand for Division (D.), Family (F.), Subdivision (Sd.) and Subclass (Sc.). Q indicates whether or not the sequence contains a 60-amino-acid region towards the C-terminal end of the Q-loop. (–) indicates that this portion of the sequence is not available. The sequence references are as follows: [a]GenBank: ECOCYD, D90713, ECAE000176, [71–73]; [b]GenBank: S63811, D90713, ECAE000176, [71,72, 87, 106]; [c]GenBank: AVICYDAB, [25]; [d]GenBank: HIU32787, [107]; [e]Actinobacillus Genome Sequencing Project, personal communication; [f]GenBank: KPCYDAB, [108]; [g]Genbank: AF001503, [17]; [h]GenBank: SYTRES, [109]; [i]These sequence data were produced by the *Y. pestis* Sequencing Group at the Sanger Centre and can be obtained from ftp.sanger.ac.uk/pub/pathogens/yp; [j]Sequence data were obtained through early release from The Institute for Genomic Research at www.tigr.org and/or through NCBI at www.ncbi.nlm.nih.gov.; [k]GenBank: PACIOAB, [22]; [l]GenBank: RPCYDB, RPZ82486, [110,111]; [m]http://capsulapedia.uchicago.edu/capsulapedia/Searches/BLAST.shtml; [j]GenBank: MTCY01B2, [112]; [n]These sequence data were produced by the *B. pertussis* Sequencing Group at the Sanger Centre and can be obtained from ftp.sanger.ac.uk/pub/pathogens/bp. [o]These sequence data were produced by the *C. jejuni* Sequencing Group at the Sanger Centre and can be obtained from ftp.sanger.ac.uk/pub/pathogens/cj; [p]These sequence data were produced by the *S. coelicolor* Sequencing Group at the Sanger Centre and can be obtained from ftp.sanger.ac.uk/pub/S_coelicolor/sequences; [q]GenBank: D83026, [113,114]; [r]GenBank: AF008220, [113,114]; [s]GenBank: STAPTSIA, [115]; [t]Chlamydia Genome Project, personal communication; [u]GenBank: D90904, [116]; [v]GenBank: AE000736, [105]; [w]GenBank: HSTBP, [117]; [x]GenBank: AF016485; [y]Genbank: AF2297, [68].

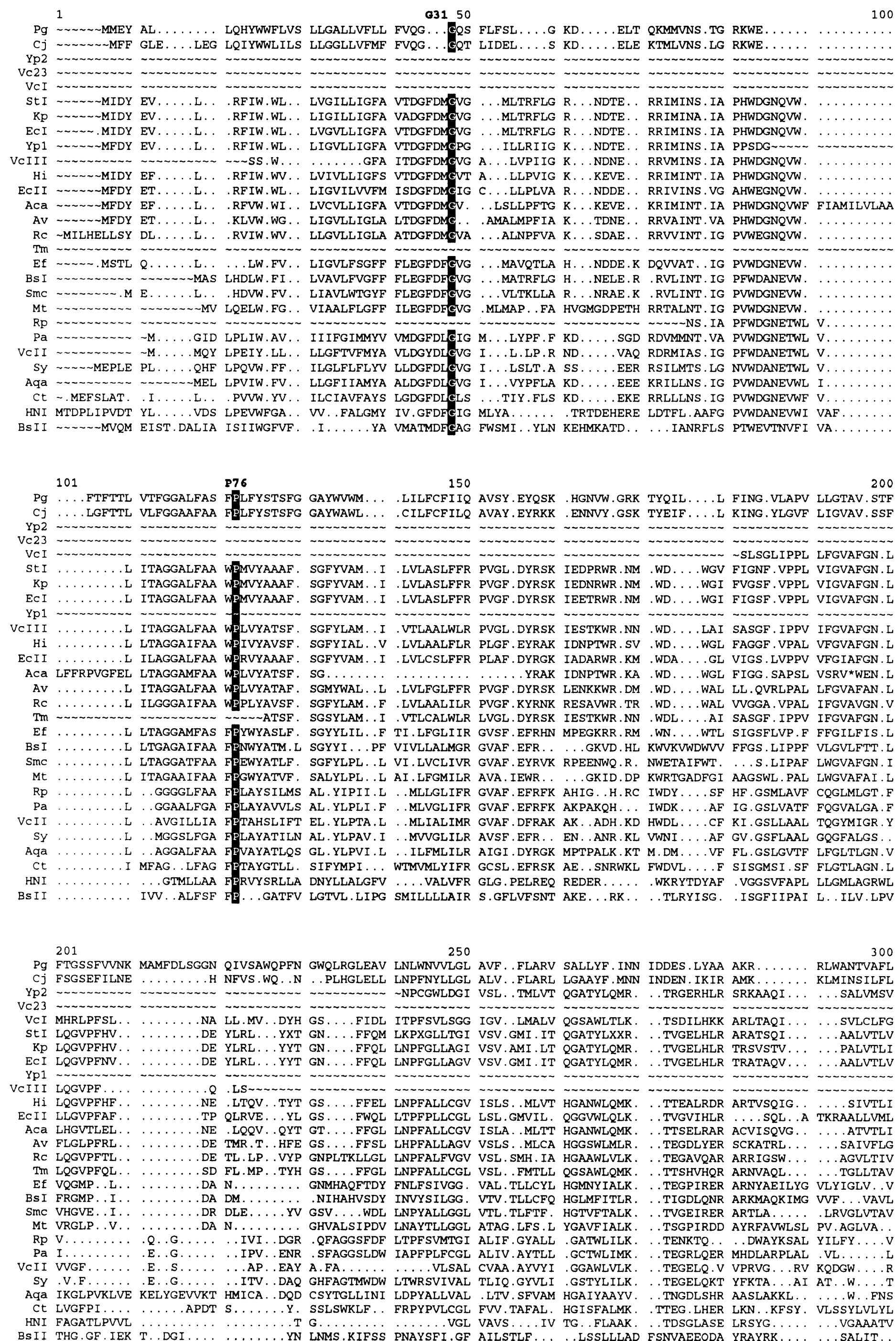


Fig. 2. Cytochrome *bd* subunit II alignment. Created using the same parameters as in Fig. 1.

portant during the transition between anaerobic and aerobic growth conditions, presumably scavenging deleterious molecular oxygen and other reactive oxygen species from the cell [16]. Other intriguing roles of cytochrome *bd* are also slowly coming to light. For example, in the alimentary tracts of young chickens, non-virulent strains of *Salmonella typhimurium* must have a functional cytochrome *bd* in order to

```
      301                                                 350                                                400
   Pg IFFLAYIAFL LTTEGFAV.. .........N PVSKEVYMEP YKYLYNFLD. MPAVLAVFL. LGVILV.LG. GI.GLTLLRK GFKRGIWLHG TGTVLTVLAL
   Cj PFFLGFLAWI FLKDGFSV.. .......DTN GV...VSMSA NLYLYNFLNQ M..IFAVLLV IGVILVLLG. MVQG.T...K GCSKAIFTLG LGTVLTVFAL
  Yp2 AFLLAGIWLV KGIDGFVITS VLD.TAAESX PMRKEVAHQA GAWLINFNKY PILWAL..PA LGVILPL..F TILLS..... RFEKGAWAF. LFSSLTIACV
 Vc23 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~A LGVSMPL..L SVLAS..... RLDKGGLAF. LTSSLGNAGV
  VcI LFIIGG.FMV HGMNGYLIVS SLD.YNAISN PLNKVVFQQA GAWLTNFERY PWMWFA..P. MGGLMMA..M GVFLS....A KWKNDAITF. VCASVTNACI
  StI CIALAGVWVM YGIDGYVVTS AIDHHTA.SN PLTKEVARET GAWLVNFNNA PILW.LV.PA LGVVLPL..L TILTSR.... .MEKGAWAF. LFSSLTLACI
   Kp CFALAGVWVY YGIDGYVVKS VMD.HTGPSN PLTKEVAREA GAWMVNFNNM PALW.AI.PA LGPWLPL..L TVISTK.... .ADKGAWAF. LFSSLTLACI
  EcI CFALAGVWVM YGIDGYVVKS TMDHYAA.SN PLNKEVVREA GAWLVNFNNT PILW.AI.PA LGVVLPL..L TILTAR.... .MDKAAWAF. VFSSLTLACI
  Yp1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
VcIII ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   Hi AFVLAGVW.L YSKDGYVVTS TIDHF.APSS PMNKEVAVET GAWFRNFNEM PILW..IFPA LAVVAALL.. .....NAAFS KANRCGFAF. FFSALTMAGV
 EcII CFLLAGYWLW VGIDGFVLLA Q.DA.NGPSN PLMKLVAVLP GAWMNNFVES PVLW..IFPL LGFFCPL..L TVMA.....I YRGRPGWGF. LMASLIQFGV
  Aca AFVLAGVWLS F.KDGFVVTS VIDH.NAPSN PIGKEVAVQA GAWFNNYKEM PILW..LFPV LAVGGALL.. .....NAIFS KANRCGFAF. LFSSLTMAGV
   Av CFFICGLWLL LGIEGQNLVD NFDPNVA.LN PLTKQVTLDN SGWQTNYVRY PL..TQFAPL LGLVG..GAL ALMGAQT... ..KRNGLAF. LGTSLAIIGA
   Rc TFALAGVWLA FGIPAYQFVT APDP.LGPSN PLLSEVVR.G GSWIDAYQTR P..WIAVAPA L.AVLGLGGA VVL.MR.... .L.WSGWAV. TASMIGIFGV
   Tm CFVAAGFWVQ H.IDGYMIVG AIDGNAA.SN PL~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   Ef .FAV....LM Y......FKT ..DFYE..KN .F..AVT... ...LI..... ....LTLAIV VLTVIA.N.V GVFKRK.... ....EMLAF. LASGLTLVVL
  BsI AFA....AL. .....SAYQT ..D....... MF....T... .......... RRGEITIPLA VLIVICFMLA AVFIRKKKDG WTFG.M.... TGAGLALT..
  Smc .LALA..FLL W......TQA ..DSGDA.KS .L...VA... ...LV..... ....VAVAAL VAALMA.NQA G....R.... ....EGWSF. ALSGVTIVAA
   Mt GFG......L WTQL..AY.. GKDW.....T WLVLAVAGCA QA........ .......AAT VL........ .VW.RRVSDG W......AF. MCTLIVVAAV
   Rp ALFMG.L.VS LSAPFLNNYI NHRWFS.... .........M PN.I..Y.YL SII..PIITV LI.FIKLI.. .....KAIKQ K..KEVKPF. IY..TILLFL
   Pa AI.IG.V.VS LWTPLAHAEI AARWFS.... .........L PN.L..FWFL P.V..PIL.V LVTFYALL.. .....RSVAN N..DHVKPF. VL..TLVLIF
 VcII R~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   Sy TLAGA.VFIT ISTPAFSEEA RAQLFT.... .........A PL.V..YIF. AAI..PLVGL LFIGL.LL.. .....RSLYL R..EENTPI. IW..TFLVFS
  Aqa ALLY..LIVH ILTMIYHPLL IKNYFT.... .......... .R.M..FIF. PAL..PI..V IVAFITII.. .....VALNR Q..KYSLAF. .WSSTAMIF.
   Ct SLLIATI.L. .GMP....QT L........G VCCRI..EGA PG.IPAY... PLI......I LLSVVTLSCC .YAEKRAV.. SIGKYGKAF. VLSC...INL
  HNI AY.LGGVVVL LGT....... .......... ....VVATDA GGAASAVLSL PV...AAVVA LSIVAGLGG. SVLARRGRYR .......AW. LASALALP.T
 BsII ....GPISLL FAV..CIMVT .......... .....MRNEA N.WLYSGMMN DFSWI.IASF ITFVIA.GIA LFLPNKSFGQ NIGKPRLALV ...AIGIQYF

      401                                                 450                                                 501
   Pg LLVAGWNDTS W.YP....ST YDLQ...... SSLTIENASS SHFTLKVM.. SYVS..LLIP FVLAYIFYA. WRALD.I... ..RKITKKEM EQDDHVY~~~~
   Cj LSSIGLGQSA F.YP....SL SDLQ...... SSLTLKNASS SYYTLSVM.. AYVS..LLVP FVLAYIIYV. WNAMDKV... ...KITREEI ANDDHAY~~~~
  Yp2 ILTAGXX... .XFPFVMPSS T......MPN VSLTMWDATS SLLTLKVMT. ..IVAIIFVP IILLYTSWCY XK....MF.. GR..IDKEFI ENNKHSLY~~~
 Vc23 IFTAGFA... .MFPFVMPSS L......NPA HSLTMWDATS SQVTLELMT. ..VVAVVMVP IILGYTIWSY YK....MF.. GR..LDDKFI EDNKNSLY~~~
  VcI ILTAGFA... .MFPFIMPSS F......EPS HSLTLWDATS SERTLNIMT. ..GVAFVMLP IILFYTAFSY RT....MF.. GR..LDKQYI ERNHHSLY~~~
  StI ILTAGIT... .MFPFVMPSS TMMN....A. .SLAMWDATS SQMTLN.LM. TWVA.AVLVP IILIYTSWCY WK....MF.. GR..IAREHI ESNTHSLY~~~
   Kp ILTAGIA... .MFPFIMPSS TAMN....A. .SLTMWDATS STLTLN.VM. TYVA.IVFVP IILAYTTWCY WK....MF.. GR..ITREDI EKNTHSLY~~~
  EcI ILTAGIA... .MFPFVMPSS TMMN....A. .SLTMWDATS SQLTLN.VM. TWVA.VVLVP IILLYTAWCY WK....MF.. GR..ITKEDI ERNTHSLY~~~
  Yp1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~~
VcIII ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~~
   Hi IITAAVS... .MFPFVMPSS S......HPE QSLLMWDSTS SELTLTLML. ..IFAVVFVV IALAYTIWSY SK....MF.. GR..LDANFI DKNKHSLY~~~
 EcII IFTAGIT... .LFPFVMPSS ....V..SPI SSLTLWDSTS SQLTLSIML. ..VIVLIFLP IVLLYTLWSY YK....MW.. GR..MTTETL RRNENELY~~~
  Aca ILTAAIA... .MFPFVMPSI S......HPE MSLLMWDATS SKLTLTLM.. .FFLSLIFVV ILLSYTIWAY YK....MF.. GR..IDSSFI EDIKNSLY~~~
   Av ILTAGFA... .CSRSVMPSS ....I..DPA SSLTIWDAVS SQKTLGIML. ..IVAIIFVP IILGYTLWCY WR....MW.. GK..LNDQTI EANPHGLY~~~
   Rc ISTVGLS... .MFPFILPSS ..LNP..SA. .SLTVWDSSS SQMTLF.IM. L.VSTVIFMP IILAYTSWVY .KV...LW.. GK..VRPEDI SRNPNA.Y~~~
   Tm ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~~
   Ef VALLFSG... .LFPRVM... ..IGS..EGF .DLLIKDATS TPYTLK.IM. TWIS.LSILP FVLAYTAWSY Y.....IF.R KR..I.SQTA VPEGY~~~~~~
  BsI VGMIFIS... .LFPRVM..V SSLHS...AY .DLTVANASS GDYSLK.VMS ..IAALTLLP FVIGSQIWSY Y.....VF.R KR...VSHKE .PMTY~~~~~~
  Smc VAMLFLT... .LFPNVMTST ..LNA..D.W .SLTVTNASS SAYTLK.IM. TWLA.VIATP VVLLYQGWTY W.....VF.R KR..IGTQHL ADASH~~~~~~
   Mt VVLLFGA... .LYPNLVPST ..LNP..Q.W .SLTIHNASS TPYTLK.IM. TWVTAFFA.P LTVAYQTWTY W.....VF.R QR..ISAERI PPPTGLARRAP
   Rp LGYLGLAI.S .IWPYIVP.. ........Y. .KVTLENAAA VPESQSLL.. .LVGALIFLP VILGYTFYCY Y.....IF.R GKSS..SQPL Y~~~~~~~~~~
   Pa LGYSGLGI.S .LWPNIIP.. ........P. .AVSIWEASA PPQSQGFM.. .LVGALFIIP FILGYTAWSY Y.....VF.R GKVK..HGDG YH~~~~~~~~~
 VcII ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~~
   Sy LSFIGLGF.I .IFPNIIP.. ........P. .SVTIYEAAA APSSLVFM.. .LTFIGFLIP ILLFYNIYNY L.....VF.R GKIV..TD~~ ~~~~~~~~~~~
  Aqa .SAIGIAFAS .IYPTLVPSI YDENPFENPV HSITVFNSAS SEKTLLTM.. .LIIALIGVP MVLTYKFFVY R.....IF.W GKVKVPKEGG Y~~~~~~~~~~
   Ct LSPI.LAYNI LLFPNLLVST VD.N..R..Y .TMTVFNAAA ETRTLQHLV. TIV..LIGLP FVVAYAVYIY ..R...VF.R GK..TDFPSI Y~~~~~~~~~~
  HNI LLTLLVAV.. LLYPTVYPPT .......... .GLLVREAVV SPLALNLV.. ....TVLGFP VLLL.VLW.Y FKFLYGVF.S G..PIEGEGY GG~~~~~~~~~
 BsII L..ASYAYGR AHLPYMI... .......YP. .DVTVMSGFT EPATFRALFA TYI.....VA FIILFPGFFF WK....MFMR DKRYIRQEE~ ~~~~~~~~~~~
```

Fig. 2 (*continued*).

suppress growth and colonization by virulent *S. typhimurium* strains [17]. Also, increased production of cytochrome *bd* in *Klebsiella pneumoniae* elevates the level of nitrogen fixation by the organism [18]. Although immunological studies have indicated for some time that cytochrome *bd* is relatively widespread among the Gram-negative bacteria [19], the lack of sequence data has hampered studies on cytochrome *bd*.

The two-dimensional topology of cytochrome *bd* has been predicted based on Kyte–Doolittle [20] and Goldman–Engleman–Steitz [21] hydropathy plots of the amino acid sequence [22–26]. Subunit I was predicted to contain seven transmembrane helices with the N-terminus located in the cytoplasm and the C-terminus in the periplasm. Subunit II was predicted to have eight transmembrane helices with both termini found in the periplasm.

Partial proteolysis [27] and monoclonal antibody binding [28] studies found that a large, hydrophilic domain on the periplasmic side of the membrane [29] is necessary for quinol oxidation. This domain, known as the Q-loop, is located in subunit I, and covalent modification with a photo-reactive quinol analogue indicates that it contributes at least in part to the quinol binding site [30]. Proximity mapping using an artificial protease has demonstrated that the Q-loop is adjacent to a portion of the subunit II polypeptide located between the first two transmembrane spans [31]. Hence, this region of subunit II must also be located on the periplasmic side of the membrane.

Other approaches have also provided information about the protein topology. Based on thiol reactivity with Ellman's reagent [32], which reacts only with cysteines that are solvent accessible [33], none of the cysteines in the protein are solvent accessible. This would be consistent with a model in which the endogenous cysteines are all located in buried transmembrane regions of the protein [31]. The gene fu-

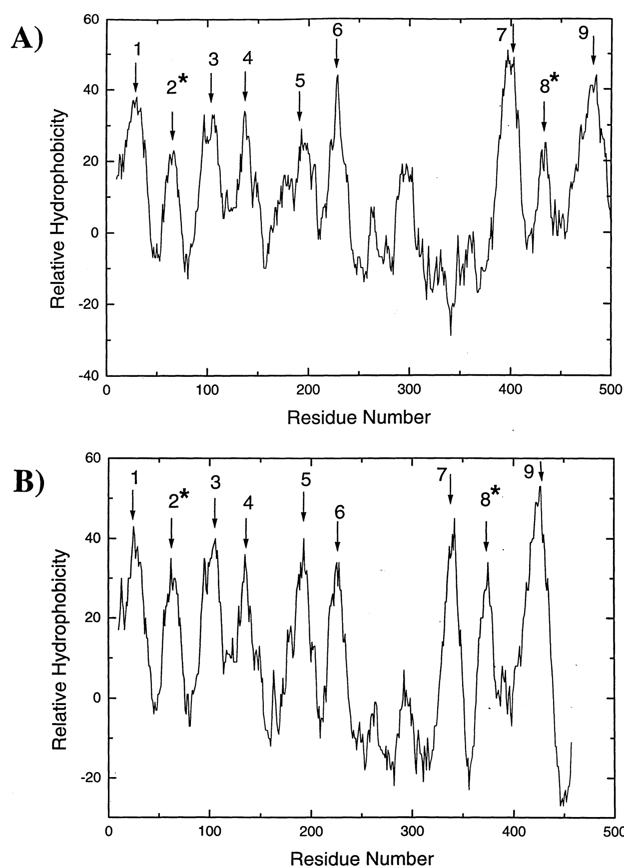


Fig. 3. Kyte–Doolittle hydropathy profiles. (A) *E. coli* cytochrome *bd*-I and (B) *B. subtilis* cytochrome *bd*-I. A window of 19 residues was used. Asterisk indicates transmembrane helices newly added to topology.

sion approach, using β-galactosidase [23] and alkaline phosphatase [26] fusions has provided considerable information about the topology of each of the two subunits. However, it has been noted that certain elements of these data do not fit the predictions based on the hydropathy profile for subunit I from *E. coli* [23,26].

The three-dimensional structure of cytochrome *bd* is known only at a very rudimentary level. It is a heterodimeric [34] integral membrane protein composed of subunits I and II, which are 58 and 43 kDa, respectively [35,36]. It contains three prosthetic groups: heme b_{558}, heme b_{595} and heme *d*. Hemes b_{558} and b_{595} are protoporphyrin IX while heme *d* is a chlorin [37].

Heme b_{558} is the initial electron acceptor from quinol [38]. Heme b_{558} is low-spin, six-coordinate and is located entirely within subunit I [39]. It has been shown that this low-spin heme has histidine/methionine ligation, with H186 in subunit I (I-H186) [40] and I-M393 [41] as the two axial ligands.

Hemes b_{595} and *d* appear to form a heme–heme binuclear center where the oxygen chemistry occurs [42,43]. Heme b_{595} is high-spin, five-coordinate and I-H19 has been proposed to be its axial ligand [40,44,45]. Heme *d* is high-spin and appears to be virtually always five-coordinate [45], even though this heme binds O_2, CO and cyanide [46–52]. A protein-based axial ligand for heme *d* has remained enigmatic. Although electron nuclear double resonance (ENDOR) studies indicate that the ligand is not nitrogenous when heme *d* is oxidized [53], electron paramagnetic resonance (EPR) work suggests that when heme *d* is reduced the ligand is nitrogenous [43]. Another EPR study, on oriented bilayers, indicated that hemes b_{558} and *d* are oriented at an angle of 90° with respect to the plane of the membrane while heme b_{595} is at an angle of 60° [54].

Progress using site-directed mutagenesis as a probe of structure and function of cytochrome *bd* has been slow because the relatively few sequences that have been available until recently, are closely homologous, leaving a large number of apparently conserved residues, the mutagenic targets of choice. Recently, however, a substantial number of sequences encoding cytochrome *bd* have become accessible from numerous bacteria as well as several archaea. In this report, these sequences are used to re-evaluate the topology of subunit I of cytochrome *bd*, assigning it nine transmembrane helices instead of the previous seven. This revised topology is compatible with most of the available structural data, and suggests that all three of the heme prosthetic groups are located near the periplasmic side of the membrane. On the periplasmic edge of one of the new proposed transmembrane spans is a newly revealed conserved region of subunit I, containing the sequence GRQPW. This is the most conserved region of the protein and it is an obvious candidate for participating in either the binding of quinol, or possibly, heme *d*. Additionally, phylogenetic analyses suggest that cytochrome *bd* has been horizontally transferred between prokaryotes a number of times, producing an evolutionary tree substantially different from the canonical one based on 16sRNA.

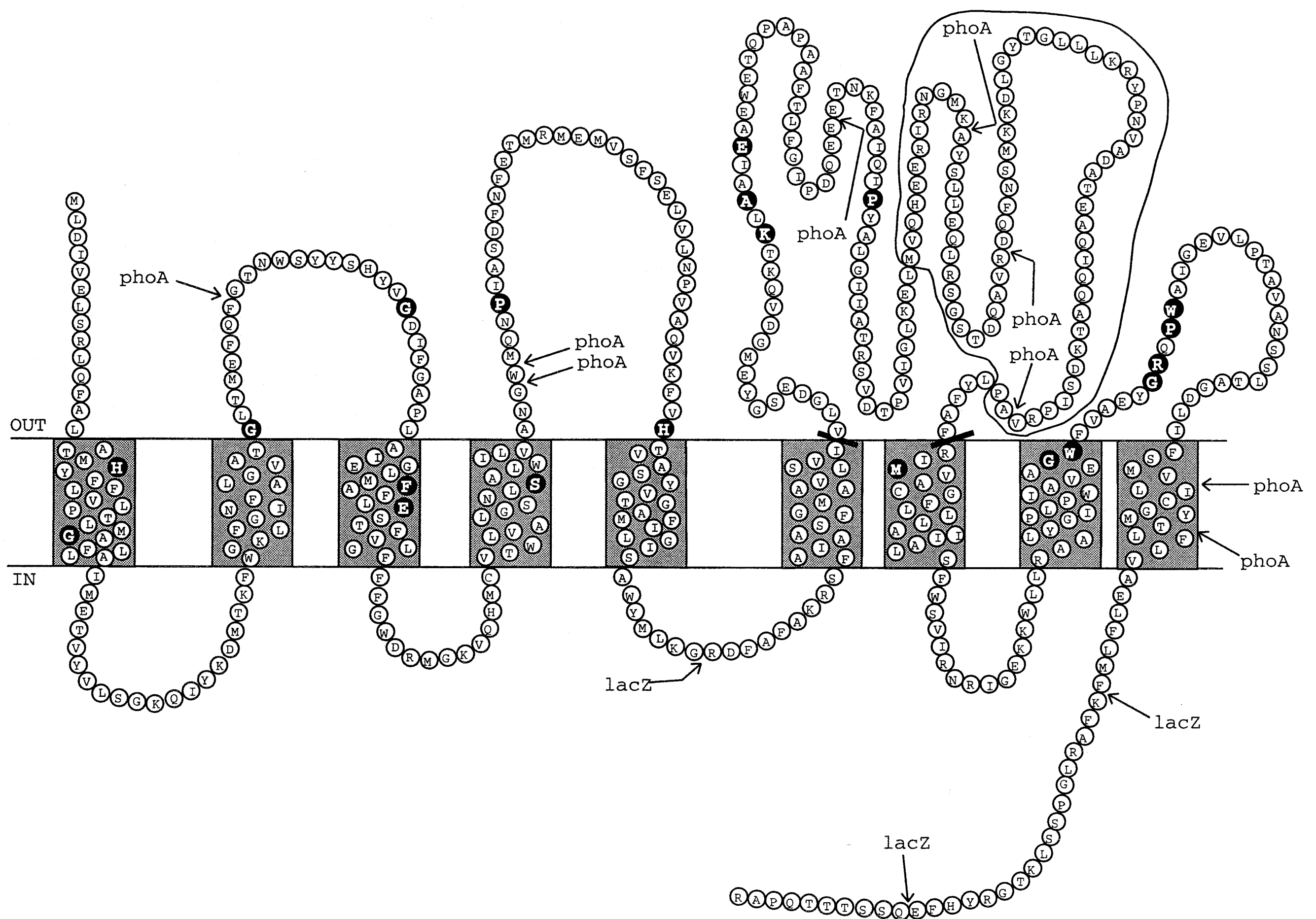


Fig. 4. Topological model of cytochrome *bd*-I subunit I from *E. coli*. The shaded boxes show the location of predicted transmembrane helices. The periplasm and cytoplasm are denoted by 'out' and 'in', respectively. Completely conserved residues are shown with reverse contrast. *Lac*Z and *Pho*A gene fusions with high activity are indicated. At the end (C-terminus) of helix VI and at the beginning of helix VII, the start and end, respectively, of the Q-loop are delineated with broad lines. The region of the Q-loop missing in several sequences is encircled.

2. Experimental procedures

2.1. Sequence analysis

Homology searches were performed with the BLAST 2.0 program [55] accessible at the National Center for Biotechnology Information (URL: http://www.ncbi.nlm.nih.gov) or using BLAST [56] in the GCG package (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI). The TBLASTN method was used for all searches. Sequences were obtained from GenBank [57], The Institute for Genomic Research (TIGR), or other services (see Table 1). All sequences were inspected for frameshift errors and corrected when possible. Multiple sequence alignments of the cytochrome *bd* sequences were performed using the program Pileup in GCG. Gap creation and extension penalties used were 3.0 and 1.0, respectively. At certain regions, the alignments had to be adjusted manually using the sequence editor in GCG. The resulting alignments were submitted to two different algorithms for predicting transmembrane helices in membrane proteins based on multiple sequence alignments: PHDtopology [58–61] (URL: http://www.embl-heidelberg.de/predictprotein/predictprotein.html) and TMAP [62]

(URL: http://www.embl-heidelberg.de/tmap/). Hydropathy profiles of single sequences were created using the Kyte–Doolittle algorithm [20] in the program GREASE [63]. Plots of amino acid location on a transmembrane helix were made using the program HelicalWheel in GCG.

To estimate the evolutionary distances, phylogenetic trees were created using four programs in the Phylip package [64,65]. SEQBOOT was used to bootstrap the sequences and create 100 data sets. Then, PROTDIST and the Dayhoff PAM matrix were used to create a distance matrix for the each randomly ordered data set. Next, NEIGHBOR was used to construct neighbor-joining trees [66] from the distance matrix. Finally, CONSENSE was used to select to determine the consensus tree. The trees were plotted with TreeView [67].

2.2. Sequencing

Sequencing and synthesis of all oligonucleotide primers used for sequencing was done by the Genetic Engineering Facility at the University of Illinois (Urbana, IL).

3. Results

3.1. Sequence analyses

The cytochrome *bd* sequences available are shown in Table 1. There are 22 complete and eight partial sequences for subunit I. For subunit II, there are 20 complete and eight partial sequences. Twenty-six different organisms are represented. Five organisms contain multiple cytochrome *bd* sequences. In addition, the complete genome of *Archaeoglobus fulgidus* contained a second subunit I sequence (Genbank: AF2296) [68] with some homology to subunit I. This sequence was not included in this analysis, however, because its C-terminus is shortened by over 100 residues and it is highly divergent. It is intriguing that no subunit II homologue was found in the *A. fulgidus* genome. Despite biophysical evidence to the contrary [69], a cytochrome *bd* sequence was not found in the complete genome of *Helicobacter pylori* [70].

The sequence for *E. coli* cytochrome *bd*-I has been reported three times. The two genomic sequences of *E. coli* recently reported [71,72] provide a sequence

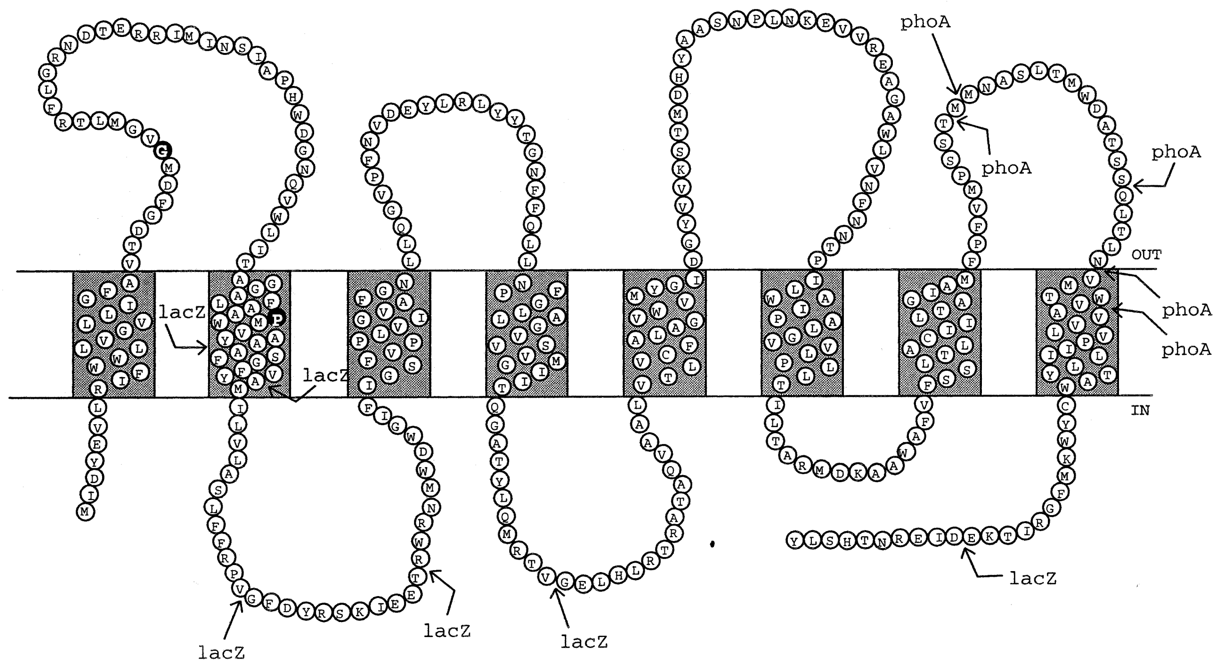


Fig. 5. Topological model of cytochrome *bd*-I subunit II from *E. coli*. See Fig. 4 for details.

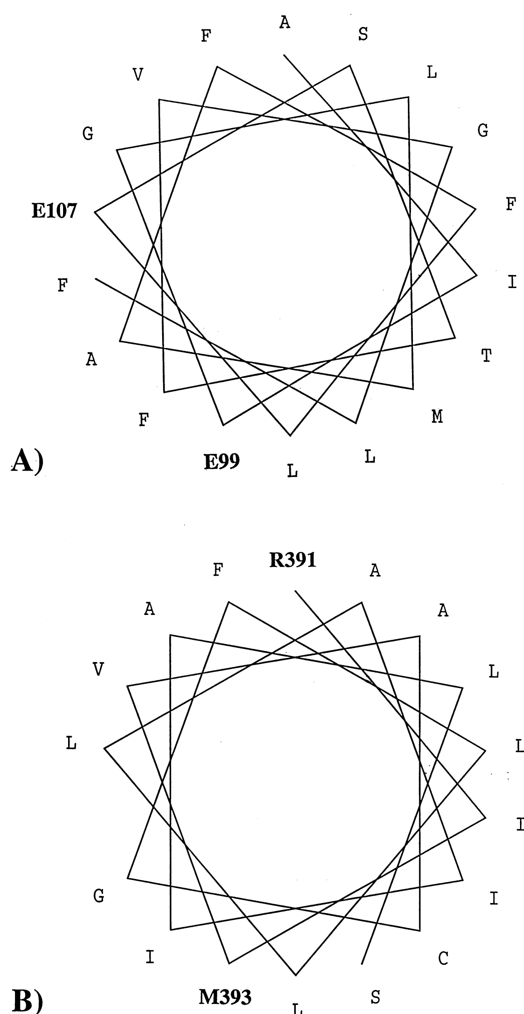


Fig. 6. Helical wheels of transmembrane regions of subunit I (*E. coli* bd-I). (A) helix III (E99, E107); (B) helix VII (R391, M393). Hydrophobic residues are boxed.

for cytochrome *bd*-I differing in three amino acids from the cloned sequence initially deposited [73]. All three amino acids are located in subunit I. With the genomic amino acid denoted on the left and the original sequence on the right followed by the residue number, they are ML121, FL213 and ML481. The cloned sequence was reexamined, confirming in part the original sequencing of these genes. Residues 121 and 481 in subunit I are, in fact, leucines. Residue 213 in subunit I, though, is a phenylalanine, indicating either that the original submission contained a sequencing error at this position or that a mutation was acquired at some point in the laboratory.

An alignment of the cytochrome *bd* sequences indicates that 20 residues in subunit I (Fig. 1) and two in subunit II (Fig. 2) are completely conserved. The heme ligands I-H19, I-H186 and I-M393 are the only conserved histidines and methionines (other than the initiating methionines). As noted previously for Pabd (*Pseudomonas aeruginosa*) [22], the C-terminal third of the Q-loop, approximately from I-L310 to I-P385, is also not present in a number of other sequences (Table 1).

The new topology prediction is based on multiple sequence alignments of the 22 complete sequences available for subunit I and the 20 for subunit II. Fig. 3 illustrates the differences in Kyte–Doolittle hydropathy profiles between individual sequences. The use of multiple sequences for the topology prediction allows the algorithm to see beyond these variations between sequences and produce a more accurate result. The algorithm predicts the location of transmembrane helices and whether loops are cytoplasmic or periplasmic. The predicted topologies for subunits I and II are shown in Fig. 4 and Fig. 5, respectively. Two new transmembrane helices are predicted in subunit I. There is a transmembrane helix beginning at I-W55 and ending with I-T69 and the N-terminus is now located in the periplasm. Also, there is a transmembrane helix beginning with I-R424 and terminating at I-W441, relocating the C-terminus of subunit I to the cytoplasmic side of the membrane. Additionally, transmembrane helix II in subunit II is highly unusual in that it is predicted to be some 32 amino acids long. It was manually truncated at 24 residues for Fig. 5.

Four highly conserved charged residues are located within transmembrane helices. I-E107 is completely conserved in all sequences and only two turns away from I-E99 on the same face of helix III (Fig. 6A). In helix VII (Fig. 6B), I-R391 is found on the opposite side of I-M393, a ligand for heme b_{558}. I-R391 is an asparagine in one sequence (Fig. 1) and histidine in two others, conserving a nitrogen group capable of hydrogen bonding. The transmembrane helices in subunit II do not contain conserved residues, except II-P76.

Unrooted phylogenetic trees were made using the largest continual stretches of good sequence alignment where data are available from the most organisms. For subunit I (Fig. 7), residues 63 to

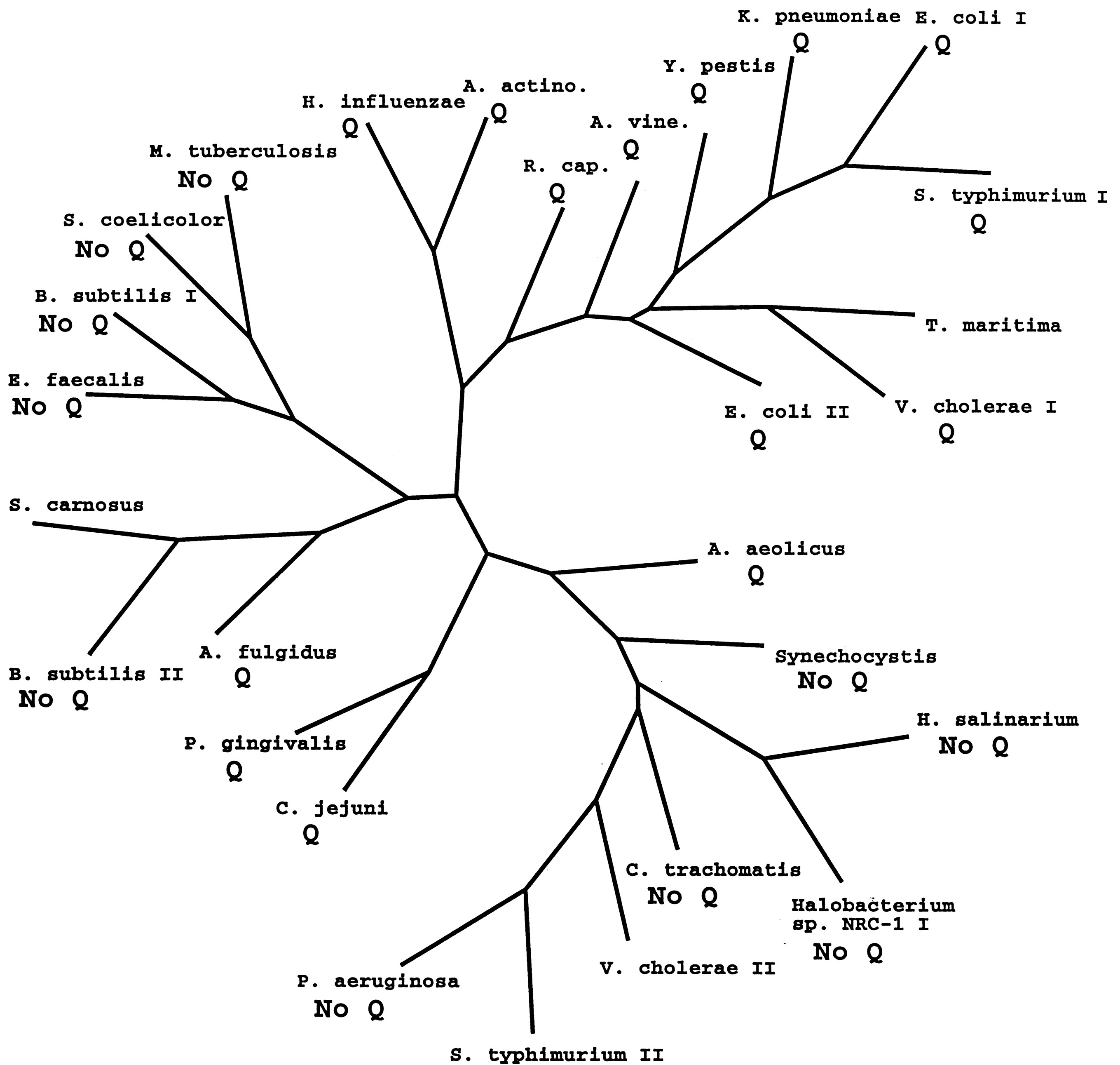


Fig. 7. Unrooted phylogenetic tree for subunit I. For sequences that are known to have or not have the C-terminal portion of the Q-loop, 'Q' or 'No Q', are respectively appended to the name. Residues 63 to 144 (*E. coli* cytochrome *bd*-I numbering) were used, the largest region of good alignment. The Q-loop region was not used for tree creation.

144 were used. For subunit II (Fig. 8), residues 51 to 214 were used. There are some expected groupings, such as that of the gamma proteobacteria clustering together, as do most of the Gram-positive bacteria, but there are also significant exceptions. For example, the Gram-negative *P. aeruginosa* and one of the *S. typhimurium* sequences are found in a completely different region. Also, the archaeon *A. fulgidus* groups with *Staphylococcus carnosus* and one of the *Bacillus subtilis* sequences. Furthermore, the extremely thermophilic bacterium *Thermotoga maritima* is found in the

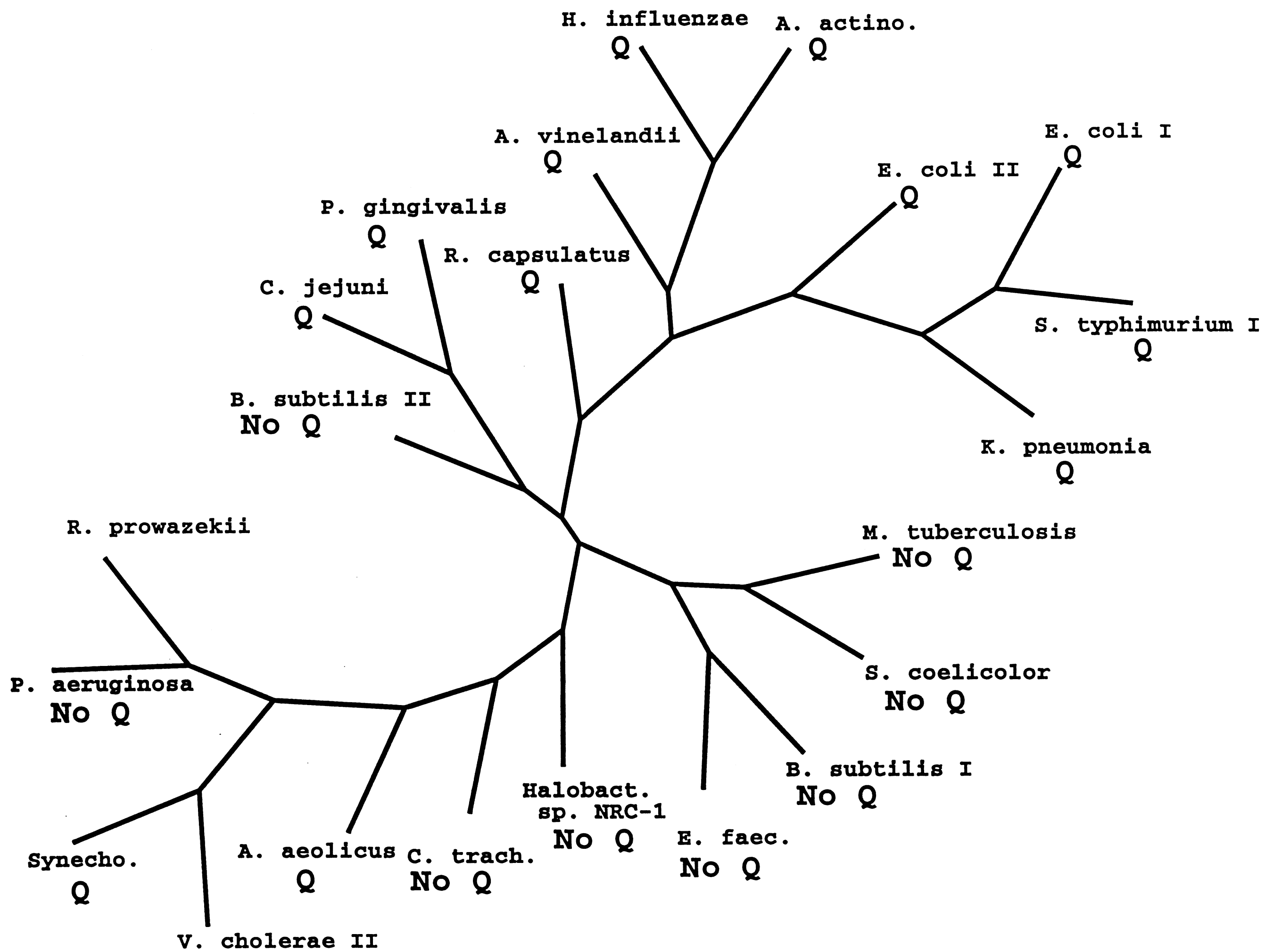


Fig. 8. Unrooted phylogenetic tree for subunit II. Residues 51 to 214 (*E. coli* bd-I numbering) were used, the largest region of good alignment. Sequences known to have or not have the C-terminal portion of the Q-loop have 'Q' or 'NO Q', respectively, appended to their name.

midst of the gamma subclass of Gram-negative bacteria.

4. Discussion

4.1. Topology

Information from *Lac*Z and *Pho*A gene fusions is complementary and defines the topology of the protein. The previously proposed topology does not agree entirely with the *Lac*Z and *Pho*A fusions [26] reported previously for subunit I of cytochrome *bd* from *E. coli*. In contrast, the revised topology proposed in the current work agrees well with the gene fusion data. The previous topology placed the C-terminus of subunit I on the periplasmic side, leaving the two active *Lac*Z fusions in the C-terminal tail unexplained. In the new topology for subunit I, the *Lac*Z fusions with high activity are found entirely on the cytoplasmic side of the membrane. In the old topology, the highly active *Pho*A fusion in the loop between helices II and III was located in the cytoplasm [26]. In the revised topology no PhoA fusions with activity are located on the cytoplasmic side of the membrane.

The revised topology remains consistent with results from chymotrypsin digestion [27] that localized the Q-loop to the periplasmic side of the membrane. It also is in accord with proximity mapping using an

artificial protease [31] which indicated that loop I–II in subunit II is close to and on the same side of the membrane as the Q-loop.

The topology prediction algorithms used are good at predicting the number and general location of transmembrane helices, with a reported accuracy of 86% [61]. It must be recognized, however, that the ends of the helices predicted are still imprecise and can shift several residues either way depending on which sequences are input to the program. In light of this, the following discussion will describe residues as roughly located in the cytoplasmic, middle or periplasmic third of the transmembrane helices.

One of the most interesting consequences of the revised topology is that it locates H19 in subunit I, the ligand for heme b_{595}, in the periplasmic third of the membrane. If H19 is in the periplasmic third of the membrane, then heme *d* must also be in this region, since heme b_{595} and heme *d* appear to share a binding pocket within the protein [42,43]. This situation, if correct, would be similar to that of cytochrome *c* oxidase [74,75], where the dioxygen-reactive site is located near the periplasmic side of the prokaryotic membrane (intermembrane space for the mitochondrial oxidase). Since the protons required in the chemistry of making water come from the opposite side of the membrane (bacterial cytoplasm), this necessitates at least one pathway for protons to reach the active site. Cytochrome *c* oxidase has at least two putative channels that allow protons access to its active-site, the heme–copper binuclear center [74–77]. In principle, the network of proton-conducting channels in cytochrome *bd* can be less complex than those in the heme–copper oxidases, since cytochrome *bd* does not pump protons.

It has been suggested [24] that a protonation site is located close to heme b_{595} which may be the immediate source for substrate protons, based on the sensitivity of the heme b_{595} EPR signal to pH [78] and the pH-dependency of its reaction with nitrite [79]. Although subunit II does not have any conserved, hydrophilic transmembrane helices, helical wheel plots of transmembrane helix III in subunit I suggest that one face of this helix has appropriately positioned protonatable residues. The completely conserved E107 (*E. coli* numbering) is two helical turns directly below E99 (which is a glutamine in only one sequence (Fig. 1)). It is conceivable that the two glutamates participate in a proton-conducting channel to the oxygen-reactive active site (heme *d*/heme b_{595}) from the cytoplasm. Also, T26 (subunit I, *E. coli* numbering), which is located in the middle of helix I below H19, is also highly conserved and could be part of a proton-conducting channel.

4.2. The quinol binding site

It has been demonstrated experimentally that the N-terminal portion of the Q-loop is somehow involved in quinol binding ([29,80]). The current sequence alignments indicate there are seven sequences of cytochrome *bd* in which 75 residues are missing at the C-terminal portion of the Q-loop. Hence, this part of the Q-loop probably is not important for quinol binding. The revised topology locates the highly conserved GRQPW region (Fig. 1), also on the periplasmic side of the membrane, close to the end of the new proposed transmembrane helix VIII in subunit I. Thus, this GRQPW region might contribute, along with the first part of the Q-loop, to a quinol oxidation site.

R391 in subunit I (*E. coli* numbering) is the only highly conserved, positively charged residue within the membrane. It is located on the opposite face of the putative transmembrane helix VII from M393 (Fig. 6B), which is the axial ligand for heme b_{558} [41]. The proximity to heme b_{558} along with the sequence location of R391 at the end of the Q-loop suggests that it could participate in quinol binding or electron transfer from quinol. Semiquinones have been shown to have a functional role in other proteins [81–83] and the positive charge of R391 could help stabilize a semiquinone anion species during turnover. A thermodynamically stable semiquinone has been observed in *E. coli* cytochrome *bd* [84]. Alternatively, this arginine might interact with the propionic acid groups of heme b_{558} as is the pattern observed in cytochrome *c* oxidase [75,85,86]. Mutagenesis experiments should clarify the role of R391.

4.3. The roles of multiple cytochrome bd oxidases

The adaptive responses of bacteria and archaea to a variety of growth conditions by using branched respiratory pathways and multiple terminal oxidases are well-documented [9,11]. Several bacteria have

been found to contain more than one cytochrome *bd* (Table 1). The second cytochrome *bd* from *B. subtilis*, although related to the other low (G+C)-content Gram-positive bacterium *S. carnosus*, is significantly divergent from the other bacterial sequences (Figs. 7 and 8). The incomplete genome from *Vibrio cholerae* reveals two partial but unique genes encoding overlapping regions of subunit I. Remarkably, *V. cholerae* contains three different incomplete genes encoding subunit II, indicating that there are three different cytochrome *bd* oxidases in *V. cholerae*. The two sequences encoding subunit I in *A. fulgidus* are located adjacent to each other in the genome and actually overlap a few bases. Although they are in different positive reading frames it is conceivable that they form a functional heterodimer, in which case the missing subunit II in *A. fulgidus* is substituted by a modified version of subunit I (AfbdI). It will be interesting to see the biophysical properties of any cytochrome *bd* oxidase from this archaeon. Relatively little is known, however, about why a microbe specifically uses multiple cytochrome *bd* oxidases [87]. Except for *Halobacterium salinarium* NRC-1 pNRC100 which has two identical sequences due to an inversion sequence [88], all these organisms contain divergent cytochrome *bd* sequences. There is a second cytochrome *bd* encoded in *E. coli* (cytochrome *bd*-II), that is also capable of oxygen reduction coupled to quinol oxidation. Cytochrome *bd*-II in *E. coli* is encoded by *App*Y [89,90]. However, expression of this protein has been achieved only under unusual conditions. Transcriptional activation of *App*Y was induced by a plasmid encoding a 3.4-kb region of DNA from the alkaliphilic *Bacillus firmus* OF4 in an *E. coli* strain lacking the respiratory oxidases that allow respiratory growth under normal conditions, (cytochrome *bd*-I and cytochrome bo_3) due to genomic deletion [87]. It is worth noting that in *E. coli*, cytochrome *bd*-II exhibits higher sensitivity to cyanide inhibition than does the well characterized cytochrome *bd*-I [87], suggesting a functional distinction. It seems likely that multiple cytochrome *bd* oxidases found in the same organism probably have different functional roles.

4.4. Phylogeny

Over time, naturally occurring single amino acid mutations tend to conserve only those amino acids essential for structure and function. Alignments of highly divergent sequences reveal conservation of residues that are indispensable. Larger changes in proteins are also possible, though, since lateral gene transfer can move even entire genes directly between organisms. In such cases, phylogenetic trees derived from those proteins will be different from the true evolutionary lineage of the organism. The 16sRNA phylogeny was the first to elucidate the three domains in the tree of life to be archaea, bacteria and eukaryotes and has been widely held as the standard phylogeny. Recent genomic data, however, has raised a considerable number of objections to partitioning the tree of life in this way [91–93]. It is therefore useful to compare the cytochrome *bd* phylogenetic trees with that of 16sRNA to further test its applicability as a general model.

It should be noted that the phylogenetic trees were created using residues outside the Q-loop (Fig. 4). It is of interest, therefore, that sequences possessing the complete C-terminal region of the Q-loop cluster entirely within two branches of the subunit I tree (Fig. 7) and almost entirely within one branch of the subunit II tree (Fig. 8). Moreover, organisms that do not have the sequence encoding the full C-terminus of the Q-loop dominate the rest of the tree. Since the phylogenetic trees were made without using the Q-loop region of the sequences, this provides independent support for the validity of the phylogenetic interpretation.

The phylogenetic analysis suggests that lateral gene transfer of cytochrome *bd* has occurred on several occasions. In support of this conclusion, a recent study of the *E. coli* strain MG1655 genome [94] predicted that the genes encoding cytochrome *bd*-II were transferred laterally into the genome. The cytochrome *bd* phylogenetic trees indicate that although most organisms of the gamma subclass contain a cytochrome *bd* similar to the two cytochrome *bd* sequences in *E. coli*, some of them (*S. typhimurium*, *V. cholerae*) also contain at least one other cytochrome *bd* sequence that is more closely related to those found in the cyanobacteria and the archaeal halobacteria. Although the phylogenetic tree based on subunit I sequences was created without using the Q-loop region of sequence, this separation of the gamma subclass is supported by the lack of the

C-terminal portion of the Q-loop in *P. aeruginosa* cytochrome *bd* (this portion of the sequence is not available for StbdII), in contrast to all the other cytochrome *bd* sequences from the gamma subclass of proteobacteria. Since the only cytochrome *bd* sequence known from *P. aeruginosa* clusters within this divergent group, this further predicts that complete sequencing of its genome will reveal a second cytochrome *bd* more closely related to the *E. coli* cytochrome *bd* sequences. It is worth noting that the *P. aeruginosa* cytochrome *bd* is proposed not to contain heme *d*, but, to utilize heme *b* instead [22]. This may prove to be a common feature of this grouping. Utilization of different heme groups for the same function has been extensively observed in the heme–copper oxidases [11].

Cytochrome *bd* sequences from the more ancient bacteria and archaea also provide ample evidence of lateral gene transfers. The ancient hyperthermophilic bacterium *T. maritima* is found in the middle of the gamma subclass of proteobacteria in the tree based on subunit I (Fig. 7). Likewise, the most ancient bacterium on the tree, *Aquifex aeolicus*, instead of clustering close to the archaea as expected, is found with *C. jejuni*, an epsilon subclass proteobacterium, and *P. gingivalis*, a member of the cytophagales. The archaeal halobacteria sequences are more related to the cyanobacteria and to some of the Gram-negative sequences than to the archaeon *A. fulgidus*. Similarly, *A. fulgidus* exhibits greater similarity to the low (G+C)-content Gram-positive bacteria than to the archaeal halobacteria sequences. This can only be explained by there having been a substantial amount of horizontal transfer [95] of cytochrome *bd* genes between domains, similar to other phylogenetic analyses of proteins that have placed archaeal branches among those of the low (G+C)-content Gram-positive bacteria [96]. The mode of horizontal gene transfer is, perhaps, more clear in the archaeal halobacteria, since it has been suggested that the plasmid on which the cytochrome *bd* sequences of *Halobacterium* sp. NRC-1 are found may be particularly susceptible to transfer between both archaea and bacteria [97].

As more sequence data are made available, though, it is becoming apparent that phylogenetic analysis of a single protein often fails to completely support the 16sRNA phylogeny, particularly if the enzyme studied is metabolic (cytochrome *bd*) or biosynthetic in function as opposed to informational (16sRNA) [96]. Often, lateral transfer of genes is predicted to have occurred and models are clearly emerging that the contents of genomes are highly dynamic. For example, in the 100 million years that *E. coli* has been evolving, approximately 1400 kb of DNA has been transferred into and 1400 kb has been lost from its genome [94,98]. The size, but not the contents, of the genome has remained relatively constant. Along these same lines, the phylogenetic trees reported here for cytochrome *bd* reveal no clear separation between the bacteria and archaea, in agreement with [91–93], but in contrast to what would be expected if the canonical 16sRNA tree were observed [99,100]. The trees are consistent, however, with the annealing theory of genomes, which states that the last common ancestor implied by the phylogenetic trees was actually a number of organisms with very high rates of horizontal gene transfer [101].

Phylogenetic analyses of the heme–copper oxidases have been used to show that a mutual ancestor of the archeal and bacterial enzymes was present before atmospheric oxygen became abundant, i.e., before photosynthesis arose [102,103]. Due to the apparent high degree of lateral gene transfer of cytochrome *bd* between organisms, however, the phylogenetic analyses presented here are unable to provide further evidence either for or against this intriguing hypothesis. It is hoped that as more sequences become available, studies of the phylogeny of cytochrome *bd* will be able to contribute to this debate.

Acknowledgements

The authors express gratitude to the following groups and individuals for making sequence data available prior to publication: The Institute for Genomic Research; Actinobacillus Genome Sequencing Project, and B.A. Roe, F.Z. Najar, S. Clifton and D.W. Dyer; Chlamydia Genome Project and R.S. Stephens, S. Kalman, C. Fenner and R. Davis; Rhodobacter Capsulatus Sequencing Project; and the *Y. pestis*, *B. pertussis*, *S. coelicolor* and *C. jejuni* Sequencing Groups at the Sanger Centre. Support for this work was provided by the NIH, HL16101 (to R.B.G.).

References

[1] P.A. Cotter, V. Chepuri, R.B. Gennis, R.P. Gunsalus, J. Bacteriol. 172 (1990) 6333–6338.
[2] C. Georgiou, T.J. Dueweke, R.B. Gennis, J. Bacteriol. 170 (1988) 961–966.
[3] C.W. Rice, W.P. Hempfling, J. Bacteriol. 134 (1978) 115–124.
[4] R. D'mello, S. Hill, R.K. Poole, Microbiology 142 (1996) 755–763.
[5] J.G. Koland, M.J. Miller, R.B. Gennis, Biochemistry 23 (1984) 445–453.
[6] M.J. Miller, R.B. Gennis, J. Biol. Chem. 260 (1985) 14003–14008.
[7] M.J. Miller, R.B. Gennis, Methods Enzymol. 126 (1986) 138–145.
[8] K.C. Minghetti, R.B. Gennis, Biochem. Biophys. Res. Commun. 155 (1988) 243–248.
[9] Y. Anraku, Annu. Rev. Biochem. 57 (1988) 101–132.
[10] Y. Anraku, R.B. Gennis, Trends Biochem. Sci. 12 (1987) 262–266.
[11] B.L. Trumpower, R.B. Gennis, Annu. Rev. Biochem. 63 (1994) 675–716.
[12] M.W. Calhoun, J.W. Thomas, R.B. Gennis, Trends Biochem. Sci. 19 (1994) 325–330.
[13] J.A. Garcia-Horsman, B. Barquera, J. Rumbley, J. Ma, R.B. Gennis, J. Bacteriol. 176 (1994) 5587–5600.
[14] A. Puustinen, M. Finel, T. Haltia, R.B. Gennis, M. Wikström, Biochemistry 30 (1991) 3936–3942.
[15] S. Jünemann, P.J. Butterworth, J.M. Wrigglesworth, Biochemistry 34 (1995) 14861–14867.
[16] B.S. Goldman, K.K. Gabbert, R.G. Kranz, J. Bacteriol. 178 (1996) 6348–6351.
[17] L. Zhang-Barber, A.K. Turner, G. Martin, G. Frankel, G. Dougan, P.A. Barrow, J. Bacteriol. 179 (1997) 7186–7190.
[18] L.F. Comaduran, F. Lara, M. Soberón, Biotech. Lett. 20 (1998) 489–493.
[19] R.G. Kranz, R.B. Gennis, J. Bacteriol. 161 (1985) 709–713.
[20] J. Kyte, R.F. Doolittle, J. Mol. Biol. 157 (1982) 105–132.
[21] D.M. Engelman, T.A. Steitz, A. Goldman, Annu. Rev. Biophys. Biophys. Chem. 15 (1986) 321–353.
[22] L. Cunningham, M. Pitt, H.D. Williams, Mol. Microbiol. 24 (1997) 579–591.
[23] C.D. Georgiou, T.J. Dueweke, R.B. Gennis, J. Biol. Chem. 263 (1988) 13130–13137.
[24] S. Jünemann, Biochim. Biophys. Acta 1321 (1997) 107–127.
[25] F. Moshiri, A. Chawla, R.J. Maier, J. Bacteriol. 173 (1991) 6230–6241.
[26] G. Newton, C.-H. Yun, R.B. Gennis, Mol. Microbiol. 5 (1991) 2511–2518.
[27] R.M. Lorence, K. Carter, R.B. Gennis, K. Matsushita, H.R. Kaback, J. Biol. Chem. 11 (1988) 5271–5276.
[28] R.G. Kranz, R.B. Gennis, J. Biol. Chem. 259 (1984) 7998–8003.
[29] T.J. Dueweke, R.B. Gennis, Biochemistry 30 (1991) 3401–3406.
[30] F.-D. Yang, L. Yu, C.-A. Yu, R.M. Lorence, R.B. Gennis, J. Biol. Chem. 261 (1986) 14987.
[31] J.B. Ghaim, D.P. Greiner, C.F. Meares, R.B. Gennis, Biochemistry 34 (1995) 11311–11315.
[32] G.L. Ellman, Arch. Biochem. Biophys. 82 (1959) 70–77.
[33] P.C. Jocelyn, Methods Enzymol. 143 (1987) 44–66.
[34] M.J. Miller, M. Hermodson, R.B. Gennis, J. Biol. Chem. 263 (1988) 5235–5240.
[35] J.B. Ghaim, P.H. Tsatsos, A. Katsonouri, D.M. Mitchell, R. Salcedo-Hernandez, R.B. Gennis, Biochim. Biophys. Acta 1330 (1997) 113–120.
[36] M.J. Miller, R.B. Gennis, J. Biol. Chem. 258 (1983) 9159–9165.
[37] R. Timkovich, M.S. Cork, R.B. Gennis, P.Y. Johnson, J. Am. Chem. Soc. 107 (1985) 6069–6075.
[38] G.N. Green, R.M. Lorence, R.B. Gennis, Biochemistry 25 (1986) 2309–2314.
[39] G.N. Green, R.G. Kranz, R.M. Lorence, R.B. Gennis, J. Biol. Chem. 259 (1984) 7994–7997.
[40] H. Fang, R.-J. Lin, R.B. Gennis, J. Biol. Chem. 264 (1989) 8026–8032.
[41] T.M. Kaysser, J.B. Ghaim, C. Georgiou, R.B. Gennis, Biochemistry 34 (1995) 13491–13501.
[42] J.J. Hill, J.O. Alben, R.B. Gennis, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 5863–5867.
[43] H. Hori, M. Tsubaki, T. Mogi, Y. Anraku, J. Biol. Chem. 271 (1996) 9254–9258.
[44] S. Hirota, T. Mogi, Y. Anraku, R.B. Gennis, T. Kitagawa, Biospectroscopy 1 (1995) 305–311.
[45] J. Sun, M.A. Kahlow, T.M. Kaysser, J.P. Osborne, J.J. Hill, R.J. Rohlfs, R. Hille, R.B. Gennis, T.M. Loehr, Biochemistry 35 (1996) 2403–2412.
[46] B.C. Hill, J.J. Hill, R.B. Gennis, Biochemistry 33 (1994) 15110–15115.
[47] M.A. Kahlow, T.M. Loehr, T.M. Zuberi, R.B. Gennis, J. Am. Chem. Soc. 115 (1993) 5845–5846.
[48] M.A. Kahlow, T.M. Zuberi, R.B. Gennis, T.M. Loehr, Biochemistry 30 (1991) 11485–11489.
[49] R.M. Lorence, R.B. Gennis, J. Biol. Chem. 264 (1989) 7135–7140.
[50] R.K. Poole, C. Kumar, I. Salmon, B. Chance, J. Gen. Microbiol. 129 (1983) 1335–1344.
[51] R.K. Poole, I. Salmon, B. Chance, J. Gen. Microbiol. 129 (1983) 1345–1355.
[52] J. Sun, J.P. Osborne, M.A. Kahlow, T.M. Kaysser, J.J. Hill, R.B. Gennis, T.M. Loehr, Biochemistry 34 (1995) 12144–12151.
[53] F.S. Jiang, T.M. Zuberi, J.B. Cornelius, R.B. Clarkson, R.B. Gennis, R.L. Belford, J. Am. Chem. Soc. 115 (1993) 10293–10299.
[54] W.J. Ingledew, R.A. Rothery, R.B. Gennis, J.C. Salerno, Biochem. J. 282 (1992) 255–259.
[55] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang,

Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. 25 (1997) 3389–3402.
[56] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, J. Mol. Biol. 215 (1990) 403–410.
[57] D.A. Benson, M.S. Boguski, D.J. Lipman, J. Ostell, Nucleic Acids Res. 25 (1997) 1–6.
[58] B. Rost, C. Sander, R. Schnwider, CABIOS 10 (1994) 53–60.
[59] B. Rost, Methods Enzymol. 266 (1996) 525–539.
[60] B. Rost, R. Casadio, P. Fariselli, in: D. States (Ed.), The Fourth International Conference Intelligent Systems for Molecular Biology, AAAI Press, Menlo Park, CA, 1996, in press.
[61] B. Rost, P. Fariselli, R. Casadio, Protein Sci. 5 (1996) 1704–1718.
[62] B. Persson, P. Argos, J. Mol. Biol. 237 (1994) 182–192.
[63] W.R. Pearson, GREASE version 2.0u63, University of Virginia, 1996.
[64] J. Felsenstein, Cladistics 5 (1989) 164–166.
[65] J. Felsenstein, PHYLIP (Phylogeny Inference Package) version 3.5c, Department of Genetics, University of Washington, Seattle, 1993, distributed by the author,.
[66] N. Saitou, M. Nei, Mol. Biol. Evol. 4 (1987) 406–425.
[67] R.D.M. Page, Comput. Appl. Biosci. 12 (1996) 357–358.
[68] H.-P. Klenk, R.A. Clayton, J.-F. Tomb, O. White, K.E. Nelson, K.A. Ketchum, R.J. Dodson, M. Gwinn, E.K. Hickey, J.D. Peterson, D.L. Richardson, A.R. Kerlavage, D.E. Graham, N.C. Kyrpides, R.D. Fleischmann, J. Quackenbush, N.H. Lee, G.G. Sutton, S. Gill, E.F. Kirkness, B.A. Dougherty, K. McKenney, M.D. Adams, B. Loftus, S. Peterson, C.I. Reich, L.K. McNeil, J.H. Badger, A. Glodek, L. Zhou, R. Overbeek, J. Gocayne, J.F. Weidman, L. McDonald, T. Utterback, M.D. Cotton, T. Spriggs, P. Artiach, B.P. Kaine, S.M. Sykes, P.W. Sadow, K.P. D'Andrea, C. Bowman, C. Fujii, S.A. Garland, T.M. Mason, G.J. Olsen, C.M. Fraser, H.O. Smith, C.R. Woese, J.C. Venter, Nature 390 (1997) 364–370.
[69] R.J. Maier, C. Fu, J. Gilbert, F. Moshiri, J. Olson, A.G. Plaut, FEMS Microbiol. Lett. 141 (1996) 71–76.
[70] J.F. Tomb, O. White, A.R. Kerlavage, R.A. Clayton, G.G. Sutton, R.D. Fleischmann, K.A. Ketchum, H.P. Klenk, S. Gill, B.A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E.F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H.G. Khalak, A. Glodek, K. McKenney, L.M. Kitzegerald, N. Lee, M.D. Adams, J.C. Venter, Nature 388 (1997) 539–547.
[71] F.R. Blattner, G. Plunkett III, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gergor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, Y. Shao, Science 277 (1997) 1453–1462.
[72] T. Oshima, H. Aiba, T. Baba, K. Fujita, K. Hayashi, A. Honjo, K. Ikemoto, T. Inada, T. Itoh, M. Kajihara, K. Kanai, K. Kashimoto, S. Kimura, M. Kitagawa, K. Makino, S. Masuda, T. Miki, K. Mizobuchi, H. Mori, K. Motomura, Y. Nakamura, H. Nashimoto, Y. Nishio, N. Saito, G. Sampei, Y. Seki, H. Tagami, K. Takemoto, C. Wada, Y. Yamamoto, M. Yano, T. Horiuchi, DNA Res. 3 (1996) 137–155.
[73] N.G. Green, H. Fang, R.-J. Lin, G. Newton, M. Mather, C. Georgiou, R.B. Gennis, J. Biol. Chem. 263 (1988) 13138–13143.
[74] S. Iwata, C. Ostermeier, B. Ludwig, H. Michel, Nature 376 (1995) 660–669.
[75] T. Tsukihara, H. Aoyama, E. Yamashita, T. Takashi, H. Yamaguichi, K. Shinzawa-Itoh, R. Nakashima, R. Yaono, S. Yoshikawa, Science 272 (1996) 1136–1144.
[76] I. Hofacker, K. Schulten, Proteins 30 (1998) 100–107.
[77] A.A. Konstantinov, S. Siletsky, D. Mitchell, A. Kaulen, R.B. Gennis, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 9085–9090.
[78] S. Jünemann, J.M. Wrigglesworth, J. Biol. Chem. 270 (1995) 16213–16220.
[79] R.A. Rothery, A.M. Houston, W.J. Ingledew, J. Gen. Microbiol. 133 (1987) 3247–3255.
[80] T.J. Dueweke, R.B. Gennis, J. Biol. Chem. 265 (1990) 4273–4277.
[81] A. Magalon, R.A. Rothery, G. Giordano, F. Blasco, J.H. Weiner, J. Bacteriol. 179 (1997) 5037–5045.
[82] C.A. Yu, S. Nagaoka, L. Yu, T.E. King, Biochem. Biophys. Res. Commun. 82 (1978) 1070–1078.
[83] C.A. Yu, S. Nagoaka, L. Yu, T.E. King, Arch. Biochem. Biophys. 204 (1980) 59–70.
[84] S.F. Hastings, T.M. Kaysser, F. Jiang, J.C. Salerno, R.B. Gennis, W.J. Ingledew, Eur. J. Biochem. 255 (1998) 317–323.
[85] S. Yoshikawa, K. Shinzawa-Itoh, R. Nakashima, R. Yaono, E. Yamashita, N. Inoue, M. Yao, M.J. Fei, C.P. Libeu, T. Mizushima, H. Yamaguchi, T. Tomizaki, T. Tsukihara, Science 280 (1998) 1723–1729.
[86] C. Ostermeier, A. Harrenga, U. Ermler, H. Michel, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 10547–10553.
[87] M.G. Sturr, T.A. Krulwich, D.B. Hicks, J. Bacteriol. 176 (1996) 1742–1749.
[88] W.-L. Ng, S. Kothakota, S. DasSarma, J. Bacteriol. 173 (1991) 1958–1964.
[89] T. Atlung, K. Knudsen, L. Heerfordt, L. Brøndsted, J. Bacteriol. 179 (1997) 2141–2146.
[90] T. Atlung, L. Brøndsted, J. Bacteriol. 176 (1994) 5414–5422.
[91] E. Pennisi, Science 280 (1998) 672–674.
[92] D.-F. Feng, G. Cho, R.F. Doolittle, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 13028–13033.
[93] E. Mayr, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 9720–9723.
[94] J.G. Lawrence, H. Ochman, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 9413–9417.
[95] J.G. Lawrence, J.R. Roth, Genetics 143 (1996) 1843–1860.
[96] J.R. Brown, W.F. Doolittle, Microbiol. Mol. Biol. 61 (1997) 456–502.
[97] S. DasSarma, P. Arora, FEMS Microbiol. Lett. 153 (1997) 1–10.
[98] J.G. Lawrence, Trends Microbiol. 5 (1997) 355–359.
[99] G.J. Olsen, C.R. Woese, R. Overbeek, J. Bacteriol. 176 (1994) 1–6.

[100] C.R. Woese, Microbiol. Rev. 51 (1987) 221–271.
[101] C. Woese, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 6854–6859.
[102] J. Castresana, M. Saraste, Trends Biochem. Sci. 20 (1995) 443–448.
[103] J. Castresana, M. Lübben, M. Saraste, D.G. Higgins, EMBO J. 13 (1994) 2516–2525.
[104] A. Balows, H.G. Trüper,, M. Dworkin, W. Harder, K.-H. Schleifer, Vol. I, 2nd ed., Springer, Berlin, 1991.
[105] G. Deckert, P.V. Warren, T. Gaasterland, W.G. Young, A.L. Lenox, D.E. Graham, R. Overbeek, M.A. Snead, M. Keller, M. Aujay, R. Huber, R.A. Feldman, J.M. Short, G.J. Olsen, R.V. Swanson, Nature 392 (1998) 353–358.
[106] J. Dassa, H. Fsihi, C. Marck, M. Dion, M. Kieffer-Bontemps, P.L. Boquet, Mol. Gen. Genet. 229 (1991) 341–352.
[107] R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.-F. Tomb, B.A. Dougherty, J.M. Merrick, K. McKenney, G. Sutton, W. KitzHugh, C.A. Fields, J.D. Gocayne, J.D. Scott, R. Shirley, L.-I. Liu, A. Glodek, J.M. Kelley, J.F. Weidman, C.A. Phillips, T. Spriggs, E. Hedblom, M.D. Cotton, T.R. Utterback, M.C. Hanna, D.T. Hguyen, D.M. Saudek, R.C. Brandon, L.D. Fine, J.L. Fritchman, J.L. Fuhrmann, N.S.M. Geoghagen, C.L. Gnehm, L.A. McDonald, K.V. Small, C.M. Fraser, H.O. Smith, J.C. Venter, Science 269 (1995) 496–512.
[108] N.S. Juty, F. Moshiri, M. Merrick, C. Anthony, S. Hill, Microbiology 143 (1997) 2673–2683.
[109] V. Dartois, O. De Backer, C. Colson, Gene 127 (1993) 105–110.
[110] J.O. Andersson, S.G.E. Andersson, Microbiology 143 (1997) 2783–2795.
[111] S.G.E. Andersson, A.S. Eriksson, A.K. Naeslund, M.S. Andersen, C.G. Kurland, Microb. Comp. Genomics 1 (1996) 293–315.
[112] W.J. Philipp, S. Poulet, K. Eiglmeier, L. Pascopella, V. Balasubramanian, B. Heym, S. Bergh, B.R. Bloom, W.R.J. Jacobs, S.T. Cole, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 3132–3137.
[113] K. Yoshida, K. Shindo, H. Sano, S. Seki, M. Fujimura, N. Yanai, Y. Miwa, Y. Fujita, Microbiology 142 (1996) 3113–3123.
[114] A. Lapidus, N. Galleron, A. Sorokin, S.D. Ehrlich, Microbiology 143 (1997) 3431–3441.
[115] D. Kohlbrecher, R. Eisermann, W. Hengstenberg, J. Bacteriol. 174 (1992) 2208–2214.
[116] T. Kaneko, S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda, S. Tabata, DNA Res. 3 (1996) 109–136.
[117] J. Soppa, T.A. Link, Eur. J. Biochem. (1998) in press.