Evolution of b-type cytochromes in prokaryotes

Vassiliki Lila Koumandou^{a,1}, Sophia Kossida^{a,2}

Supplementary figure legends:

Figure S1: Phylogenetic reconstruction of all cytochrome-b sequences used in this study, except K00241.

The tree shown is based on PhyML analysis of 472 sequences from all orthology groups (except K00241) and all species. Numerical values at the nodes of the tree indicate statistical support based on 100 bootstraps. For species name abbreviations, and full details of accession numbers for all protein sequences used, refer to Table S1. All sequences are marked with a letter from "A" to "L" at the end to distinguish different orthology groups (see Table 1), and are also color-coded accordingly, as shown in the insert. Most sequences from the "A" (red) and the "B" group (blue) cluster together; "C" group sequences (cyan) cluster with the divergent "B" sequences; "D" and "F" sequences each form a small tight cluster (green, and purple, respectively) with good bootstrap support for the clustering of these two groups; "E" sequences (orange) are split into two clusters, one containing bacterial sequences, and the other containing archaeal sequences which group with archaeal "I" sequences (khaki); all but 2 "G" sequences (brown) cluster together; all but 11 "H" sequences (magenta) end up in one of two clusters; hypothetical "I" (khaki) sequences are dispersed, although most group with hypothetical "E" and "J" sequences (as might be expected); all "J" sequences (light blue) end up in one of three clusters interspersed with "I" sequences, except for two which end up in the

"G" cluster; "K" sequences (yellow) are dispersed and quite divergent, while most "L" sequences (grey) cluster together.

Figure S2: Phylogenetic reconstruction of cytochrome-b of the $b_d f$ and the bc complex, including sequences with a cytochrome-b domain which are not part of an orthology group.

The tree shown is the best Bayesian topology based on 205 sequences total, including cytochrome b_6 of the cytochrome b_6f complex fused to subunit IV (A: K02635/7), cytochrome-b of the bc complex (B: K00410/2, C: K03887/8, D: K03891), narC (F: K15879), as well as sequences with a cytochrome-b domain which are not part of an orthology group (E & I). Numerical values at the nodes of the tree indicate statistical support (posterior probability) based on 100 bootstraps. For species name abbreviations, and full details of accession numbers for all protein sequences used, refer to Table S1. All sequences are marked with a letter at the end to distinguish different orthology groups (see Table 1), and are also color-coded accordingly, as shown in the insert. Some of the most prominent groupings discussed in the text are annotated.

Figure S3: Phylogenetic reconstruction of the cytochrome-b subunit of ubiquinolcytochrome c reductase (bc complex), formate dehydrogenase, Ni/Fe dehydrogenase, and the cytochrome b556 subunit of succinate dehydrogenase.

A: The tree shown is the best Bayesian topology of 205 sequences total, including cytochrome-b of the bc complex (B: K00410/2), formate dehydrogenase (H: K00127), and Ni/Fe hydrogenase (L: K03620). Numerical values at the nodes of the tree indicate statistical support (posterior probability) based on 100 bootstraps. All sequences are marked with a letter at the end to distinguish different orthology groups (see Table 1), and are also color-

coded accordingly, as shown in the insert. The tree is the same as that presented in Figure 2, shown in star format for easier comparison with the tree in panel B.

B: The tree shown is the best Bayesian topology of 360 sequences total, including all sequences as in panel A (marked and color-coded as above), as well as 155 sdhC sequences (K00241). The six groups of K00241 sequences are annotated I-VI as in Figure 3.

Figure S4: Multiple sequence alignment of eukaryotic and prokaryotic cytochrome b561.

The alignment, including 22 eukaryotic cytochrome b561 (K08360) and 61 prokaryotic cytochrome b561 (K12262) sequences (dataset as in Figure 4), was generated using Muscle, and visualized in Jalview. The four helices (pink) predicted by PSIpred for the E. coli sequence, along with the 6 helices based on the structure of the Arabidopsis protein, as reported in Figure S3 of (Lu et al., 2014), are shown above the alignment. Red squares, blue dots, green triangles and pink diamonds correspond to histidine residues coordinating the iron atoms, amino acids that play important roles in substrate recognition on the cytoplasmic side and the non-cytoplasmic side, and amino acids that use their side chains to form H-bonds with the heme groups, respectively, as in Figure S3 of (Lu et al., 2014). Magenta bars correspond to previously reported substrate-binding motifs, ALLVYR and SLHSW, again as annotated in Figure S3 of (Lu et al., 2014). Conserved residues are highlighted in the alignment (blue: >80% agreement, mid blue: >60% agreement, light blue: >40% agreement; only the residues that agree with the consensus residue for each column are colored). The conservation annotation histogram below the alignment reflects conservation of the physicochemical properties, and marks absolutely conserved residues (score 11) with a yellow asterisk '*', and columns where physicochemical properties are conserved (score 10) with a yellow '+'; less conserved positions are shown in darker colors with decreasing score. The quality annotation histogram reflects the likelihood of observing a mutation in any

particular column of the alignment based on the BLOSUM62 matrix scores (for each column, the sum of the ratios of the two BLOSUM62 scores for a mutation pair, and each residue's conserved BLOSUM62 score, are normalized and plotted on a scale of 0 to 1) The consensus histogram reflects the percentage of the modal residue per column, and the consensus sequence logo is shown below the sequence for conserved regions ('+' denotes non-conserved residues and '-' denotes gap residues).

Figure S5: Phylogenetic reconstruction of eukaryotic and prokaryotic cytochrome b561, including sequences containing a cytochrome-b domain but not belonging to an orthology group.

The tree shown is based on PhyML analysis of 223 sequences total, including eukaryotic cytochrome b561 (K08360), prokaryotic cytochrome b561 (K12262 and K15879) sequences, prokaryotic sequences annotated as cyt-b561 but which are not part of an orthology group, and prokaryotic sequences containing a cytochrome-b domain but not belonging to an orthology group (see Table 1). K12262 sequences are marked with a "G" at the end, K15879 sequences with an "F", and the ones not belonging to an orthology group with an "E", "I" or "J" (see Table 1 for details). Sequences are also color-coded as indicated in the insert (E: orange, F: purple, G: brown, I: khaki, J: light blue). For species name abbreviations, and full details of accession numbers for all protein sequences used, refer to Table S1.





I: cytochrome-b/b6-domain-containing proteins



10 20 30 40 50 60 70 64	80 90 100 110 120 130 S L T A G K I P D V A S A G P F K R P G V L I S A V G T H V A A V A L L AMMWR T S L T A G K I P D V A S A G P F K R P G V L I S A V G T H V A A V A L L AMMWH T V L L L G	140 150 160 170 180 190 LKKALSPEAP LLTWR TFPL - HPFLMTLAFGFLSPVAAAAWRS - Y	200 210 220 230 240 250 2 		310 320 330 340 350 350 360 360 360 360 360 360 360 36	O O O 370 380 390 400 410 420 43 LITGILSLAGRGDNVKESDVGYKAAAVRG
262 MLQDSSPSAQRNLMAEDSSDDVER 85	S LTAGK I P DVA SAGP FK R P GVL I SAVGT HVAAVAL LAMMWHT VLLLGFGKAR HESNGFLMR I LY I TY F I LLA YHPASIW LATP TAAAT SP P T G S GK ANMDP AL I N FK V LY VL T Q L C G L T M I V L VATW NGG S T K T F I A S V S L I Q V L G I I A I V L T G IW ATP T A LP Y Y V A F S Q L L G L T L V AMT G A	L K K A L S P E A P L L T W R T F P L - H P F L M T L A F G F L S P V A A A A W R S Y		K GWAVH FQ S AH SWV - G I VAL AL F L LQ	YIVGFITYFFPGMPIPIRQ-LVMPF <mark>H</mark> QMF <mark>G</mark> VLIF-I <u>F</u> VSITV <mark>A</mark> MGI-SERAAW	GDN VKESDVGYKAAAVVALLLSLLLGLLLQR AEN YTHRR RTALAAVVVWAQAFHVIAGLVSSTIVVEVPKLDTNAEEEEARLINATEHP
-41		M S K F E S G F AWN E D P D K E F N Y - H P T F M I M G M V F L F G E A L L V Y V E F N W - H P L F M T I G F I Y L Y G N S I L I Y		I VNLVSLHSWI - GL SVVILYFAQ	YIVGFITYFFPGMPIPIRQ-LVMPF <mark>H</mark> QMF <mark>G</mark> VLIF-I <u>F</u> VSIT <mark>VA</mark> MGI-SERAAW-	
MSKDADVEQATKPESGVVQVEPATPEQPQAAVPTTTSTATTVTAVTNGEKPAAATPATTATPATAEQVQPATTTACT #6 51 51 3		MGK Y R G G F AWD G S A Q P K N - HP L CMV I G L VY L Y S E AM I VY			VVACEVAELADCIDENVDI AMMDI <mark>U</mark> IVECIECE V – LAIASAIMCI TEVALE	KHICWIKEGQMCAQQATSSFVGVFTFLYTVCVLLLVLNPRWKRQSLPEEEGLHHLTSSHSM
3	LPWYVAGSQVVGLACVVITGV	MGHYRGGYAWDG SAO FENV - HPLCMVLCLVEL VCDAVLVV		F P NMY S LH SW L - G L S A V I V F S LQ		- ATKI - PATSTEPPAGVLANVIGVMTVVFGALVVTLATEPSTKKKPTPEDTALLNSSSVNE IPK TYSKFVPEGVLINTLGIVLILLAAITVFVVTNKDFQREEPTAEQVALIQNSDEK - - NLGG KYSAFEPEGVLANVLGLLLACFGGAVLYILTRADWKRPSOAEEOALSMDFKTLTE
99MATSVSS	FFARISGLLAAALVIYW	ALVFKSSFL <mark>P</mark> QSTSQEDLIYAVL <mark>H</mark> PLLMVNGFI <mark>L</mark> ISGEAILVH		I S HMY S L H SWC - GM L T F V L F F L Q	WLLGLGFFLFPWASAQLKS-WYLPLHVFFGLLLL-VMSVGSCLLGI-TEKLLF- DGCSSHLAGWLMGFMSFWHRGEMRAVRI-TVLPWHIFLGLYTY-GLAVATAETGL-LEKITFL	NIMSTYSTFAPEGILANVLGLLLVCFGVLVGYVVTKEDFRRPPNPEEEALSVHFKTLSE/ .QTKKDVSKHSPESMVVNSLGLSLALLSGIIILAAVSPGYQSFRSELTYSESKKLMT
V		A L V F K S S F L P Q S T S Q E D L V Y A V L H P L L M V I G F I L I S G E A I L V H	RWLPGSRGFKKSVHLCLQGLALACGIFGIWTKFHGNDGIVAN	FFSLHSWM-GLICISLFGAQ	LAVATAETGL-LEKITFL FFSLQVDIELCSFQ WAACEVTYWYDCCSENSEA, SLMENNYSLCLSLY, A	QTKKDASKHSSESMVVNSLGLSLALLSGIIILAAVSPKFQRDLTYSESKKMMTI
43		T V H Y R G G L A L S S D N K D H T F N V - HP V MM V T G L T L F N G E AM L A Y		KGI - ENFYSLHSWL-GL - ACLFLFAFQ		LQVN-QVIIKYSTEAMLVNIMGVLILILGGFVILGVVIP VSGKDQV LTQ
5MAAGLGVMAAGLGV	EDEDEDEEEEEDSSMTIGVEALPLTFVAHALAVAAAVMVLV	C S F R G G L A F E A D N K N L I F N V - H P V L M L I G Y I I L G S E A I M I Y	KIFPKLNHDTTKLIHLILHAIAIVLGAVGIYCAFKFHNESO KSLPLKKEVKKVIHLVLHAIAMILGSVGIAAAFKNHNESN	I ANLY S LH SWL - G I GT I S LY G I Q	LTLATAELGL-LEKLTF- WIFGFVAFFYPGGSAIIRS-ESLPWHVLFGIFVY-VLTLATAELGL-LEKLTF-	- LQ S S GLDKY GA EA FLVN FT GLVVAL F GAAVVVAAVAP AHV E EP E GY AP I P VN LE S S G I DKY GP EALLVN FT AV I T I LY GA F V I L S V L GQ AP V EDDDK
0	FIHVSSNVNLVSVFLQWKQSASRLTVITHLFGILAIILMLVW	Q L - <mark>H</mark> P F L M L F G L I I I G G E A I M S Y	KSLPLKKEVKKLI <mark>H</mark> LVLHAIAIILGCV <mark>G</mark> IAAAFKNHNESN KTISSAMIVQKSIHMFLHLIALCLGIVGICAVFRFHDN	I A N L Y S L H S W L - <mark>G</mark> I T I I S L Y G I Q	WIYGFLVFFYPKGSPTLRS-ECLPWHVLFGIFVY-I LAVGNAALGF-LEKLTF- WVFGFFTFMFPKAGKQTRA- SMLPWHVCGGRALL-Y MATCAALTGL-IEKATF-	LES SGIAKYGSEALLVNFTAVVTVLYGAFVILSVLGQAPVEDDDSYSAI
2	VVAHLFGILAFILMLVW VVAQLLAAAVLTLTLV	L L H Y R G G I E Y H S D N P D R V F N A - H P F F M F C G F I F L V G E A MMA Y	RTVLSRGTVKKSTHMTLHLTALCLGTVGTCAVFKYHDMTH RTMLGSREVKKAVHLLLHLVALAFAAVGLYAAFKFHHD	A E D V Y S L H S W I - G L V T F C L F C L Q	WVLGFSAFMFPKAAPSTRA-SMLPWHVSGGRALL-YMAICAALTGL-MEKSTFL WLVGFVYFVFPGAVMTMRA-DYAPWHILLGIVIF-LMAICTAETGLARFIF- 	N L K N H R E A R L I N F T G L F V L L F G I F V D L S V A L A R Y V
6MAERTDDGTQPDWLRSPQPTTYSATDAAATLEA/ -169MKKEAPAGMKKEAPAG	AGDV EAASSSLP FLQK I QHLAGIASILILSPISIVLVAMWASSL	G G V SWA E G D A K R V F NW - HP I LMV S G Y A LM N V G V L I F R T S G T S S Y Q S S L S T V F E G M Q T G S P Q G - V T A L A W V H V V L G M A V L L F A C W R V G L R F A R G V P	I NGT SAVAYP D SR NGK K RG I MK T I HG S I WVVV F L FG I VAM L AVTK SHNDA I SGY I AN APAF EPKTP EQG F L A FVT H E L T F F L Y L A I L AMP I L G	L Y S F H SW - GV T V L S L F T LQ F		NC SYTVD SP DL FP I SNYGK I P DT CK I SHGLG I V I LAMGVCTT LA LAK FP V L
162	LLHWAVAVLIIPQFLLGENIAE GF <mark>HWLMA</mark> ALFAWQFAGAL	A F G K W M S S E P T P F N P A V V F - H V M T G L T I C A L V I W R L V L R L R H G A P L V G G S H F T L G F T L F V L V L L R G V W G L A N L H R R P	– – – – – – A A L – – – – D S G S P VMNK V <mark>A</mark> S G V Q H T MY L V L I L M V V S <mark>G</mark> – – – – – – – – – – – – – – – – – – –	T AW F - G	LILIHVAGAV-KGQFVL- VVLGHAAMAI-LHRVML-PADLLHVWLGFALL-AVVLGHAAMAI-LHRRLW-	KNGIIR RMMKAE
186MSSKTYIWDDRSG	GFHWLMALLFLWQFASA YSLISRSLHWGMALLFAMQFVTSA	I L R V F A K D M P I Y S F F W S A - H F Q L G F A L L V L A L L R G V W G L M N I H R R P		FSFL-GMPIFAKT	AIVGHVAMAL-VHHFAL- LVLGHIAMAVGWHFAL- LVLGHIAMAVGWHFIK-	R D D T L R Y M T G H G T P N
186MSHNCSKFHGNFMNQQHTS 186MPQHAQT	LFHWLVVALLIGLITTFY	G K LM K E G T DP A T K S D T V W - H E S L G T L V F A L T V L K L T W V A F K P A A I H D L T P D T S A L H - L S S I K L - H I S L G L L I F V V V I A R I G W S R L V G K P	EPVSQWR I LQ FAGKAFHVLLNLATLLIPATAVLALGSKAHP EPVSQWR I LQ FAGKAFHVLLNLATLLIPISGFLRVASKDRS	AEFF-GLFQIPSP	LVGLHAAAAI-YHHLIL- LVGLHVAAAL-WHQYIL- LVGLHVAAAL-WHQYIL-	RDHALERMLPWARRAA
-181		FRAAAADDVATE-AAPLAFLHMSFGLAILALMVVRLFIRRRTGAP	EP I DGAP R SQQL I AGAVHLL FYVVV FGM I A SGI G MML A SGAA ELP ELP EE I GRP FR I AAHTTQ I L FY AL LL AMP V FG I L FV E AHGR T	P A I F - SR DVTALP EF	LLIAHLAAAL-YHHLR VGKNHGVQE-VFAFLHFWGGIAVI-LLLIAHVGGAV-RHELR	
<i>181</i> MAEA MAEA MAEA MAEA	R L HWWT A G L V A T G F A I A W R P A Q R R L HWWT A G L V A T G F A I A W A L HWL NA A L V L A V L P L A W	I M V A V P L R D L L V K F L - - - - - - L F Q L - <mark>H</mark> K T I <mark>G</mark> L T V F A <mark>L</mark> V - - VW <mark>R</mark> L G L R N - - R H G R P - - - - - - - - - - - - -	– – – – – – A E A – – – – D G L S I A E R R A A A A G Q A L <mark>L Y</mark> G L L L A V P V L <mark>G</mark> Y L T A C S A – – – – – P S <mark>G</mark> F – – P – – – – – – P D – – – – – P Q A P R L L T L I G R V N <mark>H</mark> Y L <mark>L Y</mark> A I F L I MP I S <mark>G</mark> Y L – – L S A – – – – – – L S <mark>G</mark> R D T P	T L F L - <mark>G</mark> V I P I P S V	LAVGHALQAI-RHH LAVGHALQAI-RHH VLLHIAGTA-WHLVIR-	– – <mark>R</mark> R G R L – L L R Q M W R A P P R – – – – – – – – – – – – – – – – –
180MQREG	I I HWV LAV S I F F L F I S S WL HW SMALL I F VA I G L G I	WM L A L P L P S E D L V Y R A F P F Q L - H K N I G I T L V L L L G L L L Y V R F K H R P R	VEISKEYPRWMHRLAVADHVILYLLIFAVCISGYLSSSFSGWG PYRRPLGHLTEMAARAGHLALYGLMLFMPLSGYLFSGAGGYS	T TWW-WT V G F P NW	VAAVHISGAL-YHAFR	N D R V V R R M L R L
193MDTDNAEVW	FLHWVLAVLVUVVUFAL	TFGTLGYEGLGERFGEDTTDTTYTW-HKSFGVLVLLLMLVKVGLKLKHGKF TFGTLGYEGLSNLVGAETAGKLFTY-HKSIGVLVLLLTILRLGWRTRGTTP	PYDPPLSGVERAVGGVTLLLLYLLSLGVPIGGMVATAVSGYP	LEFF-TLSLESP		
205		GRGFLPKGDARALVLSV-HIATGFAIAAILLLRLAWRLTLGHKPP	ALP ADMTP A ER S A S H A V H F G LY A LML A V P L I G V V L A W LR G T T	L FAWR LDVPM	LAGLHALAAI-AHHALL- LAGLHALAAI-AHHALL- LAALHAAAAL-WHHFIR-	R D D T L N R M L P L I G D R L R A P R P W T R H T E R L L R R L G I H P T R A D
-184	F L HW I MA I I F I SAW I I G F A L HW L TAV L V L G L W L V G T	Y S G N F L N Y E T D G S F K -	– – – – – – E L P – – – – A T M S P L M K K L A H L G H L A L Y V I L I A L P I S <mark>G</mark> C F – – Y S W – – – – – – S A <mark>G</mark> H – – P – – – – – – – – – – – – – V E D P G P L H R L A K A T H A A L Y V L L L V I V G L <mark>G</mark> I A – – N A F – – – – – – V R <mark>G</mark> V – – S	A P V L - Y <mark>L</mark> F N I <mark>P</mark> Q L	MVAGHILAAL-KHHFID- LALFHALAAL-VHHYLW- CLAANVLM-A LALFHALAAL-VHHYLW-	K D N V L K S M I K Q S R
-186	LIHWLIFFAVIALFASMN VIHWVMAAAMIAAFVLGK VDADAD	Y A H G L E K T D P L R - R T L Y D W - H K A V G T A V L G L A V L R I L W I K I H G A P	DLV P S P R I T E V I A R I S H G L LY L L L L A V P I T G L G M T L F S G R O A P E G P R R D V I V A K A V Q G V LY L T M L A L P V L G M T A V L T S G A T P L P P S V T C M O P T A A H S H L A LY T L L V M P V A C V V P V K A C C F P	V	VVLIHVIGAL-WHQYFR- ALVLHVAGAM-KHALIL- ALVLHVAGAM-KHALIL-	HDATLGRMVPWLRGK
181MRIKNTST	VNHWVTAVLFIGMIALGI	Y M G S L P – K G P E R – S D – – – – – L Y D L – H K S I G V I I L G L F – – F F R I L W L K – – L S P N P – – – – – – – – – – – – – – – – –		V		
191 MADERPLT	VMHWLIAVLMIGNLLLGL TLHWTIGALVLTNLILGI	L A E K L P D E D I R F A I D T - <mark>H</mark> K S I <mark>G</mark> I T V L G L V L L R I L W R L G H R P P	– – – – – – P F S – – – – T GMAAW ER G L AH LAH F G L Y L FM L WM P L T <mark>G</mark> W L H D S AW – – – – – – K AA P E I P – – – – – – P L P – – – – A S V P GW E Q R AA R G L <mark>H</mark> WT F <mark>Y</mark> A L M I AM P L T <mark>G</mark> W I – – F S S – – – – – A G K Y – – P	L	GFVMALPADRKEMLHG-LFGEMHEIGGYLLI-A LLLLHIAGAL-KHQFID- LLVGHVGAAL-KHQFID- LLVGHVGAAL-RHHFIL-	GEPEFR RMWP R
1-179	TLHWVMAAAILALWVVGH SIHWLVFLLVIAAYCAME	M I D A L P – K G P V R – S E – – – – – V I G L – H K A I G V V M L V M A – – V A R L A W R L – – I R P Q P – – – – – – F R G F F P R S D R P – – – – – – – – – L I N M I H V S C G I S I L V L M – – V V R L L L R L – – K Y P T P – – – – – – – – – – – – – – – – –	PLPSSMSPGEQLLAKAGHVALYLLMLLIPMDGVLMSQSGGRE PIIPKPKPMMTGLAHLGHLVIYLLFIALPVIGLVMMYNRGN	VSVF-GLVLPAL	ILIGHVAAAL-RHRFIL- VIGLHAAAAL-RHRFIL- SEANFERVD-SLKSWHETLANLGY-FVIGLHAAAAL-AHHYFW-	R D D VMARM L P G K
178MRNTLARQR	FFHWAVFLLVALATLA	LKGFAGKGSPAR - SL AMAA - HEWTGLLVLVLA VPRLLWR L VRGAP	PAEPG-SRWMHLAGEAMHWVLYLFIFAQPILGLLTLNAGGHVLT PAEPG-SRWMHLAGEAMHWVLYLFIFAQPILGLLTLNAGGHVLT	L P G L - G L Q I P AW	VIGLALAAL-THHFTK- VIGLALAAL-THHFTK- VGPDDALKD-QLEEIHETLGNAFY-LVIGLAMAAL-FHHYML- VIGLHAMAAL-FHHYML-	
186 MT F R K T M D .82 MNWK N T A S		L R E L F D K G S V P R - E A L K S L - H F M L G L L M F A L V W L R L AMR A I Y P A P	RPQAALPQWQAMAGKLAHLVLYAIMIGMPLLGWLMLSAAGKP LIQPALPRSQEISSKLLHLALYALMIGMPITGWLLLSASGKP	I P F F - G L A L P A L	LIGGHAAAAL-FHHYVQ- LIGGHAAAAL-FHHYVQ- LGQNKELAS-QIKEVHEFVGTAGY-YLIGLHTIAAL-YHHYIK-	R D N T L L R MWP G S R K L A T
-182	GIHWLMLLLFIAVYACIE	LRVFYPKGSELR-EGLKTW-HFMLGMLLFALVWLRLAVRLSTITP	LIRPEPEKWQDLSAKLM <mark>H</mark> LALYALMIGMPLT <mark>G</mark> WLLLSASGKP YGGKE	I P F F - G F E L P A L	LIGESKDIAK-QIKEIHELVGTAGY-YLIGMHAAAAL-YHHYWL- LIAIHALAGL-WQHYVL-	R D N T L T R M L P P R K
181MLRNTS	LFHWLCALAVIAALVFIE	L K G N F P K G D P L R - S G L G Y A - H V Q A G L I V L L L V L P R L A W R L G N P P P	PHPPAP SPAMNLLAHAGHWALYALMLALPILGIAFIQANGRE PHPPQQPAWERLAAGATHWLLYGLMIAVPVTGYLLSTADGRP	V A L F - G L A L P A F	LSILHAAAAV-WHHRFL- LSILHAAAAV-WHHRFL- LVLLHIAAAL-KHALID- CPOLENOAD-LACDLHEWLAWTLL-CLVLLHIAAAL-KHALID-	
182	GLHWLTAIAVLTLFALGV	WMV D L S Y Y D P W Y R Q G P D I - H R S V G I L L FMAM V L R L V W R L L S P P P	RPLPNHQRWEVVSAHLAHVALYLLIFVAMISGYLISTADGSA KSLPGHQRWEKLAAGLVHRLLYLMLLGILISGYLISTADGRA	I	LAGVHALGAL-KHHLID- LAGVHALGAL-KHHLID- LVTLHGLGAL-KHHFID-	
180	T L HWVTA I AVIG L FAA G L 	WM V D L N Y Y S Q W Y – K P – – – – – – V P HW – <mark>H</mark> K S V <mark>G</mark> L L L A F A T – – I F <mark>R</mark> L L WR S – – F K G H P – – – – – – – – – WMM D L S Y Y S E W Y – R T – – – – – A P HW – <mark>H</mark> K S V <mark>G</mark> L L L A G <mark>L</mark> T – – L F <mark>R</mark> L I WK A – – L S S S P – – – – – – – –	– – – – – – – – – – – – – – – – – – –	I D I F - NW F S V P A L	LVVIHAIAAL-KHHFIN- GALFENQAD-IAGEVHYYLSLSLI-A LVVIHAIAAL-KHHFIN- LAGLHAVAAL-KHHFIN-	K D N T L T K M L G V K E K
188MAQKTETLQLKDSPA -188MALPFGDEAMALPFGDEA	VNHWGVALGMIALLGSGL ARYGAVSRIAHWLTALAVIALIGLGW	LMAY GP - LTR EA - VA P I R DW - H K A L G V V V L I V G LWR V GWR L AQ G A P VMAD L P L G AQ K L E L Y A L - H K S I G A L V LMVT L AR L GWR L AQ R G P	RPA SDMP PWLARAAKAV HWG LLATVLAMP I SGILI SV YNG RA SGH A EHAPWER VMAKAAHVG LYAGLLAMP LTGWI A SS AAN FP	V T S F - GV V I P A	LLVVHVAGAL-KHHFLD- LLVVHVAGAL-KHHFLD- VAP DQALRE-AAATLHGALAWGVV-GLIVLHAAGAV-KHHLID-	HDATLRRMLRGAA
.75		Y M A G L D - R S P F R - F E L Y G W - H K S F G F L A L L L L V A R L L I K R R S P I P M V T S L A D Y S T L V A L - H R P L G I L I L L L A G L R L A N R W R H A P P W M T G L S F D N P L R - G N V Y N L - H K A L G A L T V V L L A R L A W F R V O P P		V	LMLLHLAGAL-KHRLFDD 	R P Q A D V L P R M L
-182	FLHWCMTAMIVGLIGLGFLHWCMTAMIVGLIGLGF	YMGQQP - P S P EK - Y A L Y E L - HK S L G V V T V L L L F A R F I WR A A N P P P L T M T H A G L A P L R Q F Q L Y Q W - HK S V G I T V L A L T A L R V L WR L T H R P P	ELPTAY SQLLRTASHLGHFALYVLMFAVPVSGWVMSAAYGRP PHPAGMPARERRAAAAAHHLLYLLLVGLPLTGWAVVSLSPFNIP	ASLF-GVPIPP	VIAGHTLFAL-KHHFID- VIAGHTLFAL-KHHFID-	K D G L L W R M W V R C P
-186	ALHWLIALLIVCAFALGW GLHWIIALLIFAAFGLGW	VMT D I P G F T P T K - L K Y F S W - H K WM G V T V F A L A V I R V L WR A T H V P P	S L P G G T P AWQ R A A S H G V H I L L Y V L M I V I P V T G Y L Y S S A S N I P P V A A G T P AWQ A K A A G A H H L L Y V L I I V V P I T G Y L Y S L S A G V P	V VY L - <mark>G</mark> I VP L P R L	LVSLHVVAAL-KHQLLD- VVAVHAAAAI-KHQLLD- VVAVHAAAAI-KHQVVD-	R D G L L S R M L P F A K
188MTTASAPRPAR	ALHWIMGLSIVAVICVGW ALHWLLGLAIIGSFAVGV	Y M T G L P – F S P Q R – L K – – – – – L Y NW – H K W A G V A I L A L S – – V L R L L WR L – – T H R P P – – – – – – – – – – Y M T D L P – F S P Q R – L K – – – – – L Y NW – H K WA G V V I L A L S – – A L R L L WR L – – T H R P P – – – – – – – – – – – – – – – – –	ALPAAVLAAMP RWQQAAHHGVHYLMYAL FLIVPLVGWAYSSAAGFP ALPAEVEAAMP RWQRIAHHGTHHLMYLL FFAIPLTGWAYSSAAGFP	I V F L - G L V P L P D F	LVVLHVAAAV-KHQLID- LVVLHVAAAV-KHQLID- VCVNKELAE-LIKPLHKLAAFGLG-ALVLMHIGAAL-KHQWID-	R D G L V G R M L P A R G
180 MADR MADR MADR	LLHWLIALLIFAIFPLALLHWLIALLIFAIFPLAL 	YMHGLP - LSP LK - LK LYAY - HKWAGMIVLMLA LVKIYWKV THRPP YMHGLP - LSP FK - LK IYSW - HKWAGVTVFLLV WLRLAWR F GHRPP	AMPSAMGAAMRAMAHAGHLALYLLIFIIPIIGWLMSSAKGFP ALPSAMGAAMRAMAHAGHLALYLLMIAIPLSGWLMSSAKGIQ ALV	VVYF-GVLPFPDM		
-181	VLHWLVVLGIFAAFPLGV 	YMHELP - L S P Y K - LQ L Y S Y - H KW I G V T I F L L V A L R L AWR A G H P P	ALPPLPAWQRRASAAVHGLLYLLMVAIPLSGWLMSSAKGFP ALPDNLPRWQKIASHATHQLLYVLLFVVPLSGWLMSSAKGYK	TVWF-GVLPLPDL	LVVVHVAAAL-QHHFIE- LVVVHVAAAL-QHHFIE- LVVLHIAAVL-KHRLID-	R H P F L Q R M G W G R K E L S
vation						
	05267466564555450	0 0 0 0 0 0 0 0 0 0 0 0 0 0 - 0 0 0 0 0 5 - * 2 3 4 8 4 5 7 6 7 7 4 4 3 8 6 6 6 3 2		0 0 6 7 - 1 0 0 0 1 0 0 0	6557353569-363220-	0100004540000
uality		التعد مطالح الأطالات أعلي معجلاتهم	and the second	مستعلقها الالمسيحي ال	in the second	
insus						

