Research

A phylogenetic study of cytochrome *b*561 proteins

Wim Verelst* and Han Asard†

Addresses: *Laboratory of Plant Physiology, Department of Biology, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium. †Department of Biochemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0664, USA.

Correspondence: Wim Verelst. E-mail: wverelst@ruca.ua.ac.be.

Published: 28 May 2003

Genome ***Biology*** 2003, **4:**R38

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2003/4/6/R38

Received: 27 January 2003
Revised: 10 March 2003
Accepted: 24 April 2003

Abstract

Background: As an antioxidant and cofactor to numerous metabolic enzymes, ascorbate has an essential role in plants and animals. Cytochromes *b*561 constitute a class of intrinsic membrane proteins involved in ascorbate regeneration. Despite their importance in ascorbate metabolism, no evolutionary analysis has been presented so far on this newly described protein family.

Results: Cytochromes *b*561 have been identified in a large number of phylogenetically distant species, but are absent in fungi and prokaryotes. Most species contain three or four cytochrome *b*561 paralogous proteins, and the encoding genes usually have four or five exons. At the protein level, sequence similarities are rather low between cytochromes *b*561 within a single species (34-45% identity), and among phylogenetically distant species (around 30% identity). However, particular structural features characterizing this protein family are well conserved in members from all species investigated. These features comprise six transmembrane helices, four strictly conserved histidine residues, probably coordinating the two heme molecules, and putative ascorbate and monodehydro-ascorbate (MDHA) substrate-binding sites. Analysis of plant cytochromes *b*561 shows a separation between those from monocotyledonous and dicotyledonous species in a phylogenetic tree.

Conclusions: All cytochromes *b*561 have probably evolved from a common ancestral protein before the separation of plants and animals. Their phyletic distribution mirrors the use of ascorbate as primary antioxidant, indicating their role in ascorbate homeostasis and antioxidative defense. In plants, the differentiation into four cytochrome *b*561 isoforms probably occurred before the separation between monocots and dicots.

Background

Ascorbate (vitamin C) is generally known for its detoxification of damaging reactive oxygen species during aerobic metabolism and under stress conditions [1]. Through the modulation of levels of reactive oxygen species, ascorbate is implicated in the control of cell expansion, cell division and programmed cell death [2]. As a cofactor to numerous iron- and copper-containing oxygenases, ascorbate is also involved in the biosynthesis of essential molecules, such as the plant hormones ethylene and gibberellic acid, cell-wall glycoproteins and antimicrobial agents [3,4]. The recent unraveling of the ascorbate biosynthetic pathway in plants [5,6] has

resulted in a renewed interest in this molecule, and in the recognition that much remains to be learned about its regulation and metabolism.

While our knowledge of ascorbate biosynthesis and catabolism is rapidly expanding, little is known about the mechanisms by which ascorbate is regenerated throughout the plant cell. Cytochromes *b*561 constitute a newly identified class of membrane proteins possibly implicated in replenishing ascorbate pools in plant cells, essential to maintaining the physiological functions of this important molecule [7].

Cytochromes *b*561 are intrinsic membrane proteins containing two heme molecules, and reducible by ascorbate [8,9]. They have been suggested to function as electron transporters, shuttling electrons across membranes from ascorbate to an acceptor molecule. The one-electron oxidation product of ascorbate, monodehydro-ascorbate (MDHA) has been shown, at least *in vitro*, to function as an electron acceptor for mammalian and plant cytochromes *b*561 [8,10,11]. The cytochrome *b*561-catalyzed reduction of MDHA results in the regeneration of the fully reduced ascorbate molecule.

The presence of cytochromes *b*561 in plants was first demonstrated on the basis of their biochemical properties. Ascorbate-reducible cytochromes with a wavelength maximum near 561 nm and a typically high redox potential (E'_0 around +140 mV) were found in purified plasma membrane fractions from various species [12,13]. The availability of the primary sequence of the cytochrome *b*561 from bovine adrenal gland chromaffin cells [14,15] has recently resulted in the identification of homologous sequences in plants [7,16]. Putative cytochrome *b*561-encoding genes have now been identified in nearly all organisms for which considerable genomic sequence information is available, including invertebrates (insects, nematodes, platyhelminths, tunicates), vertebrates (mammals, amphibians) and plants (both monocots, dicots and gymnosperms) (this paper and [7,17]). The presence of cytochromes *b*561 in a wide variety of species indicates the general importance of this class of proteins in eukaryotic cell physiology.

Members of the cytochrome *b*561 protein family are characterized by a number of structural features, likely to play an essential part in their function [7,17,18]. They are highly hydrophobic proteins with six transmembrane helices, four conserved His residues, possibly coordinating two heme molecules, and predicted substrate-binding sites for ascorbate and monodehydro-ascorbate (MDHA). The occurrence of cytochromes *b*561 in species phylogenetically remote as nematodes and mammals, tunicates and insects, or amphibians and plants, is intriguing. The strict conservation of essential structural features suggests that the mode of action and physiological function of these proteins may be very similar. We present here a further analysis of the similarities and differences between cytochromes *b*561 throughout the animal and plant kingdoms, at the level of genomic organization and protein structure. Also, we examine the phylogenetic relations between all cytochrome *b*561 proteins identified so far, to allow hypotheses to be made on the evolution of these ubiquitous proteins.

Results

Analysis of cytochrome *b*561-encoding genes

Upon the completion of the *Arabidopsis thaliana* genome project in December 2000, it became possible to identify all genes encoding cytochromes *b*561 in a plant. Four putative *b*561 genes were identified in *A. thaliana*: *Artb561-1* to *-4* (for *Arabidopsis thaliana* cyt *b*561) [7]. Genomic sequences possibly encoding cytochromes *b*561 were also identified in *Oryza sativa* and *Craterostigma plantagineum*. Most putative *b*561-encoding sequences from other plant species, however, are so far only represented by expressed sequence tags (ESTs).

Genomic sequences for cytochromes *b*561 are also known for a number of animal species, including *Drosophila melanogaster*, *Caenorhabditis elegans*, *Homo sapiens* (human) and *Mus musculus* (mouse). The genomes of human and mouse each encode three cytochromes *b*561: one involved in ascorbate regeneration inside chromaffin granules (Hosb561-1, Mumb561-1), one present in the duodenum (Hosb561-2, Mumb561-2) that may function as a ferric reductase in the plasma membrane of duodenal mucosa, and a third that seems to have a ubiquitous distribution in mammalian tissues, but whose physiological function is not known. This is tentatively referred to as 'ubiquitous cytochrome *b*561' (Hosb561-3, Mumb561-3) (H.A., unpublished work).

All the *A. thaliana* sequences have a similar organization, with four exons of comparable lengths (*Artb561-1* is shown in Figure 1). The intron-exon structure of cytochrome *b*561 genes identified in *O. sativa* and *C. plantagineum* is similar (Figure 1). Mammalian genes coding for the duodenal cytochrome *b*561 isoform also have a similar organisation as their plant homologues, with four exons and three introns (Figure 1). Interestingly, however, genes encoding the two other mammalian *b*561 cytochromes have five exons (Figure 1). Cytochrome *b*561 genes with a different intron-exon structure within one species are also present in *C. elegans*. One of the *C. elegans* cytochrome *b*561 proteins (F55H2.5) is encoded by a gene (*Caeb561-1*) consisting of only two exons and one intron (Figure 1), while the two other genes in this nematode (F39G3.4 and F39G3.5; *Caeb561-2* and *-3* respectively) each have four exons. Similarly, two homologous genes have been identified in the fruit fly *D. melanogaster* (CG1275 and CG8776). The first has two different splicing variants, each with four exons, and the second gene has either five or six exons, also depending on splicing.

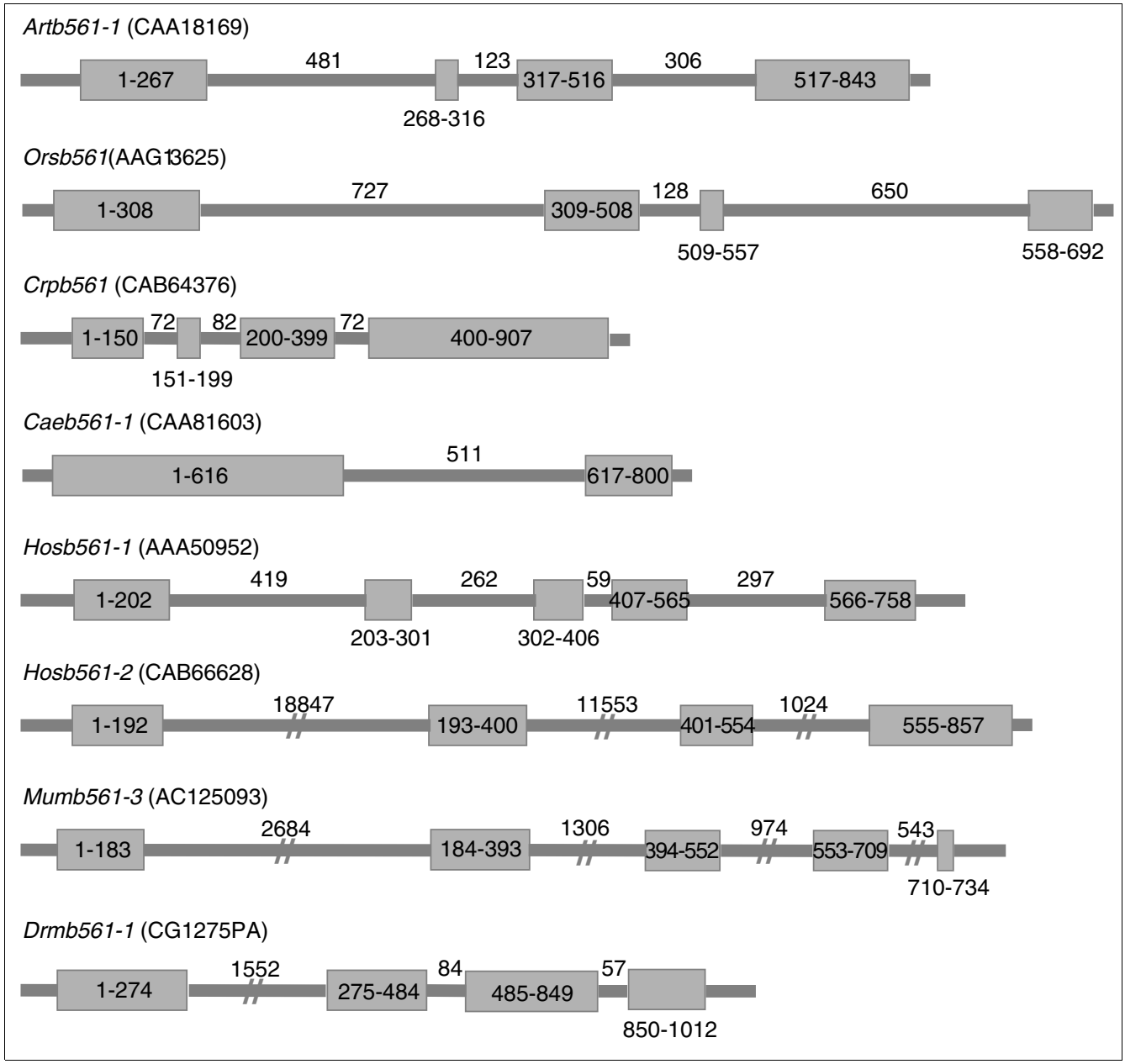

Figure 1
Genomic organization of cytochrome *b*561-encoding genes in different species. Species and genes shown are: *Arabidopsis thaliana* (*Artb561-1*), *Oryza sativa* (*Orsb561*), *Craterostigma plantagineum* (*Crpb561*), *Caenorhabdites elegans* (*Caeb561-1*), *Homo sapiens* (genes encoding chromaffin granule cytochrome *b*561 (*Hosb561-1*) and duodenal cytochrome *b*561 (*Hosb561-2*) respectively), *Mus musculus* (gene for the newly discovered 'ubiquitous' isoform, *Mumb561-3*), and *Drosophila melanogaster* (*Drmb561-1*). Boxes represent exons and lines between correspond to introns or untranslated regions. The numbers in the exons represent starting and ending nucleotide positions when introns are ignored; the numbers above the lines represent the length of introns in base-pairs.

The presence of five exons in the mammalian chromaffin granule cytochrome *b*561 has been considered to support a five transmembrane helix structure for these proteins [19]. However, the occurrence of cytochrome *b*561-encoding genes with very different gene structures indicates that this one-helix-one-exon model is probably not valid for all members of this protein family. Moreover, secondary-structure prediction routines almost invariably indicate the presence of six transmembrane helices for each of the identified putative cytochrome *b*561 genes. The correlation between predicted transmembrane structures and intron-exon structure is presented in Figure 2 for three different *A. thaliana* genes (*Artb561-1*, *Artb561-2*, *Artb561-4*), a homolog from rice (*Orsb561*) and the human chromaffin cytochrome *b*561

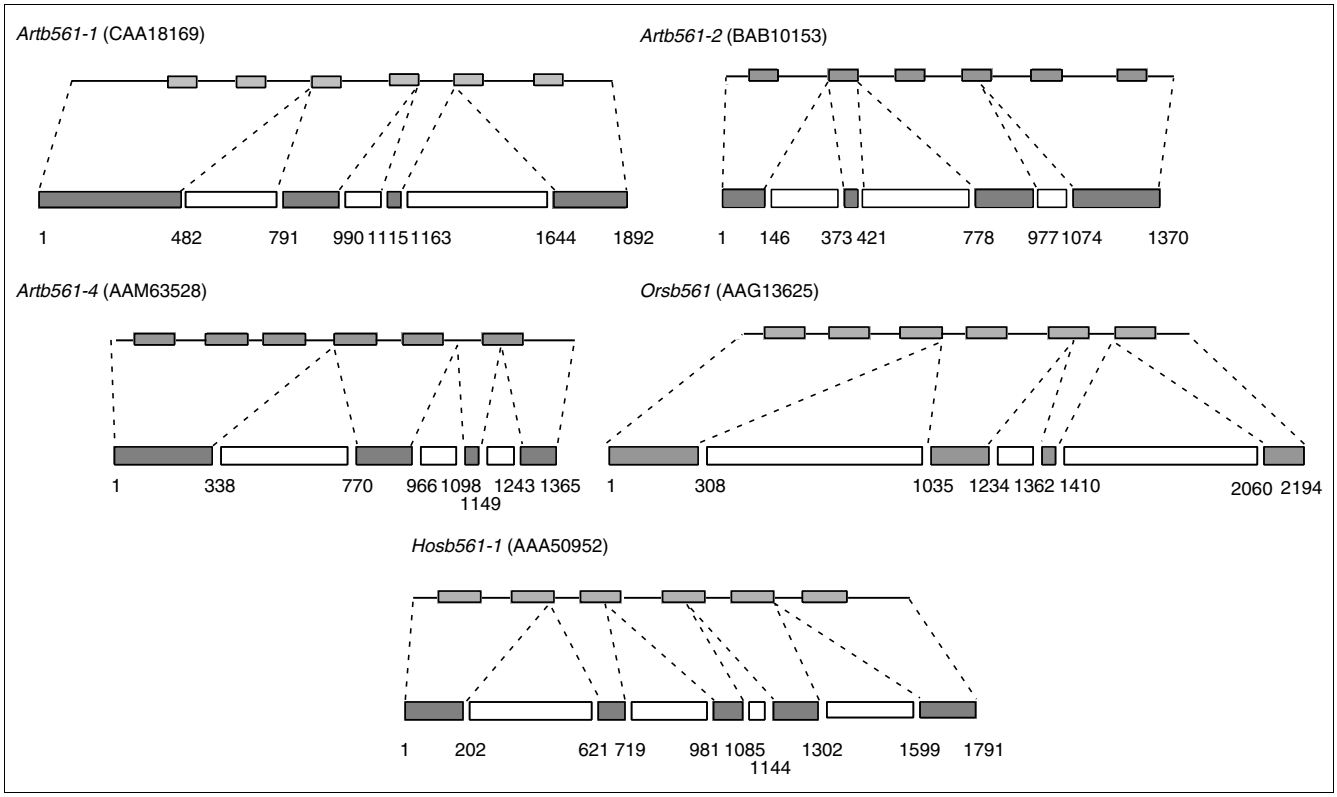

Figure 2
Correlation between transmembrane helices and exon-intron structure in plant and human chromaffin cytochrome *b*561 proteins and genes. For each protein-gene pair, the protein is shown in the top line with the transmembrane helices as gray bars, and the gene organization is shown below with exons as grey and introns as white bars. The first exon encodes one (*Artb561-2*), two (*Artb561-1* and *Hosb561-1*) or three (*Artb561-4* and *Orsb561*) transmembrane helices. Numbers represent nucleotide positions. Transmembrane regions were predicted with TMHMM [43].

(*Hosb561-1*). This analysis indicates that individual transmembrane helices may be encoded by different exons. For example, the predicted transmembrane helices 4 in Artb561-2, helix 6 from Artb561-4 and helix 4 in Hosb561-1 are each encoded by exons 3 and 4 (Figure 2). In rice, transmembrane helix 5 is partially encoded by exon 2 and exon 3. A similar situation, in which single transmembrane helices are encoded by more than one exon, was observed for the plasma membrane ATPase (BT002855) and an *A. thaliana* sodium channel (AY113938) (data not shown). The number of full transmembrane helices encoded by individual exons also varies among the genes. In the different cytochrome *b*561-encoding genes in *A. thaliana*, the first exon respectively encodes one, two or three helices (Figure 2).

Alternative splicing of cytochrome *b*561-encoding genes is observed in *D. melanogaster*. Different transcripts are found for each of the two putative genes (CG1275 and CG8776). The first gene has two transcripts (*Drmb561-1* and *-2*), each encoded by four exons. *Drmb561-1* is only different from *Drmb561-2* in that it has 86 extra amino acids at the amino terminus. This extra region (predicted to be cytoplasmic) is absent in all other cytochrome *b*561 proteins identified thus far. The second gene from *D. melanogaster* has three different transcripts, encoded by five (*Drmb561-3*) or six (*Drmb561-4* and *-5*) exons. Again, the only difference is seen at the amino terminus of the encoded proteins. Drmb561-5 is nearly identical to Drmb561-3, with only a difference in 20 amino-terminal amino acids. The sequence of the Drmb561-4 protein corresponds to that of the other two proteins, except for the lack of 80 amino acids at the amino terminus. At the protein level, Drmb561-4 thus proves to be the shortest isoform, lacking the first transmembrane helix. All six helices are present in the other two isoforms.

Comparison of putative cytochrome *b*561 proteins

The only cytochrome *b*561 protein that has so far been purified and sequenced, is the bovine chromaffin granule cytochrome *b*561 [14,15]. On the basis of sequence similarity to this protein, a large number of putative cytochrome *b*561 protein sequences have been identified from EST sequences obtained from a large variety of species [7]. The similarity between sequences from different organisms is usually not high at the DNA level. However, at the protein level well-conserved features are apparent. In Figure 3 a selection of cytochrome *b*561 protein sequences from phylogenetically diverse species is aligned. As reported by [17], and as apparent from Figure 1, the conservation in the first and sixth

transmembrane helix of the cytochromes *b*561 is limited to the overall hydrophobicity. This observation suggests that these helices may be primarily involved in protein folding and stability, rather than in the catalytic activity.

In addition to the conservation of the predicted transmembrane structures, four His residues suggested to bind the two heme groups are strictly conserved in all cytochrome *b*561 homologs. Furthermore, predicted binding sites for MDHA and ascorbate - as suggested by [18] - are conserved. A potential MDHA-binding motif - xYSLHSWxGx - with x being a hydrophobic amino acid in most proteins, is highly conserved. In the *Anopheles gambiae* (malaria mosquito) *b*561 protein, Angb561, the Ser is conservatively replaced by Thr. The Tyr in the consensus sequence is not conserved in the proteins from *C. elegans* - Caeb561-1, -2 and -3 - nor in Artb561-3, but it is present in all Artb561-3 orthologs identified in other plant species (data not shown).

The suggested ascorbate-binding site (ALLVYRVFR in the mammalian chromaffin cytochrome *b*561 [18]) is almost perfectly conserved in all mammalian proteins, but the degree of conservation is much lower in plants (Figure 3). The first three or four amino acids of this consensus sequence are generally hydrophobic, and the Tyr is found in all proteins except Artb561-4, in which it is replaced by His. The triplet codons for Tyr and His differ by only one nucleotide, suggesting that a single point mutation may be responsible for this substitution. The His at this position is found in all Artb561-4 orthologs from other plant species (data not shown), and therefore seems a consistent substitution in this isoform of plant cytochromes *b*561. The Arg from this motif is conservatively replaced by Lys in all available plant sequences. To the best of our knowledge the sequence requirement for ascorbate and MDHA binding has not been experimentally confirmed, and so it is not possible to evaluate whether the observed conservation is sufficient for the functional interaction with the ascorbate-MDHA redox couple.

In addition to the conserved structural features mentioned above, all cytochromes *b*561 show strict conservation of five Gly, two Pro, one Lys and one Gln (green in Figure 3). A number of aromatic residues are also well conserved (yellow in Figure 3). These have been suggested to take part in the electron transfer between the two heme molecules, possibly by electron tunneling [17].

Phylogeny of cytochromes *b*561

When all available cytochrome *b*561 protein sequences - including those derived from ESTs - are aligned with Clustal W [20], and subsequently organized in a phylogenetic tree (Figure 4), two main clusters containing the plant and animal cytochromes *b*561 can be discriminated. However, the position of the Artb561-3 orthologs in the plant cluster is not supported by high bootstrap values (less than 80%).

A closer examination of the plant group reveals the clustering of sequences from *Lycopersicon esculentum* (tomato, Lyeb561), *L. hirsutum* (Lyhb561), *Solanum tuberosum* (potato, Sotb561), *C. plantagineum* (resurrection plant, Crpb561), *Beta vulgaris* (beet, Bevb561-1), *Populus tremula* (poplar, Potb561), *Medicago truncatula* (barrel medic, Metb561-1), *Oryza sativa* (rice, Orsb561), *Zea mays* (maize, Zemb561), *Hordeum vulgare* (barley, Hovb561-1), *Sorghum bicolor* (sorghum, Sobb561), *Triticum aestivum* (wheat, Trab561-1) and *T. monococcum* (Trmb561) with the Artb561-1 protein, suggesting that these proteins are orthologs of Artb561-1 from *A. thaliana*. The localization of a cytochrome *b*561 from *Pinus taeda* (loblolly pine, Pitb561) within this cluster is not supported by a bootstrap value above 80% and should be interpreted with caution. Similarly, sequences from *T. aestivum* (Trab561-2), *H. vulgare* (Hovb561-2), *M. truncatula* (Metb561-2), *Glycine max* (soybean, Glmb561-2), *Zinnia elegans* (zinnia, Zieb561) and *B. vulgaris* (Bevb561-2) are clustered with Artb561-2. Orthologs of Artb561-3 have been identified in *Mesembryanthemum crystallinum* (ice plant, Mecb561-3), *M. truncatula* (Metb561-3) and *Gossypium hirsutum* (cotton, Gohb561), and Artb561-4 orthologs in *M. truncatula* (Metb561-4) and *B. vulgaris* (Bevb561-4).

Most of the plant cytochrome *b*561 sequences so far have been obtained from dicotyledonous plants, but for Artb561-1 and -2 some orthologs from monocotyledons are available. These form separate 'sub-clusters' within the clusters of Artb561-1 and -2 respectively, whereas an Artb561-1 ortholog from the gymnosperm *P. taeda* is more distantly related with these clusters. For Artb561-3 and -4 no orthologs have been identified in monocotyledonous plants so far. Within the animal group, all known chromaffin granule cytochromes *b*561 - from *H. sapiens* (human, Hosb561-1), *M. musculus* (mouse, Mumb561-1), *Sus scrofa* (pig, Susb561), *Bos taurus* (bovine, Botb561) and *Ovis aries* (sheep, Ovab561) are clustered together with the homolog from *Xenopus laevis* (African clawed frog, Xelb561). The duodenal cytochromes *b*561 from humans, mouse and rat (Hosb561-2, Mumb561-2 and Ranb561-2) form a separate group, as do the chromaffin granule cytochromes *b*561 and the third type of mammalian cytochrome *b*561 ('ubiquitous' cytochrome *b*561 from humans and mouse, Hosb561-3 and Mumb561-3). The position of the cytochrome *b*561 protein from *Ciona intestinalis* (Ciib561; sea squirt, a tunicate) is uncertain, as the bootstrap value is below 80%. The cytochrome *b*561 homologs identified in *C. elegans* and the flatworm *Dugesia japonica* could not reliably be associated with any of the clusters.

The outcome for the phylogenetic tree was essentially the same when T-Coffee [21] or POA [22] software packages were used, which are in some cases considered more reliable than Clustal W [23]. Similarly, using only the central cytochrome *b*561 'core domain' as defined by Ponting [24], instead of the full sequence, the clustering of cytochrome *b*561 homologs is

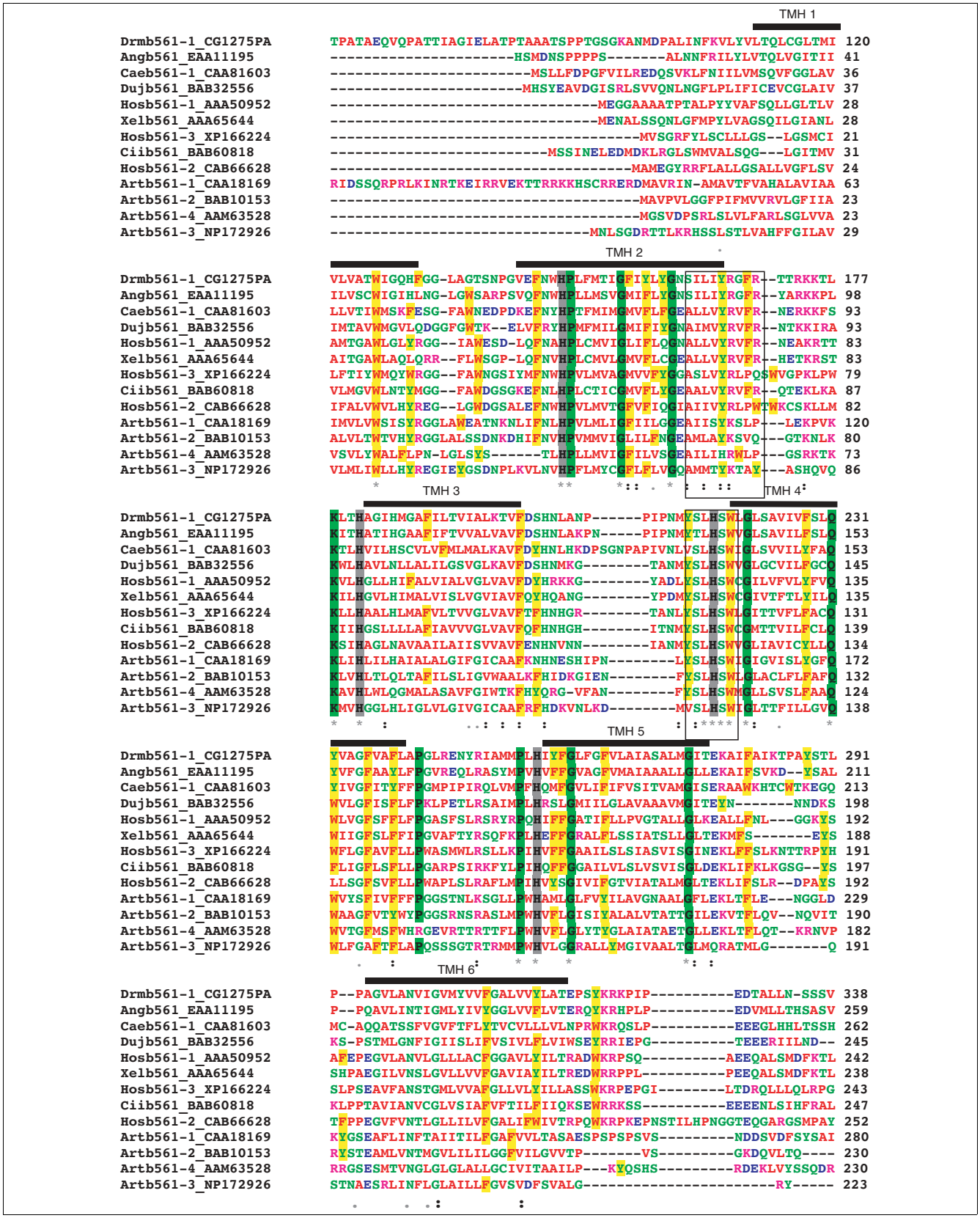

Figure 3 *(see legend on next page)*

Figure 3 *(see previous page)*
Alignment of cytochrome *b*561 protein sequences from plant and animal species. Conserved features are marked: TMH, transmembrane helices; conserved histidine residues, gray shading; conserved aromatic residues, yellow shading; and predicted MDHA-binding site (YSLHSW) and ascorbate-binding site (ALLVYRVFR) in boxes. Other conserved residues are marked in green. Red, small hydrophobic amino acids; green, hydroxyl or amino basic side chains; blue, acidic; purple, positively charged. Conservative changes at a specific position are marked with : under the alignment, while * indicates the perfect conservation at a position in all aligned proteins. The sequence from Hosb561-3 is unpublished (H.A.). Transmembrane helices were predicted with TMHMM software [43].

identical to that presented in Figure 4, and the overall appearance of the resulting tree is not altered (Figure 5).

To examine the selective pressure on cytochrome *b*561-encoding genes after various gene duplication events, a Li93 analysis was performed using the DAMBE software package [25,26], based on pairwise comparisons. The amount of synonymous substitutions (K_s in Table 1) is apparently saturated when comparing *Artb561-3* to the three other paralogous genes from *A. thaliana*, indicating that *Artb561-3* may represent an early gene duplication event, and has accumulated a large number of 'silent' substitutions. Nonsynonymous substitutions (that is, codon alterations leading to amino acid replacements, K_a in Table 1) for these genes range from 0.49 to 0.75 (substitutions/position), which is much lower than the K_s values. *Artb561-2* and *-4* have the lowest K_s (0.47), suggesting that they originated from a more recent gene duplication in *A. thaliana*. Comparable values are found for the orthologs of *Artb561-2* and *-4* in *B. vulgaris* and *M. truncatula*, indicating that this last gene duplication had occurred well before the radiation within the dicotyledons (data not shown). In mouse, K_s values are equally high (and probably saturated) for the genes encoding the three cytochrome *b*561 isoforms. K_a values are also comparable to those of the different plant isoforms (Table 1).

Comparison of cytochromes *b*561 from dicotyledonous species to their orthologs from monocotyledons reveals that K_s and K_a values are generally both around 0.3 to 0.5. Knowing that both clades diversified 130 million years ago [27], the average rate of synonymous and nonsynonymous substitutions is calculated to be around 3×10^{-9} substitutions per position per year (Table 1). This is comparable to the synonymous evolution rate of other nuclear-encoded plant genes ($5\text{-}30 \times 10^{-9}$ substitutions/position/year [28]). Among dicotyledons (using estimated divergence times for different families, derived from Figure 2 in [29]) the average synonymous substitution rate is higher ($12.63 \pm 2.41 \times 10^{-9}$ substitutions/position/year), but the nonsynonymous rate is very similar ($2.38 \pm 1.18 \times 10^{-9}$ substitutions/position/year). For the resurrection plant (*C. plantagineum*), however, which has adapted to life in an extremely dry environment, a particularly high nonsynonymous substitution rate was found as compared to the ortholog from potato (13.25×10^{-9} substitutions/position/year), whereas the synonymous rate was the same as for the orthologs in other dicots (Table 1).

K_s and K_a were also calculated for animal cytochrome *b*561 genes. Comparing the gene from the tunicate *C. intestinalis* (*Ciib561*), which branched off the vertebrate lineage around 540 million years ago [30], with the chromaffin granule cytochrome *b*561 genes from vertebrates (*X. laevis*, *H. sapiens* and *M. musculus*) gives a K_s of around 0.65 and a K_a of 0.8. This is remarkably low, considering the large phyletic distance between tunicates and vertebrates, and implies synonymous and nonsynonymous substitution rates of $1\text{-}1.5 \times 10^{-9}$ substitutions/position/year.

The Gu99 likelihood ratio test in the DIVERGE software package [31] was used to examine whether proteins from different branches in the phylogenetic tree have different functional constraints, and whether they may have functionally diverged. The Theta ML value, indicative for the level of functional divergence between proteins [31], is relatively low (0.213) according to [32], when monocotyledonous and dicotyledonous Artb561-1 orthologs are compared. This suggests that these orthologs are likely to have similar biochemical functions. The Gu99 likelihood assay also indicated that a functional adaptation may have occurred at only five amino acids between these orthologs, supporting their functional similarity. A similar analysis of Artb561-1 and -2 orthologs yielded a theta ML value of 0.533, and 22 residues that may have functionally adapted. These numbers are still rather low [32]. However, four of these residues (A, I, S and P) are situated in the predicted ascorbate-binding site (AIISYKSLP in Artb561-1). A similar analysis for the mammalian cyts *b*561 (chromaffin cyts *b*561 versus the cluster with the other two mammalian cyts *b*561) resulted in a very low theta ML of 0.031, and no residues likely to have an altered function.

Discussion

At the gene level, conservation among cytochrome *b*561-encoding genes is not very high [7]. The genomic organization usually comprises four or five exons in plants and mammals (Figure 1). In addition to genes with four or five exons, invertebrates contain homologous genes with only two exons (*C. elegans* F55H2.5), or with six exons (*D. melanogaster*). Splice variants are observed for *D. melanogaster* CG1275 and CG8776. This alternative splicing results in a cytochrome *b*561 protein with an additional amino-terminal region (Drmb561-1), and in a protein that lacks the first of the six

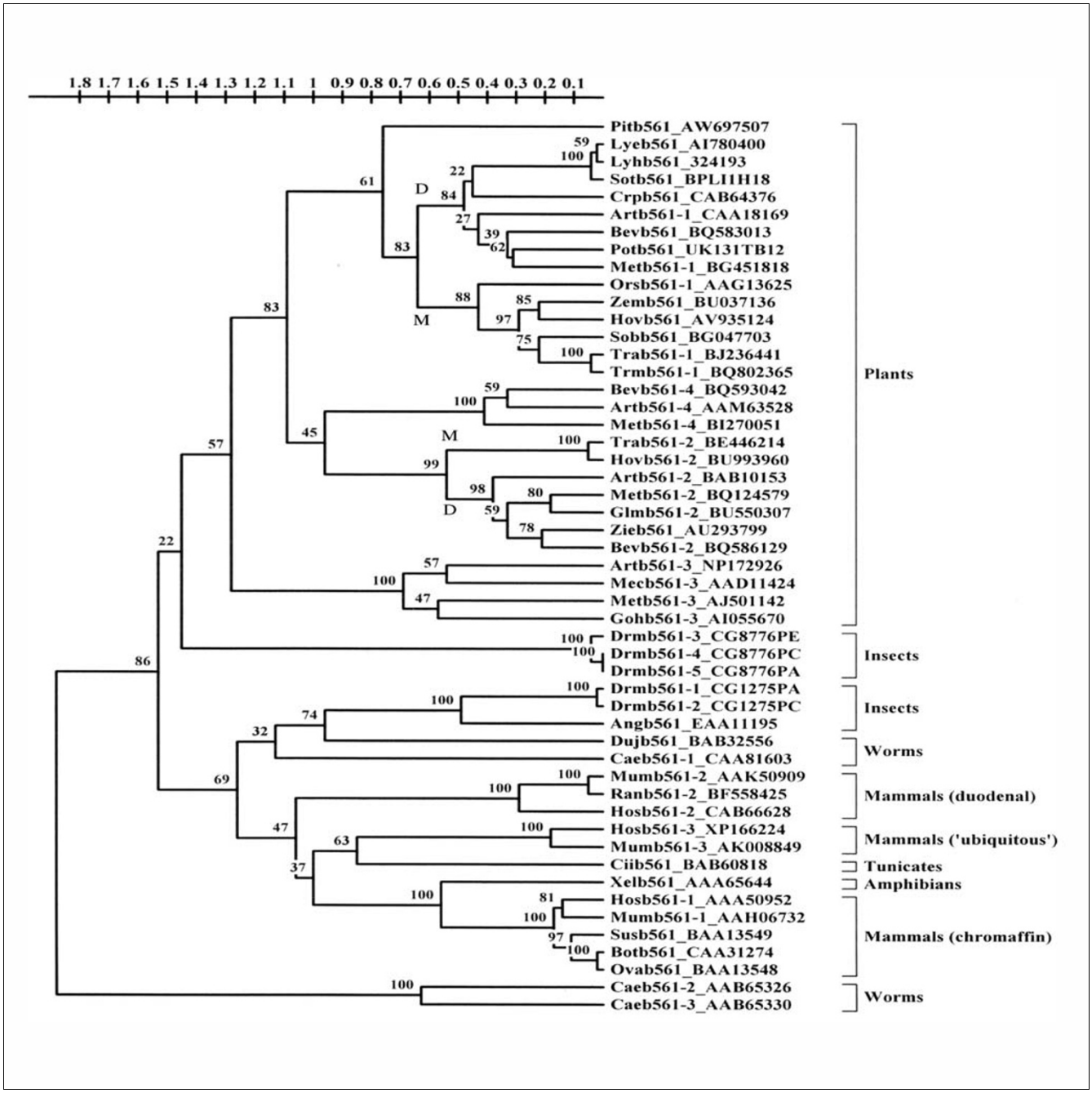

Figure 4
Unrooted phylogenetic tree including all known cytochrome *b*561 proteins from plants and animals. The tree was derived from a Clustal W [20] alignment of the amino-acid sequences, created with Treecon software [42]. The distance scale above the tree represents the number of substitutions per site, and bootstrapping values are shown at each branch point (percentage of 200 bootstrap samples). Dicot and monocot clusters are marked as D and M, respectively, among the orthologs of *Artb561-1* and *Artb561-2*.

transmembrane helices (Drmb561-4), commonly present in plant and mammalian cytochromes *b*561. The physiological implication of these altered cytochrome *b*561 structures is unclear. Interestingly, the alternative splice products have in each case retained the cytochrome *b*561 'core structure' as defined by Ponting [24], that is, four transmembrane helices containing the conserved heme-ligating His residues and the predicted substrate-binding sites.

Despite the availability of extensive genomic sequence information from fungi and prokaryotes, the presence of cytochrome *b*561-like sequences is restricted to animals and

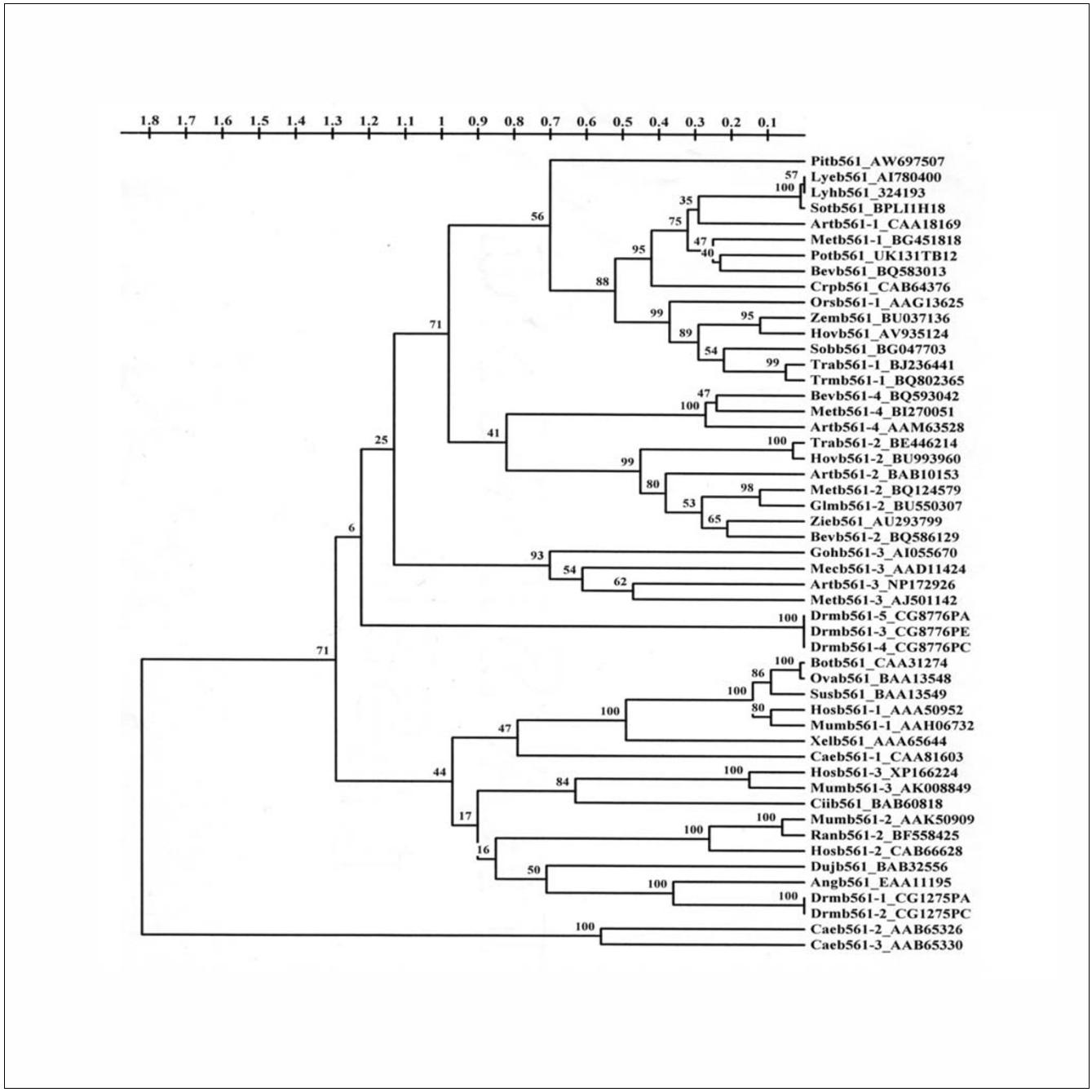

Figure 5
Unrooted phylogenetic tree of the CB domains from all known cytochrome *b*561 proteins from plants and animals. The tree was derived from a Clustal W [20] alignment of the amino-acid sequences, created with Treecon software [42] in the same way as that in Figure 4.

plants. Some fungal species have been demonstrated to synthesize the ascorbate analogs D-arabo-ascorbate (*Penicillium*) or erythro-ascorbate (*Candida albicans, Saccharomyces cerevisiae*) instead of ascorbate [33,34]. However, these compounds occur at very low concentrations, and they probably have only limited importance - if any - as antioxidants [34]. Ascorbate is apparently completely absent in prokaryotes [33]. The absence of cytochromes *b*561 in fungi and prokaryotes may therefore be related to the absence of L-ascorbate as a major antioxidant.

Our phylogenetic analysis supports the hypothesis that all cytochrome *b*561 proteins have probably evolved from a single protein, present in the common ancestor of plants and animals. This conclusion is supported by the perfect conservation of the 'core structural features' in cytochrome *b*561

Table 1

Li93 analysis of a selection of cytochrome *b*561-encoding genes

		K_s	K_a	K_a/K_s	Synonymous rate	Nonsynonymous rate
Plant paralogs						
Artb561-1	Artb561-2	2.00	0.49	0.245		
Artb561-1	Artb561-3	2.90	0.57	0.197		
Artb561-1	Artb561-4	1.75	0.55	0.314		
Artb561-3	Artb561-4	2.93	0.75	0.256		
Artb561-2	Artb561-4	0.47	0.63	1.340		
Mouse paralogs						
Mumb561-1		1.9	0.61	0.321		
Mumb561-1	Mumb561-3	1.71	0.51	0.298		
Mumb561-2	Mumb561-3	2.01	0.5	0.249		
Dicot versus monocot						
	Average	0.37 ± 0.09	0.43 ± 0.06	1.22 ± 0.38	2.88 ± 0.69	3.33 ± 0.45
Within dicots						
	Average	-	-	-	12.63 ± 2.41	2.38 ± 1.18
Exception						
Crpb561	Sotb561	0.48	0.53	1.10	11.75	13.25
Tunicate versus vertebrates						
	Average	0.65 ± 0.13	0.78 ± 0.12	1.26 ± 0.45	1.21 ± 0.24	1.45 ± 0.22
Human versus mouse						
Hosb561-1	Mumb561-1	0.12	0.19	1.583	1.20	1.90
Hosb561-2	Mumb561-2	0.47	0.16	0.340	4.70	1.60
Hosb561-3	Mumb561-3	0.48	0.07	0.146	4.80	0.70

K_s, synonymous substitutions; K_a, nonsynonymous substitutions. The rate of synonymous and nonsynonymous substitutions (right-hand columns) is expressed as 10^{-9} substitutions/position/year. The estimated divergence times between different branches used for these calculations are 130×10^6 years for monocots-dicots [27], 12×10^6 years for tomato-potato [44], 540×10^6 years for tunicate-vertebrates [30], and 100×10^6 years for human-mouse [45]. Divergence times for different dicot families are estimated from Figure 2 in [29]: 90 million years for Fabales (*M. truncatula*), Caryophyllales (*B. vulgaris*, *M. crystallinum*) and Malvales (*G. hirsutum*), 40 million years for Lamiales (*C. plantagineum*) and Solanales (*S. tuberosum*), and 20 million years for the radiation within the Fabales (*M. truncatula* and *G. max*).

proteins in a range of phylogenetically very distinct species (Figure 3). It should be noted that the conservation of the putative ascorbate-binding sites is high in the mammalian proteins, but considerably less in the plant sequences (Figure 3). Although an ascorbate-reducible cytochrome *b*561 has been demonstrated in several plant species [12,35], this theoretically leaves open the possibility that other substrates may function as electron donors to the plant cytohromes *b*561. The Gu99 likelihood analysis indicates that four amino-acid residues in the ascorbate-binding site may show functional adaptation, supporting the possibility of different substrate-binding site affinities among cytochromes *b*561 in *A. thaliana*.

Plant and animal cytochromes *b*561 generally separate into two clusters in a phylogenetic tree (Figure 4), indicating a diversification early in evolution. The pairwise similarity between each of the *A. thaliana* cytochromes *b*561 and the animal homologs is comparable (data not shown), suggesting that they diversified within an evolutionarily short time span from a single ancestral protein, after the separation between plants and animals. Cytochrome *b*561 proteins from invertebrates (insects, nematode and flatworm) are not tightly linked to the plant or mammalian cluster (Figure 4).

Within the mammalian cluster of cytochromes *b*561, the three paralogous proteins (from chromaffin granules, duodenal

and 'ubiquitous') form separate, monophyletic groups. The presence of a cytochrome *b*561-like protein in *C. intestinalis* is interesting. Tunicates, the most primitive chordates, are considered the direct ancestors of vertebrates [36], suggesting that the protein from *C. intestinalis* might represent an ancestral form of the mammalian cytochromes *b*561. Synonymous substitution rates for the cytochrome *b*561 genes from *C. intestinalis* and different vertebrates are rather low (1-1.5 × 10^{-9}/position/year, Table 1), as compared to other nuclear-encoded mammalian genes (for example, 4.61 × 10^{-9}/position/year [37]). Nonsynonymous substitution rates are in the same range, indicating a good conservation at the protein level.

In addition to conclusions on the evolutionary relationship between cytochromes *b*561 in different species, the cladogram points to interesting relations between cytochrome *b*561 isoforms within a single species. The four paralogous proteins from *A. thaliana* (Artb561-1, -2, -3 and -4) form separate clusters with their respective orthologs from other plant species (Figure 4). It is thus likely that gene duplication events had already occurred in algae or primitive plants. The identification of a cytochrome *b*561 in the gymnosperm *P. taeda* (Pitb561) supports this suggestion.

The clusters containing the *A. thaliana* isoforms Artb561-1 and Artb561-2 also contain homologous sequences from both dicotyledons and monocotyledons. In both cases, the proteins from monocots and dicots form separate sub-clusters. This observation suggests that the four different cytochromes *b*561 have evolved separately in dicots and monocots, and hence that the diversification between these proteins had already occurred in their common ancestor. Substitution rates in cytochrome *b*561 genes from monocots and dicots are comparable to those in other plant nuclear-encoded genes (Table 1). The *Artb561-1* ortholog in the drought-resistant resurrection plant (*Crpb561*), seems to have a remarkably high rate of nonsynonymous substitution (13.25/position/year). Interestingly, the Artb561-2 and -4 proteins tend to group in the same cluster (Figure 4), suggesting that they diversified more recently in evolution, which is supported by our Li93 analysis (Table 1).

Cyts *b*561 are possibly involved in the regeneration of ascorbate through transmembrane electron transport [8,10,11,18]. This functional conservation is supported by our likelihood ratio test (Gu99 test). As ascorbate is present in different subcellular organelles in plants and animals, it is not surprising to find members of the cytochrome *b*561 protein family in different organelles. In humans, the chromaffin tissue cytochrome *b*561 is present in the membrane of a subcellular secretory vesicle [8,10,38], whereas the duodenal cytochrome *b*561 is a plasma membrane protein [39]. The subcellular localization of the third human isoform (Hosb561-3) is not yet known. In *A. thaliana* and several other plant species, at least one of the cytochrome *b*561 isoforms is present in the plasma membrane [7,12]. The subcellular localization of the other isoforms remains to be determined. The association with different membranes raises the question on the subcellular localization of the ancestral cytochrome *b*561. The apparent absence of a cytochrome *b*561 in the plasma membrane of algae [7] can be taken as an indication that the primitive form of the protein was located in internal membranes.

Materials and methods

Database searches used BLAST [40] on the website of the National Center for Biotechnology Information (NCBI [41]). Multiple sequence alignments were made with Clustal W [20], and confirmed with POA [22] and T-Coffee [21], which have been reported to give more accurate alignments [23]. These data were converted into a cladogram using Treecon software [42]. Distances were calculated with Poisson correction, and for tree topology, complete linkage clustering was used. The Li93 analysis was performed with DAMBE [26], after aligning the genes pairwise with Clustal W and T-Coffee. The maximum likelihood analysis (Gu99) was performed with the DIVERGE software [31], using the phylogenetic tree from Figure 4. DIVERGE calculates a theta ML value indicative of the level of functional divergence between proteins in different clusters of the tree, and a posterior probability to trace the amino-acid positions that are likely to be responsible for the functional divergence between proteins in both clusters. Transmembrane structures were predicted with TMHMM [43].

Acknowledgements

The authors thank Etsuko Moriyama (UNL) for valuable advice and for critically reading the manuscript. This research was supported by grants from the Fund for Scientific Research (FWO) and the University of Antwerp; W.V. is Research Assistant of the Fund for Scientific Research - Flanders (Belgium) (FWO). A contribution of the University of Nebraska Agricultural Research Division, Lincoln, NE 68583. Journal Series No. 14061.

References

1. Noctor G, Foyer CH: **Ascorbate and glutathione: keeping active oxygen under control.** *Annu Rev Plant Physiol Plant Mol Biol* 1998, **49:**249-279.
2. Mittler R: **Oxidative stress, antioxidants and stress tolerance.** *Trends Plant Sci* 2002, **7:**405-410.
3. Arrigoni O, De Tullio MC: **The role of ascorbic acid in cell metabolism: between gene-directed functions and unpredictable chemical reactions.** *J Plant Physiol* 2000, **157:**481-488.
4. Davey MW, Van Montagu M, Inze D, Sanmartin M, Kanellis A, Smirnoff N, Benzie IJJ, Strain JJ, Favell D, Fletcher J: **Plant L-ascorbic acid: chemistry, function, metabolism, bioavailability and effects of processing.** *J Sci Food Agricult* 2000, **80:**825-860.
5. Wheeler GL, Jones MA, Smirnoff N: **The biosynthetic pathway of vitamin C in higher plants.** *Nature* 1998, **393:**365-369.
6. Agius F, González-Lamothe R, Caballero JL, Muñoz-Blanco J, Botella MA, Valpuesta V: **Engineering increased vitamin C levels in plants by overexpression of a D-galacturonic acid reductase.** *Nat Biotechnol* 2003, **21:**177-181.
7. Asard H, Kapila J, Verelst W, Bérczi A: **Higher-plant plasma membrane cytochrome *b*561: a protein in search of a function.** *Protoplasma* 2001, **217:**77-93.
8. Wakefield LM, Cass AEG, Radda GK: **Functional coupling between enzymes of the chromaffin granule membrane.** *J Biol*

Chem 1986, **261:**9739-9745.
9. Asard H, Horemans N, Preger V, Trost P: **Plasma membrane *b*-type cytochromes.** In *Plasma Membrane Redox Systems and Their Role in Biological Stress and Disease.* Edited by: Asard H, Bérczi A, Caubergs RJ. Dordrecht: Kluwer Academic Publishers; 1998:1-31.
10. Njus D, Kelley PM, Harnadek GJ, Pacquing YV: **Mechanism of ascorbic acid regeneration mediated by cytochrome b_{561}.** *Ann NY Acad Sci* 1987, **493:**108-119.
11. Horemans N, Asard H, Caubergs RJ: **The role of ascorbate free radical as an electron acceptor to cytochrome *b*-mediated trans-plasma membrane electron transport in higher plants.** *Plant Physiol* 1994, **104:**1455-1458.
12. Asard H, Venken M, Caubergs R, Reijnders W, Oltmann FL, De Greef JA: ***b*-Type cytochromes in higher plant plasma membranes.** *Plant Physiol* 1989, **90:**1077-1083.
13. Askerlund P, Larsson Ch, Widell S: **Cytochromes of plant plasma membranes. Characterisation by absorbance difference spectrophotometry and redox titration.** *Physiol Plant* 1989, **76:**123-134.
14. Perin MS, Fried VA, Slaughter CA, Südhof TC: **The structure of cytochrome *b*561, a secretory vesicle-specific electron transport protein.** *EMBO J* 1988, **7:**2697-2703.
15. Tsubaki M, Nakayama M, Okuyama E, Ichikawa Y, Hori H: **Existence of two heme B centers in cytochrome *b*561 from bovine adrenal chromaffin vesicles as revealed by a new purification procedure and EPR spectroscopy.** *J Biol Chem* 1997, **272:**23206-23210.
16. Asard H, Terol-Alcayde J, Preger V, Del Favero J, Verelst W, Sparla F, Pérez-Alonso M, Trost P: ***Arabidopsis thaliana* sequence analysis confirms the presence of cyt *b*-561 in plants. Evidence for a novel protein family.** *Plant Physiol Biochem* 2000, **38:**905-912.
17. Bashtovyy D, Bérczi A, Asard H, Páli T: **Structure prediction for the di-heme cytochrome b-561 protein family.** *Protoplasma* 2003, **221:**31-40.
18. Okuyama E, Yamamoto R, Ichikawa Y, Tsubaki M: **Structural basis for the electron transfer across the chromaffin vesicle catalyzed by cytochrome *b*561: analyses of DNA nucleotide sequences and visible absorption spectra.** *Biochim Biophys Acta* 1998, **1383:**269-278.
19. Srivastava M: **Genomic structure and expression of the human gene encoding cytochrome *b*561, an integral protein of the chromaffin granule membrane.** *J Biol Chem* 1995, **270:**22714-22720.
20. Higgins DG, Thompson JD, Gibson TJ: **Using CLUSTAL for multiple sequence alignments.** *Methods Enzymol* 1996, **266:**383-402.
21. Notredame C, Higgins DG, Heringa J: **T-Coffee: a novel method for fast and accurate multiple sequence alignment.** *J Mol Biol* 2000, **302:**205-217.
22. Lee C, Grasso C, Sharlow MF: **Multiple sequence alignment using partial order graphs.** *Bioinformatics* 2002, **18:**452-464.
23. Lassmann T, Sonnhammer ELL: **Quality assessment of multiple alignment programs.** *FEBS Lett* 2002, **529:**126-130.
24. Ponting CP: **Domain homologues of dopamine hydroxylase and ferric reductase: roles for iron metabolism in neurodegenerative disorders?** *Hum Mol Genet* 2001, **10:**1853-1858.
25. Li W-H: **Unbiased estimation of the rates of synonymous and nonsynonymous substitution.** *J Mol Evol* 1993, **36:**96-99.
26. Xia X, Xie Z: **DAMBE: Data analysis in molecular biology and evolution.** *J Hered* 2001, **92:**371-373.
27. Crane PR, Friis EM, Pedersen KR: **The origin and early diversification of angiosperms.** *Nature* 1995, **374:**27-33.
28. Wolfe KH, Li W-H, Sharp PM: **Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs.** *Proc Natl Acad Sci USA* 1987, **84:**9054-9058.
29. Magallón S, Sanderson MJ: **Absolute diversification rates in angiosperm clades.** *Evolution Int J Org Evolution* 2001, **55:**1762-1780.
30. Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, *et al.*: **Resolution of the early placental mammal radiation using bayesian phylogenetics.** *Science* 2001, **294:**2348-2351.
31. Gu X, Vander Velden K: **DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family.** *Bioinformatics* 2002, **18:**500-501.
32. Wang Y, Gu X: **Functional divergence in the caspase gene family and altered functional constraints: statistical analysis and prediction.** *Genetics* 2001, **158:**1311-1320.
33. Arrigoni O, De Tullio MC: **Ascorbic acid: much more than just an antioxidant.** *Biochim Biophys Acta* 2002, **1569:**1-9.
34. Spickett CM, Smirnoff N, Pitt AR: **The biosynthesis of erythroascorbate in *Saccharomyces cerevisiae* and its role as an antioxidant.** *Free Radic Biol Med* 2000, **28:**183-192.
35. Bérczi A, Lüthje S, Asard H: ***b*-Type cytochromes in plasma membranes of *Phaseolus vulgaris* hypocotyls, *Arabidopsis thaliana* and *Zea mays* roots.** *Protoplasma* 2001, **217:**50-55.
36. Satoh N, Jeffery WR: **Chasing tails in ascidians: developmental insights into the origin and evolution of chordates.** *Trends Genet* 1995, **11:**354-359.
37. Li W-H, Tanimura M: **The molecular clock runs more slowly in man than in apes and monkeys.** *Nature* 1987, **326:**93-96.
38. Njus D, Wigle M, Kelley PM, Kipp BH, Schlegel HB: **Mechanism of ascorbic acid oxidation by cytochrome *b*561.** *Biochemistry* 2001, **40:**11905-11911.
39. McKie AT, Barrow D, Latunde-Dada GO, Rolfs A, Sager G, Mudaly E, Mudaly M, Richardson C, Barlow D, Bomford A, *et al.*: **An iron-regulated ferric reductase associated with the absorption of dietary iron.** *Science* 2001, **291:**1755-1759.
40. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25:**3389-3402.
41. **National Center for Biotechnology Information - BLAST** _[http://www.ncbi.nlm.nih.gov/blast]
42. Van de Peer Y, De Wachter R: **TREECON: a software package for the construction and drawing of evolutionary trees.** *Comp Appl Biosci* 1993, **9:**177-182.
43. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol* 2001, **305:**567-580.
44. Moniz de Sá M, Drouin G: **Phylogeny and substitution rates of angiosperm actin genes.** *Mol Biol Evol* 1996, **13:**1198-1212.
45. Skrabanek L, Wolfe KH: **Eukaryote genome duplication - where's the evidence?** *Curr Opin Genet Dev* 1998, **8:**694-700.