

# Mitochondrial ATP Synthase: A Bioinformatic Approach Reveals New Insights About the Roles of Supernumerary Subunits *g* and A6L

Sangjin Hong<sup>1</sup> and Peter L. Pedersen<sup>1,2</sup>

Received September 1, 2004; accepted September 30, 2004

The mitochondrial ATP synthase is a membrane protein complex which couples the proton gradient across the mitochondrial inner membrane to the synthesis of ATP from ADP + P<sub>i</sub>. The complex is composed of essential subunits for its motor functions and supernumerary subunits, the roles of which remain to be elucidated. Subunits *g* and A6L are supernumerary subunits, and the specific roles of these subunits are still matters of debate. To gain insight into the functions of these two subunits, we carried out the alignment and the homolog search of the protein sequences of the subunits and found the following features: Subunit *g* appears to have isoforms in animals, and the transmembrane domain of the animal subunit *g* contains a completely conserved acidic residue in the middle of a helix on the conserved side of the transmembrane helix. This finding implicates the conserved acidic residue as important for the function of subunit *g*. The alignment of A6L protein sequences shows a conserved aromatic residue at the N-terminal domain with which the N-terminal MPQL sequence comprises a unique MPQLX<sub>4</sub>Ar motif that can signify the protein A6L. The conserved aromatic residue may also be important for the function of A6L.

**KEY WORDS:** Bioinformatics; mitochondria; ATP synthase; supernumerary subunits; subunit *g*; subunit A6L.

## INTRODUCTION

The mitochondrial ATP synthase (Fig. 1) is responsible for the generation of ATP driven by an electrochemical gradient of protons across the inner membrane. The mitochondrial ATP synthase is composed of at least 17 or 20–21 subunits in animals and fungi, respectively (Hong and Pedersen, 2003; Pedersen *et al.*, 2000). Compared to the bacterial ATP synthase, which seems to contain the minimal composition required for the activity of ATP synthase, the mitochondrial ATP synthase includes additional subunits, called supernumerary subunits. Subunits *g* and A6L that belong to the supernumerary subunits are found in both the animal and fungal mitochondrial ATP synthase. These subunits are integral membrane proteins

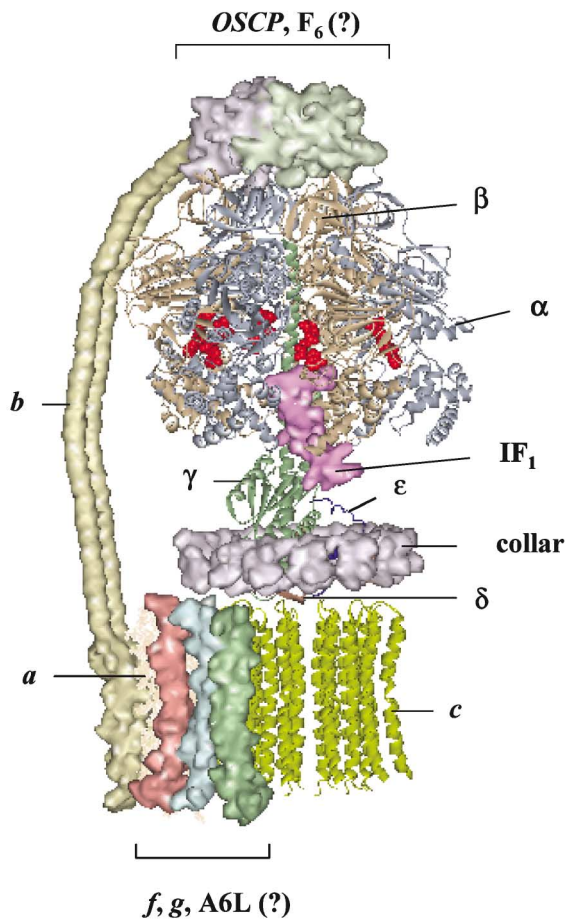
with a single transmembrane helix. The N-termini of subunits *g* and A6L are on the matrix and cytosolic sides of the membranes, respectively (Belogradov *et al.*, 1996). The functions of the subunits are still matters of debate although some roles of the subunit *g* have been reported in yeast (Arnold *et al.*, 1998; Boyle *et al.*, 1999). To gain insight into the potential functions of the subunit *g* and A6L, we aligned the protein sequences of the subunits and searched for the homologous proteins in the genomes of which the whole sequences are available. Here we found some interesting features of these subunits that may be of functional significance.

## MATERIALS AND METHODS

The protein sequences were retrieved by BLAST database search (Altschul and Koonin, 1998; Altschul *et al.*, 1997) with the default parameters (BLOSUM62 matrix, 0.005 as an *E*-value threshold, no low complexity filtering in the query sequence) on the nonredundant (nr)

<sup>1</sup> Department of Biological Chemistry, Johns Hopkins University, School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205-2185.

<sup>2</sup> To whom correspondence should be addressed; e-mail: ppederse@jhmi.edu.



**Fig. 1.** Current view of the structure of mitochondrial ATP synthase. For the generation of the model, yeast mitochondrial ATP synthase (1qol) was used as a template and the coordinates of each subunit of the template structure were replaced by superimposition with the coordinates from high resolution structures ( $\alpha$ ,  $\beta$ , and IF<sub>1</sub> subunits from 1OHH;  $\gamma$ ,  $\delta$ , and  $\epsilon$  subunits from 1e79;  $a$  and  $c$  subunits from 1c17; transmembrane part of subunit  $b$  from 1b9u). The other subunits in the model were constructed manually using Quanta. No positions are assigned to the subunits  $d$ ,  $e$ , and factor B, one or more of which is a candidate for the collar.

protein database. The sequences collected were aligned using ClustalW (Thompson, 1994) and Macaw (Schuler *et al.*, 1991) followed by manual adjustment. The prediction of transmembrane (TM) region was performed using SOSUI (Hirokawa *et al.*, 1998).

Full species names are (1) FUNGI: *C. glabrata*—*Candida glabrata*; *E. nidulans*—*Emericella nidulans*; *K. lactis*—*Kluyveromyces lactis*; *N. crassa*—*Neurospora crassa*; *P. anserina*—*Podospira anserina*; *S. pombe*—*Schizosaccharomyces pombe*; Yeast—*Saccharomyces cerevisiae*; *Y. lipolytica*—*Yarrowia lipolytica* (2) FUNGI/METAZOA (incertae sedis): *A. parasiticum*—*Amoebidium parasiticum* (3) METAZOA (Animal): *A.*

*pectinifera*—*Asterina pectinifera*; Atlantic cod—*Gadus morhua*; Atlantic salmon—*Salmo salar*; Blue whale—*Balaenoptera musculus*; Bovine—*Bos taurus*; Cat—*Felis catus*; *C. elegans*—*Caenorhabditis elegans*; Cheetah—*Acinonyx jubatus*; Chicken—*Gallus gallus*; Coelacanth—*Latimeria chalumnae*; Dog—*Canis familiaris*; Donkey—*Equus asinus*; Earthworm—*Lumbricus terrestris*; European sea bass—*Dicentrarchus labrax*; Fruit fly—*Drosophila melanogaster*; Geoffroy's cat—*Oncifelis geoffroyi*; Goldfish—*Carassius auratus*; Gorilla—*Gorilla gorilla*; Guinea pig—*Cavia porcellus*; Harbor seal—*Phoca vitulina*; Hippopotamus—*Hippopotamus amphibius*; Honeybee—*Apis mellifera*; Horse—*Equus caballus*; Human—*Homo sapiens*; Ibex—*Capra ibex*; Jaguarundi—*Herpailurus yaguarondi*; Lancelet—*Branchiostoma lanceolatum*; Little-red-flying-fox—*Pteropus scapulatus*; Meadow vole—*Microtus pennsylvanicus*; Mouse—*Mus musculus*; New Zealand long-tailed bat—*Chalinolobus tuberculatus*; Orangutan—*Pongo pygmaeus*; Ostrich—*Struthio camelus*; Pig—*Sus scrofa*; Puma—*Puma concolor*; Purple sea urchin—*Strongylocentrotus purpuratus*; Rabbit—*Oryctolagus cuniculus*; Rat—*Rattus norvegicus*; Rhinoceros—*Rhinoceros unicornis*; Sea lamprey—*Petromyzon marinus*; Sea urchin—*Paracentrotus lividus*; Sheep—*Ovis aries*; Spiny dogfish—*Squalus acanthias*; Spotted catshark—*Scyliorhinus canicula*; Squirrel—*Sciurus vulgaris*; Wallaroo—*Macropus robustus* (4) MALAWI-MONAS: *M. jakobiformis*—*Malawimonas jakobiformis* (5) OCHROMONADALES: *O. danica*—*Ochromonas danica* (6) LAMINARIALES: *L. digitata*—*Laminaria digitata* (7) RECLINOMONAS: *R. americana*—*Reclinomonas americana* (8) MESOSTIGMATALES: *M. viride*—*Mesostigma viride* (9) COLEOCHAETALES: *C. globosum*—*Chaetosphaeridium globosum* (10) EMBRYOPHYTA (Plant): *A. thaliana*—*Arabidopsis thaliana*; Beet—*Beta vulgaris*; Carrot—*Daucus carota*; Evening primrose—*Oenothera berteriana*; Radish—*Raphanus sativus*; Rice—*Oriza sativa*; Sunflower—*Helianthus annuus*.

## RESULTS AND DISCUSSION

### Isoforms of Animal Subunit *g*

From the search for the subunit *g* homologs in the protein sequence database, we found that there are sequence homologs of subunit *g* in animals, which show significant homologies to the identified subunits *g*. In humans, we retrieved three different protein sequences of subunit *g* from the BLAST search. After carefully examining the sources of the sequences and excluding differences

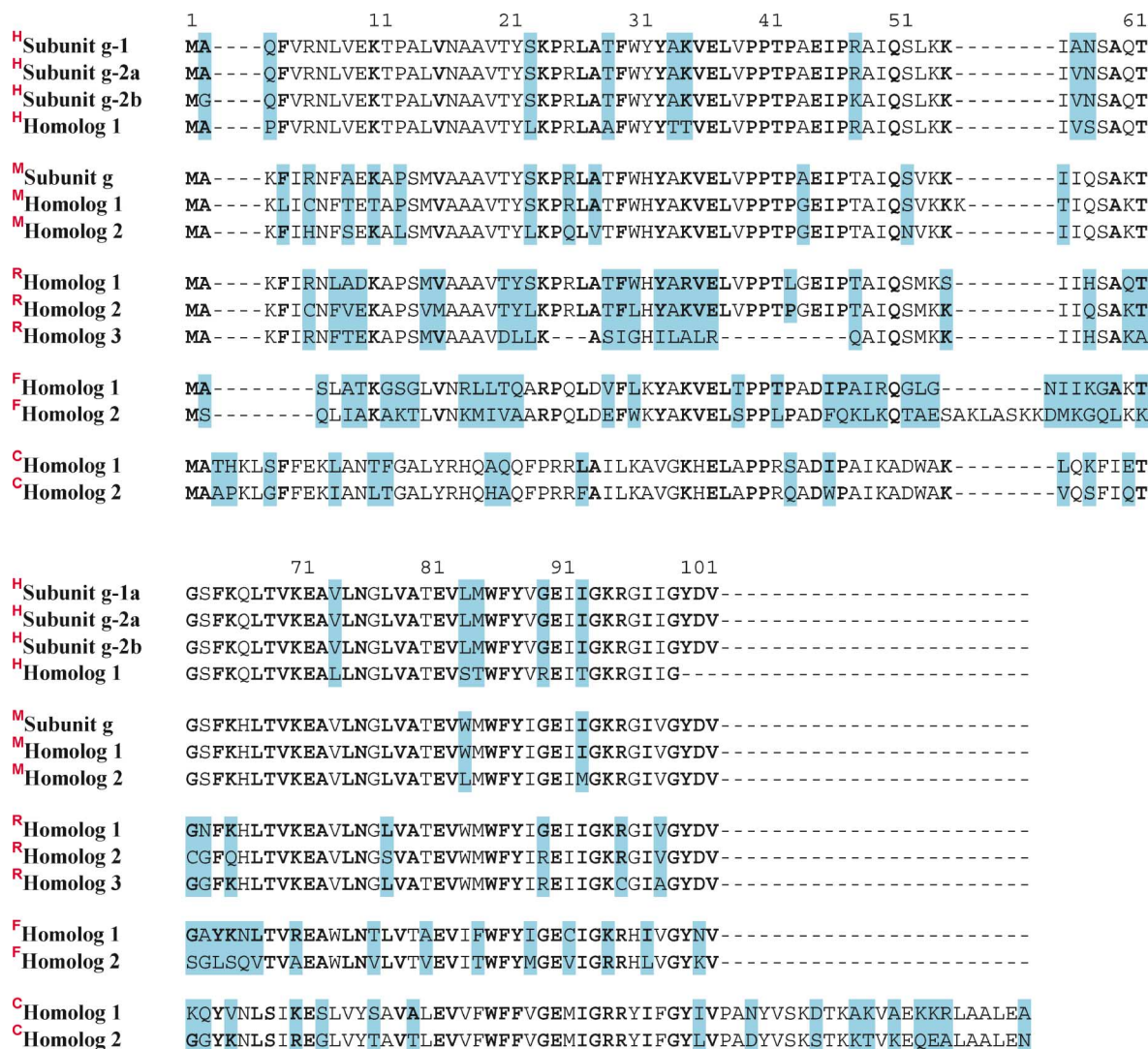


Fig. 2. Sequence alignment of subunit *g* and the isoforms of subunit *g* from animals. Conserved residues are in bold, and the residues that show variations within a given species are highlighted in blue. H, M, R, F, and C in superscript denote Human, Mouse, Rat, Fruit fly, and *C. elegans*, respectively.

possibly derived from single-nucleotide polymorphism, we found that there are two different subunits *g* of which the coding regions are localized to different chromosomes; one (named here as subunit *g*-1) localized to chromosome 3 and the others (named as subunit *g*-2a and b) to chromosomes 11. The protein sequences of the different subunits *g* are almost identical; 97–99% identity and 98–99% similarity (Fig. 2). In addition, we found a sequence homolog of the subunits *g* in humans. Except for the lack of three C-terminal amino acid residues and the replacement of some amino acid residues, the subunit *g* homolog shows very high homology to the reported human subunits *g*; 88% identity and 90% similarity.

Careful examination of the database of another animal, fruit fly, also led us to find two different potential subunits *g* (gene products of l(2)06225 and CG7211). The two potential subunits *g* were reported separately as components of the proton-transporting ATP synthase complex. The two proteins show 51% identity and 63% similarity. In addition, in the search of subunit *g* homologs in the mouse genome database, we found two protein sequences which are significantly homologous to the subunit *g* of mouse and whose expressions are identified. They show 98–99% identity and 99–100% similarity to the murine subunit *g*.

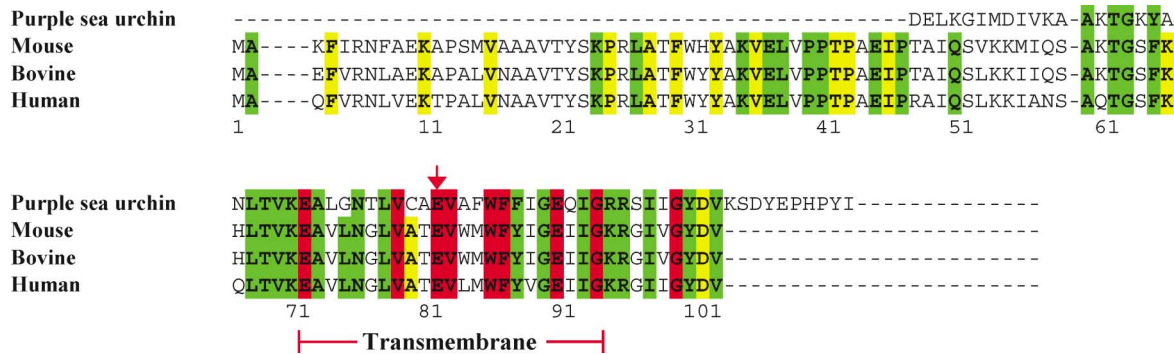
We further carried out BLAST search on the genomes of other animals of which the whole sequences are

**Table I.** Subunit *g* and Isoforms of Subunit *g*

Organism	Name	Location		Expression	Accession number
		Chromosome	Gene		
Human	Subunit <i>g</i> -1	11	ATP5L	Gene with protein product	O75964 (Strausberg <i>et al.</i> , 2002) and NP_006467 (Zhang <i>et al.</i> , 2000)
	Subunit <i>g</i> -2	3	DKFZp566G013 and AF092124.1:74..385	Supported by alignment with mRNA	CAB43378 or T08727 (Wiemann <i>et al.</i> , 2001), and AAC61597
	Subunit <i>g</i> Homolog 1	n/a	AF092923.1:328..630	Supported by alignment with mRNA	AAP97217
Fruit fly	Subunit <i>g</i>	2L	l(2)06225	Gene with protein product, a component of the proton-transporting ATP synthase complex	AAF53041
	Subunit <i>g</i> homolog 1	2L	CG7211	Gene with protein product, a component of the proton-transporting ATP synthase complex	AAF52549 and NP_609142 (Adams <i>et al.</i> , 2000)
Mouse	Subunit <i>g</i>	9	Atp51	Gene with protein product	NP_038823 (Jeon <i>et al.</i> , 1999)
	Subunit <i>g</i> homolog 1	X	LOC236764	Supported by alignment with mRNA	XP_125392
	Subunit <i>g</i> homolog 2	10	LOC231046	Supported by mRNA and EST alignments	XP_124482
Rat	Potential subunit <i>g</i> 1	8	LOC300677	Supported by alignment with ESTs	XP_217128
	Subunit <i>g</i> homolog 2	6	LOC298752	Ab initio	XP_233814
	Subunit <i>g</i> homolog 3	15	LOC305891	Supported by alignment with ESTs	XP_224187
<i>C. elegans</i>	ATP synthase <i>g</i> homolog ASG-1	I	asg-1	Gene with protein product	NP_492352 (Fraser <i>et al.</i> , 2000)
	ATP synthase <i>g</i> homolog ASG-2	X	asg-2	Gene with protein product	NP_509152 (Kamath <i>et al.</i> , 2003)

available. From the search of the subunit *g* homologs in the genomes of rat and *C. elegans* in which the sequences of subunit *g* have not been identified yet, we found two or more subunit *g* homologs in these organisms (Table I). In rat, three subunit *g* homologs with significant homologies and with similar peptide length to the human and murine

subunits *g* have been found from the BLAST search in the rat genome database. They showed 66–82% identity and 69–88% similarity among them. Interestingly, another protein composed of 233-amino acid residues whose expression was identified by EST and whose function has not been assigned, was found by a BLAST search to have



**Fig. 3.** Sequence alignment of animal subunit *g*. Conserved residues in the sequence alignment of animal subunit *g* and the isoforms of subunit *g* are shown in bold colored with red (completely conserved), green (highly conserved; more than 80% identity), and yellow (weakly conserved; more than 70%, but less than 80% identity). Only identified subunit *g*s of animals are shown in this figure. The conserved acidic residue in the middle of transmembrane helix is indicated by an arrow.

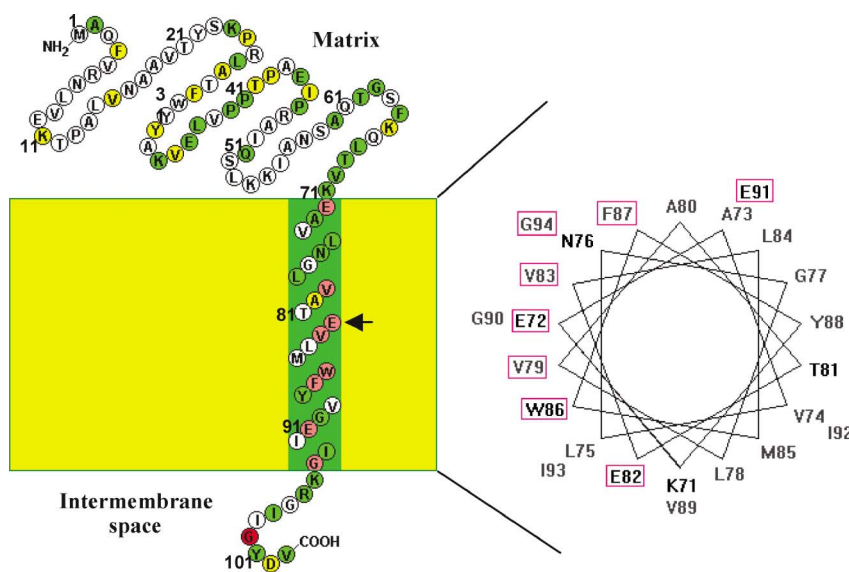
a C-terminal sequence highly homologous to the subunit *g* homologs of rat. The C-terminal sequence of the protein shows 84% identity and 91% similarity to the subunit *g* homolog-1, which is the most homologous among the rat subunit *g* homologs to the identified subunits *g* of human and mouse. On the other hand, in *C. elegans*, we found two subunit *g* homologs by scanning the genome. The pairs of homologs show 76% identity and 83% similarity.

No subunit *g* isoforms seem to exist in fungi as a BLAST search for subunit *g* homologs in the genomes of *S. cerevisiae* and *S. pombe* found no additional protein sequences which have significant homology to the subunit

*g*. Taken together, we suggest that subunit *g* does have isoforms in animals.

### A Conserved Acidic Residue in the Middle of Transmembrane Helix of Animal Subunit *g*

The sequence alignment of animal subunit *g* shows that the transmembrane domain is generally well conserved relative to the N- and C-terminal domains (Fig. 3). From careful examination of the transmembrane domain, we found that the completely conserved residues in the sequence alignment of animal subunit *g* are localized on one



**Fig. 4.** Orientation of the completely conserved residues in the transmembrane domain of subunit *g*, and the location of a conserved acidic residue in the middle of transmembrane helix. In the helical wheel diagram, the completely conserved residues are highlighted with boxes.

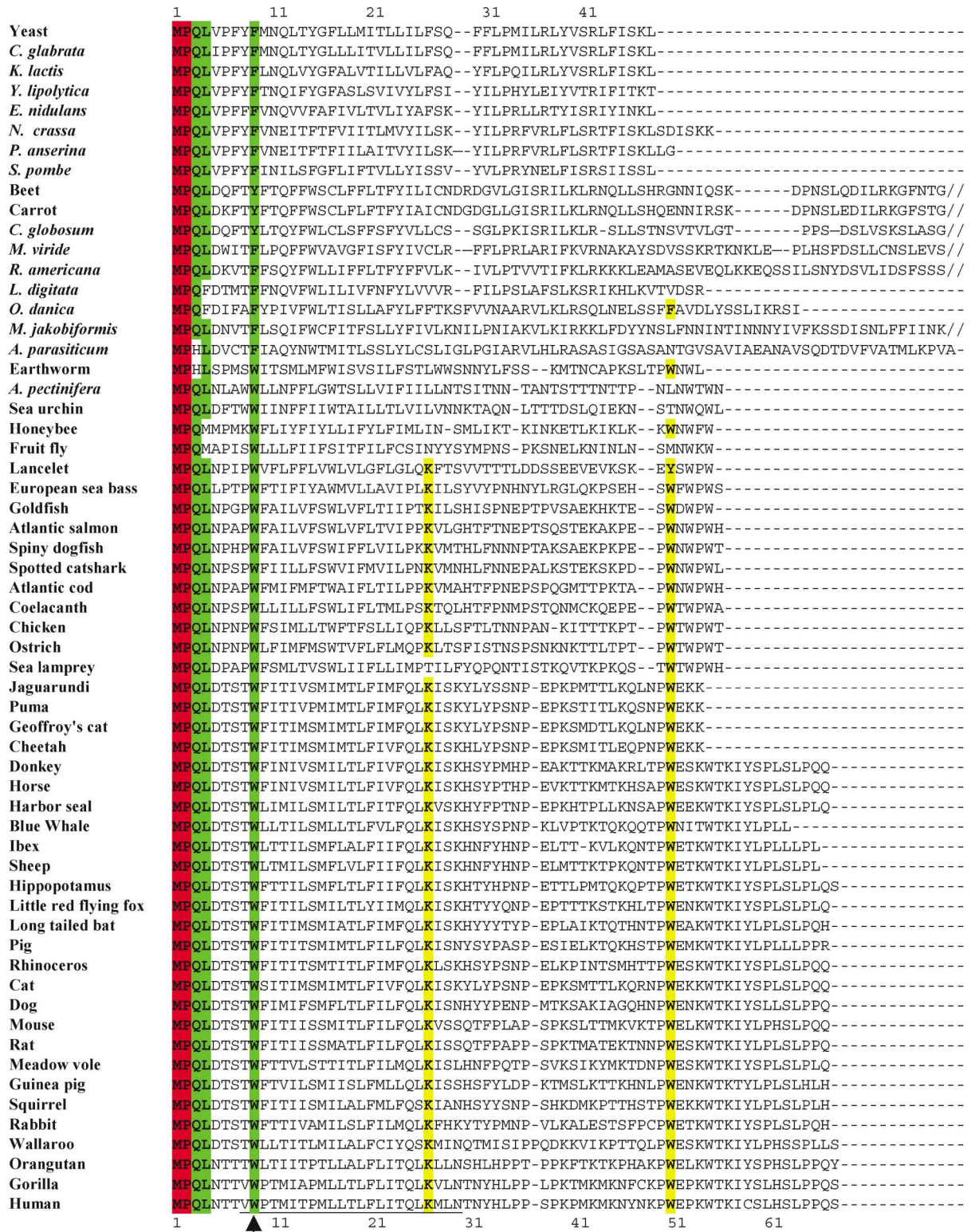


Fig. 5. Sequence alignment of A6L. The 60 out of 76 aligned A6L sequences are shown in this figure. Conserved residues are highlighted the same as in Fig. 3. The conserved aromatic residue in the MPQLX<sub>4</sub>r motif is marked by an arrow. The A6L from some species of which the sequences are much longer than most other A6L sequences are truncated at their C-termini (marked with //).

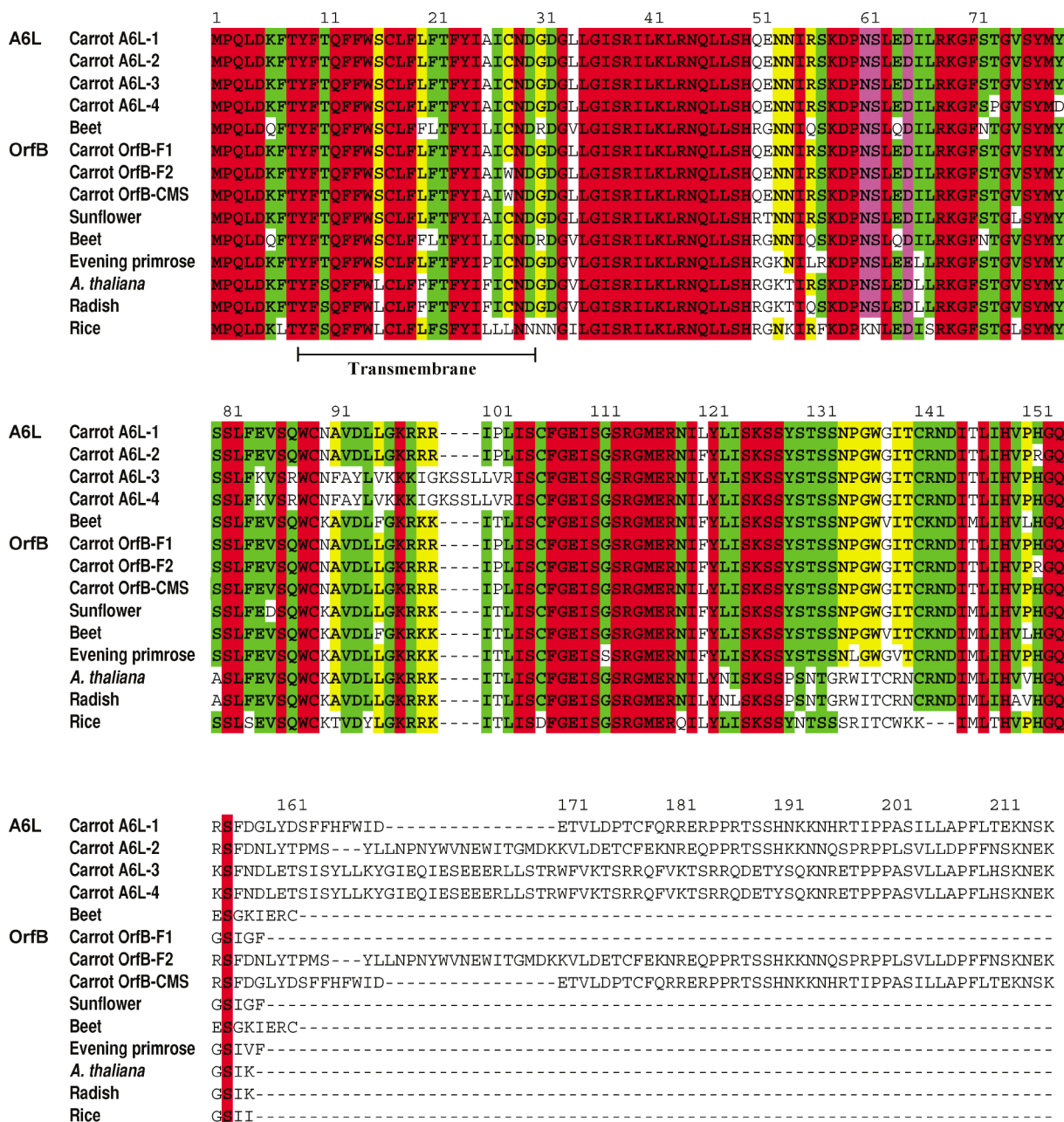


Fig. 6. Sequence alignment of plant A6L and OrfB. Conserved residues are highlighted the same as in Fig. 3.

side in the transmembrane helix (Fig. 4). The conserved side of the helix includes completely conserved two aromatic residues (Trp86 and Phe87 in the human sequence), and three acidic residues (Glu72, 81, and 91 in the human sequence). The conserved side is likely to play a major role in the interaction of subunit *g* with adjoining subunits since it contains most of the completely conserved residues of subunit *g*.

In the sequence alignment of animal subunit *g*, we found also in the transmembrane domain that three conserved aromatic residues (completely conserved Trp86 and Phe87, and highly conserved Tyr88 in the human sequence) are clustered and that a conserved acidic residue, Glu82 in the human sequence, is positioned in the middle of the transmembrane helix (Fig. 4). The location of the acidic residue attracts attention as its presence in the

middle of a transmembrane helix is uncommon. Interestingly, a rare case is found in another subunit in the ATP synthase complex, subunit *c*. In this subunit that comprises the central part of the  $F_0$  motor together with subunit *a*, a conserved acidic residue is located in the middle of transmembrane helix, and plays a crucial role in the proton translocation across the membrane. No other subunits of the ATP synthase are found to contain a conserved acid residue in the middle of a transmembrane helix. The role of the acidic residue conserved in the middle of a transmembrane helix of subunit *g* in animals needs further study, and is suggested here to play a significant role in the function of this subunit.

### A Conserved MPQLX<sub>4</sub>Ar Motif at the N-terminal Domain of the Subunit A6L

The supernumerary subunit A6L has been reported previously to contain a highly conserved MPQL motif from sequence alignment (Gray *et al.*, 1998; Stephens *et al.*, 2003). To determine whether other functionally important conserved motifs might be present, we retrieved and aligned three times more sequences than had been aligned previously. The careful examination of the 76 aligned sequences of A6L led us to discover that an aromatic residue is conserved at the fifth residue after the conserved MPQL motif at the N-terminal domain (Fig. 5). The other domains, i.e. the transmembrane and the C-terminal domains showed no residues or motifs that were significantly conserved.

Recently, it has been suggested based on the experimental data that the OrfB, which is involved in cytoplasmic male sterility in plants, is the plant equivalent of A6L (Heazlewood *et al.*, 2003; Sabar *et al.*, 2003). On the basis of this information, we retrieved and aligned the protein sequences of A6L and OrfB from plants. From the sequence alignment, as shown in Fig. 6, we found that the two proteins, A6L and OrfB are highly homologous, and in some plants where both the A6L and OrfB sequences are available, the sequences of the two proteins are the same (e.g., A6L and OrfB from beet), or almost identical (e.g., A6L-1 and OrfB-CMS, and A6L-2 and OrfB-F2 from carrot). As seen in the sequence alignment of A6L in Fig. 5, the lengths of the C-terminal domain of A6L are quite variable, and in plants, the C-terminal domain is remarkably elongated compared to fungi and animals. The long additional sequence at the C-terminal domain found in plants is highly conserved among plants except at the end region of the domain. This suggests that this sequence may be associated with a particular function that is unique to plants. Despite the variations in the A6L and OrfB sequences in animals, fungi, and plants, an aromatic

acid residue is always found conserved at position 9. The MPQLX<sub>4</sub>Ar motif that includes the conserved aromatic residue, seems to be the only conserved motif in A6L and plant OrfB, and is likely to play a role in the function of A6L.

### ACKNOWLEDGMENT

This work was supported by NIH Grant No. CA 10951 to PLP.

### REFERENCES

- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., and Ibegwam, C. (2000). *Science* **287**, 2185–2195.
- Altschul, S. F., and Koonin, E. V. (1998). *Trends Biochem. Sci.* **23**, 444–447.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). *Nucl. Acids Res.* **25**, 3389–3402.
- Arnold, I., Pfeiffer, K., Neupert, W., Stuart, R. A., and Schagger, H. (1998). *EMBO J.* **17**, 7170–7178.
- Belogradov, G. I., Tomich, J. M., and Hatefi, Y. (1996). *J. Biol. Chem.* **271**, 20340–20345.
- Boyle, G. M., Roucou, X., Nagley, P., Devenish, R. J., and Prescott, M. (1999). *Eur. J. Biochem.* **262**, 315–323.
- Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. (2000). *Nature* **408**, 325–330.
- Gray, M. W., Lang, B. F. R. C., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., Brossard, N., Delage, E., Littlejohn, T. G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, Y., and Burger, G. (1998). *Nucl. Acids Res.* **26**, 865–878.
- Heazlewood, J. L., Whelan, J., and Millar, A. H. (2003). *FEBS Lett.* **540**, 201–205.
- Hirokawa, T., Boon-Chieng, S., and Mitaku, S. (1998). *Bioinformatics* **14**, 378–379.
- Hong, S., and Pedersen, P. L. (2003). *J. Bioenerg. Biomembr.* **35**, 95–120.
- Jeon, D., Kim, H., and Baik, M. (1999). *Biosci. Biotechnol. Biochem.* **63**, 767–768.
- Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D. P., Zipperlen, P., and Ahringer, J. (2003). *Nature* **421**, 231–237.
- Pedersen, P. L., Ko, Y. H., and Hong, S. (2000). *J. Bioenerg. Biomembr.* **32**, 325–332.



- Sabar, M., Gagliardi, D., Balk, J., and Leaver, C. J. (2003). *EMBO Rep.* **4**, 381–386.
- Schuler, G. D., Altschul, S. F., and Lipman, D. J. (1991). *Proteins* **9**, 180–190.
- Stephens, A. N., Khan, M. A., Roucou, X., Nagley, P., and Devenish, R. J. (2003). *J. Biol. Chem.* **278**, 17867–17875.
- Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Altschul, S. F., Zeeberg, B., Buetow, K. H., Schaefer, C. F., Bhat, N. K., Hopkins, R. F., Jordan, H., Moore, T., Max, S. I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A. A., Rubin, G. M., Hong, L., Stapleton, M., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Scheetz, T. E., Brownstein, M. J., Usdin, T. B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S. S., Loquellano, N. A., Peters, G. J., Abramson, R. D., Mullahy, S. J., Bosak, S. A., McEwan, P. J., McKernan, K. J., Malek, J. A., Gunaratne, P. H., Richards, S., Worley, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hulyk, S. W., Villalon, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Young, A. C., Shevchenko, Y., Bouffard, G. G., Blakesley, R. W., Touchman, J. W., Green, E. D., Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M., Butterfield, Y. S., Krzywinski, M. I., Skalska, U., Smailus, D. E., Schnerch, A., Schein, J. E., Jones, S. J., and Marra, M. A. (2002). *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16899–16903.
- Thompson, J. D. (1994). *Nucl. Acids Res.* **22**, 4673–4680.
- Wiemann, S., Weil, B., Wellenreuther, R., Gassenhuber, J., Glasl, S., Ansorge, W., Bocher, M., Blocker, H., Bauersachs, S., Blum, H., Lauber, J., Dusterhofs, A., Beyer, A., Kohrer, K., Strack, N., Mewes, H. W., Ottenwalder, B., Obermaier, B., Tampe, J., Heubner, D., Wambutt, R., Korn, B., Klein, M., and Poustka, A. (2001). *Genome Res.* **11**, 422–435.
- Zhang, Q. H., Ye, M., Wu, X. Y., Ren, S. X., Zhao, M., Zhao, C. J., Fu, G., Shen, Y., Fan, H. Y., Lu, G., Zhong, M., Xu, X. R., Han, Z. G., Zhang, J. W., Tao, J., Huang, Q. H., Zhou, J., Hu, G. X., Gu, J., Chen, S. J., and Chen, Z. (2000). *Genome Res.* **10**, 1546–1560.