Minireview

Sequence anatomy of mitochondrial anion carriers

Petr Ježek*, Jan Ježek

Department of Membrane Transport Biophysics, No. 75, Institute of Physiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, CZ 14220 Prague, Czech Republic

Received 31 October 2002; revised 27 November 2002; accepted 3 December 2002

First published online 11 December 2002

Edited by Vladimir Skulachev

Abstract Two hundred and eighty-four genes of eight eukaryotic genomes for mitochondrial anion carriers were sorted into 43 (+18 single protein) subfamilies. Subfamilies differ by the number, nature, and locations of charges and polar residues in the *trans*membrane α-helices. Consequently, these residues and the rarely unique residues of the matrix and cytosolic segments most likely determine the different molecular phenotypes (functions). 'Common ancestral hydrophilic segments' were found in matrix and cytosolic segments, with interchangeable polar residues. Thus the hydrophobic microstructures of hydrophilic carrier parts are supposed to predetermine structure/conformation, whereas polar and charged microstructures should predetermine function, namely in the *trans*membrane spanning α-helices.

Key words: Mitochondrial anion carrier gene family; Mitochondrial carrier; Membrane transport; *Trans*membrane α-helix

1. Introduction

When the first few mitochondrial anion carrier proteins (MACPs)[1] were sequenced, it was recognized that they form a unique gene family of homologous proteins with quite striking features. Usually, their sequences are about 300 amino acid residues (AAR) long and are composed of threefold sequence repeats (yet imperfect) of ~100 AAR [1–9]. Predictions of their *trans*membrane folding identified six *trans*membrane α-helices, each two belonging to one of three triads. Although various lengths of α-helices were predicted [10,11], the model of Klingenberg for *trans*membrane folding[2] (shorter odd and longer even helices [12]) is preferable, since as we show here, it reflects a common architecture of all known carriers (genes, Table 1). Even its longer α-helices form structures theoretically shorter in length than the presumable 40 Å membrane width [13]. The most amazing feature of MACPs is that on the interface between the odd α-helices and the proximal matrix segments, unique specific sequences are located, commonly called the MACP signatures [10,14]: P-n/ϕ/OH-E/D-x-n/OH-K/R-x-K/R-x, where ϕ is an aromatic and n a neutral residue. Some irregularities are well recognized in Fig. 1. The MACP signatures probably predetermine formation of the matrix segments [13,15]. Concerning known carrier functions, more phenotypes were known in the 1980s than sequences [14,16–18]. With the advent of genome sequencing, 35 MACPs were found in *Saccharomyces* yeast [9,10] and novel phenotypes were identified and annotated for yeast [9,19–21]. Identification of 32 MACPs in *Caenorhabditis elegans* followed [7,22], as well as 43 in *Drosophila melanogaster* [23], and ~46 MACPs in the human genome[3]. Plant MACPs can be found in databases of the *Arabidopsis thaliana* genome (Table 1).

We have selected eight available genomes with 284 MACP genes (Table 1), defined as sequences containing MACP signatures. The Jotun–Hein type of a homology-based phylogeny tree was constructed (Clustal method of the Lasergene 99 Megalign program, using the Dayhoff PAM 250 matrix). The tree was tentatively sorted into at least 43 subfamilies (Table 1) plus 18 other, mostly single-gene-containing, subfamilies. For the purpose of this work, the uncoupling proteins (UCPs [13,15]) were left in one subfamily, although it was possible to sort them into the BMCP subfamily, the subfamily of UCP4s and of UCP1,2,3. Note that similarity of UCPs with other carriers is in line with the proposed function of UCPs as fatty acid anion carriers [13,15]. Overall, 22 sub-

*Corresponding author. Fax: (420)-2-96442488.
E-mail address: jezek@biomed.cas.cz (P. Ježek).

Abbreviations: AAC, ADP/ATP carrier; AAR, amino acid residues; BMCP, brain mitochondrial carrier protein (or UCP5); Ca-dep.per., calcium-dependent peroxisomal (carrier); MACP, mitochondrial anion carrier protein; UCP, uncoupling protein

[1] The term 'mitochondrial *anion* carrier protein family' is preferred here, even if some members translocate zwitterions at physiological pH (e.g. carnitine carrier). The reason lies in the assumption that mitochondrial cation carriers, such as the K^+/H^+ antiporter, Na^+/H^+ antiporter, K_{ATP}channel, Ca^{2+} carriers, etc., form another one or several other distinct gene families.

[2] The Klingenberg model [12] defines the three odd α-helices to be terminated by the second residue of each of the three MACP signatures. The odd α-helices are hence formed by the first two MACP signature residues (the first always being Pro) and by the preceding 16 AAR −18 for the fifth α-helix. The even α-helices are defined by the relative coordinates from the well conserved Gly residues located as the fourth, fifth, and fourth AAR (counting steps from the matrix interface) of the second, fourth, and sixth α-helix. Cytosolic and matrix segments are then given by the connecting parts between helices. Only in these connecting segments, we allowed for deletions/insertions.

[3] See http://drnelson.utmem.edu/mitocarriers.html.

0014-5793/02/$22.00
doi:10.1016/S0014-5793(02)03779-1

Table 1
Resulted sorting of mitochondrial anion carriers into subfamilies

SUBFAMILY	PHENOTYPE	Genes (proteins), used in the data set, fitting into the given subfamily: Human 51: 41 plus 10 in sole Hu subfamilies	Drosophila 45:	42 plus3	C.elegans 33:	32 plus 1	Yeast S.c. 35:	34 plus 1	S.pombe	22 Candida	38 Arabidopsis	22	Dictyostelium	38	SUBFAMILY
UCPs	FA- uniport	HuUCP1 HuUCP2 HuUCP3 HuUCP4 HuBMCP	DrAC009216 ,14955	DrAC017377 ,b	CeAF003384						Ara51418.t00050	Ara67268.t00007 Ara67885.t00017	DdJAXa74g07.r1		UCPs
AAC	ADP/ATPantiport	HuANT1 HuANT2 HuANT3	DrAC012929 ,b		CeAF003141	CeZ82059 CeU64842	YMR056c YBL030c YBR085w		SpAL023634.1	Contig6-2508			DdAF039211.1		AAC
plant AAC	ADP/ATPantiport	-			CeZ49207	(CeZ68882 in AACs)					Ara60034.t00004	Ara67221.t00010 Ara67921.t00019			plant AAC
Phosphate	Pi.H+symport	HuPhosphate HuP04	DrAC018185	DrAC020205	CeZ74028	CeX76113 CeU49830	YER053c		SpAL136536.1	Cntg6-2403b Contig6-2197			DdC94418	*DdVSD164*	Phosphate
citrate	citrate/diCOOant	HuCitrate	DrAC015137		CeZ22180		YBR291c=CTP1		SpZ97209	Contig6-1765 Contig6-2519					citrate
dicarboxylate	diCOOPi/diCOOan	HuNM_012140	DrAC020252, 009385	DrAC013100 @	CeU23525		YLR348c=DIC1			Contig6-2365					dicarboxylate
oxaloacetate	oxalAc/diCOOant	HuAC008053.5 HuAF125531.1	DrAC017347 ,b,c				YKL120w=OAC1		SpAL035085	Contig6-2514			DdCHR2.0.4869		oxaloacetate
oxodicarb.	oxoglu/oxoadip	-	DrAC020102	DrAC008182	CeZ81577		YOR222w YPL134c=ODC1		SpAL360054.1	Contig6-2027			DdJC1a211b04.s1		oxodicarb.
oxogl/malate	oxoglu/malate	HuOxoglutarate-malate	DrAC017782	DrAC014153 ,b	CeU80836										oxogl/malate
succ./fum.	succ/fumarate	-					YJR095w=SFC1	*YFR045w*		Contig6-2021	Ara67571.t00005		DdAU053402.1 *DdVSE757 DdCntg5352 DdIIAFP1D67915*		succ./fum.
Ca dep. perox.a		HuAC016331.2,w HuAL360268 HuAL359258 dtto b	DrAC020327		CeZ72511	CeZ50873									Ca dep. perox.a
Ca dep. perox.b									SpAL035085.1	Cntg6-2403c					Ca dep. perox.b
Ca dep.	HuNM_003705	HuAK000766.1 HuAC027760.1 HuAK023106.1	DrAC012788	DrAC018265 ,b	CeU00052	CeU58750 CeZ78542	YPR021c			Contig6-2169			DdContig9875 DdC90247.1 *DdIIBJP1D1394*	Ca dep.	
carnitine	carn/acetylcarn	HuZ28872	DrAC018177	DrAC019974	CeX76115		YOR100c=CRC1			Contig6-2313					carnitine
Flx1a							YEL006w YIL006w		SpAL133156.1 SpAL021766	Contig6-2076 Contig6-2403					Flx1a
Flx1b	FAD import ?	HuFLX1	DrAC019525				YPR128c			Contig6-2519c Contig6-2081			DdCHR2.0.42096		Flx1b
Flx1c	FAD import	HuAC012213.3		DrAC017981	CeZ49068		YIL134w=Flx1			Contig6-2498					Flx1c
Mrs	RNAsplicing	HuAC018410.3 HuAL353719.5			CeZ66521		YJL133w YKR052c		SpZ99168.1	Contig6-2443 Contig6-2443	Ara61269.t00031	Ara22505.t00002	DdIIAFP1T24509	*DdAU034372.1*	Mrs
ornithine-like	ornithine/H+ant?	-	DrAC014984		CeU23412						Ara67799.t00018				ornithine-like
Ort1	ornithine/H+ant						YBR104w YPR058w YOR130c=Ort1		SpZ69727.1	Contig6-2519b Contig6-2488,-2065,-2047					Ort1
Pet8		HuPet8	DrAC010580		CeZ68160		YNL003c=Pet8		SpZ70721.1 SpAL034353.2	Contig6-2512	*Ara51062.t00022*				Pet8
Rim2		HuNM_018155.1 HuAC073194.4	DrAC017153		CeX76116		YBR192w=Rim2		SpAL355632.1	Contig6-2321			DdCHR2.0.7192	*DdC90142*	Rim2
Ydl119c-a		HuAC026309.15								Contig6-2238			DdCHR2.0.42277	Ddcarrier G	Ydl119c-a
Ydl119c-b							YDL119c		SpAL355013.1	Contig6-1946 Contig6-1829					Ydl119c-b
Ydr470c							YDR470c		SpAL133157.1	Contig6-2093					Ydr470c
Ygr096w		HuAF182404.1	DrAC005709	DrAC013977	CeAF043695		YGR096w		*SpAL034433.2*	Contig6-2434	Ara60613.t00013	Ara67814.t00008	DdCHR2.0.52211		Ygr096w
Ygr257c		HuAC019152.5	DrAC010671,b	DrAC018172	CeZ46787		YHR257c		SpAL136535.1	Contig6-2291			DdCHR2.0.80699		Ygr257c
Yhm2a		HuAC022956.2 HuAL138752.5	DrAC012986		CeZ66220										Yhm2a
Yhm2b							YMR241W=Yhm2		SpAL035536.1	Contig6-2154					Yhm2b
Yhr002w	acc.CoAintomatrix						YHR002W=LEU5		SpZ98597.1	Contig6-2363					Yhr002w
Ymr166c							YMR166c		SpAL031966.1	Contig6-2135	Ara60016.t00014		DdM24569.1		Ymr166c
Ynl083w	Ca binding						*YNL083w*				Ara67572.t00002	Ara68068.t00012	DdIIAFP1D16968 *DdIIAFP1D17023 DdAU075923.1 DdADP1d1847*		
Ypr011c							YPR011c			Contig6-2125			DdIIBPP1D00123		Ypr011c
Yhm1							YDL198c=Yhm1		SpAL031525.1	Contig6-1664 Contig6-2272					Yhm1
Graves dis.ant.p	sole human	HuAC037456.6													Graves dis.ant.p
Subfamily 1	subfamily of :	HuAB007915.2	DrAC019755		CeZ70310	*CeZ71259*									Subfamily 1
Subfamily 2	only *Dictiostelium*												DdContig4406	DdIIADP1d1847	Subfamily 2
Subfamily 3	subfamily of :	HuTHC160221	DrAC020238		CeZ79755										Subfamily 3
Subfamily 4	only *Arabidopsis*										Ara52132.t00014	Ara158.t00028			Subfamily 4
Subfamily 5	only *Arabidopsis*										Ara60540.t00021	Ara67190.t00009			Subfamily 5
Subfamily 6	only *Dictiostelium*												DDC90619	DdJC1c25a09	Subfamily 6
Subfamily 7	*Candida&S.pombe*								SpAL022299.1	Contig6_2226					Subfamily 7
Subfamily 8	subfamily of :	HuZ58467.1	DrAC020113	DrAC007817	CeZ75955	CeZ75532					Ara60024.t00006		DdContig1152		Subfamily 8
Subfamily 9	*Dictio&S.pombe*			@=ArAC005445						Contig6_2503			DdCHR2.0.3104	DdContig16299	Subfamily 9
Sole Human subfamilies: 8															
Subfamily 10		*(AF177333.1~ornithine;AC008579.45)*													
Subfamily 11		*AL157871.2 gene1,2*													
Subfamily 12		*AL079303.3*													
Subfamily 13		*AC026346.3*													
Subfamily 14		*AC069303.2*													
Subfamily 15		*AF223466.1*													
Subfamily 16		*AC011427*													
Subfamily 17		*AF123303*													
Sole *Drosophila* subfamilies			*DrAC014256 DrosoColtgeneY12495 DrAC019385*												
Sole C.elegans subfamilies					*CeU80931*										
Sole *Sacharomyces* subfamily							*YJR077c ~MIR1~phosphate carr.*								
Sole *Arabidopsis* subfamily											*Ara67039.t00007*				
Sole *Dictyostelium* subfamilies													*DdCHR2.0.56803 Dd93974.1 DdJC2a30b10.s1*		

Altogether 284 MACP genes from the eight selected genomes were used for construction of the Jotun–Hein type of a homology-based phylogeny tree (not shown) by the Clustal method of the Lasergene 99 Megalign program, using the Dayhoff PAM 250 matrix. The tree was sorted into the 43 main subfamilies listed in each line of the table. The MACP genes (proteins) fitting to each subfamily are organized in columns according to the origin of the species (genomes). Nineteen genes forming sole branches (10 human with two closely paired) were left ungrouped, as they represent 17 other, mostly single-gene-containing, subfamilies. The 22 identified 'panphylogenetic' subfamilies, existing in most phylogenetic clades, are displayed without a background as well as other 'mixed' subfamilies. Eight yeast-specific subfamilies (plus one *S. pombe* and *C. albicans*-specific and the sole *Saccharomyces*-specific subfamily) have rows between the lines, whereas seven subfamilies of 'higher' eukaryotes not existing in yeast and nine sole human subfamilies have gray shadowing.

The gene names written in italics were not used for analysis such as displayed in Figs. 1 and 2, since either they belong to the single-gene-containing subfamilies or not all MACP signatures and/or helices were recognized in them, and consequently, the analysis would be ambiguous. The total number of sequences used for the phylogeny tree from each genome (species) is listed in the first line; with the number falling to the main subfamilies plus number of genes fitting to single-gene subfamilies, which cannot be analyzed for conserved residues. Note that Subfamilies 1 and 2 are the closest relatives of Ygr096w and Yhr002w, respectively. Subfamilies 3, 4, and 5 are closest to Ypr011c and the Ca-dep.per. subfamily. Subfamily 6 is closest to Flx1b, while Subfamilies 8 and 9 belong to the ornithine/Ort/carnitine cluster of the tree.

families were 'panphylogenetic', since they contained MACPs of species covering phylogenesis from the slime mold *Dictyostelium discoideum* to humans including all three yeasts and one plant (*Arabidopsis*) analyzed. Eight yeast-specific (plus one *Schizosaccharomyces pombe*- and *Candida albicans*- and one sole *Saccharomyces cerevisiae*-specific) subfamilies were found besides seven subfamilies of 'higher' eukaryotes (not existing in yeast) and nine sole human subfamilies (Table 1).

2. Characteristics of MACP subfamilies

The subfamilies were analyzed separately for *trans*membrane α-helices defined according to Klingenberg[2] [12] and for presumably water-exposed cytosolic or matrix segments. For α-helices, only the subfamily-conserved AAR at given positions are depicted by color coding (Fig. 1). For cytosolic and matrix segments all residues existing at given positions[4] for all MACPs of the given subfamily are depicted (Fig. 2). Hence, if one position can be occupied by six different AAR types, six colors appear at this position.

2.1. Main distinction between transmembrane α-helices and matrix or cytosolic segments

Strikingly distinct features of α-helices vs. matrix or cytosolic segments were revealed (Figs. 1 and 2). In α-helices, charged and polar residues were more frequently conserved at the given relative coordinates within the given subfamily (Fig. 1), whereas in cytosolic and matrix segments rather neutral (n), Pro, aromatic (ϕ), and Gly residues were conserved at the given locations. Such locations formed hydrophobic patches within the identified (Fig. 2) 'common hydrophilic sequences' at the terminal parts of matrix (in both ends of cytosolic, Fig. 3) segments, where the charged and polar residues are highly interchangeable at given coordinates[4,5]. Thus finding a negative charge in one member, a positive charge in another member, or polar or neutral AAR in the third subfamily member is not uncommon. Such common hydrophilic sequences probably evolved from one ancestral sequence of the MACP prototype and should not contain any specific features of subfamilies. They terminate the matrix and form the start and ends of cytosolic segments. The MACP signatures are actually superconserved sequences of this type. Recognizing these features, two hypotheses can be stated.

Hypothesis I: Subfamilies differ in the number, nature, and relative locations of charges and polar residues in the *trans*membrane segments. Neutral AAR (except of Gly and Pro determining conformation/mobility) represent only a hydrophobic milieu not contributing to the main interactions of the carriers with their transport substrates and regulatory ligands. These interactions are contributed by charged (polar) residues spread on α-helices by the pattern characteristic for each subfamily (Table 2). This pattern represents a substantial part of the molecular phenotype for the given subfamily (together with rarely unique AAR of matrix and cytosolic segments). It also defines the given subfamily (Table 2; together with other unique subfamily-conserved AAR) and may be used for future annotation of other subfamily members from new genomes.

Hypothesis II: In contrast, the cytosolic and matrix segments (presumably exposed into an aqueous microenvironment) could possess at most of the positions (except of the MACP signatures and E/D-G matrix doublets) any charged or polar residues, which just predetermine the hydrophilic nature of these polypeptide parts. Phylogenesis apparently did not conserve any special character of these AAR and what was only conserved is their polar or charged character. However, in presumably ancestral common hydrophilic sequences at certain positions[5] the MACP-conserved hydrophobic AAR (n, ϕ, G, P) are spread. They represent hydrophobic patches of a hydrophilic chain and determine the conformation of these water-exposed protein parts. They could also relay conformational changes, including those which cause the physical movement of the transported substrate.

In conclusion, hydrophobic microstructures are supposed to predetermine structure/conformation of carrier water-exposed parts (they may predetermine function by forming cavities, e.g. substituting Ile for Gly between subfamilies, etc.), whereas polar and charged microstructures should predetermine function in the *trans*membrane spanning α-helices. Figs. 1 and 2 demonstrate the pattern supporting the above two hypotheses. In the further text, we describe some detailed features.

3. Characteristic features of *trans*membrane α-helices in MACPs

3.1. First α-helix

The used model[2] [12] considers the 16 AAR prior to Pro (P1) of the first MACP signature as forming the first α-helix. Among 43 evaluated subfamilies (Fig. 1), nine possess a negative charge located[4] at positions −5, −6, −7 from P1 (at position 0). We call them as having the type I first α-helix. Fourteen subfamilies contain a positive charge (type II) at position −5. As Fig. 1 shows, a further 13 subfamilies contain only Gln/Asn or His (type III), another four only OH groups (Ser/Thr; type IV), whereas the remaining three contain conserved residues of neutral character only (type V).

In type I subfamilies, additional His are found in the Yhm1 (+2 OH), Ymr166c, succinate/fumarate, and Mrs subfamilies; the two Gln/Asn exist in the Yhm2-b subfamily (+2 OH); two OH/SH groups are found in some UCPs; one OH in the citrate subfamily. Type II subfamilies contain one to three OH groups in different positions. Among type III subfamilies, only four should not have additional OH groups (subfamilies 8, 9, Ort1, and Ydr470c). Subfamily specificities (Table 2) also include conserved His at −5 (shared also by Mrs) and Cys at −9 for phosphate and Cys at −7 for the carnitine subfamily (also in Graves'). Unique Gln/Asn at −10 exist for Yhm2-b and at −6 for ornithine-like and Yhm2-a subfamilies. OH groups at −11 are unique for the Yhm1 subfamily; OH at −1 for Ygr257c and Ypr011c; OH at −7 for Subfamily 6; OH at −9 for Subfamilies 1 and 7; OH at −12 for ornithine-like and Subfamily 6; OH at −14 for Subfamily 5.

[4] We have introduced relative coordinates of AAR in α-helices. In each triad, the MACP signature proline is set at zero position, following residues are numbered 1, 2, 3, etc., and preceding AAR as −1, −2, −3 (Pro coordinates). For even α-helices, however, well conserved glycines are set as zero (helix Gly coordinates).

[5] Due to the existence of variable parts, we use two kinds of relative coordinates. The first are Pro coordinates, the same as for helices. The second are (matrix) Gly coordinates, used for both backward and forward numbering. They begin from the conserved Gly at the end of each matrix segment. Note that these are different Gly than those employed for numbering of α-helices.

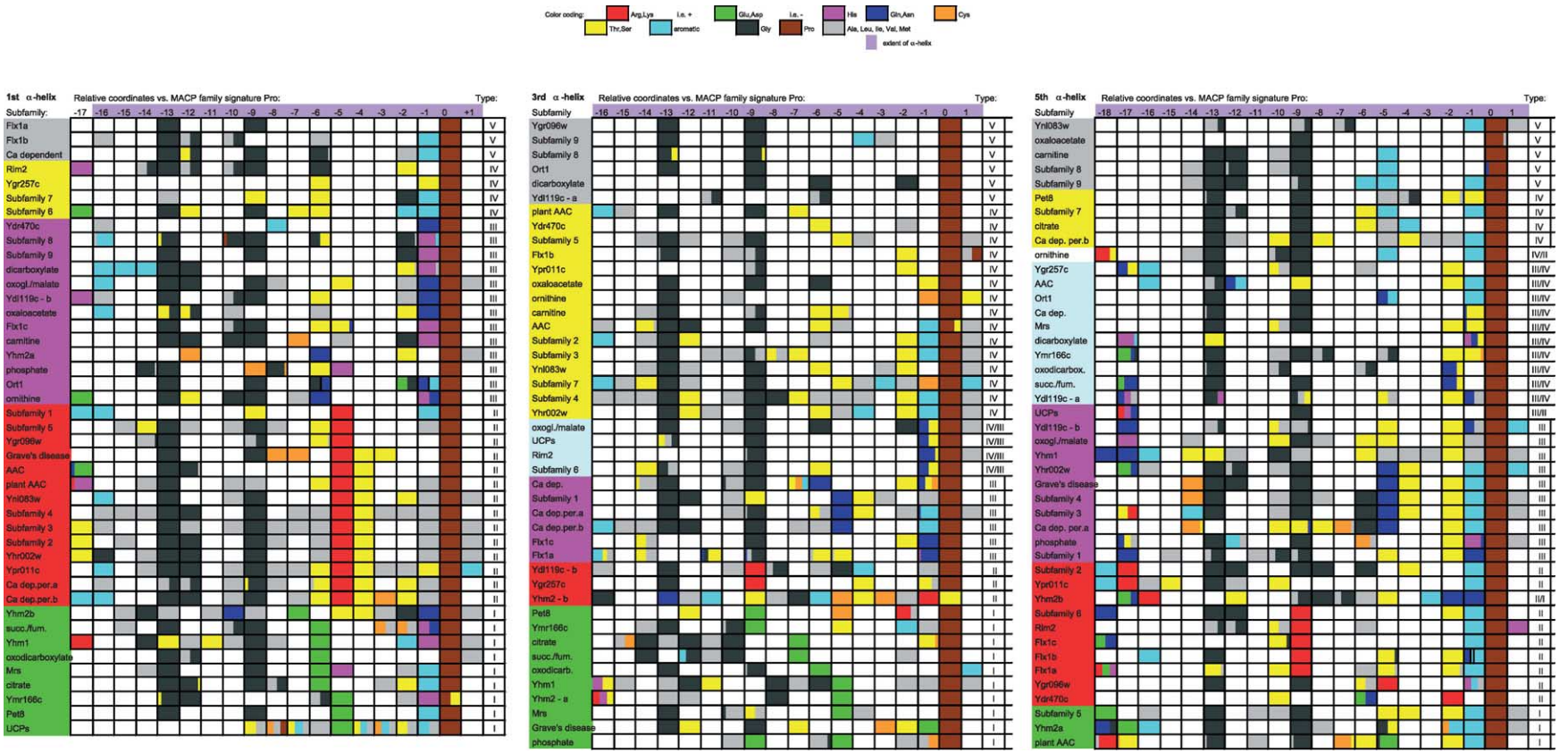

Fig. 1. Subfamily-conserved residues in *trans*membrane α-helices of mitochondrial anion carriers. The representative sequences of MACP subfamilies (only residues conserved within a given subfamily are depicted by color coding) are ordered for each α-helix according to polarity (least polar are placed at the top of each panel). When a gray color is displayed at a certain position, it represents an exact neutral AAR conserved (otherwise white spaces representing non-conserved AAR stand for neutral ones quite frequently). The typology of α-helices is given by the similar color-coded background around the subfamily name: type I (green) for α-helices with negative charges; type II (red) for α-helices with positive charges; type III (magenta) for α-helices with Asn, Gln and His; type IV (yellow) for α-helices with OH groups (Ser, Thr); and type V (gray) for α-helices with only subfamily-conserved neutral or aromatic residues or subfamily-conserved Gly or Pro. Some subtypes are recognized, such as a light blue background for the third and fifth α-helices, depicting alternations between NH_2 and OH groups (type III/IV); an orange background for the second α-helix, pointing out subfamilies having only a single positive charge in this helix (in addition to the one at position −1); and a pink background for the fourth α-helix, showing subfamilies of a zwitterionic type (having the positive–negative charge pair in addition to the positive charge at position −1). In the fourth α-helix, Subfamily 6, Flx1a and Rim2 contain a deletion at position +19, hence the first negative charge usually conserved as the first residue of the third cytosolic segment has the relative coordinate 19. Consequently, we do not consider Flx1a type I.

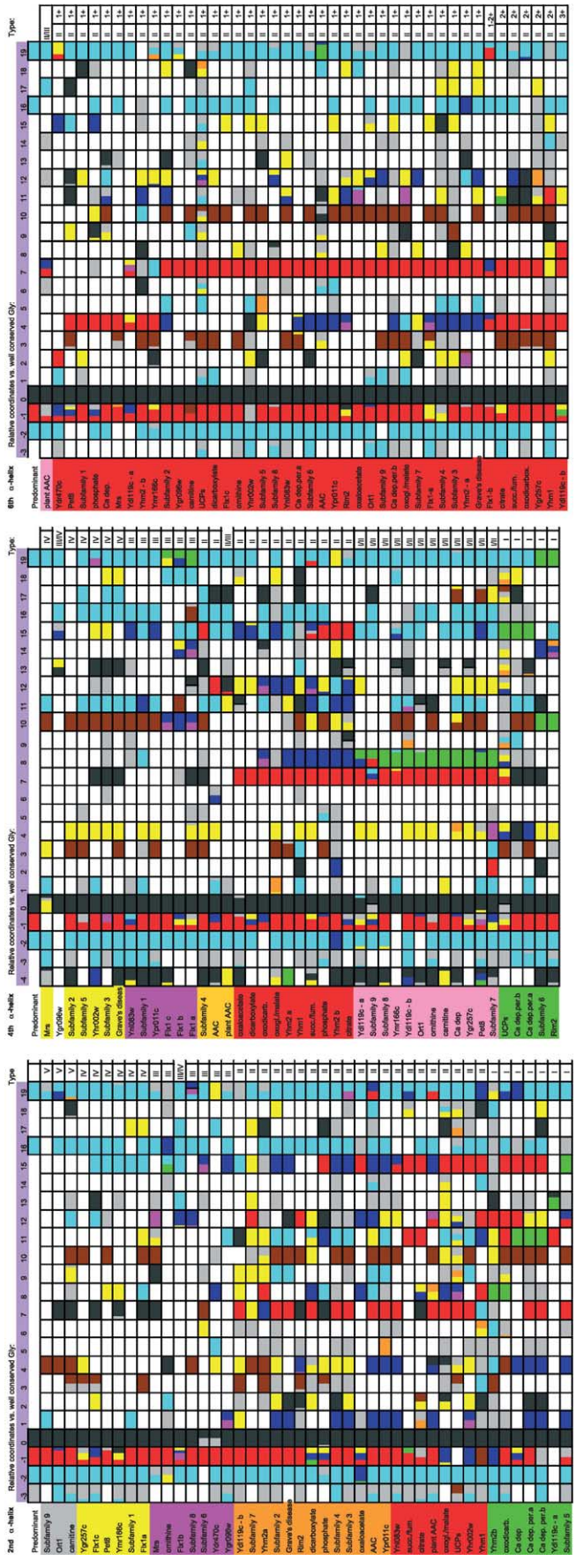

Fig. 1 (*Continued*).

3.2. Third α-helix

Twenty-six and 19 subfamilies contain the conserved Gly at positions −13 and −9, respectively, relative to the second MACP signature Pro (P3). Ten subfamilies contain 'type I' third α-helix, i.e. with negative charges. Only the phosphate subfamily contains two negative charges (at positions −9 and −5). Interestingly, seven subfamilies with negative charges in the first α-helix possess them also in the third α-helix. Those

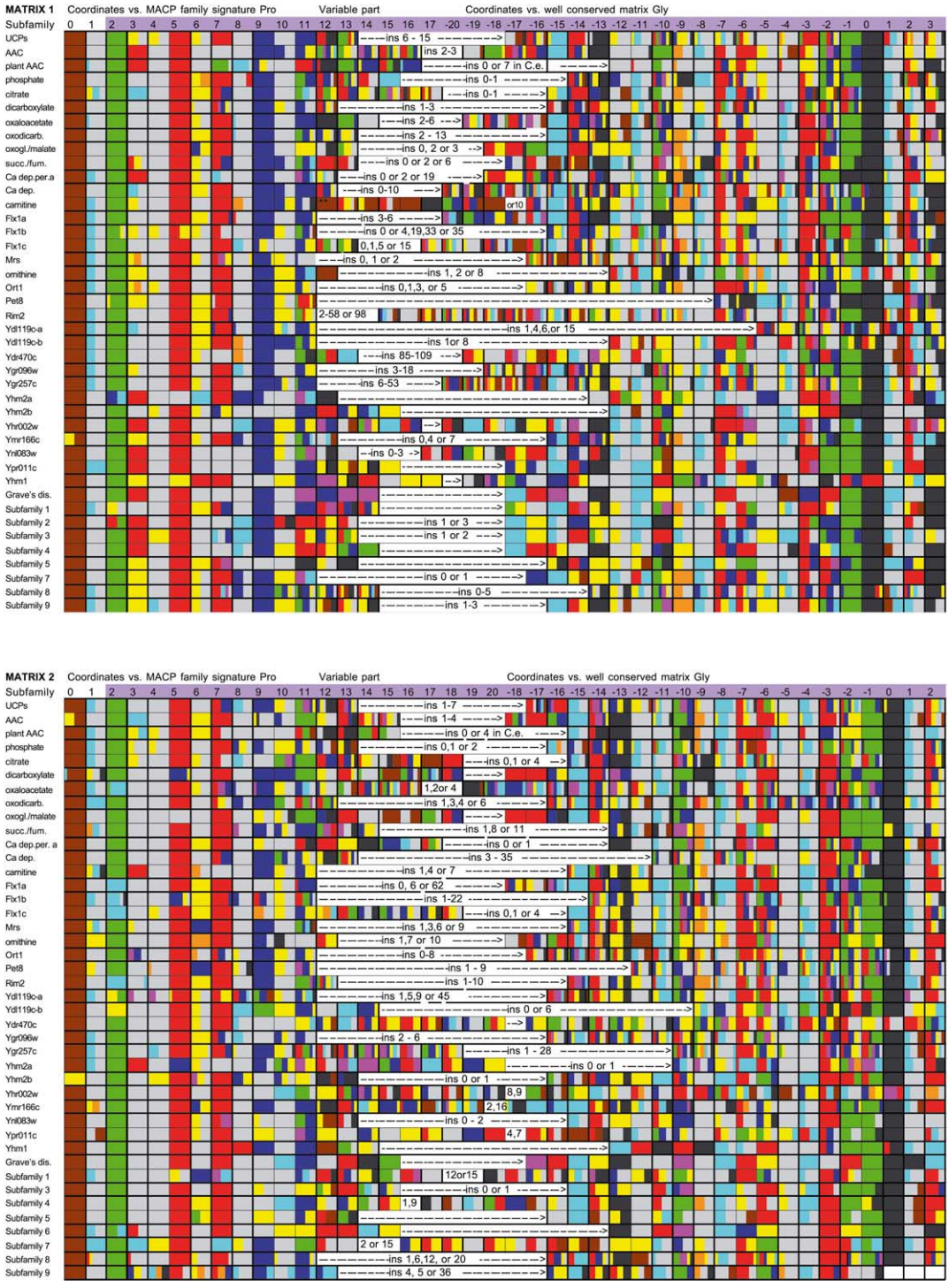

Fig. 2. Pattern of subfamily sequences in the matrix segments for mitochondrial anion carriers. The sequences of MACP subfamilies for matrix segments 1–3 are depicted by the same color coding as in Fig. 1. All residues of all proteins (genes) considered in this work from the given subfamily are represented in each position by at least one/sixth color strip. Relative areas of each color were made proportionally (when possible) to the relative abundance of the given type of residue (not for MACP signatures). Variable regions were put for each subfamily so that either no AAR are missing for the shortest member of each subfamily (or maximum two or three AAR of the shortest member are missing). The numbers following 'ins' (an insert) at arrows then describe the number of AAR in the variable part existing in all subfamily members. Where no number is displayed, the sequence is continuous. Missing subfamilies have either no well recognized segment or some were omitted for maintaining figure resolution (e.g. Ca-dep.per.-b).

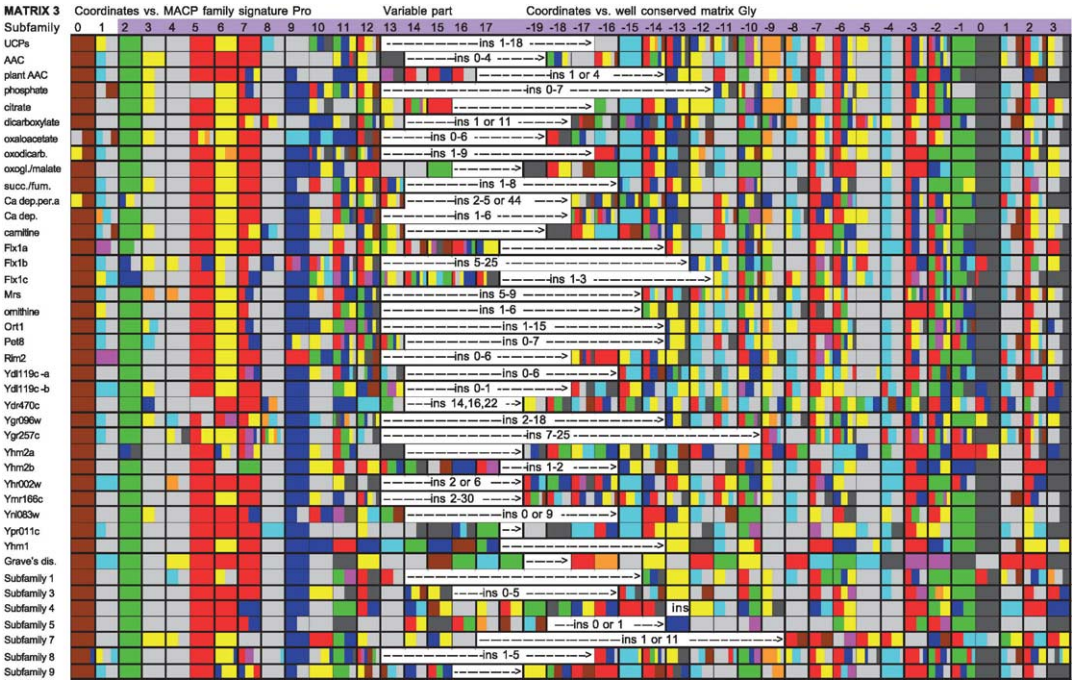

Fig. 2 (*Continued*).

are the Ymr166c (E/D at −9), Pet8 (−9), citrate (−7), succinate/fumarate (−7), oxodicarboxylate (−6), Mrs (−5), and Yhm1 (−5) subfamilies. Except for oxodicarboxylate and succinate/fumarate all other type I subfamilies contain one additional OH or SH group (Mrs and phosphate with degeneration). Some carriers of the Pet8 subfamily are 'zwitterionic' since they also contain the positive charge at position −2.

Only three subfamilies have one positive charge (Ygr257c and Ydl119c-b at position −9; Yhm2-b at −1). The latter also contains two Cys at −5, −3; one Gln/Asn at −13; and three OH groups (one instead of the MACP signature Pro). Ygr257c has up to two OH groups. Six subfamilies contain one NH_2/NH-bearing AAR (type III), 15 contain only OH groups (type IV) and four (UCPs, oxoglutarate/malate, Rim6, Subfamily 6) are of a mixed type III/IV due to degeneration. Subfamily 4 and Yhr002w contain three OH groups; Subfamilies 2, 3, 7, or Ynl083w have two OH groups. Six type V subfamilies exist. Unique features are represented by Gln/Asn at −6 in the Ca-dependent subfamily and at −13 in Yhm2-b; OH at −11 in Yhm1; Cys at −1 in the ornithine and Subfamily 7; at −3 in Yhm2-b and Graves' subfamily; Cys at −5 in Yhm2-b and Pet8. Yhm2-b contains a unique motif starting at position −7: G-ϕ-C-OH-C-n-K/R-OH.

3.3. Fifth α-helix

The fifth α-helix contains 18 AAR preceding the third MACP signature Pro (P5) [12]. Only three subfamilies have the type I fifth α-helix: Subfamily 5, the subfamily of 'plant' ADP/ATP carriers (AACs) (not its *C. elegans* carrier), and Yhm2-a. In 10 type II subfamilies the positive charge 'floats' from position −1 (Ydr470c); around positions −5 (Ygr096w), −9 (Flx1a,b,c, Rim2, and Subfamily 6), to positions −16 (Yhm2b) and −17 (Ypr011c and Subfamily 2). There are only 11 subfamilies of type III (several with degeneration) and a further 10 of mixed type III/IV. Four type IV subfamilies exist, plus the ornithine-like subfamily with alternating positive charge at −18. The carnitine subfamily has strictly conserved Gly at positions −12, −13, and −9. The positions −13 and −9 are occupied by Gly and −1 by ϕ in many other subfamilies. Uniquely conserved are at −17 His in the oxoglutarate/malate subfamily; OH groups at −15 for Ypr011c; at −14 for Yhm1; at −12 for Yhm2-b; and at −8 for the Ca-dependent peroxisomal (Ca-dep.per.)-a,b subfamily. Cys at −14 is a quite specific feature for the Graves' disease carrier and Subfamily 4.

3.4. Second α-helix

The second α-helix includes three residues prior to the conserved Gly (G2, position 0), G2 alone, and 19 following AAR[2], the last being the conserved aromatic residue. It begins with a conserved n-ϕ-K/R-G motif at positions −3, −2, −1, and 0. Consequently, we do not consider the positive charge at −1 in our typology. Overall, the second α-helix is rich in positive charges. Even if seven subfamilies with one negative charge can be classified as having the type I second α-helix, all of them also contain at least one or two positive charges and the oxodicarboxylate subfamily even has three positive charges. Four type I subfamilies have the positive charge at position 15, which is also held by eight type II subfamilies. Other positive charges are usually located at positions 7, 11, 12 in both type I and type II subfamilies. Twenty-one subfamilies exist with the type II second α-helix (eight having at least two additional positive charges, besides K/R at −1) and only three with a type V (neutral) second α-helix (carnitine with alternating OH group at +9). Unique residues are Cys at +5 for Ypr011c; His at +12 for Mrs; Gln/Asn at positions 7 and 16 for Yhm2a and the ornithine subfamily, respectively; and OH groups at positions 1, 6 and 19 for Subfamily 6, Ydl119c-b, and Ydr470c, respectively. Gln/Asn exist only in two subfamilies at position 11 (phosphate and Yhm1), as well

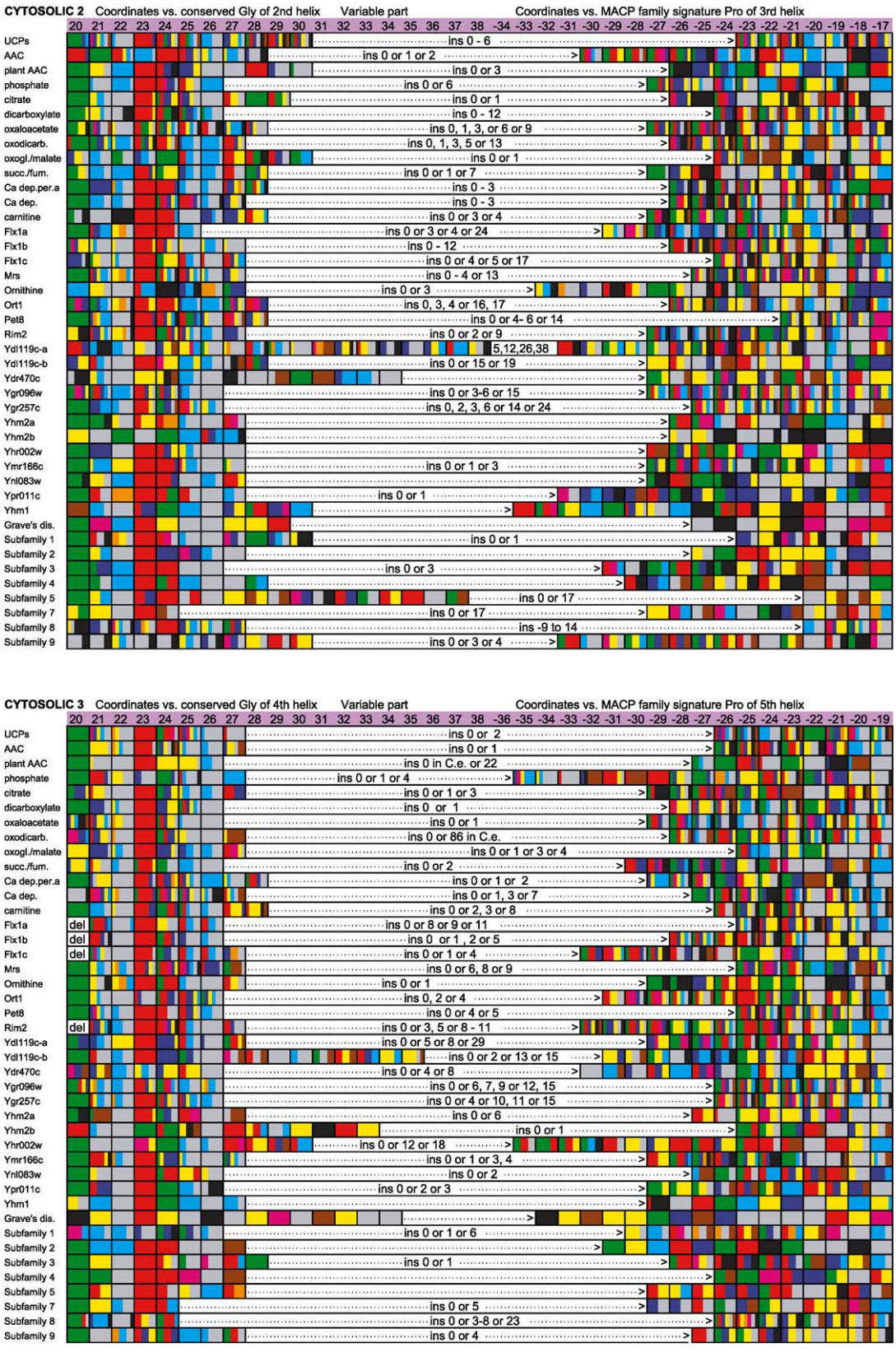

Fig. 3. Pattern of subfamily sequences in the cytosolic segments for mitochondrial anion carriers. The sequences of MACP subfamilies for cytosolic segments 2 and 3 are depicted by the same color coding as in Fig. 1. All residues of all proteins (genes) considered in this work from the given subfamily are represented in each position by at least one/sixth color strip. Relative areas of each color were made proportionally (when possible) to the relative abundancy of the given type of residue (not for MACP signatures). Variable regions were put for each subfamily so that either no AARs are missing for the shortest member of each subfamily. The numbers following 'ins' (an insert) at arrows then describe the number of AARs in the variable part existing in all subfamily members. Where no number is displayed, the sequence is continuous. Missing subfamilies have either no well recognized segment or some were omitted for maintaining figure resolution (e.g. Ca.dep.perox.-b).

Table 2
Characteristic and unique features of subfamilies of mitochondrial anion carriers

	Helix 1	Helix 2	Helix 3	Helix 4	Helix 5	Helix 6	
COMMON FEATURES	G-9 G-13	ϕ-2R-1G0 P10 ϕ16,19	G-9 G-13	ϕ-2R-1G0 ϕ19	G-13	G0 R-1 ϕ–2 ϕ19	
UCPs	I E-5	II R7,15 H/N8	IV N/OH-1	I E15	II/III R/H/N-17	II R7 N/H/OH12 OH/C18 P3 ϕ5	UCPs
AAC #	II R-5 OH-4	II R7 N1,4,12,15 OH11	IV OH-6	II R12 OH4	IV OH-2,-9 N/ϕ-12	II R7 N4 G11 **ϕ18** E/N19	AAC
plant AAC	II R-5 OH-4,-6	II R7 R/H12 N1,15	IV OH-7	II/I\ R/G12 OH11	I E-5 OH-2,-6,C/OH-7	II/III R/N7 **ϕ17**	plant AAC
Phosphate	III **H-5 C-9** OH-6	II R15 **N11** OH4 **G12**	I **E-5,-9**	II R7R/N15 N8 N/OH12	III N/H-1 OH-2	II R4 **OH10 N15**	Phosphate
citrate	I E-5 OH-2	II R11,15	I **E-7 G-10-14**-12	II R7,15 N8,11 N/OH12	IV OH-6	II2+ R4,7 E/OH11 P3	citrate
dicarboxylate	IIIH/A-1OH/A-2 **ϕ-15,-14**	II R7 N8 OH2,10,12	V **G-2**,-6	II R7 N/OH15 OH12	IV OH-13 H/ϕ-17	II R7 OH12 ϕ1 P3,10	dicarboxylate
oxaloacetate	III N-1 OH-2	II R15 N8,12	IV OH-1 G-6	II R7 N15 OH12	V common (G-9)	II R7 H/ϕ11 OH12 **ϕ18** P10	oxaloacetate
oxodicarb.	I E-5	IZ **E8** R11,12,15 P4	I **E-6**	II R7 N/H8,12	IV H/OH-2	II2+ R4,7 G11,**12** P3,10	oxodicarb.
oxogl/malate	III N-1 OH-5	II R7,15 N8 **OH13**,12,10	V N/OH-1 **G-2**,-6	II R7 N12,15 OH4,**18**	III **H-17** OH-2,-5,-6,-10	II R7 **H11 ϕ4 G9** OH2,12 P3 ϕ1	oxogl/malate
succ./fum.	I E-5	II R11,15	I **E-7 G-10,-14**	II R7R/N15 N8,11OH4,10	IV H/OH-2 E/N-17	II2+ R4,7 G11 N12 P3,10	succ./fum.
Cadepperox.a, b	II R-5OH-2,-4,-6**C-3inb**	IZ E11R7,15N1,4**OH18inb**	III N-5 OH-4 for a	I E15 N4 in a T18 inb	III N-**5OH-8**-2-9-10C-14-7	II R7 N4 OH15 in b P3(10inb)	Ca dep. perox.a, b
Ca dep.	V not conserved	IZ E11 R12,15	III **N-6** OH-2	IZ E8 R7 C/OH4 OH12	III N/OH-1	II R4 **ϕ12** P10	Ca dep.
carnitine	III **C-7** H-1	V P4	IV OH-6	IZ E8 R7 **OH1**	V G-12	II R7 **ϕ9 G18** P10	carnitine
Fix1a	V common (G-13-9)	IV OH11,17 P3	III N/H-1 OH/n-14,-6,-5	III N/H10,14 P3 E19	IIR-9 OH-10OH/n-13-5-2	II R7 N/H4 OH15 P/OH3 P10	Fix1a
Fix1b	V ϕ-1	III N/OH12 **ϕ13**	IV OH-2	III N10 N/OH14	II R-9 OH-2 OH/G-5	II2/1+ R/N4,7	Fix1b
Fix1c	III H-1 OH-6	IV OH2 **G13**	III N-1 OH-2	III N/H10	II R-9 OH/G-10	II R7 OH15 P10	Fix1c
Mrs	I E-5 **H-5**	III **H12 G9**	I E-5	IV OH3	III N/OH-1 OH-2	II R4	Mrs
ornithine	III **N-6 OH-12**	III **N16 ϕ2 G4**	IV **C-1**	IZ E8 R7 OH4	II/IV R/OH-18	II R7 OH8 P3	ornithine
Ort1	III/V	V P4	V common (G-9-13)	IZ E/n8 R7 OH4	III/V N/ϕ-5	II R7 N/OH12 P10	Ort1
Pet8	I E-5	IV OH8	I E-9 **C-5** OH-12	IZ E/N8 R7 N15 OH12	IV OH-1	II R4 OH9 ϕ15 **G11,17**	Pet8
Rim2	IV OH-2	II R12 N/H4	IV N/OH-1	I **E10** OH4	II R-9 H1	II R7 N/H4 N11 OH18 P10	Rim2
Ydl119c a	not conserved	IZ E/G13 R11	V G/OH-11	IZ E/N8 R7 OH4,12	IV OH-2 N/n-1 N/H/n-17	II R/OH4 R/H/OH7	Ydl119c a
Ydl119c b	III N-1 OH-6	II R12**OH1**,4,8,9 P3**G11**	II **R-9** ϕ-4	IZ E8 R7 OH4	III N-17N/n-1 OH-2,-5	II3+ R4,7, **R8** OH11	Ydl119c b
Ydr470c	III N-1 **ϕ-8**	III N8 **OH19**,4 ϕ5,17	IV OH-6	not recognized	II R-2 E/N-6 OH-10	II **R2 N15** ϕ1	Ydr470c
Ygr096w	II R-5	III N/H1N15	V common (G-9-13)	III/IV N/n15 OH/G13	II R-5	II R7	Ygr096w
Ygr257c	IV **OH-1**,-6	IV OH4	II **R-9** OH-4	IZ E8 R7 OH12	IV OH-2 ϕ-16 N/OH-17	II2+ R4,7 OH17 **C12** P10	Ygr257c
Yhm2a	III **N-6 C-12** OH-2	II **R8 N7** P3,4 OH9,11	I E-5	II R7 N8,12 OH11	I E-17	II R7 N4 **N16** OH8 **ϕ12** G13 P/H2	Yhm2a
Yhm2b	I **E-7** N-1, **N-10**	IZ**E8**R12P3OH10,11**G13ϕ4,6**	II**R-1C-3,-5** N-13 OH-4,-9	II R7,15 **N2**,8,11OH10,14	II R-16 N-1,-2**OH-12**-5	II R4 N11 OH12 **ϕ10 G6,8** P3	Yhm2b
Yhr002w	II R-5 OH-4	II R7,15 N1 OH2,4 **G18**	IV OH-2,-6,-12 ϕ-4	IV OH4,15	III N-5 OH-4 E/N-17	II R7 OH2,15 **ϕ17** P10 ϕ5	Yhr002w
Ymr166c	I E-5 H-1	IV OH8	I E-9 OH-5 **ϕ-2**	IZ E/ϕ8 R/OH7	IV H/OH-2E/N-17C/OH-1	II R4 N/OH12 **ϕ7**	Ymr166c
Ynl083w	II R-5 OH-2,-4	II R7 R/N15 N1,4OH12	IV OH-2,-6	III N15	V common (ϕ-1)	II R7 OH13 N/OH11 P3,10	Ynl083w
Ypr011c	II R-5 **OH-1**,-4,-6	II R7 **C5** N1,4,15 OH12	IV OH-2	III N15 OH4	II R-17 **OH-15**,-10	II R7 N4 OH11,15 **ϕ6**	Ypr011c
Yhm1	I E-6H-1**OH-11,-13ϕ-2**	II **R8**,12,15 **N11 ϕ3,7,14**	I E-5 **OH-11** G-6	II R7 N8 OH4,12	III N-18-17OH-2-6-13-**14**	II2+ R4,11 OH2,7,8,15 **ϕ14 G9** P3,10	Yhm1
Graves dis.ant.p	II R-5 OH-3,-4	II R7 N1,15 **G12**,2ϕ9,11	I **E-1** OH-2,-7,-12 **C-3**	IV OH4,18 P3,10 G7,13	III N-5OH-2,-4G-12 C-14	II R7 N4,12 OH11,17,18 P10 ϕ5	Graves dis.ant.p
Subfamily 1	II R-5 **OH-9**	V **ϕ1** OH17	III N-5 OH-4,-9	III N11 OH4	III N-1 OH-2,-5	II R4 OH2,12 **G18**	Subfamily 1
Subfamily 2	II R-5 OH-4	II R7 N1,15 **OH18**,2,4	IV OH-2,-12	IV OH4 P3	II R-17 OH-4 **G-8**-12	II R7 N4,11 G13 OH12,18 P10 ϕ5	Subfamily 2
Subfamily 3	II R-5 OH-4	II R7 N1,15 OH2,4	IV OH-2,-7	IV OH4,15,**18**	III N-5 OH-2,-4R/OH-17	II R7 N4 **G8 P9** OH2,15,17,18 ϕ5	Subfamily 3
Subfamily 4	II R-5 OH-4	II R7 N1,15 OH4	IV OH-2,-7,-12	II R15	III N-5 OH-2,-4C-14G-6	II R7 N4 G15 OH2,11,17 P10 ϕ5	Subfamily 4
Subfamily 5	II R-5 **OH-14**,-6	IZ **E15** N1,4 R7	IV OH-5	IV OH4	I E-17 OH-2,-4,-5	II R7 **C5** OH4,15 P10 G2 G13	Subfamily 5
Subfamily 6	IV **OH-7**,-6-12 **ϕ-2**	III N/H15 **OH6 P7**	IV N/OH-1 OH-6,-14	I **E10**,19 **G2 N14** OH4	II R-9 OH-2 G-12 N-18	II R7 N4 G2 P10	Subfamily 6
Subfamily 7	IV **OH-9**,-6	II R7OH7,9,11,**13**,15,17P4	IV OH-5,-14 **C-1 ϕ-3**	IZ E8 R7,2 **H4** OH12,**17**	IV OH-6	II R7 N12 OH4,9 G2	Subfamily 7
Subfamily 8	III H-1	III N/H12 P4	V common (G/S-9-13)	IZ E8 R7	V G-12	II R7 N/OH12	Subfamily 8
Subfamily 9	III H-1	V P4 **ϕ1**	V ϕ-4	IZ E/R8 R/N/ϕ7 OH4	V G-12 **ϕ-6**-5	II R7 N12 OH8 P3,10 ϕ1	Subfamily 9

@ only one sequence known # CeU64842 is not included Ca dep per b is IV

	Matrix 1	Matrix 2	Matrix 3	Cytosolic 2	Cytosolic 3	
COMMON FEATURES:	MACP signatures & common ancestral hydrophilic sequences			Common hydrophil.s.	Common hydrophil.s.	
UCPs	common	common	ϕ-15		E-22	UCPs
AAC	R-6,E-10,G-13, ϕ-15	E11 R-17 ϕ-15 G-13	**G13** ϕ-15			AAC
plant AAC	common	R-7 G-13	common	R28		plant AAC
Phosphate	**ϕ15**	OH10	common		**R-29**	Phosphate
citrate	common	E11 **R18 P17** G-13	**R15 OH-12 C-9** ϕ-15	**E28 G-25**		citrate
dicarboxylate	common	**E11,-10 R-18**,-17,-6 ϕ-15 **G-9**	common			dicarboxylate
oxaloacetate	**G12**, ϕ-15	common	E-10 ϕ9-15			oxaloacetate
oxodicarb.	ϕ-15 R-3	common	E-2 R-3 ϕ-15			oxodicarb.
oxogl/malate	E-16,R-14,ϕ-15,OH11,-12	E11,-2 R13-**18**,-17,-**6** ϕ-15	**R12**,-6,-3 **E15,-16**,-10 **G-19**			oxogl/malate
succ./fum.	G-13, ϕ-12	common	E-2			succ./fum.
Ca dep. perox.a, b	G-1 E-2	**R12 OH13 G14** G-13	common	N21 E-18 both for a	P-22	Ca dep.perox.a,b
Ca dep.	common	common	ϕ-15			Ca dep.
carnitine	OH-5	common	**G14=-18** R-7 **ϕ-16**	G21		carnitine
Fix1a	E-2,R-3,OH-6-9,G-10-13,ϕ–15	ϕ–8-15	common			Fix1a
Fix1b	common	common	common			Fix1b
Fix1c	common	ϕ-15	common			Fix1c
Mrs	common	common	common			Mrs
ornithine	**P12**	**C6 P-12** R-7	R-3	G24 N-17 E-27	**G-21**	ornithine
Ort1	R-7	common	R-7			Ort1
Pet8	G-1, ϕ1-5	common	common			Pet8
Rim2	common	common	**H1 R9,-16**			Rim2
Ydl119c a	common	common	common			Ydl119c a
Ydl119c b	R-7	**E9** OH10 ϕ13	common	N-17 OH-27		Ydl119c b
Ydr470c	**ϕ13**	does not have common seq.	ϕ12	**E30 P31**OH23,-23,-17 OH-22		Ydr470c
Ygr096w	common	common	common			Ygr096w
Ygr257c	common	common	common			Ygr257c
Yhm2a	**G12 N8,-5** R-3	**H14 N19 OH20 ϕ 17**	**G13** R-3 **R-1** no com.seq	**E-20**		Yhm2a
Yhm2b	**OH15 R12** R-7,-**2**	**OH-12** ϕ-11,-8	**N16 N-1**	**G-17-20-24**	R20,27,32G31OH30,33	Yhm2b
Yhr002w	**ϕ15**,-**19**,-4 G-13	**OH12** ϕ10 no common seq.	ϕ-15	**N21 E-22**	R-21 R-28	Yhr002w
Ymr166c	common	ϕ-15-17	common			Ymr166c
Ynl083w	E-2 R-16,-3	**N-10** OH-13	**R-14** E-10 ϕ-15	E-25		Ynl083w
Ypr011c	**OH15**,-9 E-2 ϕ-11,-15 G-13	**R20 N19 OH12,16 P-15**	**E17 G15,-18** OH-13,-6 **ϕ-17**,-15	N-23		Ypr011c
Yhm1	**H12 N-15 OH16,17-6** R14,-10,0	**P12 R-12-10 ϕ-13 G-11-9**	**E14 R11 N12,15,-6 ϕ13** OH-13	**P20R-20G-25**-21E-29R-33ϕ-26-30; N-23 R-27		Yhm1
Graves dis.ant.p	**H10,13,14-15 OH-10** R-2-10-16	P-4; **E15** R13-**16 H-10-17 OH-14-6**;	**E16,18 R-17-12-8 C-16 H-3-2**	OH27,28 R29 H-18,-20 G-21-**24**		Graves dis.ant.p
Subfamily 1	R-3 ϕ–15	R-7	OH-13 E-2	**E28**		Subfamily 1
Subfamily 2	**R11 N-2-14** OH-13-16 ϕ-11-17	G-10; n.d	@ n.d.	R-23 E-24	G-24N-27R-28OH-30	Subfamily 2
Subfamily 3	R-14-10 OH-5-16 ϕ-4-17	R13,-6 ϕ-15-8	OH10,-13			Subfamily 3
Subfamily 4	**R11**-16-6 **E14** ϕ-17 G-13	**E-10 N-18** G-13 ϕ10-15	**E11,19 R23** OH-12	E-17 **P-20** G-21	H25 H-24	Subfamily 4
Subfamily 5	**G11,13** E-6 OH-13 ϕ-7	common	**N-13 G11,15**	R35 OH27		Subfamily 5
Subfamily 6	common @	**R12,13,-11 H-10** N-7 **OH10,11,15**,-6	n.d.			Subfamily 6
Subfamily 7	**H-16** OH-13 C-9 ϕ-15	E11,-6 **OH-12-13** ϕ7	R10 no common seq.	OH-27		Subfamily 7
Subfamily 8	common	common	common			Subfamily 8
Subfamily 9	OH-5 C-9 ϕ-15 G-13	deletion of 0,1,2	E-10 **R-16 OH-19**,10 ϕ-15			Subfamily 9

Grave's in cyt3: H29OH28,32,P31 H-19R-21N-26-28P-27-32OH-30

Legend of Table 2 on next page.

as OH at 13 (oxoglutarate/malate and Subfamily 7); and OH at 18 (Ca-dep.per.-b and subfamily 2).

3.5. Fourth α-helix

The fourth α-helix includes four AAR preceding the conserved Gly (G4), G4 alone, and 19 following AAR, the last one being the aromatic residue. It begins with conserved ϕ-K/R-G at positions −2, −1 and 0. Twenty subfamilies contain here a 'central Arg' at position +7. Among them 12 are zwitterionic (five with degeneration), having one negative charge at +8. Five subfamilies possess the 'true' type I fourth α-helix with a sole negative charge (in UCPs OH alternate with the positive charge) at +10 (Subfamily 6, Rim2) or +15 (Ca-dep.-per.-a,b, UCPs). Type I to IV subfamilies also quite frequently have OH groups at positions +4 or +12; Pro at +10; ϕ at +15, +16. Yhm2b or the citrate subfamily contains two positive charges at +7, +15; (degeneration exists at +15 in the phosphate and succinate/fumarate subfamilies), plus four (or three) NH_2-bearing groups. Yhm1, Yhm2a, and four 'carboxylate' subfamilies have only one positive charge (at +7). Subfamily 4, AACs, or plant AACs do not contain a positive charge at +7, but at +15 and +12, respectively (degenerated to Gly in plant AACs). The unique Asn/Gln at +2 can be recognized for the Yhm2-b subfamily; at +4 for Ca-dep.per.-a; at +14 for Subfamily 6; His at +4 for Subfamily 7. Unique OH groups at +17 exist in Subfamily 7 and at +14 in Yhm2-b; and at +1 in the carnitine subfamily.

3.6. Sixth α-helix

The sixth α-helix begins three residues prior to the MACP-conserved Gly (G6 set at 0), terminates at position +19 [12] by aromatic AAR, common also at +16, and possesses the most common architecture, starting with x-ϕ-K/R-G at −2 to 0. Position +10 is usually occupied by Pro. Most conserved are the positive charges. In addition to one at position −1, all but one subfamily have at least one additional positive charge; seven subfamilies share two additional charges at +4 and +7 (+4, +11 in Yhm1; one rarely degenerated in dicarboxylate, both degenerated in Flx1-b or Ydl119c-a); the eighth has three (Ydl119c-b; but the fourth one at −1 is degenerated). Another 27 subfamilies contain the additional positive charge at +7; others (cluster of eight subfamilies) have it at +4. The Ydr470c subfamily has it at +2. *Arabidopsis* AACs contain only Asn/Gln at +7, whereas *C. elegans* AAC has a positive charge here. Rather unique are His at +11 for the oxoglutarate/malate subfamily (degenerated in oxaloacetate); Cys at +12 in Ygr257c; Gln/Asn at +16 in Yhm2-a or at +15 in the phosphate and Ygr470c subfamilies. Unique is also OH at +10 in the phosphate subfamily. OH or NH_2 groups are scattered between +8 and +19 in many subfamilies, but usually shared by several of them.

4. Detailed features of matrix segments in MACPs

Matrix segments begin by the third residue of the MACP signature [12]. Consequently, the first seven residues (up to position +8)[5] of each matrix segment are well conserved. The common motif, however, also includes the following Q/N-n/OH, conserved in the first and second matrix segments or Q/N in the third matrix segment. After these nine residues and one variable AAR (two in the third matrix segment), a variable part begins at position +12 (+13 for the third matrix segment). This means variable even within the same subfamily – in some carriers no residues exist, whereas in others very long sequences are found. The following (third) part bears the above-described common hydrophilic sequences.

4.1. The first matrix segment

The first matrix segment extends from[5] Pro coordinate 2 up to Gly coordinate 3 (corresponding to −7 coordinate from Gly of the second α-helix). One can recognize the conserved n/Pro/ϕ at Gly coordinate 1, following the well conserved E/D-G (not found in only four subfamilies); n/ϕ at position −4; n/OH at −5; n/ϕ at −8; n/Cys/OH at −9; n/OH at −11, −12; and frequent aromatic at −15. Consequently, the sequence between −15 and 3 (Gly coordinate) probably represents the ancestral common hydrophilic sequence. The positive charges frequently occur at positions 2, −3, −6 and −7 of this common sequence, but are interchangeable even with negative charges (Fig. 2). The shortest segment, containing only one AAR between Gly coordinate −5 and Pro coordinate 11, exists in a carrier of the Ydl119c-a subfamily. But long segments can be found such as one having 109 AAR between 13 and −19 (from the Ydr470c subfamily) or 98 AAR between 11 and −23 (from the Rim2 subfamily). With such an anatomy, specificities are unlikely. Some unique residues existing in one or at most in two subfamilies are listed in Table 2.

4.2. The second matrix segment

The second matrix segment extends from[5] Pro coordinate 2 up to Gly coordinate 2 (corresponding to −7 coordinate from Gly of the fourth α-helix). The common hydrophilic (likely ancestral) sequence again contains n/ϕ (OH in a few cases) at positions 1, −4, −5, −11 and −12; n/Cys at −9, and Gly/OH/N-bearing AAR at −13. Frequent aromatic AAR exists at −15. Hence the ancestral sequence extends between 2 and −15 of Gly coordinates. Its specific feature is the frequent negative charge at −2. Positive charges are found at 2, −3, −6, and −7. Variable parts exist between 12 and −9. The example of the shortest segment is given by a carrier of the Ydl119c-b subfamily having no residues between 14 and −9 (note that another carrier of the close Ydl119c-a subfamily

←

Legend of Table 2.

The table presents results of analysis when all main subfamilies were compared to each other separately for α-helices of all three matrix and two central cytosolic segments. Residues conserved in all proteins of the given subfamily are listed in bold as unique residues at given positions[4,5], when existing in one or two subfamilies. The first number in columns stands for typology of α-helices (see legend to Fig. 1). Other features which are shared by three or more subfamilies are listed in plain text. R stands for a positive charge (K, R); E for a negative charge (E, D); N for Asn/Gln; OH for Ser, Thr; ϕ for aromatic (Phe, Tyr, Trp); G, P, and H have their usual meaning (Gly, Pro, His). The numbers indicate the relative coordinates introduced here[4,5] (see also Figs. 1 and 2). Description as 'common' for matrix segments refers to the existence of both MACP signature and 'common hydrophilic sequence', but it indicates that no conserved residues exist otherwise. Note that Graves' disease antigen protein was tentatively analyzed together with main subfamilies.

has 45 AAR between 11 and −17); the long carrier such as from the Flx1a subfamily contains 62 AAR between 11 and −18. For unique features see Table 2.

4.3. The third matrix segment

The third matrix segment follows the same principles as the other two. It extends from[5] Pro coordinate 2 up to Gly coordinate 3 (corresponding to −7 coordinate from Gly of the sixth α-helix). The common (ancestral) hydrophilic segment, extending between −15 (with common aromatic AAR) and 3, quite frequently contains n/Pro/ϕ at 1; n/ϕ (OH groups in a few cases) at −4, −8, −11, −12; n/OH at −5; n/Cys/OH at −9. Positive charges are frequent (but interchangeable) at 2, −3, −6, and −7. Short segments just with no residues between 12 and −11 exist in some phosphate carriers; the longest carrier (from the Ca-dep.per. subfamily) contains 44 AAR between 13 and −17. Table 2 again lists some unique residues.

Acknowledgements: This work was supported by Grants A5011106 from the Internal Grant Agency of the Academy of Sciences of the Czech Republic; 301/02/1215 from the Grant Agency of the Czech Republic; ME389 from the Czech Ministry of Education (Kontakt Program); and by the research project AVOZ5011922.

References

[1] Saraste, M. and Walker, J.E. (1982) FEBS Lett. 144, 250–254.
[2] Aquila, H., Misra, D., Eulitz, M. and Klingenberg, M. (1982) Hoppe-Seyler's Physiol. Chem. 363, 345–349.
[3] Aquila, H., Link, T.A. and Klingenberg, M. (1985) EMBO J. 4, 2369–2376.
[4] Aquila, H., Link, T.A. and Klingenberg, M. (1987) FEBS Lett. 212, 1–9.
[5] Runswick, M.J., Powell, S.J., Nyren, P. and Walker, J.E. (1987) EMBO J. 6, 250–254.
[6] Runswick, M.J., Walker, J.E., Bisaccia, E., Iacobazzi, V. and Palmieri, F. (1990) Biochemistry 29, 11033–11040.
[7] Runswick, M.J., Philippides, A., Lauria, G. and Walker, J.E. (1994) DNA Sequence 4, 281–291.
[8] Nelson, D.R., Lawson, J.E., Klingenberg, M. and Douglas, M.G. (1993) J. Mol. Biol. 230, 1159–1170.
[9] Nelson, D.R., Felix, C.M. and Swanson, J.M. (1998) J. Mol. Biol. 277, 285–308.
[10] El Moualij, B., Duyckaerts, C., Lamotte-Brasseur, J. and Sluse, F.E. (1997) Yeast 13, 573–581.
[11] Belenki, R., Haefele, A., Eisen, M.B. and Wohlrab, H. (2000) Biochim. Biophys. Acta 1467, 207–218.
[12] Klingenberg, M. (1990) Trends Biochem. Sci. 15, 108–112.
[13] Ježek, P. and Urbánková, E. (2000) IUBMB Life 49, 63–70.
[14] Walker, J.E. and Runswick, M.J. (1993) J. Bioenerg. Biomembr. 25, 435–446.
[15] Hanák, P. and Ježek, P. (2001) FEBS Lett. 495, 137–141.
[16] La Noue, K.F. and Schoolwerth, A.C. (1979) Annu. Rev. Biochem. 48, 871–922.
[17] Krämer, R. and Palmieri, F. (1992) in: Molecular Mechanisms in Bioenergetics (Erstner, L., Ed.), pp. 359–384, Elsevier Science, Amsterdam.
[18] Palmieri, F. and van Ommen, B. (1999) in: Frontiers of Cellular Bioenergetics (Papa et al., Eds.), pp. 489–519, Kluwer Academic/Plenum, New York.
[19] Kakhniashvili, D., Mayor, J.A., Bremze, D.A., Xu, Y. and Kaplan, R.S. (1997) J. Biol. Chem. 272, 4516–4521.
[20] Palmieri, L., Lasorsa, F.M., Vozza, A., Agrimi, G., Fiermonte, G., Runswick, M.J., Walker, J.E. and Palmieri, F. (2000) Biochim. Biophys. Acta 1549, 363–369.
[21] Palmieri, L., Runswick, M.J., Fiermonte, G., Walker, J.E. and Palmieri, F. (2000) J. Bioenerg. Biomembr. 32, 67–77.
[22] Fiermonte, G., Palmieri, L., Dolce, V., Lasorsa, F.M., Palmieri, F., Runswick, M.J. and Walker, J.E. (1998) J. Biol. Chem. 273, 24754–24759.
[23] Adams, M.D. and Venter, J.C. et al. (2000) Science 287, 2185–2195.