The EMBO Journal vol.5 no.6 pp.1335-1342, 1986

Mitochondrial targeting sequences may form amphiphilic helices

Gunnar von Heijne

Research Group for Theoretical Biophysics, Department of Theoretical Physics, Royal Institute of Technology, S-100 44 Stockholm, Sweden

Communicated by G.Schatz

Twenty three mitochondrial targeting sequences have been analysed with regard to their potential for forming amphiphilic helices. It is shown that most if not all of these sequences can be expected to form helices with high hydrophobic moments in a suitable environment. In the few cases studied so far, the segments of maximal hydrophobic moment coincide closely with 'critical' regions defined by deletions and point mutations.

Key words: mitochondrial import/targeting/topogenic sequence/amphiphilic helix

Introduction

Targeting of proteins to different cellular locations is often mediated by N-terminal 'topogenic sequences'. In particular, this is true for proteins destined for the secretory pathway and for nuclearly encoded mitochondrial and chloroplast proteins. In both groups, translocation across at least one membrane is necessary before the final destination is reached.

It would seem that such highly specific sorting events should require highly conserved targeting sequences; this turns out not to be the case, however. Secretory topogenic sequences (signal peptides) are extremely variable both in length and amino acid sequence, but they do share a common basic design: a positively charged N-terminal region, a central hydrophobic region and a more polar C-terminal region that defines the cleavage site (von Heijne, 1985). The mitochondrial targeting sequences likewise are extremely variable, and the only common theme recognized so far is a preponderance of basic and hydroxyl-carrying residues, and a corresponding paucity of acidic residues (Reid, 1985). One would suspect, however, that some common structural elements hide behind this baffling variation in primary sequence also in this case.

A first clue to what one such common element may be is provided in the accompanying paper by Roise *et al.* (1986), where it is shown that a synthetic peptide representing the yeast cytochrome *c* oxidase subunit IV (Cox IV) pre-sequence (or even the first 15 residues of this sequence) is surface active and has lytic properties qualitatively similar to those of lytic peptides such as melittin and δ-hemolysin. Lytic peptides readily form so-called amphiphilic helices at membrane-water interfaces, and, as shown by Roise *et al.* (1986), the Cox IV pre-sequence presumably also binds to membranes as an amphiphilic helix with one highly charged and one hydrophobic face. Also, Ito *et al.* (1985) have recently demonstrated that repeating synthetic model peptides with Arg or Lys residues spaced 3-5 residues apart can block import of some mitochondrial protein *in vitro*.

Given these results, it is obviously important to ask if the amphiphilic-helix concept can be extended to other mitochondrial targeting sequences. From the theoretical point of view this question must be approached with some care, since it is deceptively easy to find what may look like amphiphilic helices almost everywhere, even in random sequences (Flinta *et al.*, 1983).

In this paper, I present statistical results indicating that most if not all of the published mitochondrial targeting sequences can be folded into helices with a high hydrophobic moment (Eisenberg *et al.*, 1984a), but that the non-polar faces of these helices are not always as hydrophobic as those in lytic or surface-seeking peptides. The segments of maximal hydrophobic moment identified here also coincide quite accurately with the 'critical' segments identified in deletion studies of the Cox IV, yeast 70-kd, and human ornithine transcarbamylase (OTC) pre-sequences (Hurt *et al.*, 1985a, 1985b; Hase *et al.*, 1984; Horwich *et al.*, 1986).

Results

Mitochondrial targeting sequences are enriched for Arg, Leu and Ser (but not for Lys), and have few Asp, Glu, Val and Ile

The overall amino acid composition of the first 23 mitochondrial targeting sequences listed in Table I, including only residues 1-22 in cytochrome *c* peroxidase and residues 1-39 in cytochrome c_1, is given in Table II. Comparing this distribution of amino acids with the distribution in the 40 N-terminal residues in a sample of 132 eukaryotic cytosolic proteins (see Materials and methods), it is seen that Arg, Leu and Ser are significantly enriched in the mitochondrial sample, whereas Asp, Glu, Val and Ile are sgnificantly lower in incidence ($P < 0.01$).

Mitochondrial targeting sequences can form helices with a higher hydrophobic moment than controls, but the non-polar face does not seem to be strongly selected for high hydrophobicity

The hydrophobic moment (Eisenberg *et al.*, 1984a; see Materials and methods) is a convenient measure of the helical amphiphilicity of a sequence. The maximal mean hydrophobic moment per residue over an 11-residue moving window $<\mu_H>$, and the mean hydrophobicity per residue, $<H>$, has been used to characterize lytic or surface-seeking helices (Eisenberg *et al.*, 1984a). With the Eisenberg criteria (i.e. $<\mu_H> \geq 0.603 - 0.382<H>$), 10 of the first 23 sequences in Table I are predicted to be surface-seeking, an additional seven [marked (+)] are very close to the surface region, and only six map far from the typical surface-active peptides.

This analysis suffers from two shortcomings, however: it uses a moving window of fixed length (11 residues), and thus cannot find the segment of maximal total μ_H, irrespective of window length and not normalized on a per-residue basis; and, second, $<H>$ is an aggregate measure of the hydrophobicity of both the non-polar and the polar faces. For a membrane-interacting peptide, one would rather like to know the hydrophobicity of the non-polar face alone.

Repeating the analysis, but now searching instead for the segment of maximum total μ_H, one finds that the optimal length of the moving window is between 17 and 26 residues for a sample of surface-seeking peptides, and between 12 and 26 residues for the mitochondrial sample (data not shown). To simplify the

Table I. Mitochondrial targeting sequences

Protein	Segment	μ_H	H_{max}	Surf(E)	Surf(vH)	Reference
ATP synthase IX (*Neurospora*)	20–37	12.5	5.2	+	+	Viebrock *et al.* (1982)
Cytochrome oxidase IV	3–20	11.8	5.1	+	+	Maarse *et al.* (1984)
Rieske 2Fe-3S protein (*Neurospora*)	9–26	11.1	4.7	+	+	Harnisch *et al.* (1985)
Cyt-*c* reductase 14 kd	1–18	10.3	6.8	+	+	de Haan *et al.* (1984)
EF-Tu	6–23	10.2	4.5	+	+	Nagata *et al.* (1983)
OTC (human)	14–31	10.0	4.3	(+)	(+)	Horwich *et al.* (1984)
Cyt-*c* peroxidase	1–18	9.6	5.6	+	+	Kaput *et al.* (1982)
OTC (rat)	14–31	9.6	4.3	–	(+)	McIntyre *et al.* (1984)
ATP synthase β (*Nicotiana*)	7–24	9.3	5.9	(+)	+	Boutry and Chua (1985)
Cytochrome b_2	4–21	8.7	5.5	–	+	Guiard (1985)
Citrate synthase	11–28	8.7	2.8	–	–	Suissa *et al.* (1984)
Cytochrome oxidase VI	12–29	8.7	4.7	–	+	Wright *et al.* (1984)
Threonine deaminase	2–19	8.3	5.2	(+)	+	Kielland-Brandt *et al.* (1984)
Adrenodoxin (bovine)	4–21	8.2	5.2	–	+	Okamura *et al.* (1985)
Carbamyl phosphate synthase I (rat)	22–39.	7.9	2.7	+	–	Nyunoya *et al.* (1985)
Carbamyl phosphate synthase I (alt.)	10–27	6.9	5.3	(+)	–	
Cytochrome oxidase IV (bovine)	4–21	7.7	5.1	+	+	Lomax *et al.* (1984)
Cytochrome c_1	8–25	7.6	3.7	–	–	Sadler *et al.* (1984)
MSS 51	28–45	7.4	4.8	+	+	Faye and Simon (1983)
MSS 51 (alt.)	3–20	6.6	4.8	–	–	
Cytochrome P-450 (SCC) (bovine)	12–29	7.3	6.8	(+)	+	Morohashi *et al.* (1984)
Alcohol dehydrogenase	1–18	7.3	4.8	(+)	+	Young and Pilgrim (1985)
Aspartate aminotransferase (porcine)	7–24	7.0	6.6	+	(+)	Joh *et al.* (1985)
Mn-superoxide dismutase	7–24	6.0	3.7	–	–	Marres *et al.* (1985)
Cytochrome oxidase V	1–18	5.1	3.7	–	–	Seraphin *et al.* (1985)
Cyt-*c* reductase 17 kd	1–18	5.7	6.0	(+)	–	van Loon *et al.* (1984)
Cytochrome *c*	32–49	6.3	5.4	–	–	Smith *et al.* (1979)
70-kd protein	1–18	3.3	4.8	–	–	Hase *et al.* (1983)
Porin	2–19	9.6	6.0	+	+	Mihara and Sato (1985)

The segment of maximal hydrophobic moment (18 residues moving window) is given, together with its hydrophobic moment (μ_H), the hydrophobicity H_{max} of its non-polar face (seven neighboring residues), and two different predictions of its surface-seeking properties [according to Eisenberg (1984a) and as described in this work]. (alt.) indicates an alternative segment with high hydrophobic moment. Unless indicated, all sequences are from yeast.

Table II. Total amino acid composition of the first 23 targeting sequences in Table I (including only residues 1–22 of cyt *c* peroxidase and 1–39 of cyt c_1), and the expected number of each kind of residue based on the composition of a sample of the 40 N-terminal residues from 132 eukaryotic cytosolic proteins

Residue	Number of residues
Ala	99/84.4
Cys	8/12.3
Asp	8/49.0[a]
Glu	4/55.3[a]
Phe	32/27.5
Gly	54/57.7
His	18/17.3
Ile	26/45.8[a]
Lys	44/55.0
Leu	118/66.6[a]
Met	30/(23.9)
Asn	29/33.8
Pro	31/43.5
Gln	34/37.7
Arg	95/46.4[a]
Ser	95/54.7[a]
Thr	64/49.0
Val	47/64.4[a]
Trp	4/ 6.6
Tyr	12/21.2

Since only some of the cytosolic proteins have an N-terminal Met, the figure for Met is uncertain.
[a]Significant differences ($P < 0.01$ by χ^2 analysis).

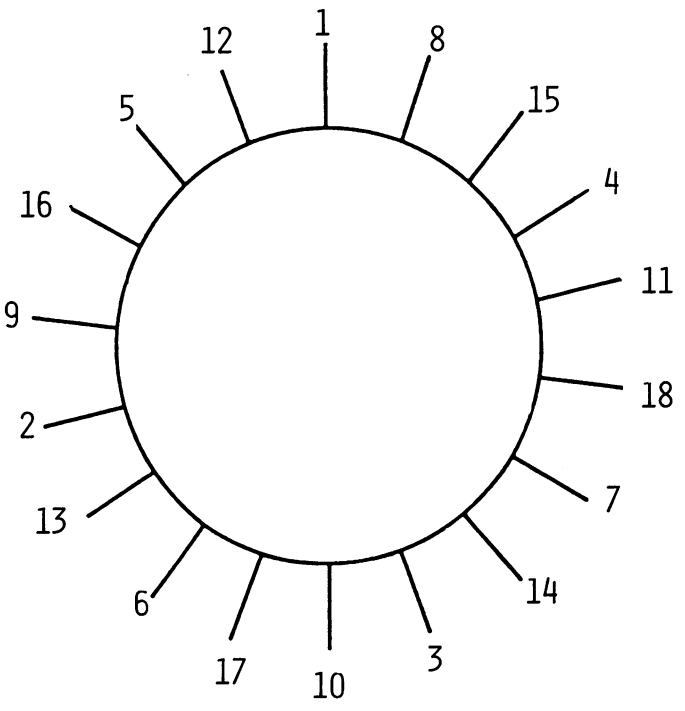

Fig. 1. Helical wheel projection, i.e. an end-on view of an 18-residue α-helix with the positions of the residues marked.

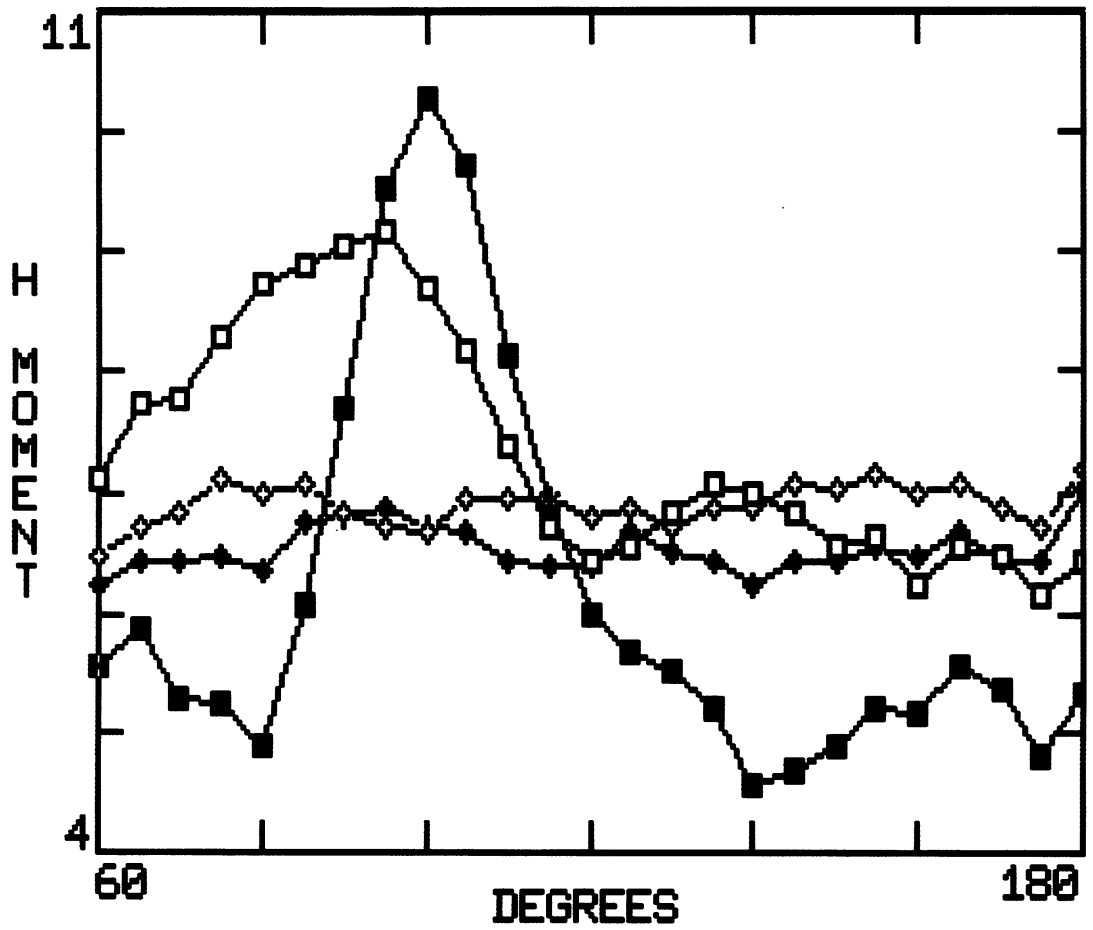

Fig. 2. The mean generalized hydrophobic moment of the $\mu_H(\delta)$-max segments as a function of δ for the mitochondrial sample (sequences 1–23 in Table I, open squares), the randomly 'scrambled' sample (open diamonds), the cytosolic sample (solid diamonds) and the sample of surface-seeking peptides (solid squares).

analysis, an 18-residue moving window (corresponding exactly to five full helical turns) will be used throughout to calculate μ_H (this is also close to the lengths of the shortest targeting sequences, e.g. the cytochrome oxidase subunit V sequence which is 20 residues long). Helical plots of known surface-seeking peptides such as calcitonin analogs (Moe and Kaiser, 1985) indicate that a hydrophobic face encompassing seven neighboring residues on an 18-residue helix (e.g. residues 2, 3, 6, 9, 10, 13 and 17, Figure 1) is sufficient to form a membrane-binding domain, and in the following analysis we will calculate the maximal hydrophobicity of the hydrophobic face, H_{max}, by finding those seven neighboring residues on the previously identified 18-residue μ_H-max helix that have the highest summed hydrophobicity. Similarly, we will calculate the hydrophobicity of the polar face, H_{min}, by finding those seven (or five, see below) residues that have the lowest summed hydrophobicity.

With this method, surface-seeking peptides are found to have $\mu_H \geq 7.3$ and $H_{max} \geq 4.5$. In the mitochondrial sample (entries 1–23 in Table I), 15 sequences map in this region (including Cox IV from yeast), and three more are very close [it should be noted that some of the calculated values may not be entirely correct since proline and glycine residues in particular can introduce hard-to-predict distortions in the helices (Terwilliger and Eisenberg, 1982; Sundaralingam *et al.*, 1985)].

Using another hydrophobicity scale conceivably could give very different results; however, with a more coarse-grained scale where Arg, Lys, Asp and Glu have the value −1, Leu, Ile, Val, Phe and Met have +1 and all other residues have a hydrophobicity value equal to zero, the same μ_H-max segments as in Table I were predicted in all but four of the 23 cases (not shown).

Thus, by these criteria, most of the mitochondrial targeting sequences seem to be good candidates for surface-seeking peptides. This contention is further strengthened by the results of a comparison between these sequences and three different control samples: one sample consisting of the 40 N-terminal residues from each of 132 cytosolic proteins (control I), one sample of 230 40-residues long randomly generated sequences with the same mean amino acid composition as the mitochondrial sequences (control II), and one sample with 10 randomly 'scrambled' versions of each of the 23 mitochondrial sequences (control III; see Materials and methods). Compared with the 15 out of 23 targeting sequences (65%) that map in the surface region, only 34% of the cytosolic sequences and 25% of the random sequences have μ_H-max and H_{max} values in this region.

Above, the hydrophobic moment has been calculated for an ideal α-helix with $\delta = 100°$ between successive residues. By varying δ, the secondary structure of the segment in question can be more thoroughly investigated (Eisenberg *et al.*, 1984b). Plots of the mean value of μ_H-max as a function of δ for the mitochondrial, surface-seeking, cytosolic and randomly scrambled samples are shown in Figure 2. As expected, the last two samples have no strong components (the curve for control sam-

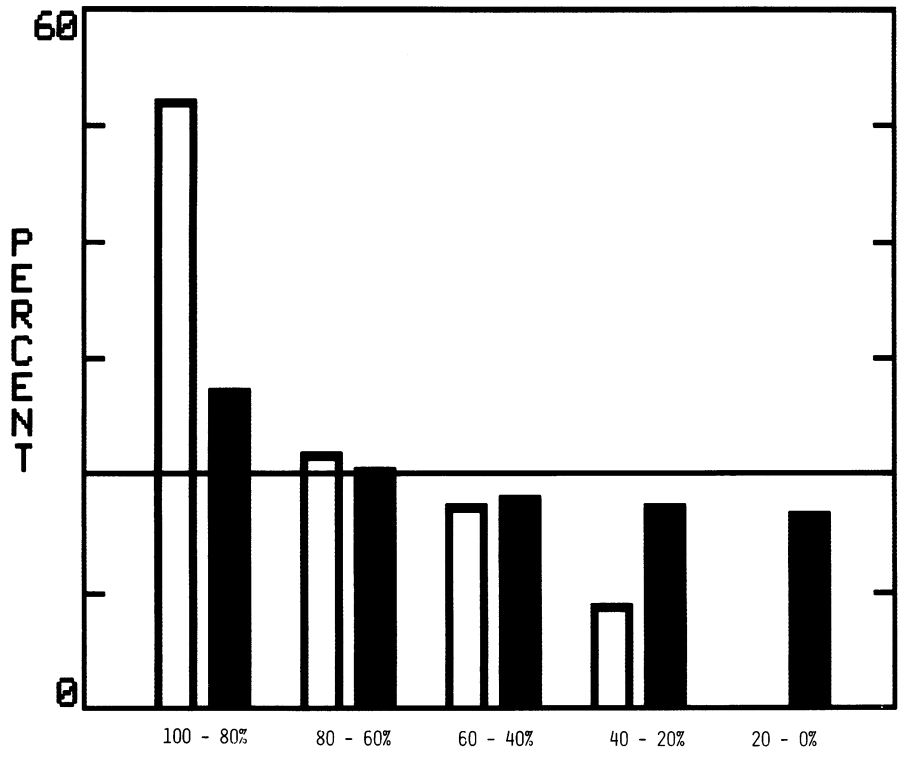

Fig. 3. An illustration of how the hydrophobic moments of individual mitochondrial (open bars) and cytosolic (solid bars) sequences compare with the values calculated for 100 scrambled copies of each wild-type sequence. The percentage of the wild-type sequences that have a μ_H value falling within a given 20% range of the μ_H distribution for its own scrambled copies are shown, i.e. 52% (12 out of 23) of the mitochondrial targeting sequences have μ_H values that are larger than 80% of the μ_H values generated by scrambling the wild-type sequence 100 times; none has a value that is less than the values for 80% or more of its scrambled versions. The horizontal line is the value expected (20%) for each of the five percentage ranges in a sample of random sequences.

ple II is similar, not shown). The surface-seeking helices have a maximum at $\delta = 100°$, again as expected, whereas the mitochondrial μ_H-max segments have a broader and slightly displaced maximum at $\delta = 95°$. This indicates that the most polar and/or most non-polar residues on a 'typical' targeting sequence tend to spiral around the helix in a clockwise direction. Since helix–helix and helix–sheet packings generally involve spiralling 'ridges' on the helical surface (Chothia *et al.*, 1977; Janin and Chothia, 1980), this shift in the δ-max value may reflect some such packing constraints.

Even for a normal helix with $\delta = 100°$, the mean μ_H-max value for the mitochondrial sequences is significantly higher ($P < 0.01$) than for the controls (I–III). Also, the mean H_{min} values calculated for seven or five neighbouring residues (polar face) are significantly lower than for the controls ($P < 0.01$), but the differences between the mean H_{max} values (non-polar face) are not significant (data not shown). This is not simply an effect of the biased amino acid composition of the mitochondrial sequences, since the mean μ_H and H_{min} values for the randomized control samples (II and III) also differ from those of the wild-type group. Thus, the composition and residue distribution may be more constrained on the polar than on the non-polar face of the mitochondrial targeting sequences.

In the calculation of mean values only the group properties of the targeting sequences are considered; however, one can also compare the maximum μ_H value for each individual sequence with the values for scrambled versions of the same sequence. The results of such an analysis is shown in Figure 3, where the wild-type value for each sequence in the mitochondrial and cytosolic samples has been compared with the μ_H-max values for 100 scrambled sequences. It is clear that the individual mitochondrial wild-type sequences tend to have higher hydrophobic moments than their respective randomized controls.

Finally, it is noteworthy that the N-terminal regions of both the outer membrane 70-kd protein and cytochrome *c* have low μ_H and H_{max} values; these two proteins are probably not imported via the same pathway as the proteins of the inner mitochondrial compartments (Reid, 1985). Yeast porin, another protein of the outer mitochondrial membrane does have an N-terminal segment of high μ_H, but this may well be related to its pore-forming abilities rather than to the import process *per se* (there are two acidic Asp residues in this porin segment, which is atypical).

Segments of maximal hydrophobic moment agree with data from deletion and point mutation studies

Three mitochondrial targeting sequences have been subjected to deletion and mutation studies so far: those from the yeast Cox IV and 70-kd proteins, and the human OTC pre-sequence (Hurt *et al.*, 1985a, 1985b; Hase *et al.*, 1984; Horwich *et al.*, 1986). The relevant constructions are shown in Table III. In the context of the present study, the following points can be made.

Cox IV. Residues 1–16 and 1–12 are sufficient to direct an attached cytosolic protein, mouse dihydrofolate reductase (DHFR), into the mitochondrial matrix. Residues 1–9 have no import activity. As shown in Table III, the two longer sequences both have high values of μ_H whereas the third one does not. Since the predicted μ_H-max segment in the wild-type protein starts at residue 3 (Table I), it is also interesting to note that a deletion of three residues from the N terminus of the Cox IV pre-sequence has an adverse effect on import (Hurt, personal communication).

Table III. Deletions and point mutations of the targeting sequence from yeast cytochrome *c* subunit IV, yeast 70-kd outer membrane protein, and human OTC

		Segment	μ_H	H_{max}
Cox IV				
W.t.	M L S L R Q S I R F F K P A T R T L C S S R Y L L	3–20	11.8	5.1
1–16	– – – – – – – – – – – – – – – – S G I M V R P L N C I V A V S Q N M G I G K	4–21	11.0	5.4
1–12	– – – – – – – – – – – – R . . . S G I M V R – – – – – – – – – – – – – – – –	4–21	10.7	5.2
1–9	– – – – – – – – – . . . . . S G I M V R – – – – – – – – – – – – – – – –	4–21	3.8	4.2
70-kd				
W.t.	M K S F I T R N K T A I L A T V A A T G T A I G A Y	1–18	3.3	4.8
Δ12–106	– – – – – – – – – – – – S N Q F F R N K K Y D D A I K	1–18	8.2	4.6
Δ15–157	– – – – – – – – – – – – – – – A L E L K P D Y S K V L L	1–18	3.2	5.5
1–12 + DHFR	– – – – – – – – – – – – G S G I M V R P L N C I V A	1–18	4.7	5.5
OTC				
W.t.	M L F N L R I L L N N A A F R N G H N F M V R N F R C G Q P L Q	14–31	10.0	4.3
Mutations	R → G	14–31	10.0	4.3
	N → G	14–31	7.8	4.3
	V → G	14–31	8.1	4.3
	R → G	14–31	7.2	4.3

The segments of maximal hydrophobic moment (18 residue moving window) and their μ_H and H_{max} values are also given (see footnotes to Table I). Dashes indicate identical residues, and dots show deletions.

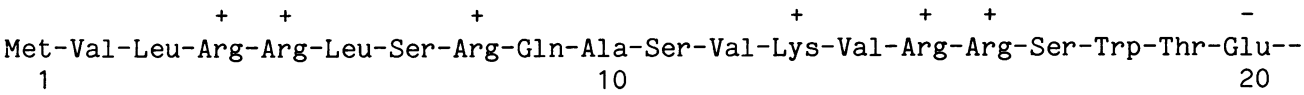

Fig. 4. N-terminal reigon of the agnoprotein from SV40. Segment 2–19 has μ_H = 10.6 and H_{max} = 4.3, and also maps close to the 'surface region' defined by Eisenberg *et al.* (1984a).

70-kd protein. This protein is located in the outer mitochondrial membrane, and is probably transported via a pathway different from that leading to the inner compartments (Reid, 1985). Deletion and fusion studies have shown that the 41 N-terminal residues are sufficient to direct proteins to the outer membrane. Surprisingly, two deletions involving the major part of the targeting sequence, Δ12–106 and Δ15–157, and a fusion between the first 12 amino acids and DHFR induce misrouting of a fraction (29%, 10% and 40%, respectively) of the protein into the matrix. The 20 N-terminal residues in the normal 70-kd targeting sequence have a very low μ_H value, but the Δ12–106 deletion fortuitously brings together two chain segments rich in basic residues, such that the resulting sequence has a high hydrophobic moment. The 70-kd/DHFR fusion also is rich in basic residues but has a lower μ_H (the basic residues are farther apart on the polar face) and a higher H_{max}. The Δ15–157 deletion, finally, has close to wild-type values but lacks the penultimate putative transmembrane segment 11–25 normally present.

OTC. Segment 14–31 has the highest μ_H value in the human OTC pre-sequence. By systematically substituting glycine for the wild-type arginines at positions 6, 15, 23 and 26, Horwich *et al.* (1986) have recently shown that Arg_{23}, and, to a lesser extent, Arg_{15} are critical for the import function. The other arginines can be substituted with no deleterious consequences. Thus, substitution of arginines outside (Arg_6) or at one end of (Arg_{26}) the identified segment does not affect import, whereas the central Arg residues seem to be important. The magnitude of μ_H and H_{max} for the Arg_{15} and Arg_{23} mutants are still within the surface region, however; thus, Gly, being a strong helix-breaker (Chou and Fasman, 1978), may perhaps act to destablize the amphiphilic helix (Horwich *et al.*, 1986).

Discussion

The statistical study presented here indicates that a common feature of mitochondrial targeting sequences may indeed be the possession of segments with a high helical moment. The few that do not have any such segments (or only segments with low H_{max} values) either have lysines (hydrophobicity = −1.5) rather than arginines (hydrophobicity = −2.53) on their polar face resulting in a smaller μ_H (Mn-superoxide dismutase, carbamyl phosphate synthase), or have a number of serines and threonines on their non-polar face reducing μ_H and H_{max} (cytochrome c_1, cytochrome oxidase V) – Ser and Thr can make hydrogen bonds to the helical backbone, and thus may not be as unfavorable as their hydrophobicity values indicate. Citrate synthase, finally, has a low H_{max} because of a glutamine on its non-polar face. In general, the non-polar faces of the mitochondrial $\mu_H$max helices are not as strongly hydrophobic as those found in surface-seeking peptides. It is also intreresting to note that the Arg/Lys bias on the polar face (Arg/Lys = 2.2, Table II) has its counterpart in the C-terminal (but not in the N-terminal) region of secretory signal peptides from eukaryotic proteins, where Arg/Lys = 4.3 (von Heijne, 1986). Possibly, Arg residues may translocate more easily than Lys residues across membranes (see below).

Compared with control samples of N-terminal regions from cytosolic proteins and randomly generated sequences, the mitochondrial targeting sequences analysed here show a consistent pattern of being able to form helices with large hydrophobic moments, but there is a considerable overlap in the distribution of μ_H values between the samples, and individual cytosolic sequences cannot be distinguished from the mitochondrial ones on the basis of μ_H alone (78% of the cytosolic proteins have

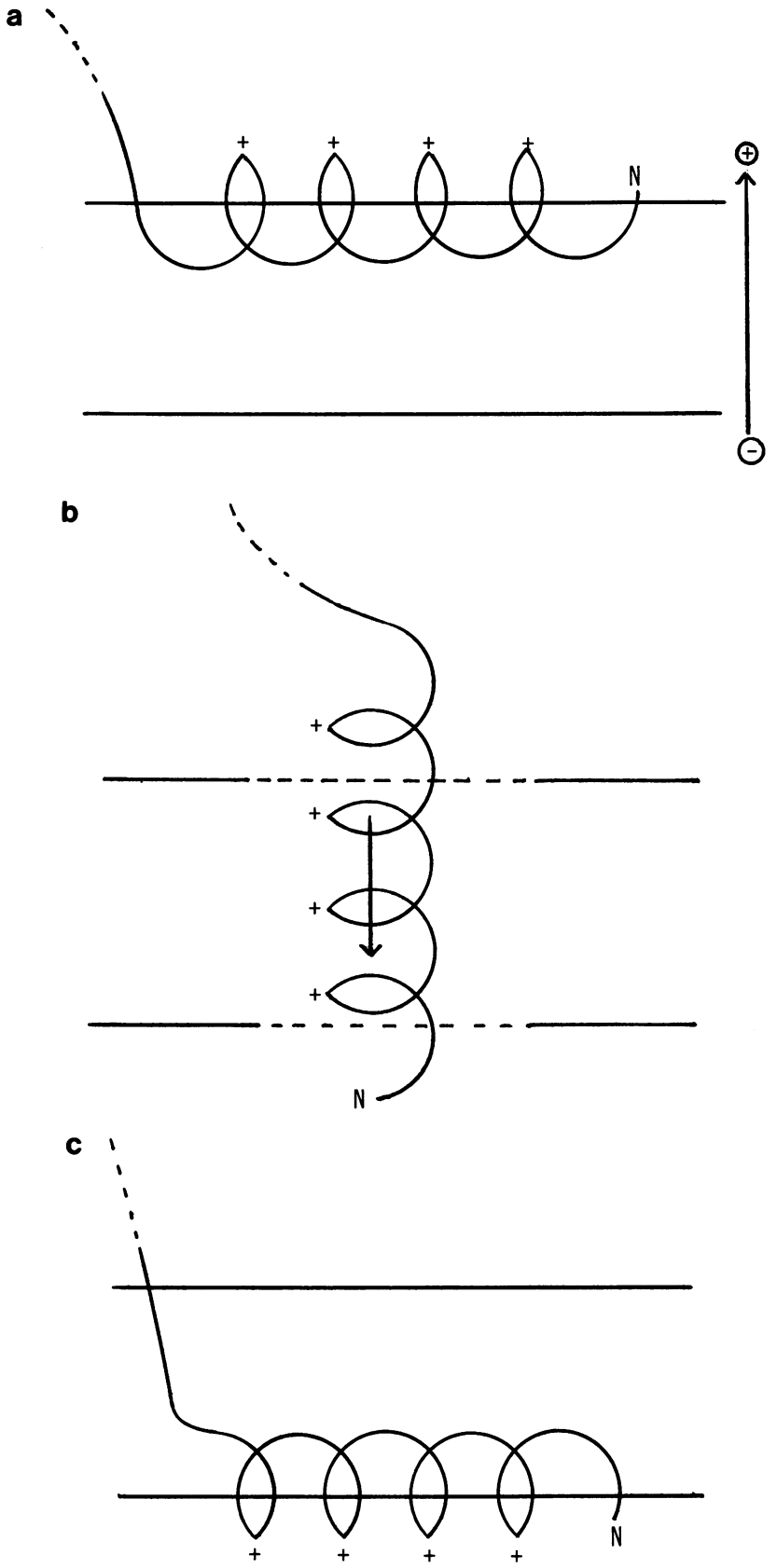

Fig. 5. A model for the interaction of the targeting sequence with the inner mitochondrial membrane, see text.

segments with $\mu_H \geq 5$, 39% have $\mu_H \geq 7$). If, however, one imposes the additional constraint that the N-terminal sequence up to and including the segment of maximum hydrophobic moment must be free of acidic residues, a clear-cut separation is achieved; only one sequence out of the original 132 in the cytosolic sample remains. This is the so-called agnoprotein from

SV40 (Figure 4), a small basic protein that may have a role in DNA−protein interactions (Jay *et al.*, 1981). It should be interesting to determine the sub-cellular location of this protein, and to check whether its N-terminal 20 residues may have some mitochondrial targeting activity.

In conclusion, and given the observation by Roise *et al.* (1986) that an electrical membrane potential $\Delta\Psi$ (+ outside, − inside) potentiates the surface-active properties of the Cox IV targeting sequence [just as $\Delta\Psi$ is needed for import across the inner membrane *in vivo* (Hay *et al.*, 1984) and *in vitro* (Pfanner and Neupert, 1985)], we will try to estimate the energetics of various imaginable dispositions of an amphiphilic targeting sequence relative to the inner membrane [or an outer−inner membrane junction, c.f. Schleyer and Neupert (1985)]. Figure 5a shows the initial interaction: the non-polar face is buried in the outer leaflet of the membrane, and the basic residues on the polar face interact with negatively charged phospholipid headgroups. The electrical component of the membrane potential will make a translocation across the membrane energetically favourable (each positive charge brought across corresponds to an energetic gain of $e^*\Delta\Psi$, or 100 meV = 2.3 kcal/mol for $\Delta\Psi$ = 100 mV), and will drive the targeting sequence through the membrane provided that the activation energy barrier is not too high. It is not inconceivable that the joint action of (i) a local perturbation or destabilization of the bilayer induced by the binding of the amphiphilic helix as in Figure 5a, and (ii) the interaction of the membrane potential with the 'helical dipole' (Hol, 1985) resulting from the alignment of the individual peptide dipoles along the long axis of the amphiphilic helix might facilitate this process. A 'transition state' such as depicted in Figure 5b with the helical dipole aligned with the electric field will be stabilized by an electrostatic term $\mu^*\Delta\Psi/d$ (= 1.0 kcal/mol for an 18-residue helix in a 30 Å thick membrane with $\Delta\Psi$ = 100 mV) where μ is the helical dipole moment and d is the membrane thickness, and by the contribution from partly translocated positively charged residues (cf. above). Whether this favourable electrostatic energy is sufficient to compensate for the free energy loss associated with bringing the charged residues into the membrane can only be speculated upon at present, but it should be noted that an analogous potential-dependent trans-membrane movement of charged residues seems to occur with the surface-active peptide melittin (Kempf *et al.*, 1982) when a threshold value $\Delta\Psi \geq 40$ mV − similar to that required for mitochondrial import *in vitro* (Pfanner and Neupert, 1985) − is reached.

Materials and methods

A total of 27 mitochondrial targeting sequences have been analysed in this study, Table I (when there is no known cleavage site, 50 N-terminal residues have been kept). However, when the group behaviour of these sequences has been sought, only the first 23 entries have been included. The cytochrome *c* reductase 17-kd subunit has been excluded because of its highly unusual amino acid composition (10 acidic and only one basic residue in the first 32 residues), porin and the 70-kd protein because they are outer membrane proteins that probably integrate into the mitochondrion via a route not utilized by the inter-membrane, inner membrane, and matrix proteins (Reid, 1985), and cytochrome *c* because it has its own mitochondrial receptor and is transported into the inter-membrane space in a co-factor-dependent manner (Hennig and Neupert, 1981).

A sample of 13 surface-seeking peptides [mellitin from *mollifera* and *florea*, synthetic mellitin-like peptide, cecropins A and B, δ-haemolysins I and II, calcitonin, synthetic calcitonin-like peptide (MCT-I), corticotropin-releasing factor, synthetic peptides ^{18}Aa and ^{18}As, and Cox IV (1−25) pre-sequence; see Eisenberg *et al.* (1984a, 1984b), Kanellis *et al.* (1980), Kaiser and Kezdy (1984) and Lau *et al.* (1983) for references] has also been analysed.

As controls, a sample of 132 N-terminal sequences from cytosolic eukaryotic proteins (first 40 residues) (Flinta *et al.*, 1986), one randomly generated sample of 230 40-residue sequences with mean amino acid composition equal to the mean amino acid composition of the first 23 entries in Table I (including only residues 1−22 of cytochrome *c* peroxidase and residues 1−39 of cytochrome c_1), and one sample consisting of 10 randomly scrambled versions of each of these 23 sequences have also been analysed.

The sequences have been characterized in four ways according to (i) their maximal mean or total hydrophobic moments ($<\mu_H>$ and μ_H) using a moving window 11 or 18 residues long, (ii) the mean hydrophobicity $<H>$ of the segment of maximal $<\mu_H>$, (iii) the maximal hydrophobicity H_{max} of seven neighbouring residues on the face of the μ_H-max helix, and (iv) the minimal hydrophobicity H_{min} of five or seven neighbouring residues on the face of the μ_H-max helix.

The hydrophobic moment is calculated according to Eisenberg *et al.* (1984a) as:

$$\mu_H = \{[\Sigma H_n \sin(\delta n)]^2 + [\Sigma H_n \cos(\delta n)]^2\}^{1/2}$$

where H_n is the hydrophobicity of the nth residue and $\delta = 100°$ for an α-helix. The hydrophobicity scale used is the 'normalized consensus scale' of Eisenberg *et al.* (1984a).

Mean values have been compared using Student's t-test, and the amino acid distribution has been analyzed with the χ^2 test.

Acknowledgements

This work was supported by a grant from the Swedish Natural Sciences Research Council.

References

Boutry,M. and Chua,N.-H. (1985) *EMBO J.*, **4**, 2159−2165.
Chothia,C., Levitt,M. and Richardson,D. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 4130−4134.
Chou,P.Y. and Fasman,G.D. (1978) *Annu. Rev. Biochem.*, **47**, 251−276.
De Haan,M., Van Loon,A.P.G.M., Kreike,J., Vaessen,R.T.M.J. and Grivell, L.A. (1984) *Eur. J. Biochem.*, **138**, 169−177.
Eisenberg,D., Schwarz,E., Komaromy,M. and Wall,R. (1984a) *J. Mol. Biol.*, **179**, 125−142.
Eisenberg,D., Weiss,R.M. and Terwilliger,T.C. (1984b) *Proc. Natl. Acad. Sci. USA*, **81**, 140−144.
Faye,G. and Simon,M. (1983) *Cell*, **32**, 77−87.
Flinta,C., von Heijne,G. and Johansson,J. (1983) *J. Mol. Biol.*, **168**, 193−196.
Flinta,C., Persson,B., Jörnvall,H. and von Heijne,G. (1986) *Eur. J. Biochem.*, **154**, 193−196.
Guiard,B. (1985) *EBMO J.*, **4**, 3265−3272.
Harnisch,U., Weiss,H. and Sebald,W. (1985) *Eur. J. Biochem.*, **149**, 95−99.
Hase,T., Riezman,H., Suda,K. and Schatz,G. (1983) *EMBO J.*, **2**, 2169−2172.
Hase,T., Muller,U., Riezman,H. and Schatz,G. (1984) *EMBO J.*, **3**, 3157−3164.
Hay,R., Böhni,P. and Gassner,S. (1984) *Biochim. Biophys. Acta*, **779**, 65−87.
Hennig,B. and Neupert,W. (1981) *Eur. J. Biochem.*, **121**, 203−212
Hol,W.G. (1985) *Prog. Biophys. Mol. Biol.*, **45**, 149−195.
Horwich,A.L., Fenton,W.A., Williams,K.R., Kalousek,F., Kraus,J.P., Doolittle,R.F., Konigsberg,W. and Rosenberg,L.E. (1984) *Science*, **224**, 1068−1074.
Horwich,A.L., Kalousek,F., Fenton,W.A., Pollock,R.A. and Rosenberg,L.E. (1986) *Cell*, **44**, 451−459.
Hurt,E.C., Pesold-Hurt,B., Suda,K., Oppliger,W. and Schatz,G. (1985a) *EMBO J.*, **4**, 2061−2068.
Hurt,E.C., Muller,U. and Schatz,G. (1985b) *EMBO J.*, **4**, 3509−3518.
Ito,A., Ogishima,T., Ou,W., Omura,T., Aoyagi,H., Lee,S., Mihara,H. and Izumiya,N. (1985) *J. Biochem.*, **98**, 1571−1582.
Janin,J. and Chothia,C. (1980) *J. Mol. Biol.*, **143**, 95−128.
Jay,G., Nomura,S., Anderson,C.W. and Khoury,G. (1981) *Nature*, **291**, 346−349.
Joh,T., Nomiyama,H., Maeda,S., Shimada,K. and Morino,Y. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6065−6069.
Kaiser,E.T. and Kezdy,F.J. (1984) *Science*, **223**, 249−255.
Kanellis,P., Romans,A.Y., Johnson,B.J., Kercret,H., Chiovetti,R., Allen,T.M. and Segrest,J.P. (1980) *J. Biol. Chem.*, **255**, 11464−11472.
Kaput,J., Goltz,S. and Blobel,G. (1982) *J. Biol. Chem.*, **257**, 15054−15058.
Kempf,C., Klausner,R.D., Weinstein,J.N., Van Renswoude,J,. Pincus,M. and Blumenthal,R. (1982) *J. Biol. Chem.*, **257**, 2469−2476.
Kielland-Brandt,M.C., Holmberg,S., Peterson,J.G.L. and Nilsson-Tillgren,T. (1984) *Carlsberg Res. Commun.*, **49**, 567−575.
Lau,S.H., Rivier,J., Vale,W., Kaiser,E.T. and Kezdy,F.J. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7070−7074.
Lomax,M.I., Bachman,N.J., Nasoff,M.S., Caruthers,M.H. and Grossman,L.I. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6295−6299.
Maarse,A.C., Van Loon,A.P.G.M., Riezman,H., Gregor,I., Schatz,G. and Grivell,L.A. (1984) *EMBO J.*, **3**, 2831−2837.

Marres,C.A.M., Van Loon,A.P.G.M., Oudshoorn,P., Van Steeg,H., Grivell, L.A. and Slater,E.C. (1985) *Eur. J. Biochem.*, **147**, 153–161.
McIntyre,P., Graf,L., Mercer,J., Peterson,G., Hudson,P. and Hoogenraad,N. (1984) *FEBS Lett.*, **177**, 41–46.
Mihara,K. and Sato,R. (1985) *EMBO J.*, **4**, 769–774.
Moe,G.R. and Kaiser,E.T. (1985) *Biochemistry*, **24**, 1971–1976.
Morohashi,K., Fujii-Kuriyama,Y., Okada,Y., Sogawa,K., Hirose,T., Inayama,S. and Omura,T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4647–4651.
Nagata,S., Tsunetsugu-Yokota,Y., Naito,A. and Kaziro,Y. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6192–6196.
Nyunoya,H., Broglie,K.E., Widgren,E.E. and Lusty,C.J. (1985) *J. Biol. Chem.*, **260**, 9346–9356.
Okamura,T., John,M.E., Zuber,M.X., Simpson,E.R. and Waterman,M.R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 5705–5709.
Pfanner,N. and Neupert,W. (1985) *EMBO J.*, **4**, 2819–2825.
Reid,G.A. (1985) *Curr. Top. Membr. Transp.*, **24**, 295–336.
Roise,D. Horvath,S.J., Richards,J.H., Tomich,J.M. and Schatz,G. (1986) *EMBO J.*, **5**, 1327–1334.
Sadler,I., Suda,K., Schatz,G., Kaudewitz,F. and Haid,A. (1984) *EMBO J.*, **3**, 2137–2143.
Schleyer,M. and Neupert,W. (1985) *Cell*, **43**, 339–350.
Seraphin,B., Simon,M. and Faye,G. (1985) *Curr. Genet.*, **9**, 435–439.
Smith,M., Leung,D.W., Gillam,S., Astell,C.R., Montgomery,D.L. and Hall,B.D. (1979) *Cell*, **16**, 753–761.
Suissa,M., Suda,K. and Schatz,G. (1984) *EMBO J.*, **3**, 1773–1781.
Sundaralingam,M., Drendel,W. and Greaser,M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7944–7947.
Terwilliger,T.C. and Eisenberg,D. (1982) *J. Biol. Chem.*, **257**, 6016–6022.
Van Loon,A.P.G.M., De Groot,R.J., De Haan,M., Dekker,A. and Grivell,L.A. (1984) *EMBO J.*, **3**, 1039–1043.
Viebrock,A., Perz,A. and Sebald,W. (1982) *EMBO J.*, **1**, 565–571.
von Heijne,G. (1985) *J. Mol. Biol.*, **184**, 99–105.
von Heijne,G. (1986) *Nucleic Acids Res.*, **14**, in press.
Wright,R.M., Ko,C., Cumsky,M.G. and Poyton,R.O. (1984) *J. Biol. Chem.*, **259**, 15401–15407.
Young,E.T. and Pilgrim,D. (1985) *Mol. Cell. Biol.*, **5**, 3024–3034.

Received on 7 March 1986