Volume 278, number 1, 1–3 FEBS 09310 January 1991

ADONIS 001457939100007S

Hypothesis

Chloroplast transit peptides

The perfect random coil?

Gunnar von Heijne[1] and Ken Nishikawa[2]

[1] *Department of Molecular Biology, Karolinska Institute Center for Biotechnology, NOVUM, S-141 52 Huddinge, Sweden* and [2] *Protein Engineering Research Institute, 6-2-3, Furuedai, Suita, Osaka 565, Japan*

Received 12 October 1990; revised version received 8 November 1990

Structural analysis of chloroplast transit peptides (cTPs), including secondary structure prediction and analysis of 'cTP-like' peptides of known 3D structure, suggests that cTPs are essentially flexible peptides devoid of regular secondary or tertiary structure. It is proposed that cTPs may be designed to interact with a succession of different chaperones on the chloroplast protein-import pathway.

Chloroplast; Transit peptide; Chaperone; Random coil; Protein targeting

1. INTRODUCTION

Chloroplast transit peptides (cTPs) serve to target nuclear encoded proteins to chloroplasts with high specificity [1,2]; yet, compared to other targeting peptides such as secretory signal peptides or mitochondrial transit peptides, they lack obvious structural consensus features beyond a general enrichment for serine and threonine residues, a relatively low level of acidic residues, and a semi-conserved pattern of residues immediately surrounding the cleavage site for the stromal processing protease [3,4]. They also vary greatly in length, from around 30 residues up to more than 80 residues, Fig. 1, and, although it has been postulated that some consensus secondary structure must exist [5], no suggestions as to the most likely conformation(s) have been forthcoming.

This has led us to consider the obvious alternative, namely that cTPs are designed to be *devoid* of any regular secondary or tertiary structure. In this paper, we show that results both from secondary structure predictions and from an analysis of the conformations of short 'cTP-like' peptides in proteins of known 3D structure support this possibility. If cTPs indeed are 'perfect random coils', a series of interactions with cytosolic and chloroplastic chaperones could be all that is required for chloroplast targeting and import.

Correspondence address: G. von Heijne, Department of Molecular Biology, Karolinska Institute Center for Biotechnology, NOVUM, S-141 52 Huddinge, Sweden

2. METHODS

45 non-homologous higher-plant cTPs with known cleavage sites for the stromal processing protease were extracted from a collection of about 150 published sequences (a listing is available from G.v.H.). The highly conserved N-terminal Met–Ala dipeptide [3], as well as 3 semi-conserved residues from the C-terminus [4] were removed, and the secondary structure of the remaining sequences was predicted using a 'joint' prediction method [6] that achieves an improved prediction accuracy by combining the output from 5 widely used methods [7–11]. In addition, the Brookhaven Protein Data Bank of 3D protein structures was screened for 10-residue peptides containing at least 8 hydroxylated (Ser, Thr, Asn, Gln) or Pro residues, and the secondary structure of these segments as classified by the Kabsch–Sander method [12] was recorded.

3. RESULTS AND DISCUSSION

cTPs have an extremely high content of hydroxylated amino acids, Ser and Thr in particular, and a significantly lowered content of acidic residues [3]. Positively charged residues are found, but only at about the same frequencies as they appear in soluble proteins in general. In addition, the N-terminal 5–10 residues in most cases lack both positively charged residues as well as Gly and Pro [3], and the 3 C-terminal residues often fit a semi-conserved consensus motif [4], possibly as part of a slightly longer segment with a certain potential for forming an amphiphilic β-strand [3]. Beyond this, no clear structural features have been identified so far; in particular, stretches of apolar amino acids (as found in secretory signal peptides) or segments that can be folded into an amphiphilic α-helical structure (as found in mitochondrial targeting peptides) do not seem to be present [3].

ATP synthase, γ-subunit
MACSLSFSSSVSTFHLPTTTQSTQAPPNNATTLPTTNPIQC+ANLRELRDRI

GapB
MASHAALAPSRIPASTRLASKASQQYSFPTQCSFKRLDVADFSGLRSSNSVTFTAREASFHDVIAAQLTTKPTGAAPVRGETVA+KLKVAINGF

nitrate reductase
MASLPVNKIIPSSTTLLSSSNNNRRRNNSSIR+CQKAVSPAAE

Fig. 1. Three cTPs from spinach, chosen to demonstrate their wide range of variation in sequence. Sequence 1 [21] lacks charged residues altogether; sequence 2 [22] is 84 residues long; and sequence 3 [23] is only 32 residues long. The cleavage site for the stromal processing peptidase is indicated by '+'.

Some typical cTPs are shown in Fig. 1. Note that positively charged residues can be completely absent (sequence 1), and that hydroxylated and turn- or coil-promoting residues seem to be distributed rather uniformly along the entire length of the peptides. In fact, we have been unable to detect any significant clusters of the more abundant residues, either alone or in combination (data not shown), suggesting a more or less random distribution throughout the major part of the cTPs.

In order to look for less obvious structural features, we have performed secondary structure prediction on 45 non-homologous cTPs with known cleavage sites for the stromal processing protease (2186 residues), extracted from a collection of some 150 published sequences. Secondary structure was predicted by 5 different algorithms (see Section 2), which allows more reliably (i.e. unanimously) predicted regions to be distinguished from regions of lower reliability [6]. The results are shown in Table I: for 39% of the residues, all 5 methods give the same prediction totalling 5% helix, 6% sheet, and 89% coil; for an additional 26% of the residues 4 out of 5 methods agree and predict 12% helix, 16% sheet, and 72% coil. On a large control sample of globular proteins, the average coil prediction is around 55% (our unpublished results). Thus, two-thirds of all residues can be predicted with reasonable confidence (the expected accuracy is 65–75% [13]), and the overall result is 82% coil, with only small amounts of helix and sheet.

Table I

Secondary structure prediction

Conformation	Percent (5/5)	Percent (4/5)	Percent (5/5 + 4/5)
Helix	5	12	8
Sheet	6	16	10
Coil	89	72	82

3D structure analysis

Structural class	Percent
H	4
G	2
E	13
B	6
T	19
S	17
Irregular	39

We have also sought for structural correlates by scanning proteins of known 3D structure for peptides with an overall amino acid composition similar to the cTPs. A rather stringent screen requiring at least 8 residues out of 10 to be Ser, Thr, Asn, Gln, or Pro (in our current cTP sample, Pro is significantly enriched except in the N-terminal 10 positions – $f_{Pro} = 0.069$ compared to $f_{Pro} = 0.48$ in mature imported proteins; $P < 10^{-4}$) produced 16 non-homologous segments shown in Fig. 2. Their secondary structure was classified according to the Kabsch–Sander method [12], Fig. 2 and Table I. Out of 175 residues, only 33 (19%) are parts of repeating secondary structures (helices or β-strands), the remaining 142 residues (81%) are classified as either isolated residues making hydrogen bonds to β-sheets, turns, bends, or irregular structure. In large samples of globular proteins of known 3D structure, the corresponding percentages are around 50% helix + sheet and 50% non-regular structure [14] (and our own unpublished results).

The failure to find good candidates for membrane-interacting structures or other regular structures by such a diverse set of methods (hydrophobicity and hydrophobic-moment analysis [3], secondary structure predictions, and analysis of model peptides of known 3D structure) leads us to suggest that cTPs are in fact designed to be flexible peptides with a minimal content of regular secondary or higher-order structure, i.e. random coils. This would also be consistent with their large variation in length, and their relative tolerance to both deletions and insertions in their central region [15].

If our hypothesis is correct, it could have interesting implications for the mechanism of chloroplast protein import. Polar peptides of diverse sequences have been shown to bind efficiently to chaperones such as BiP and hsc70 [16], and it thus appears possible that cTPs would similarly have a strong affinity for other members of the hsp70-family. Although the sequence specificity of these chaperones is not known, we note that all but one of the 4 highest-affinity peptides studied by Flynn et al. [16] lack acidic residues (as do most cTPs), whereas all of the 4 lowest-affinity peptides contain one or more acidic residues. Recently, 3

```
1ALC  α-LACTALBUMIN              62 KSSQSPQSRNI
                                    B TT TT   T

1F19  R19.9 FAB FRAGMENT          3 QMTQTTSSLSAS
                                     EE  S  B  T

                                119 FPPSSQQLTSG
                                    B   SSSTTTT

1FB4  IMMUNOGLOBULIN FAB          5 TQPPSASGTP
                                    B  SEEEE T

                                165 TTKPSKQSNN
                                    E  EE TTS

1HMG  HAEMAGGLUTININ              184 HPSTNQEQTSL
                                      E SSTTHHHHH

1MEV  MEV COAT PROTEIN            147 PTGTPTKPTTQ
                                      TTS

1PRC  PHOTOSYNTHETIC
      REACTION CENTER               3 EPPPATTTQTG
                                      S  EE   S

1RBB  RIBONUCLEASE B              15 SSTSAASSSN
                                     TTSSS  STT

1SBC  SUBTILISIN                  154 NSGNSGSTNT
                                      S   BTTB

1TGB  TRYPSINOGEN                 74 PSYNSNTLNN
                                     TT BTTTTBT

2APP  ACID PROTEINASE             127 SSINTVQPQSQTTF
                                      GGG   BSS    H

2CGA  CHYMOTRYPSINOGEN            216 GSSTCSTSTPG
                                      SS  SSSEE

2EBX  ERABUTOXIN                    4 FNHQSSQPQTTKTCSP
                                      E   TTS   EEE  T

2PRK  PROTEINASE K                13 ISSTSPGTSTY
                                     HT SSSS   E

3SGB  PROTEINASE B                53 TTSGSSFPNN
                                     EEEEEE SBS
```

Fig. 2. 16 segments of known 3D structure with a high content of Ser, Thr, Asn, Gln and Pro residues extracted from the Protein Data Bank (see section 2). The PDB-code is shown with the protein name. The Kabsch–Sander secondary structure assignments are given below each sequence. H = α-helix, G = 3_{10}-helix, E = β-sheet, B = β-bridge (i.e. isolated residue making a hydrogen bond to a β-strand), T = turn, S = bend, blank = irregular structure.

hsp70 homologues have been detected in chloroplasts: two are located in the stroma, and one is tightly associated with the outer membrane but is not exposed on the surface of the organelle [17]. One might thus imagine an import pathway where the cTP is first bound to a cytoplasmic hsp70, then transferred to the membrane-bound chloroplastic hsp70, and finally to the stromal hsp70 proteins.

In such a model, targeting specificity would result from the preferential binding of the cTP to the membrane-bound hsp70, i.e. this chaperone would serve as a 'cTP receptor'. Discrimination against other unfolded polypeptides might at least in part be based on the near-absence of acidic residues in cTPs; in addition, the uncharged N-terminal region and the semi-conserved region around the cleavage site might also be involved in 'recognition' of the cTP by the membrane-bound hsp70. So far, the best candidate for a cTP receptor was identified using an approach based on anti-idiotypic antibodies [18]. However, recent work suggests that this receptor is identical to the phosphate-3-phosphoglycerate-phosphate translocator, see [19], suggesting a more indirect role in targeting.

Finally, all but one (MEV coat protein) of the polar segments found in our screen of the Protein Data Bank, Fig. 2, come from secreted proteins or the extracellular domains of cell-surface proteins, consistent with the observation that extracellular proteins tend to be rich in hydroxylated residues and prolines [20]. If cTPs are indeed designed to bind efficiently to chaperones, short 'cTP-like' peptides may be expected to be used also in proteins that, like secretory proteins, need to interact with chaperones during their biosynthesis.

Acknowledgements: This work was supported by a grant from the Swedish Natural Sciences Research Council, and by a travel grant from the Protein Engineering Research Institute, Osaka to G.v.H.

REFERENCES

[1] Boutry, M., Nagy, F., Poulsen, C., Aoyagi, K. and Chua, N.H. (1987) Nature 328, 340–342.
[2] Whelan, J., Knorpp, C. and Glaser, E. (1990) Plant Mol. Biol. 14, 977–982.
[3] Von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1989) Eur. J. Biochem. 180, 535–545.
[4] Gavel, Y. and Von Heijne, G. (1990) FEBS Lett. 261, 455–458.
[5] Keegstra, K. (1989) Cell 56, 247–253.
[6] Nishikawa, K. and Noguchi, T. (1990) in: Protein Engineering, Protein Design in Basic Research, Medicine, and Industry (Ikehara ed.) pp. 29–34, Japan Scientific Societies Press, Tokyo.
[7] Gibrat, J.F., Garnier, J. and Robson, B. (1987) J. Mol. Biol. 198, 425–443.
[8] Ptitsyn, O.B. and Finkelstein, A.V. (1983) Biopolymers 22, 15–25.
[9] Nagano, K. (1977) J. Mol. Biol. 109, 251–274.
[10] Qian, N. and Sejnowski, T.J. (1988) J. Mol. Biol. 202, 865–884.
[11] Nishikawa, K. and Ooi, T. (1986) Biochim. Biophys. Acta 871, 45–54.
[12] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.
[13] Nishikawa, K. and Noguchi, T. (1990) Methods Enzymol., in press.
[14] Kabsch, W. and Sander, C. (1983) FEBS Lett. 155, 179–182.
[15] Smeekens, S., Geerts, D., Bauerle, C. and Weisbeek, P. (1989) Mol. Gen. Genet. 216, 178–182.
[16] Flynn, G.C., Chappell, T.G. and Rothman, J.E. (1989) Science 245, 385–390.
[17] Marshall, J.S., Derocher, A.E., Keegstra, K. and Vierling, E. (1990) Proc. Natl. Acad. Sci. USA 87, 374–378.
[18] Pain, D., Kanwar, Y.S. and Blobel, G. (1988) Nature 331, 232–237.
[19] Meyer, D.I. (1990) Nature 347, 424–425.
[20] Nishikawa, K., Kubota, Y. and Ooi, T. (1983) J. Biochem. 94, 997–1007.
[21] Mason, J.G. and Whitfeld, P.R. (1990) Plant Mol. Biol. 14, 1007–1018.
[22] Brinkmann, H., Cerff, R., Salomon, M. and Soll, J. (1989) Plant Mol. Biol. 13, 81–94.
[23] Back, E., Burkhart, W., Moyer, M., Privalle, L. and Rothstein, S. (1988) Mol. Gen. Genet. 212, 20–26.