

**DUPLICACIONES EN TÁNDEM DE UN DOMINIO DEGENERADO DE UNIÓN AL
GTP EN EL ORIGEN DE LOS RECEPTORES GTPasa Toc159 Y SRP TILACOIDAL**

**TANDEM DUPLICATIONS OF A DEGENERATED GTP-BINDING DOMAIN AT THE
ORIGIN OF GTPase RECEPTORS Toc159 AND THYLAKOIDAL SRP**

MÓNICA ALEXANDRA ARIAS MALDONADO

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE BIOLOGÍA
BUCARAMANGA
2006**

**DUPLICACIONES EN TÁNDEM DE UN DOMINIO DEGENERADO DE UNIÓN AL
GTP EN EL ORIGEN DE LOS RECEPTORES GTPasa Toc159 Y SRP TILACOIDAL**

**TANDEM DUPLICATIONS OF A DEGENERATED GTP-BINDING DOMAIN AT THE
ORIGIN OF GTPase RECEPTORS Toc159 AND THYLAKOIDAL SRP**

Proyecto de Grado

MÓNICA ALEXANDRA ARIAS MALDONADO

Dirección

JORGE HERNÁNDEZ TORRES, Ph.D

Codirección

JACQUES CHOMILIER, Ph.D

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE BIOLOGÍA
BUCARAMANGA
2006**

TANDEM DUPLICATIONS OF A DEGENERATED GTP-BINDING DOMAIN AT THE ORIGIN OF GTPase RECEPTORS Toc159 AND THYLAKOIDAL SRP

Mónica Alexandra Arias Maldonado **

Key words: Translocon, chloroplast, signal recognition particle, evolution, HCA.

Abbreviations used: AA: amino acid; cp: chloroplast; HCA: Hydrophobic Cluster Analysis; SRP: Signal recognition Particle; Toc: Translocon at the Outer-envelope-membrane of chloroplast.

ABSTRACT

Sequences of proteins constituting the Toc complexes, allowing for translocation of proteins through the chloroplast membrane, have been compared on three species, *A. thaliana*, *P. sativum* and *M. truncatula*. We establish that a GTP binding domain has been tandem repeated to give rise to the three commonly referred A, G and M domains. They have also been compared to GTPase involved in the SRP system of *A. thaliana* and a distant homology has been revealed by means of HCA analysis. It resulted from this study a putative common ancestor for these proteins and a new domain definition has been proposed, based on these facts.

** Facultad de Ciencias, Escuela de Biología, Laboratorio de Biología Molecular, CINBIN, Jorge Hernández Torres, Protein Structure Prediction, IMPMC, Universités Paris 6 and Paris 7, Paris, France, Jacques Chomilier.

DUPLICACIONES EN TÁNDEM DE UN DOMINIO DEGENERADO DE UNIÓN AL GTP EN EL ORIGEN DE LOS RECEPTORES GTPasa Toc159 Y SRP TILACOIDAL

Mónica Alexandra Arias Maldonado **

Palabras claves: Translocon, cloroplasto, partícula de reconocimiento de señal, evolución, HCA.

Abreviaturas usadas: AA: aminoácidos; cp: cloroplasto; HCA: Análisis de agregados hidrofóbicos; SRP: partícula de reconocimiento de señal; Toc: Translocon en la membrana externa del cloroplasto

RESUMEN

La secuencia de las proteínas que constituyen el Complejo Toc, quienes permiten la translocación de proteínas a través de la membrana del cloroplasto, han sido comparadas en tres especies, *A. thaliana*, *P. sativum* y *M. truncatula*. Nosotros establecimos que un dominio de unión al GTP se repitió en tándem dando origen a los tres dominios llamados comúnmente A, G y M. Ellos también se compararon con GTPasas involucradas en el sistema SRP de *A. thaliana* y se reveló una homología distante por medio del análisis HCA. A partir de los resultados de este estudio se propone un ancestro putativo común para estas proteínas y una nueva definición de dominios, basados en estos hechos.

** Facultad de Ciencias, Escuela de Biología, Laboratorio de Biología Molecular, CINBIN, Jorge Hernández Torres, Protein Structure Prediction, IMPMC, Universités Paris 6 and Paris 7, Paris, France, Jacques Chomilier.

INTRODUCTION

The GTPse family of Translocons of Outer Chloroplast Membrane is a complex of several proteins that contain three domains, A, G, and M. The two last one are fairly conserved throughout the family and some structures have been determined. G domain (roughly 250 AA) is a GTPase and M is the domain (of around 400 AA) that is anchored to the outer membrane, although the details of binding are not yet very well understood since it has no predicted hydrophobic transmembrane domain. This complex has been found in *Arabidopsis thaliana*, *Pisum sativum* and *Medicago*. In the case of *A. thaliana*, it comprises Toc159, Toc132, Toc120, Toc75 and Toc34. *A. thaliana* sequences of Toc159, Toc132 and Toc120 have been aligned [1] and the G and M domain share around 65% identity. The largest diversity occurs for the A domain that can be as long as 800 AA in Toc159 up to 35 AA in Toc34 (this last one also actually lacks the M domain). One must mention that Toc75 has never reported as a related to the other members of the family on an evolutionary point of view [2] [3].

In this paper, by means of thorough sequence analysis, we propose a model of evolution for Toc proteins, by incorporation and successive duplication of a GTP binding domains, which concerns A, G and M domains. Besides, this domain is also found in thylakoidal signal recognition particles of *A. thaliana* and this will be discussed in terms of evolution.

METHODS

HCA is a method of designing a sequence on a scaffold corresponding to an alpha helix. It is then projected in the 2D space, duplicated in order to keep local environment for each amino acid, and hydrophobic residues neighbors in this plot are then clustered. HCA basically realises tightening of the sequence around hydrophobic amino acids. It has been statistically demonstrated that shapes of the clusters correspond to the centers of regular secondary structures [4]. Besides, the shapes of the clusters are keen indication of the nature of the secondary structure [5].

RESULTS

We started our study with the G domains as they are the most conserved among species, and we could also rely on the existence of a reported structure (PDB code 1h65, the Toc34 from pea). The G domains of the various *A. thaliana* are represented on figure 1, for Toc159, Toc90 and Toc34. The motifs representative of a GTP binding function are numbered from G1 to G5; they are fairly conserved in Toc proteins, also they can be significantly shifted along the sequence. Domain separation is a rather tedious task, and we used the three dimension structure of Toc34 in order to precisely define the limits of the G domain [6].

The G domain of *A. thaliana* toc159 (851-1071) has been analysed by means of HCA plots. It results that it can be aligned with three parts of the A domain, that we will call A1 (162-372), A2 (383-608), and A3 (631-839). The corresponding plots are represented in figure 2. The level of sequence identities relative to G domain are around 12%, for A1, A2 or A3. One can thus argue from this alignment that A domain from Toc159 originates from several duplications of the G domain. The GTP binding motifs, corresponding to a local signature of a few conserved

residues, are numbered from G1 to G5, and they are also represented in figure 2, some of them missing in several domains.

Among the three previously determined domains of Toc proteins, G and M domains are the most conserved [7]. Their HCA plots are aligned in figure 3 in the case of *A. thaliana*, and it also results that they might be distantly related. The level of sequence identity between these two proteins is 13%, sufficient to suggest a common origin.

From there, we have been searching for other proteins known for their GTPase activity, in order to check internal duplication of domains. This observation of a GTP binding domain which is repeated has been already made for the Der protein, although one still does not know if both are active or if one has a regulating function [8]. By means of HCA plot, thylakoidal Signal Recognition Particle 54 (cpSRP54) and its receptor from *A. thaliana* (cpFtsY) have shown internal repeats, as it is represented on figure 4. Therefore, we propose to split cpSRP54 into three domains, GN (N terminal, from 1 to 172), G (GTP binding domain, from 180 to ca. 370) and GC (C terminal domain, from 370 to 564). CpFtsY is formed by two domains, domain GN (1-161) and domain G (161-365). The mean level of sequence identity among these domains is, as in the case of Toc, of the order of 12%. The G domains of cpSRP54 and cpFtsY are highly homologous, sharing an identity of 45%.

From this observation that SRP and Toc were constituted of repeated GTP binding domains, one tried to align these domains by HCA. Figure 5 is an alignment of the G domain of cpSRP54 on the A2 domain of Toc159, with a level of identity among them of still 12%. Therefore the global distribution of the domains has also been reported in figure 1.

DISCUSSION

From sequence analysis methods, we have investigated the domain distribution in GTPase proteins of the Toc complex. As the A domain is the most divergent in the family, we paid more attention to this particular one. First, we carefully determine the length of the G domain, referring to the existence of a structure in Toc34. Hence, duplications of this domain were searched for in the A domain, which is actually composed of three such domains, that largely evolved from a common ancestor. Signal Recognition Particle receptors also present repeats of GTP binding domains and we showed by HCA alignment that their constituting domains are related, thus providing from a common ancestor. Such domains are found in a very diverse set of functions, such as proliferation, protein translation, signal transduction, cell growth, protein targeting. The common factor in all these cellular processes is GTP hydrolysis as a proofreader system [9]. This diversity of GTPases leads us to propose that, during evolution, the G domain gene was subject to easily recombine, which would have helped the occurrence of events such as of exon duplication [10]. These results allow conceiving that the molecular mechanism that resulted in tandem repeat is probably unique for all proteins of the complexes studied here.

Among the arguments in favour of a common ancestor between Toc complex and SRP complex, one can advance that both system actually act as receptors for the intra membrane transportation of protein precursors. During this process, GTP is hydrolysed. A1 to A3 domains of Toc have acquired new function devoted to protein protein interaction highly specific for the binding to their ligands. It is thus probable that A1 to A3 domains do not keep their GTPase activity, but some memory of the fold that can be recycled as a receptor. The fact that both Toc complexes and SRP complexes act as heterodimers, e.g. Toc159 with Toc34 [11] or homodimers, e.g. cpSRP54 [12] is a complementary argument in favor of a

common ancestor. Both complexes have homologues in all bacteria, and this might mean that their ancestor was present at the origin of endosymbiosis. One may guess that each complex gained new functions that have been adapted to transport in some organelles, such as endoplasmic reticulum or thylakoids [13]. Finally, the fact that SRP54 is associated to RNA in endoplasmic reticulum and bacteria, but not in chloroplast, is an argument in favor of a flexible SRP as long as adaptation to new functions is concerned. This might also have occurred for the Toc complex.

In this paper, we have proposed, on the basis of sequence homology, a model for the evolution of Toc and SRP complexes. As SRP is rather ubiquitous, one might guess that it has been recruited during evolution by the organelles. One can thus speculate that for chloroplast it was more advantageous to modify an existing machinery rather than building a new one from scratch.

ACKNOWLEDGEMENTS

This work has benefited of a grant from ECOS-Nord French and Colombian program, under n°A0502.

FIGURE CAPTIONS

Figure 1. Schematic domain representation of some Toc proteins from *A. thaliana*. They are related both to chloroplastic cpSRP54 and cpFtsY in *A. thaliana*. The five GTP binding motifs are indicated as G1 to G5.

Figure 2. HCA plots of the A1, A2 and A3 domains, aligned with the G domain of *A. thaliana* Toc159. The domains are defined as follows: A1 (162-372), A2 (383-608), and A3 (631-839) and G (851-1071). Conserved hydrophobic clusters are grey

shaded. On the same figure are represented the GTP binding motifs, indicated by G1 to G5.

Figure 3. HCA plots of G domain aligned with M domain from *A. thaliana* Toc159.

Figure 4. HCA plots of *A. thaliana* cpSRP54 receptor showing the internal repetition of a GTP binding domain. Also reported on the same figure, is the HCA plot of cpFtsY, which also presents an internal duplication. G1 to G5 allow to localize GTP binding sites.

Figure 5. HCA plots of domain G of *A. thaliana* cpSRP54, aligned with the A2 domain of *A. thaliana* Toc159.

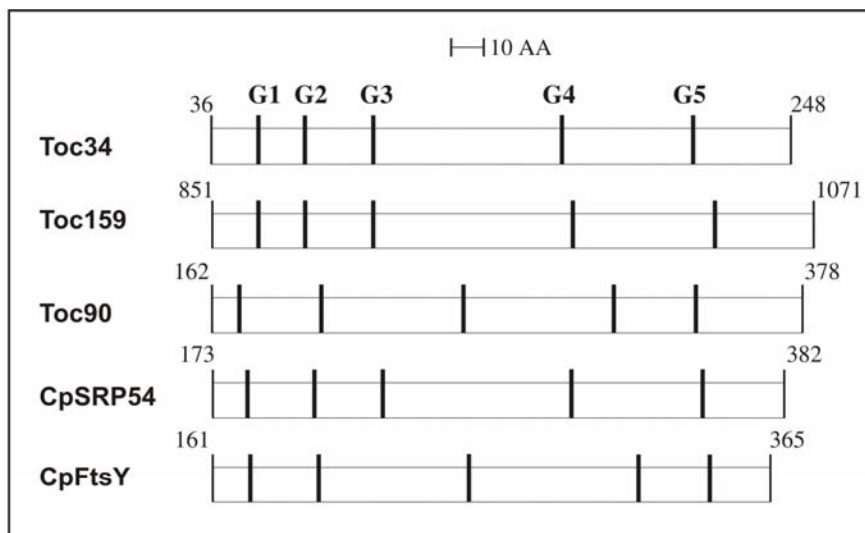


Figure 1. Schematic domain representation of some Toc proteins from *A. thaliana*. They are related both to chloroplastic cpSRP54 and cpFtsY in *A. thaliana*. The five GTP binding motifs are indicated as G1 to G5.

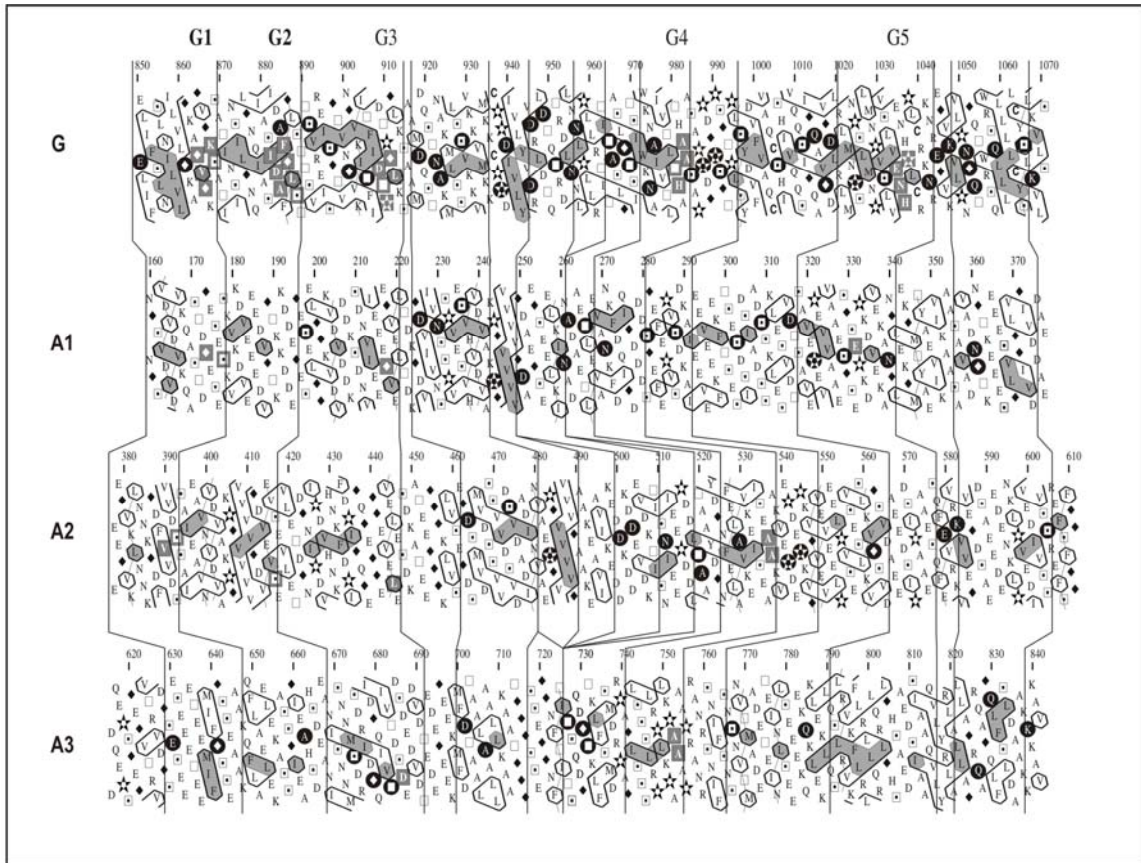


Figure 2. HCA plots of the A1, A2 and A3 domains, aligned with the G domain of *A. thaliana* Toc159. The domains are defined as follows: A1 (162-372), A2 (383-608), and A3 (631-839) and G (851-1071). Conserved hydrophobic clusters are grey shaded. On the same figure are represented the GTP binding motifs, indicated by G1 to G5.

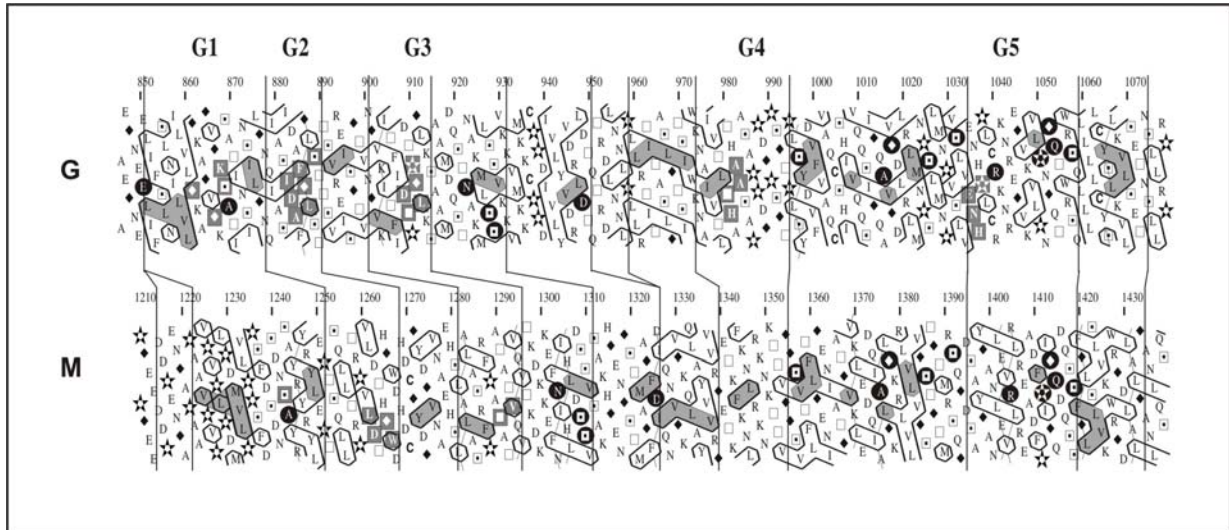


Figure 3. HCA plots of G domain aligned with M domain from *A. thaliana* Toc159.

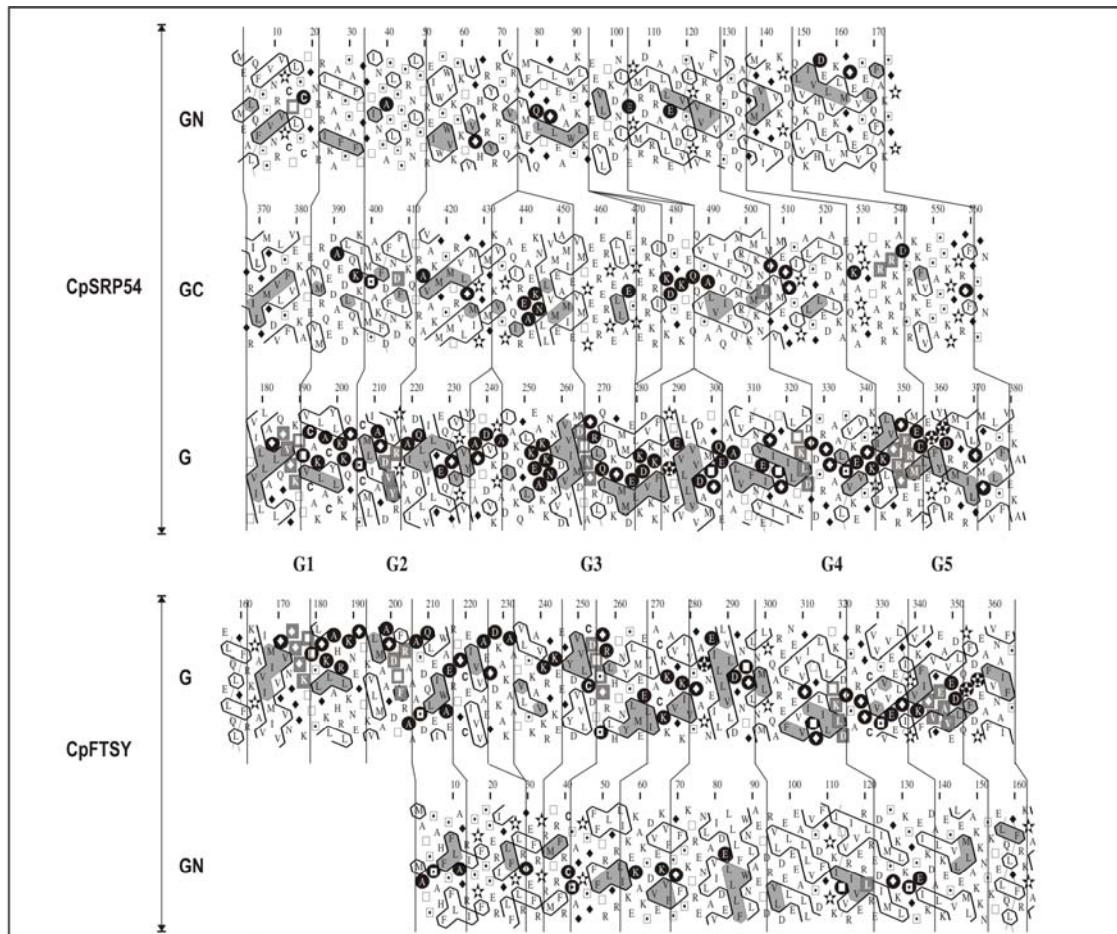


Figure 4. HCA plots of *A. thaliana* cpSRP54 receptor showing the internal repetition of a GTP binding domain. Also reported on the same figure, is the HCA plot of cpFTSY, which also presents an internal duplication. G1 to G5 allow to localize GTP binding sites.

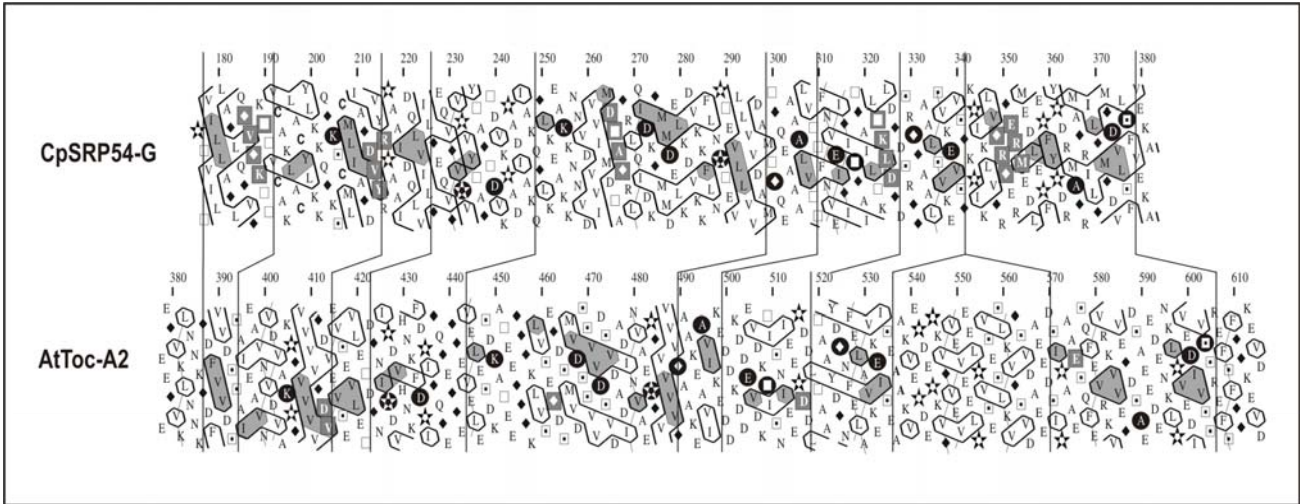


Figure 5. HCA plots of domain G of *A. thaliana* cpSRP54, aligned with the A2 domain of *A. thaliana* Toc159.

Bibliography

- 1 Bauer, J., Chen, K., Hiltbrunner, A., Wehrli, E., Eugster, M., Schnell, D. & Kessler, F. (2000). The major protein import receptor of plastids is essential for chloroplast biogenesis. *Nature* **403**, 203-207.
- 2 Sveshbujiva, N., Grimm, R., Soll, J. & Schleiff, E. (2000). Topology studies of the chloroplast protein import channel Toc75. *Biol. Chem.* **381**, 687-693.
- 3 Inoue, K. & Potter, D. (2004). The chloroplastic protein translocation channel Toc75 and its paralog OEP80 represent two distinct protein families and are targeted to the chloroplastic outer envelope by different mechanisms. *Plant. J.* **39**, 354-365.
- 4 Woodcock, S., Mornon, J. P. & Henrissat, B. (1992). Detection of secondary structure elements in proteins by hydrophobic cluster analysis. *Protein Eng.* **5**, 629-632.
- 5 Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B. & Mornon, J. P. (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol Life Sci* **53**(8), 621-45.
- 6 Sun, Y., Forouhar, F., Li, S., Yeh, S. T., Kao, S., Shr, H., Chou, C., Chen, C. & Hsiao, C. (2002). Crystal structure of pea Toc34, a novel GTPase of the chloroplast protein translocon. *NSB* **9**, 95-100.
- 7 Ivanova, Y., Smith, M., Chen, K. & Schnell, D. (2004). Members of the Toc159 import receptor family represent distinct pathways for protein targeting to plastids. *Mol Cell Biol* **15**, 3379-3392.
- 8 Hwang, J. & Inouye, M. (2001). An essential GTPase, Der, containing double GTP binding domains from *Escherichia coli* and *thermotoga maritime*. *J Biol Chem* **276**, 31415-31421.
- 9 Sprang, S. (1997). G protein mechanisms: insights from structural analysis. *Annu Rev Biochem* **66**, 639-678.

- 10 Letunic, I., Copley, R.R. and Bork, P. (2002). Common exon duplication in animals and its role in alternative splicing, *Hum. Mol. Genet.* **11**, 1561-1567.
- 11 Weibel, P., Hilbrunner, A., Brand, L. & Kessler, F. (2003). Dimerization of Toc-GTPases at the chloroplast protein import machinery. *J Biol Chem* **278**, 37321-37329.
- 12 Groves, M., Mant, A., Kuhn, A., Koch, J., Dubel, S., Robinsion, C. & Sinning, I. (2001). Functional characterization of recombinant chloroplast signal recognition particle. *J Biol Chem* **276**, 27778-27786.
- 13 Dalbey, R. & Kuhn, A. (2000). Evolutionarily related insertion pathways of bacterial, mitochondrial, and thylakoid membrane proteins. *Annu Rev ell dev Biol* **16**, 51-87.