# Comparison of Toc and Tic homologs found in the genome of *Paulinella chromatophora* endosymbionts with their higher plant counterparts

## Toc12

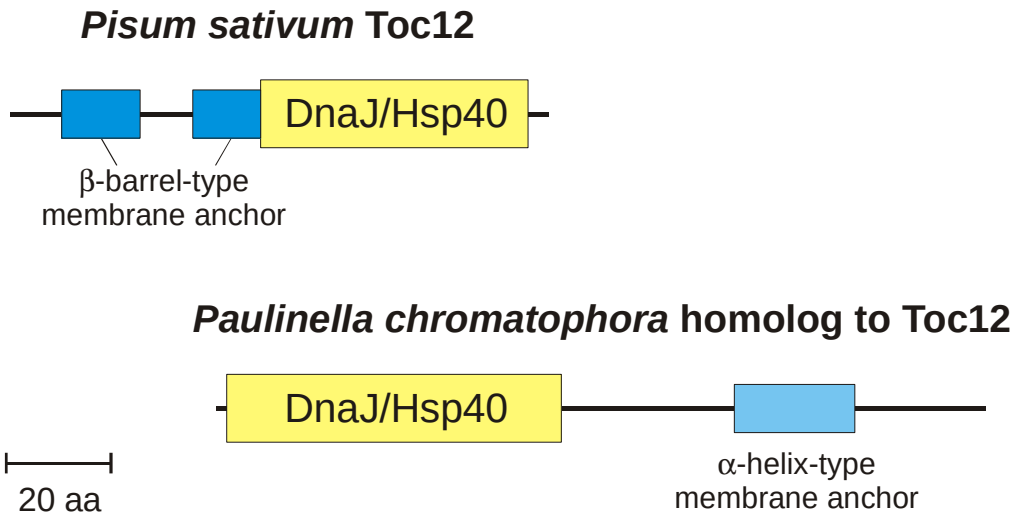


Fig. S1. Domain organization of *Pisum sativum* and *Paulinella chromatophora* Toc12 homologs. The *Pisum* Toc12 protein is characterised by a DnaJ (Hsp40) domain composed of 103 aa (Becker *et al.* 2004). The same domain was identified in the *Paulinella* homolog (147 aa), indicating its role as an Hsp70-binding protein; however, in the higher plant protein, this domain is located toward the C-terminus (Becker *et al.* 2004), whereas in *Paulinella* Toc12 it occupies an N-terminal position. The N-terminal region of higher plant Toc12 is predicted to fold into a β-barrel conformation, functioning as a transmembrane domain (Becker *et al.* 2004). We found a potential transmembrane region in the *Paulinella* homolog as well; however, it probably folds into an α-helix and is located toward the C-terminal end. It appears that these proteins, despite of their 'inverted' structures, could be functionally comparable, both anchored in envelope membranes and stimulating the ATPase activity of Hsp70-type chaperones. In the case of *Pisum sativum*, it was demonstrated that Toc12 is inserted into the outer plastid membrane with the DnaJ domain exposed into the intermembrane space (Becker *et al.* 2004). We suggest that the *Paulinella* Toc12 homolog could be anchored in either the outer or inner envelope membrane, or possibly even in both, so that its catalytic domain is oriented respectively to the intermembrane space, the matrix or even to both these spaces (see Fig. 4).

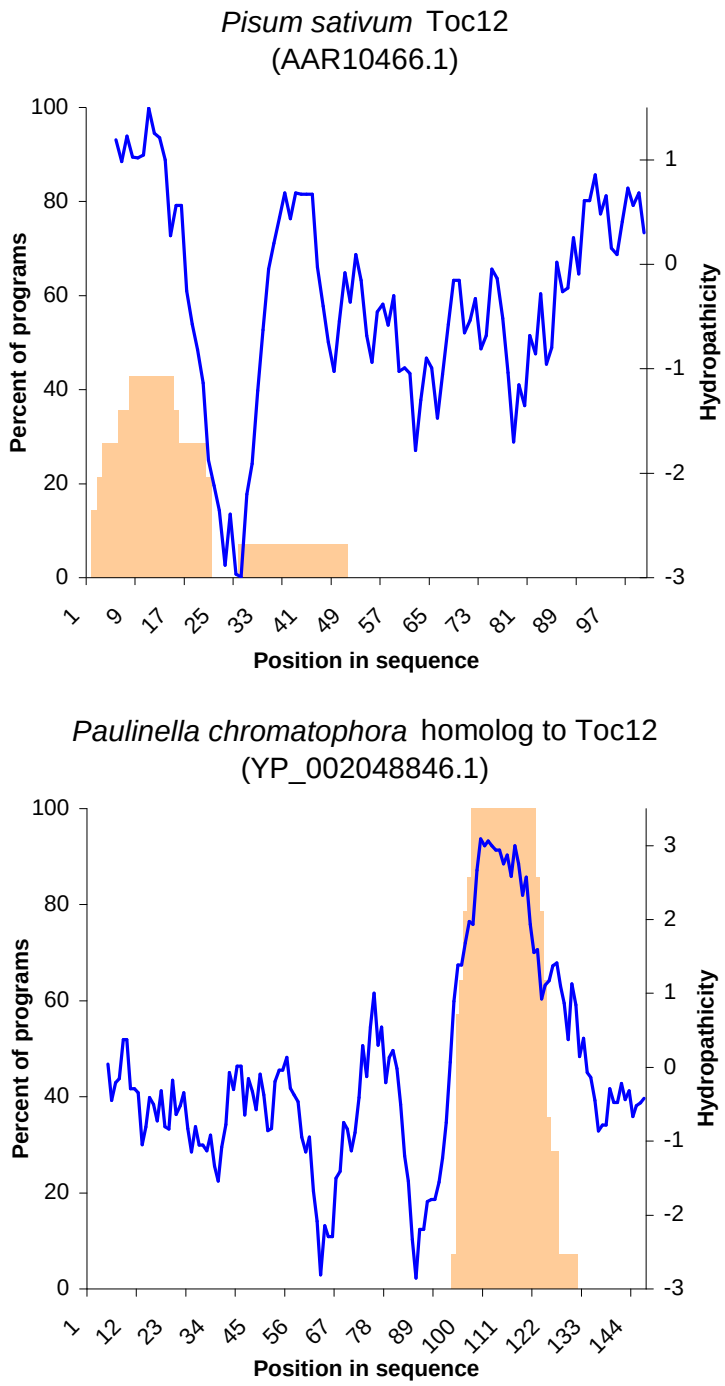


Fig. S2. Transmembrane α-helical predictions for the *Pisum sativum* Toc12 sequence and its homolog from *Paulinella chromatophora*. Left vertical axis and orange histograms show percent of programs indicating that a given site in the sequence belongs to a transmembrane α-helix. Fourteen algorithms were used in these predictions: ConPredII (separate models for prokaryotes and eukaryotes) (Arai *et al.* 2004), DAS (with cutoff 1.7) (Cserzo *et al.* 1997), HMMTOP (Tusnády & Simon 2001), MEMSAT3 (Jones 2007), PHDhtmm (Rost *et al.* 1996), PHOBIUS (Käll *et al.* 2004), PRED-TMR (Pasquier *et al.* 1999), SOSUI (Hirokawa *et al.* 1998), TMHMM (Krogh *et al.* 2001), TMPRED (Hofmann & Stoffel 1993), TUPS (Zhou & Zhou 2003), waveTM (Pashou *et al.* 2004), and WHAT (Zhai & Saier 2001). Right vertical axis and blue line represent hydropathy profiles according to the Kyte-Doolittle scale (Kyte & Doolittle 1982), assuming a sliding window length of 11 residues.

Table S1. Results of InterPro database searches using *Pisum sativum* Tic12 and its homolog from *Paulinella chromatophora* as queries.

*Pisum sativum* Toc12 (gi|38049670|gb|AAR10466.1)

| IPR001623 **Heat shock protein DnaJ, N-terminal** | | | |
|---|---|---|---|
| GENE3D | G3DSA:1.10.287.110 | *DnaJ_N* | 2.7E-11 [47-98]T |
| PFAM | PF00226 | *DnaJ* | 2.0E-18 [48-96]T |
| SMART | SM00271 | *DnaJ* | 2.4E-14 [47-103]T |
| PROFILE | PS50076 | *DNAJ_2* | 0.0 [48-103]T |
| SUPERFAMILY | SSF46565 | *DnaJ_N* | 4.6E-16 [40-96]T |

| IPR015609 **Molecular chaperone, heat shock protein, Hsp40, DnaJ** | | | |
|---|---|---|---|
| PANTHER | PTHR11821 | *Hsp40/DnaJ_Rel* | 4.6E-12 [47-98]T |

| **unintegrated** | | | |
|---|---|---|---|
| PANTHER | PTHR11821:SF79 | *PTHR11821:SF79* | 4.6E-12 [47-98]T |

*Paulinella chromatophora* homolog to Toc12 (gi|194476667|ref|YP_002048846.1)

| IPR001623 **Heat shock protein DnaJ, N-terminal** | | | |
|---|---|---|---|
| GENE3D | G3DSA:1.10.287.110 | *no description* | 5.9e-16 [1-74]T |
| PFAM | PF00226 | *DnaJ* | 2.5e-28 [2-65]T |
| SMART | SM00271 | *DnaJ* | 1.5e-18 [1-60]T |
| PROSITE | PS00636 | *DNAJ_1* | NA [45-64]T |
| PROFILE | PS50076 | *DNAJ_2* | 16.9 [2-68]T |
| PROSITE | PS00636 | *DNAJ_1* | 8e-5 [45-64]T |
| SUPERFAMILY | SSF46565 | *Chaperone J-domain* | 5.4e-18 [1-127]T |

| IPR015609 **Molecular chaperone, heat shock protein, Hsp40, DnaJ** | | | |
|---|---|---|---|
| PANTHER | PTHR11821 | *DNAJ/HSP40* | 8e-21 [2-65]T |

| **unintegrated** | | | |
|---|---|---|---|
| PANTHER | PTHR11821:SF52 | *DNAJ-RELATED* | 8e-21 [2-65]T |
| TMHMM | tmhmm | *transmembrane_regions* | NA [99-121]? |

## Toc64

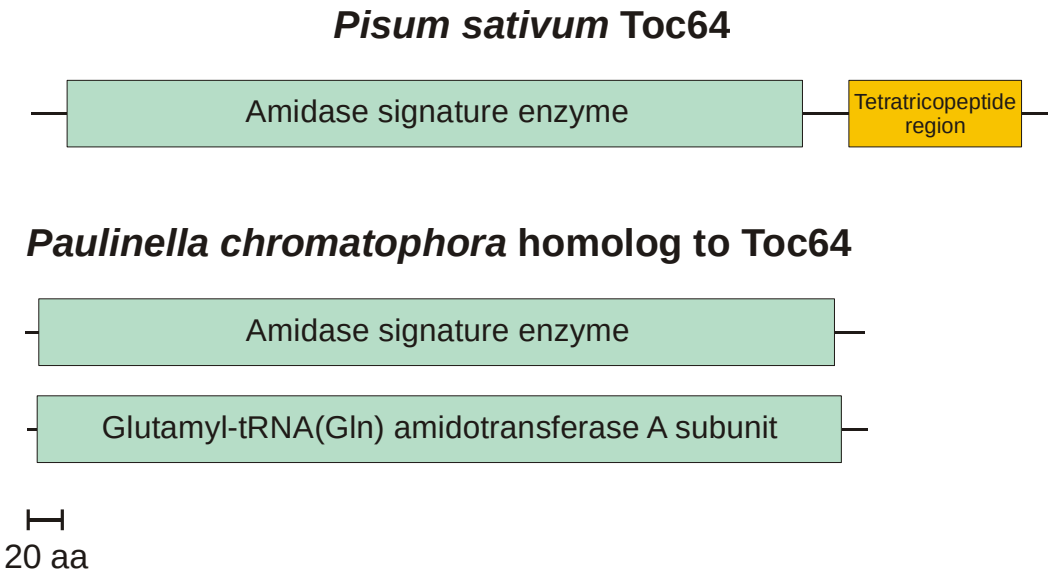


Fig. S3. Domain organization of *Pisum sativum* and *Paulinella chromatophora* Toc64 homologs. The higher plant Toc64 protein is characterized by the presence of two distinct domains: (i) amidase signature enzyme and (ii) tetratricopeptide domains (Sohrt & Soll 2000; Qbadou *et al.* 2007). The latter domain contains three tetratrico-peptide motifs forming a docking/recognition site for a cytosolic guidance complex composed of Hsp90s and associated plastid proteins (Qbadou *et al.* 2006, 2007). In contrast to *Pisum* Toc64, the *Paulinella* homolog is shorter (490 versus 593 aa) and possesses only the amidase signature domain. This suggests that *Paulinella* Toc64 fulfills functions other than assistance in protein translocation across the endosymbiont membranes. In accordance with this hypothesis, domain searches found a glutamyl-tRNA(Gln) amidotransferase A subunit domain in the *Paulinella* sequence. This domain overlaps the amidase signature domain and is not detected in the higher plant homolog.

Table S2. Results of InterPro database searches using *Pisum sativum* Toc64 and its homolog from *Paulinella chromatophora* as queries.

*Pisum sativum* Toc64 (gi|7453538|gb|AAF62870.1|AF179282_1)

| IPR000120 **Amidase signature enzyme** | | | |
|---|---|---|---|
| GENE3D | G3DSA:3.90.1300.10 | *Amidase_sig_enz* | 1.2E-70 [28-445]T |
| PANTHER | PTHR11895 | *Amidase* | 0.0 [32-454]T |
| PFAM | PF01425 | *Amidase* | 1.5E-27 [13-444]T |
| SUPERFAMILY | SSF75304 | *Amidase_sig_enz* | 1.6E-74 [11-455]T |

| IPR001440 **Tetratricopeptide TPR-1** | | | |
|---|---|---|---|
| PFAM | PF00515 | *TPR_1* | 0.0032 [511-544]T<br>0.0013 [545-578]T |

| IPR011990 **Tetratricopeptide-like helical** | | | |
|---|---|---|---|
| GENE3D | G3DSA:1.25.40.10 | *TPR-like_helical* | 1.3E-34 [475-593]T |

| IPR013026 **Tetratricopeptide region** | | | |
|---|---|---|---|
| SMART | SM00028 | *TPR* | 0.0082 [477-510]T<br>1.5E-4 [511-544]T<br>2.0E-4 [545-578]T |
| PROFILE | PS50005 | *TPR* | 0.0 [477-510]T<br>0.0 [511-544]T<br>0.0 [545-578]T |
| PROFILE | PS50293 | *TPR_REGION* | 0.0 [477-578]T |

| IPR013105 **Tetratricopeptide TPR2** | | | |
|---|---|---|---|
| PFAM | PF07719 | *TPR_2* | 0.67 [477-510]T |

| **unintegrated** | | | |
|---|---|---|---|
| PANTHER | PTHR11895:SF6 | *PTHR11895:SF6* | 0.0 [32-454]T |
| SUPERFAMILY | SSF48452 | *SSF48452* | 2.8E-29 [477-581]T |

*Paulinella chromatophora* homolog to Toc64 (gi|194476713|ref|YP_002048892.1)

| IPR000120 **Amidase signature enzyme** | | | |
|---|---|---|---|
| GENE3D | G3DSA:3.90.1300.10 | *no description* | 2.2e-159 [1-471]T |
| PANTHER | PTHR11895 | *AMIDASE* | 3.1e-225 [11-481]T |
| PFAM | PF01425 | *Amidase* | 2.9e-210 [19-463]T |
| PROSITE | PS00571 | *AMIDASES* | NA [147-178]T |
| PROSITE | PS00571 | *AMIDASES* | 8e-5 [147-178]T |
| SUPERFAMILY | SSF75304 | *Amidase signature (AS) enzymes* | 1.7e-160 [1-474]T |

| IPR004412 **Glutamyl-tRNA(Gln) amidotransferase A subunit** | | | |
|---|---|---|---|
| TIGRFAMs | TIGR00132 | *gatA: glutamyl-tRNA(Gln) and/or aspartyl-tR* | 3.3e-209 [6-475]T |

| **unintegrated** | | | |
|---|---|---|---|
| PANTHER | PTHR11895:SF7 | *GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT A* | 3.1e-225 [11-481]T |

## Tic21

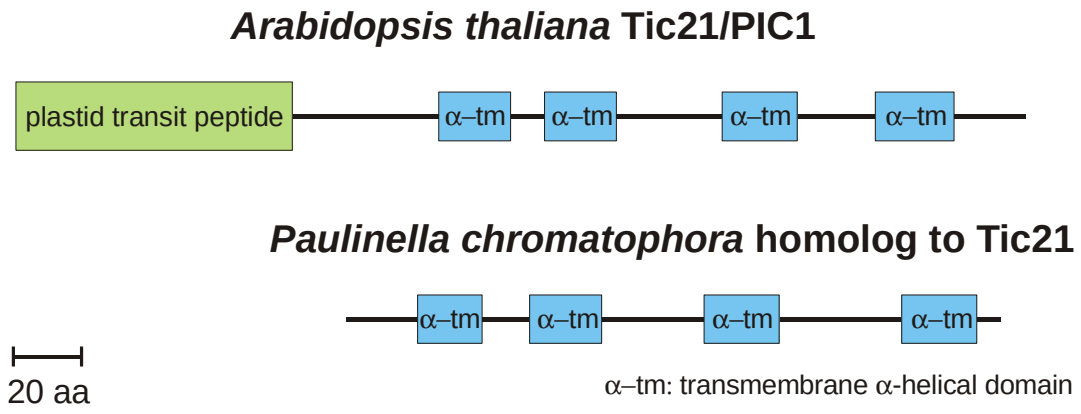


Fig. S4. Domain organization of *Arabidopsis thaliana* and *Paulinella chromatophora* Tic21 homologs. Both these proteins are predicted to have four α-helical transmembrane domains (see Fig. S5), which could form a protein-conducting channel in the inner plastid/endosymbiont membrane (Teng *et al.* 2006). The higher plant homolog is over 100 aa longer than its *Paulinella* counterpart (296 versus 192 aa). This difference results from the presence of a cleavable 81 aa-transit peptide functioning as a plastid-targeting signal in the *Arabidopsis* protein (Duy *et al.* 2007).

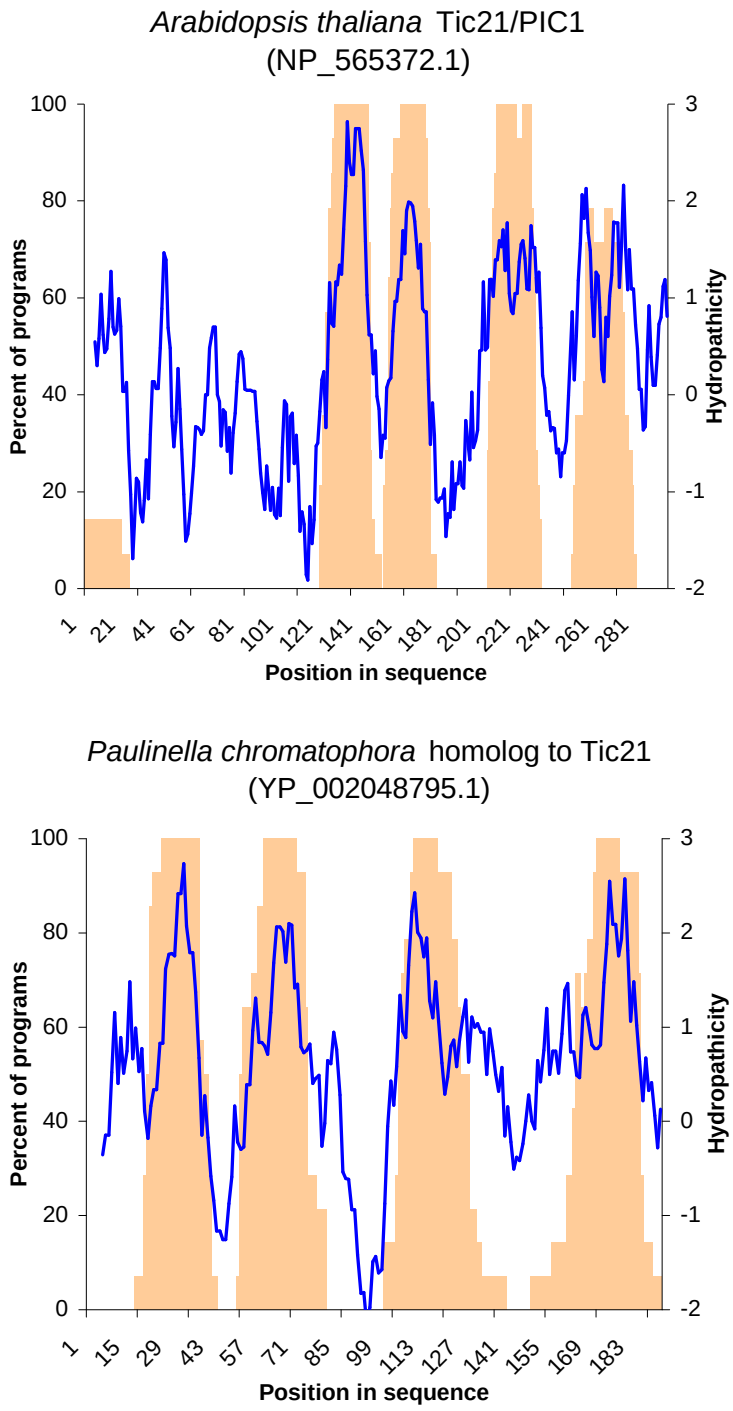


Fig. S5. Transmembrane α-helical predictions for the *Arabidopsis thaliana* Tic21 sequence and *Paulinella chromatophora* homolog. Left vertical axis and orange histograms show percent of programs indicating that a given site in the sequence belongs to a transmembrane α-helix. Fourteen algorithms were used in these predictions: ConPredII (separate models for prokaryotes and eukaryotes) (Arai *et al.* 2004), DAS (with cutoff 1.7) (Cserzo *et al.* 1997), HMMTOP (Tusnády & Simon 2001), MEMSAT3 (Jones 2007), PHDhtmm (Rost *et al.* 1996), PHOBIUS (Käll *et al.* 2004), PRED-TMR (Pasquier *et al.* 1999), SOSUI (Hirokawa *et al.* 1998), TMHMM (Krogh *et al.* 2001), TMPRED (Hofmann & Stoffel 1993), TUPS (Zhou & Zhou 2003), waveTM (Pashou *et al.* 2004), and WHAT (Zhai & Saier 2001). Right vertical axis and blue line represent hydropathy profiles according to the Kyte-Doolittle scale (Kyte & Doolittle 1982), assuming a sliding window length of 11 residues.

Table S3. Results of InterPro database searches using *Arabidopsis thaliana* Tic21 and its homolog from *Paulinella chromatophora* as queries.

*Arabidopsis thaliana* Tic21/PIC1 (gi|18397775|ref|NP_565372.1)

| **unintegrated** | | | |
|---|---|---|---|
| TMHMM | tmhmm | *transmembrane_regions* | NA [125-145]?<br>NA [155-175]? |

*Paulinella chromatophora* homolog to Tic21(gi|194476616|ref|YP_002048795.1)

| **unintegrated** | | | |
|---|---|---|---|
| SIGNALP | SignalP | *signal-peptide* | NA [1-29]? |
| TMHMM | tmhmm | *transmembrane_regions* | NA [22-42]?<br>NA [52-74]?<br>NA [109-129]?<br>NA [165-185]? |

## Tic32

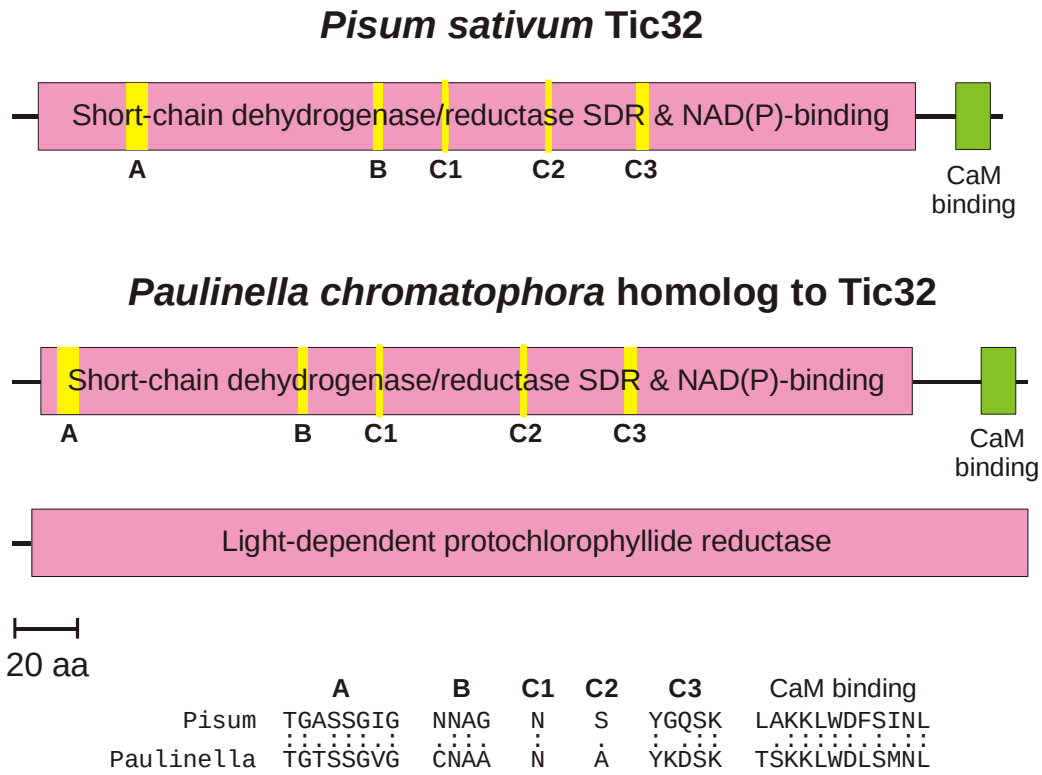


Fig. 6S. Domain organization of *Pisum sativum* and *Paulinella chromatophora* Tic32 homologs. Similar to its higher plant counterpart (Hörmann *et al.* 2004), *Paulinella* Tic32 contains three related and overlapping domains: (i) short-chain dehydrogenase/reductase (SDR), (ii) glucose/ribitol dehydrogenase, and (iii) NAD(P)-binding domains. In addition, both proteins possess a calmodulin (CaM)-binding domain. Consequently, they have similar lengths, 324 aa (*Paulinella*) and 316 aa (*Pisum*). The characteristic feature of *Paulinella* Tic32 is the presence of a light-dependent protochlorophyllide reductase domain, which overlaps the whole protein and was not detected in the higher plant homolog. The SDR domain of higher plant Tic32 proteins contains several characteristic sequence motifs, such as (i) the TGXXXGXG motif responsible for NAD($P^+$) binding (in A), (ii) the conserved NNAG motif stabilizing the central β-sheet (in B), and (iii) the active site tetrad consisting of Asn (in C1), Ser (in C2), Tyr, and Lys residues (in C3) (Hörmann *et al.* 2004). The CaM-binding domain also has a characteristic sequence motif (Chigri *et al.* 2006). Interestingly, all these motifs and residues are strongly conserved in the *Paulinella* Tic32 homolog. This suggests that *Paulinella* Tic32 is engaged in regulation of protein import via sensing the plastid redox state, as found for higher plant plastids (for a review see Stengel *et al.* 2007).

Table S4 Results of InterPro database searches using *Pisum sativum* Tic32 and its homolog from *Paulinella chromatophora* as queries.

*Pisum sativum* Tic32 (gi|42725482|gb|AAS38575.1)

| IPR002198 **Short-chain dehydrogenase/reductase SDR** | | | |
|---|---|---|---|
| PRINTS | PR00080 | *SDRFAMILY* | 1.6E-6 [108-119]T<br>1.6E-6 [164-172]T<br>1.6E-6 [199-218]T |
| PANTHER | PTHR19410 | *ADH_short_C2* | 4.7E-125 [8-288]T |
| PFAM | PF00106 | *adh_short* | 1.3E-7 [31-171]T |

| IPR002347 **Glucose/ribitol dehydrogenase** | | | |
|---|---|---|---|
| PRINTS | PR00081 | *GDHRDH* | 4.1E-18 [32-49]T<br>4.1E-18 [108-119]T<br>4.1E-18 [158-174]T<br>4.1E-18 [222-239]T |

| IPR016040 **NAD(P)-binding** | | | |
|---|---|---|---|
| GENE3D | G3DSA:3.40.50.720 | *NAD(P)-bd* | 3.6E-50 [27-241]T |
| SUPERFAMILY | SSF51735 | *NAD(P)-bd* | 1.6E-51 [27-284]T |

| **unintegrated** | | | |
|---|---|---|---|
| PANTHER | PTHR19410:SF98 | *PTHR19410:SF98* | 4.7E-125 [8-178]T<br>4.7E-125 [8-288]T<br>4.7E-125 [194-288]T |

*Paulinella chromatophora* homolog to Tic32 (gi|194477341|ref|YP_002049520.1)

| IPR002198 **Short-chain dehydrogenase/reductase SDR** | | | |
|---|---|---|---|
| PANTHER | PTHR19410 | *SHORT-CHAIN DEHYDROGENASES/REDUCTASE FAMILY MEMBER* | 6.6e-101 [9-151]T<br>6.6e-101 [187-287]T |
| PFAM | PF00106 | *adh_short* | 3.5e-07 [9-129]T |

| IPR002347 **Glucose/ribitol dehydrogenase** | | | |
|---|---|---|---|
| PRINTS | PR00081 | *GDHRDH* | 1.7e-06 [10-27]T<br>1.7e-06 [84-95]T<br>1.7e-06 [217-234]T |

| IPR005979 **Light-dependent protochlorophyllide reductase** | | | |
|---|---|---|---|
| TIGRFAMs | TIGR01289 | *LPOR: light-dependent protochlorophyllide r* | 1.4e-187 [6-324]T |

| IPR016040 **NAD(P)-binding** | | | |
|---|---|---|---|
| GENE3D | G3DSA:3.40.50.720 | *no description* | 9.5e-34 [9-240]T |
| SUPERFAMILY | SSF51735 | *NAD(P)-binding Rossmann-fold domains* | 1.4e-36 [9-287]T |

| **unintegrated** | | | |
|---|---|---|---|
| PANTHER | PTHR19410:SF100 | *PROTOCHLOROPHYLLIDE REDUCTASE* | 6.6e-101 [9-151]T<br>6.6e-101 [187-287]T |

## References


Arai M., Mitsuke H., Ikeda M., Xia J.-X., Kikuchi T., Satake M., Shimizu T. (2004) ConPred II: a consensus prediction method for obtaining transmembrane topology models with high reliability. *Nucleic Acids Research*, **32**, W390-W393.

Becker T., Hritz J., Vogel M., Caliebe A., Bukau B., Soll J., Schleiff E. (2004) Toc12, a novel subunit of the intermembrane space preprotein translocon of chloroplasts. *Molecular Biology of the Cell*, **15**, 5130-5144.

Chigri E., Hörmann F., Stamp A., Stammers D.K., Bölter B., Soll J., Vothknecht U.C. (2006) Calcium regulation of chloroplast protein translocation is mediated by calmodulin binding to Tic32. *Proceedings of the National Academy of Sciences of the united States of America*, **103**, 16051-16056.

Cserzo M., Wallin E., Simon I., von Heijne G., Elofsson A. (1997) Prediction of transmembrane alpha-helices in procariotic membrane proteins: the Dense Alignment Surface method. *Protein Engineering*, **10**, 673-676.

Duy D., Wanner G., Meda A.R., von Wirén N., Soll J., Philippar K. (2007) PIC1, an ancient permease in *Arabidopsis* chloroplasts, mediates iron transport. *Plant Cell*, **19**, 986-1006.

Hirokawa T., Boon-Chieng S., Mitaku S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, **14**, 378-379.

Hofmann K., Stoffel W. (1993) TMbase – a database of membrane spanning proteins segments. *Biological Chemistry Hoppe-Seyler*, **374**, 166.

Hörmann F., Kchler M., Sveshnikov D., Oppermann U., Li Y., Soll J. (2004) Tic32, an essential component in chloroplast biogenesis. *Journal of Biological Chemistry*, **279**, 34756-34762.

Jones D.T. (2007) Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics*, **23**, 538-544.

Käll L., Krogh A., Sonnhammer E.L.L. (2004) A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, **338**, 1027-1036.

Krogh A., Larsson B., von Heijne G., Sonnhammer E.L.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, **305**, 567-580.

Kyte J., Doolittle R.F. (1982) A simple method for displaying the hydropathic character of protein. *Journal of Molecular Biology*, **157**, 105-132.

Qbadou S., Becker T., Mirus O., Tews I, Soll J., Schleiff E. (2006) The molecular chaperone Hsp90 delivers precursor proteins to the chloroplast import receptor Toc64. *EMBO Journal*, **25**, 1836-1847.

Qbadou S., Becker T., Bionda T., Reger K., Ruprecht M., Soll J., Schleiff E. (2007) Toc64 – a preprotein-receptor at the outer membrane with bipartide function. *Journal of Molecular Biology*, **367**, 1330-1346.

Pashou E.E., Litou Z.I., Liakopoulos Th.D., Hamodrakas S.J. (2004) waveTM: Wavelet-based transmembrane segment prediction. *In Silico Biology*, **4**, 0012.

Pasquier C., Promponas V.J., Palaios G.A., Hamodrakas J.S., Hamodrakas S.J. (1999) A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the SwissProt database: the PRED-TMR algorithm. *Protein Engineering*, **12**, 381-385.

Rost B., Fariselli P., Casadio R. (1996) Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Science*, **7**, 1704-1718.

Sohrt K., Soll J. (2000) Toc64, a new component of the protein translocon of chloroplasts. *Journal of Cell Biology*, **148**, 1213-1221.

Stengel A., Soll J., Bölter B. (2007) Protein import into chloroplasts: new aspects of a well-known topic. *Biological Chemistry*, **388**, 765-772.

Teng Y.S., Su Y.S., Chen L.J., Lee Y.J., Hwang I., Li H.M. (2006) Tic21 is an essential translocon component for protein translocation across the chloroplast inner envelope membrane. *Plant Cell*, **18**, 2247-2257.

Tusnády G.E., Simon I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics*, **17**, 849-850.

Zhai Y., Saier M.H. Jr. (2001) A web-based program (WHAT) for the simultaneous prediction of hydropathy, amphipathicity, secondary structure and transmembrane topology for a single protein sequence. *Journal of Molecular Microbiology and Biotechnology*, **3**, 501-502.

Zhou H., Zhou Y. (2003) Predicting the topology of transmembrane helical proteins using mean burial propensity and a hidden-Markov-model based method. *Protein Science*, **12**, 1547-1555.