

Supplementary material

Article

Challenges in the heterologous production of furanocoumarins in *Escherichia coli*

Joana L Rodrigues ^{1,2*}, Daniela Gomes ^{1,2} and Lígia R Rodrigues ^{1,2}

¹ CEB – Centre of Biological Engineering, University of Minho, 4710-057 Braga, Portugal; daniela.gomes@ceb.uminho.pt (DG); lrmr@deb.uminho.pt (LRR)

² LABBELS – Associate Laboratory, Braga/Guimarães, Portugal.

* Correspondence: joanarodrigues@ceb.uminho.pt; joana.joanalucia@gmail.com (J.L.R.); Tel.: +351-253604423

Table S1. Gene sequence of *p*-coumaroyl-CoA 2'-hydroxylase (C2'H), prenyltransferase 1 (PT1) and psoralen synthase (PS) with codon optimization.

Gene	Sequence
C2'H (<i>Ipomoea batatas</i>)	<p>ATGCCGTCGACCACTCTGAGCACTGTTCTTAGTGATATTAATGACTTCGTCGTGAAAC AAGGCCACGGTGTTAAAGGCCTGTCCGAGCTGGGCTTACAAACCTTGCCGAACCAG TATGTGCATCCGCCAGAGGAGCGTTTAAAGCAGTATGGATGTGGTTTCCGATGACAGC ATCCCGGTGATCGATGTCAGCAATTGGGAAGACCCGAAAGTGGCCAAATTGATCTG CGACGCAGCCGAAAAGCGCGGTTTCTTTTCAGATTGTGAATCATGGCATTCCGATTGA AATGCTGGAGAAAGCAAAGGCCGCGACCTACCGTTTTTTTCGTGAACCAGCGGAGG AAAAGAAAAAGTACAGCAAGGAAAAGTGTCCAACCTCCACGTGCGCTACAGTACT AGTTTTCTTCCGCAGATTGAGAAGCGCTGGAGTGGAAGATCACCTGTCCATGTTT TATGTGTCCGATGAGGAAGCGGCCAGTACTGGCCGCCGTCGTGTCGCGATGATGCG GTTGAATACCTGAAATCGTGTGAAATGGTTAGTCGTAACTGTTAGAAGCACTTATG CAGGGTCTGAACGTTAATCAGATTGACGATTCCAAGGAGAGTCTGTTAATGGGCAGT CGTCGCATTAATATTAATATTATCCAAAATGCCCGAACCCAGATTGACCGTTGGT GTGGGTCGCCATAGCGATATCAGCACCTGACCCTGCTTTACAGGACGACATCGGC GGTTTATATGTTTCGTAAGCTGGAACATGAGGCCTGGTCCCATGTGCCGCCGGTGAAA GGTGCGTTAGTCATTAATATCGGCGATGCACTCCAGATTATGTCCAACGGTCGCTAT AAAAGTATTGAACACCGCGTGATGGCGAACGAAAGTAATGATCGTATCTCGGTTCC GGTGTGTTGTGAACCCGCGCCCGAATGATATTGTGGGCCCGCTGCCGGAGGTTTTGGC GTCGGGCGAAAAGCCAGTTTATAAACCAGTGCTGTATTCCGACTACGCAAAGCATT TTACCGTAAAGCGCATAACGGCAAAGATACGATCGCATTTCGCGCGCATTGAATAA</p>
PT1 (<i>Pastinaca sativa</i>)	<p>ATGGCGCAGACCATTATGCACAGCCGCTTGAGCTCCGTTTCCTCCATCTGCAGCGTG ACAAGGGTTTTTCGCACCCTGCCAACTCAGCGTCGTCATGCCAAAGTTGTTAACGGCG ATCAAGAAATTCGCGTTTCGTGTGGTGAGCTGCGATAAAAAATTTAGACTCCACTAAAA ATTTTAGTGGCAGCTGTGAAAAACCGATTTCGCACCCATACTAATAAACTGCTGCAAA CCATCAGCGCAACCAGCGATCGTGAAGCCATCATCCAACCGAAGGATGATTATGAA GCCCCGTGGCAAAATACGCTGCGTCGTAAATGGGATGCGTTCTGTACGTTTGCCGT CCGATTTCGCGATTGTTGACGATTATCGGCATTAGCTCGGTGAGTCTGTTACCATTGA CTTCGGTTAAAGACTTCAGCGCACCATAATTTGTGCGCTTGCTGCAGGCCCTGATTCC ATTCCTGTGCGCCAATATTTACACCAGCGGCATCAATCAGCTGGTCGATGTGGACAT CGACAAAATTAATAAGCCATACCTGCCGTTGGTTTCGGGTGAGTTTAGTCTGGGTGA AGGCCGCGCCATTGTGTCGGCCCTTGCGTTTATGTGCCTTGCGGTGGGCATCCTTAGT CACAGCACGCCACTGTTTGTGGCGTGCTGGTTTACTTCCTGATCGGTACTGCATACA GTGTGGAACCTGCCGTTACTGCGTTGGAAAACTAAACCGGCGATGGCGGCGTTTAGCA TGGCAGGTTTGATGGGCCTGACGATTCAACCAGCAGTGTGTTTATCACATTCAAAACG CCCTGGGCAAACCGATGGTTTTTTTCCAAGACCGTCGCTTTGCAACGATTTTCTTCAG</p>

CGTGTTCGCGCGGTCCTTGGTGCAATTAAGATGTCCCGGACGTTGAGGGTGATAC
 TGC GTTTGGTAACCGTACGTTTAGCGTGCGTTATGGTCAGGAGAAAGTTTTAGTGTT
 TGTCTGAACATTCTCCTCTTAGCCTACGGCTTTGCAGTCGTGGTCGGCGCGAGTAGCA
 GCTTCCTGTTATGCAAAATCGTGAGCGTTATGGGCCATACGACGTTAGCGTCGTTGTT
 ACTTCTGCGCGCCAAAAGTACGAATCCAAAAGATCCGGAGAGCACCCAGAGTTTCT
 ACATGTTTCTGTTCAAACCTTCTCTATGCAGAGTATGTC
 CTGATCCACTTCATGCGTTAA

PS
 (Ammi
 majus)

ATGAAAATGCTTGAACAGAACCCGCAGTACCTGTACTTTTTTTCCTTGTTTCTGGTTA
 CCATCTTTCTGTATAAATGGCTGACTCTCAAGAAAACCTCCGCTGAAAAATTTACCG
 CCGAGCCCCGCCGAATACCCGATTATCGGTAATCTGCACCAGATTGGTCCGGACCC
 GCAAGCATCCTTACGTGATTTGGCCAGAAAGTATGGTCCGCTGATGTTCTGAAATT
 TGGCACGGTGCCGTTCTGGTTGTTTCGTCCGCGGACGCGGCACGCGAGGCGCTGA
 AAACCCACGATCTGGTGTTTGC GGATCGTCCGTACAGTAGCGTTGCAAAATAAATT
 TTTTATAACGGCAAAGATATGGTTTTTTCGCGCTATACGGAATACTGGCGTCAGGTT
 AAGAGTATTTGCGTTACCCAGCTGTTGAGTAATAAACGCGTTAACAGCTTTCATTAT
 GTGCGTGAGGAAGAAGTTGATCTCCTGGTGAGAATCTGGAGAATAGTCATAGCA
 AAGTGGCGAACTTGACGGAATTACTGATTGAAAGTTACGGGCAATGTCGTGTGTCG
 GTTAGTGTGGGCAGTGGCGATAAAGTCGATAGTTATAAAATTCTTATTCTGGAGAT
 CATGGACATGCTGGGTTATAGCCGCAGCATCGAAGACTTTTTCCCATTACTGGGTTG
 GGTGATTGGCTGACCGGTCTGCGCGGTAAGTGGCGGAAGCGGCGAAAGGTGTG
 GATACGTTCTTGAGGGCGTTTTGAAGGAACACCTGAGCACGACCGGCTCCAAATA
 CAATGATTTTCGTTAGCATTTTACTGGAAATTCAGGAAGCAGATGCGGGCAGCAGTA
 TGGATAACGAGTGCAATTAAGAGCCTGATTTGGGACATGCTGGGCGCGGGTACGGA
 AACCATTAGCACGGCGCTTGAATGGACGCTGGCCGCACTGATCAAGAATCCAGAC
 GCGATGTTTAAACTCCAGAACGAAGTGC GCGAAATCGGCAAAGGCAAAAGCAAA
 ATCAGCGAGGCGGACCTGGTGAAAATGAATTATCTGCAGGCGGTGATGAAGGAAA
 GTATGCGCTTGATTTTACGGCGCCGCTTCTGGTGCCGCGCGAAGCGCGCCAGGAC
 ATCAAATTTATGGGCTATGACATCAGCAGTGGCACCCAAGTCCTGATTAATGCGTG
 GGCAATCGCCCGCGATCCGCTGCTGTGGGACAAACCGGAAGAATTCGCCCCGAA
 CGTTTTTTGAATTCGCCGATCGACTACAAAGGTTTTCACTATGAATTCCTGCCATT
 GGTGCAGGCCCGCGCGGTTGCCAGGCATTAGTTCGCAATGTGCATCAACGAACT
 GGTGTTGCGAACCTGGTGCAAAATTCAATTTTGAGCTGCCAGATGGTAAACGTC
 TGGAAGACCTGGATATGACCGCGGCGAGTGGCATTACTCTTCGCAAAAAATCCCCA
 CTGCTCGTTGTGGCCCGTCCGCACGTGTAA

Table S2. Primers for PCR amplification of the genes of the biosynthetic pathway (forward (FW) and reverse (REV) primers) and sequencing.

Primer name	Primer sequence ¹	Restriction enzyme site
C2'H_pET_FW/ C2'H_pRSF_FW	ACATGCCATGGAAATGCGCTCGACCACTCT	NcoI
C2'H_pET_REV/ C2'H_pRSF_S-tag_REV	TTTTTCTCGAGTTCAATGCGCGCAATG	XhoI
C2'H_pRSF_REV/ C2'H_pRSF_His_REV	AACCCAAGCTTTTATTCAATGCGCGCAAT	HindIII
C2'H_pRSF_His_FW	AACGCGGATCCGATGCGCTCGACCACTCT	BamHI
C2'H_pRSF_S-tag_FW	AAAAAACATATGCGCTCGACCACTCT	NdeI
PcPT_pETDuet_FW	AAAAAACATATGAGCCAGACCTGAT	NdeI
PcPT_pETDuet_REV	AAAAACTCGAGTTAACGCATGAAGTG	XhoI
PcPT_pRSF/pCDF_FW	AAAAAGGATCCGATGAGCCAGA	BamHI
PcPT/Δ48PcPT_pRSF/pCDF/pET28_REV	AAAAAAGCTTTTAAACGCATGAAGTG	HindIII
PcPT_pET28_FW	AAAAACCATGGAAATGAGCCAGACCTGAT	NcoI
Δ48PcPT_pRSF/pCDF_FW	AAAAAGGATCCGATGGCGTGCCACAAAAACGTTGATT	BamHI
Δ48PcPT_pET28_FW	AAAAACCATGGAAATGGCGTGCCACAAAAACGTTGATT	NcoI
PsPT_pET28_FW	AAAAACCATGGAAATGGCGCAGACCAT	NcoI
PsPT_pET28_REV	AAAAAAGCTTACGCATGAAGTGGATCA	HindIII
PsPT_pRSF_FW	AAAAAGGATCCAATGGCGCAGACCATTATG	BamHI
PsPT_pRSF_REV	AAAAAAGCTTTTAAACGCATGAAGTGGATCA	HindIII
Δ48PsPT1_FW	AAAAACATATGGCGTGCGATAAAAAATTTAGACTCCA	NdeI
Δ48PsPT1_REV	AAAAACTCGAGACGCATGAAGTGGATCAGG	XhoI
PS_pRSF_FW	AAAAAGGATCCAATGAAAATGCTTGAACAGAA	BamHI
PS_pRSF_REV	AAAAAAGCTTTTACACGTGCGGACGG	HindIII
PS_pET_FW	AAAAACCATGGAAATGAAAATGCTTGAAC	NcoI
PS_pET_REV	AAAAAAGCTTCACGTGCGGACGGG	HindIII
MA-Δ37PS_FW	AAAAACATATGGCGCCGCCGAGCCCCGCCG	NdeI
MA/8RP/28tag/2C3-Δ37PS_REV	AAAAACTCGAGCACGTGCGGACGGGCC	XhoI
8RP-Δ37PS_FW	AAAAACATATGGCTCTGTTATTAGCAGTTTTTCCGCCGAGCCCCCG	NdeI
28tag-Δ37PS_FW	AAAAACATATGGAATTATCACAAGTTTGTACAAAAAAGGAGGCTGGCGCCGGCACCAATTCAGTGGACTGGATTCAAGAAGGCGATATTACCCGCCGAGCCCCGCCG	NdeI
2C3-Δ37PS_FW	AAAAACATATGGCTAAAAAAACCAGCAGCAAAGGTAAACCGCCGAGCCCCGCC	NdeI
GroES_FW	AAAAAGGATCCGATGAATATTCGTCCATTGCATGATCGC	BamHI
GroEL_REV	AAAAAGTCGACTTACATCATGCCGCCCATGCG	SalI
MS_FW	AAAAACATATGGATATTTTACCTCCTTAC	NdeI
MS_REV	AAAAACTCGAGCTAATGCTTTGTGCGGCAC	XhoI
8RP-Δ25MS_FW	AAAAACATATGGCTCTGTTATTAGCAGTTTTTTTCTCTAAACACAAAAGGC	NdeI
Seq_MCS1_FW	GGATCTCGACGCTCTCCCTT	-
Seq_MCS2_FW	GTACACGGCCGCATAATCG	-
Seq_REV	CTAGTTATTGCTCAGCGGT	-

¹ Start and stop codons in **bold**; restriction sites in *italic*; N-terminal modifications are underlined. In order for the sequence to remain in frame one or two bases were occasionally added between the restriction site and the gene start codon.

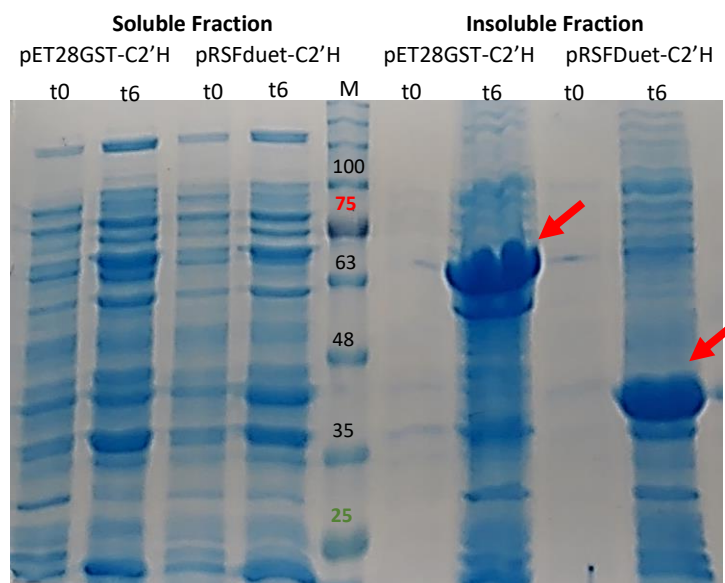


Figure S1. Protein SDS gels showing *p*-coumaroyl-CoA 2'-hydroxylase (C2'H) expression using different plasmids (pRSFDuet-C2'H and pET28GST-C2'H) at time zero (t0) of induction and after 6 h (t6) of induction. C2'H is expected around 40.41 kDa and GST+C2'H at 67.65 kDa. Arrows indicate where it is possible to observe the bands of interest. M: marker (NZYColour Protein Marker II – NZYTech).

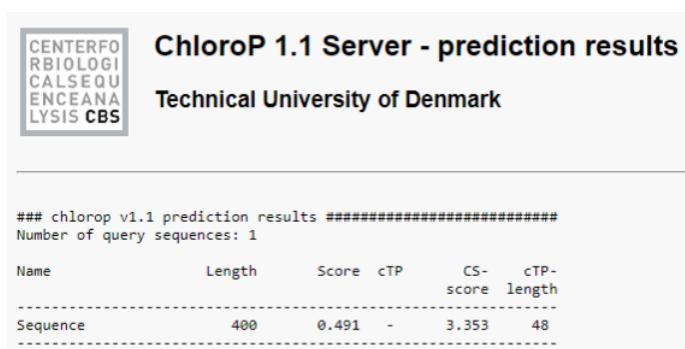


Figure S2. ChloroP results for PT from *Petroselinum crispum*.

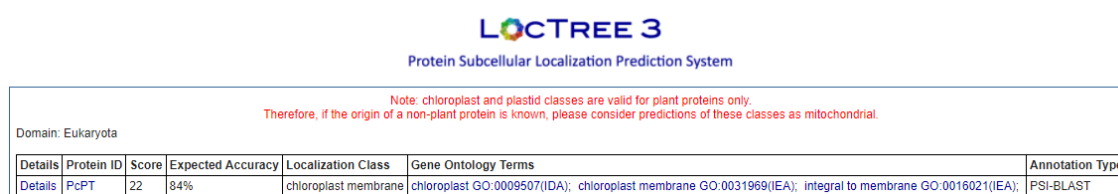


Figure S3. LocTree3 results for PT from *Petroselinum crispum*.

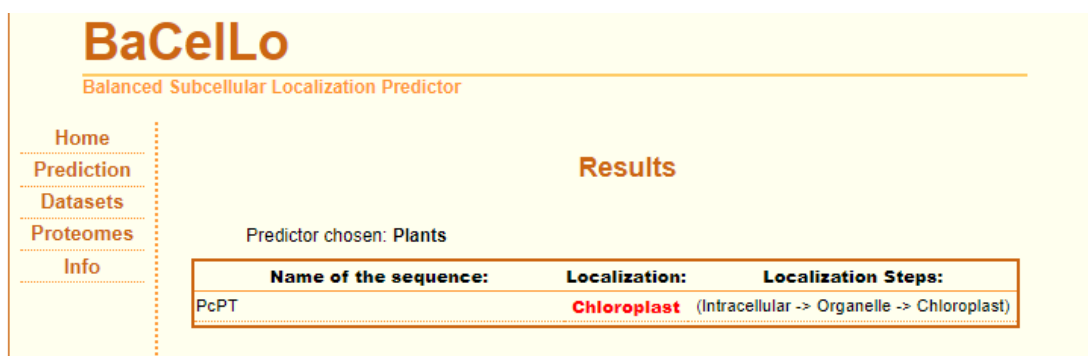


Figure S4. BaCellO results for PT from *Petroselinum crispum*.

iPSORT Prediction

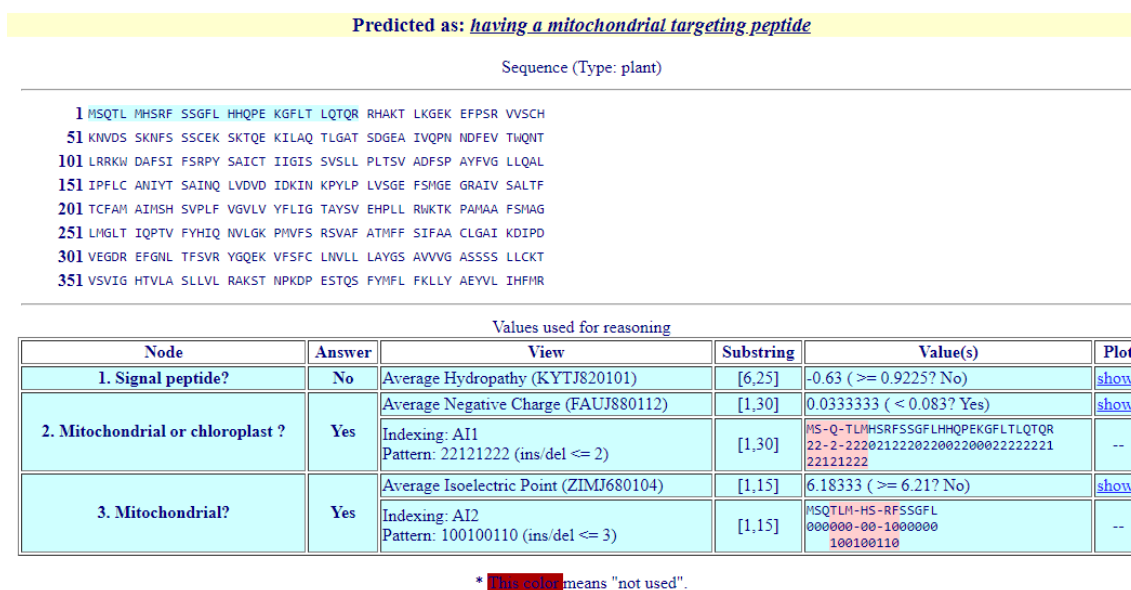


Figure S5. iPSORT results for PT from *Petroselinum crispum*.

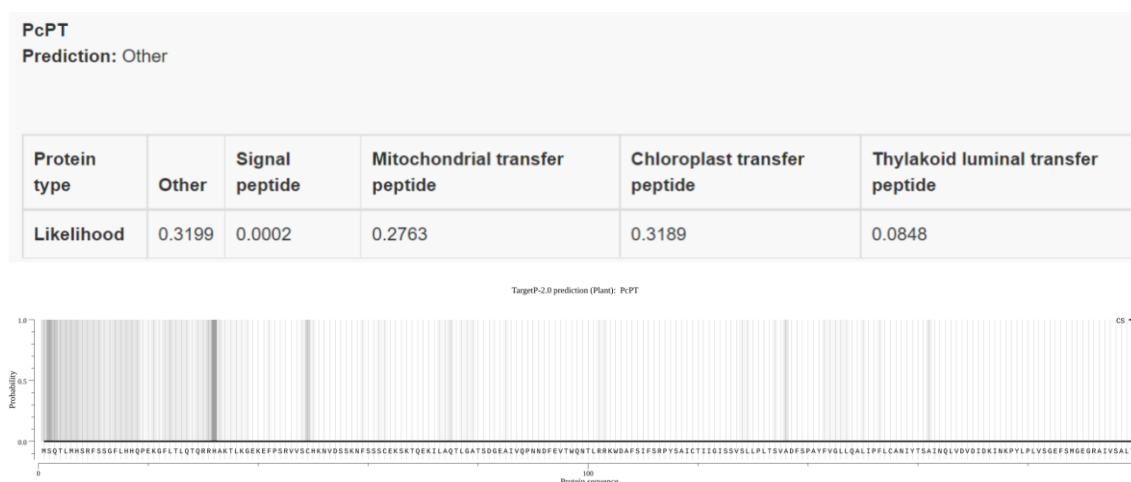


Figure S6. TargetP-2.0 results for PT from *Petroselinum crispum*.

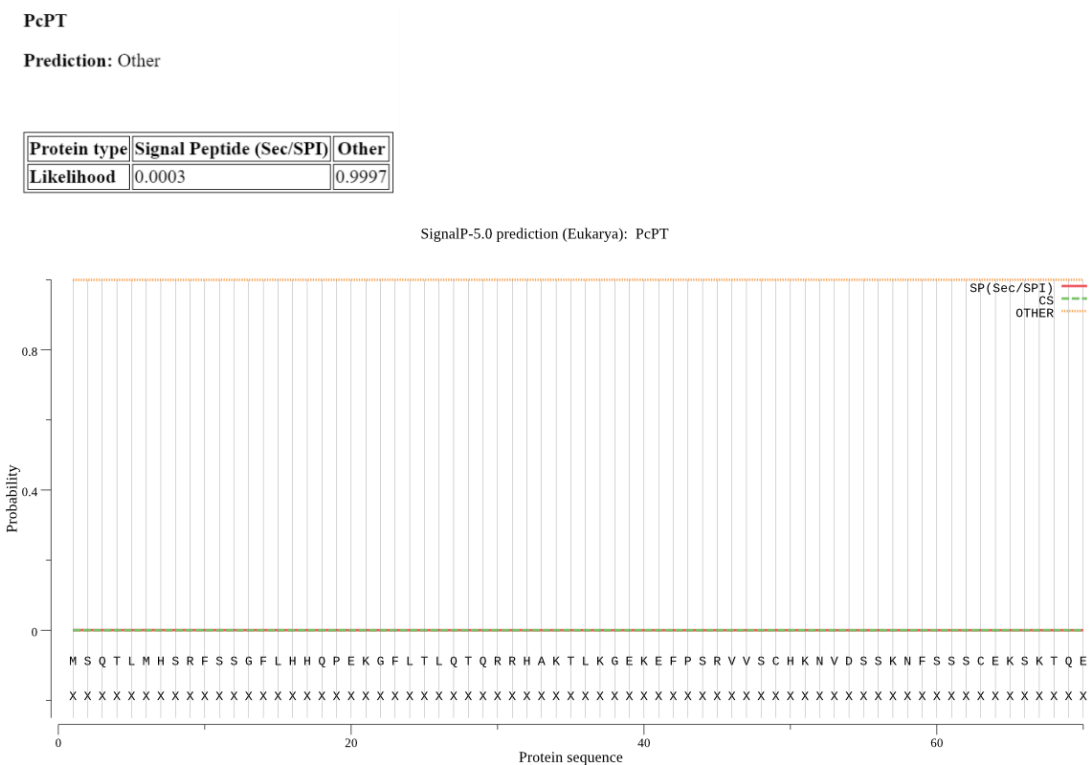


Figure S7. SignalP-5.0 results for PT from *Petroselinum crispum*.

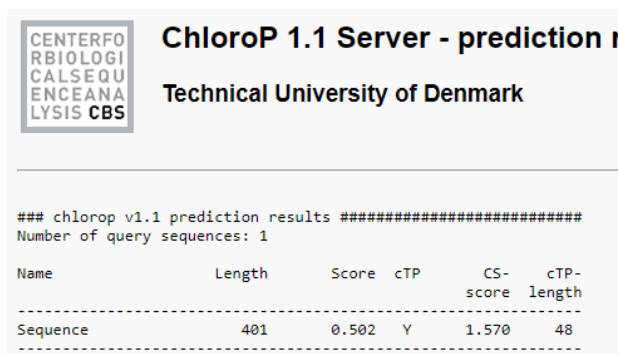


Figure S8. ChloroP results for PT1 from *Pastinaca sativa*.

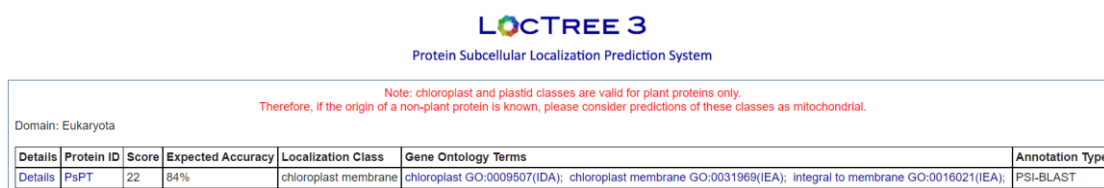


Figure S9. LocTree3 results for PT1 from *Pastinaca sativa*.

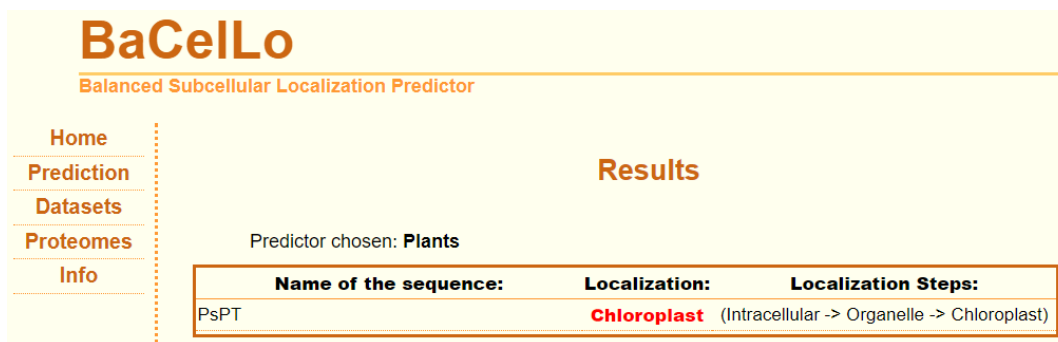


Figure S10. BaCelLo results for PT1 from *Pastinaca sativa*.

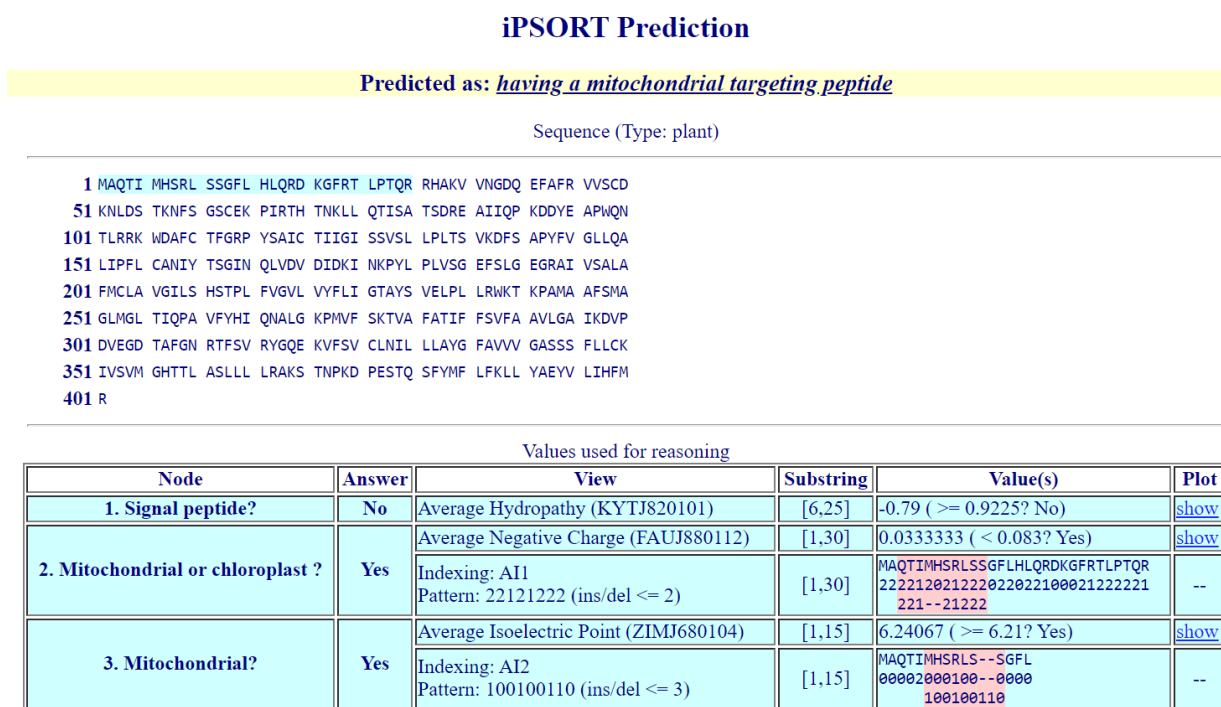


Figure S11. iPSORT results for PT1 from *Pastinaca sativa*.

Sequence

Prediction: Mitochondrial transfer peptide

CS pos: 32-33. RRH-AK. Pr: 0.7081

Protein type	Other	Signal peptide	Mitochondrial transfer peptide	Chloroplast transfer peptide	Thylakoid luminal transfer peptide
Likelihood	0.3569	0.001	0.5376	0.0989	0.0056

TargetP-2.0 prediction (Plant): Sequence

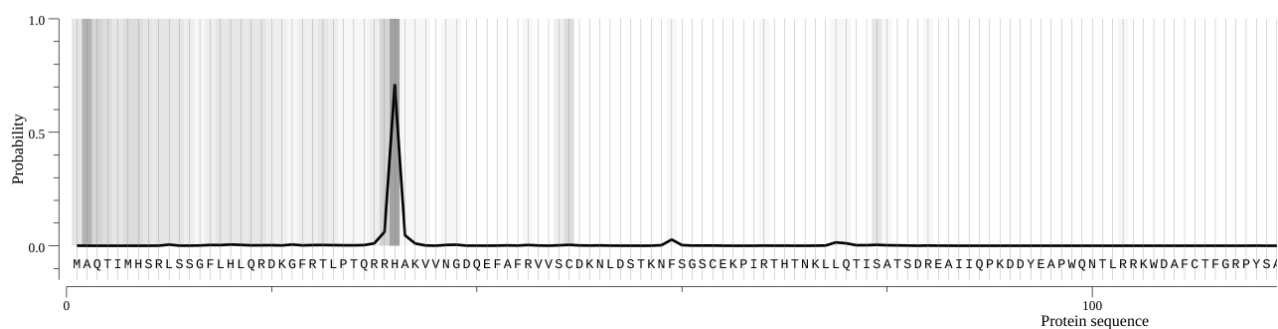


Figure S12. TargetP results for PT1 from *Pastinaca sativa*.

Prediction: Other

Protein type	Signal Peptide (Sec/SPI)	Other
Likelihood	0.0006	0.9994

SignalP-5.0 prediction (Eukarya): Sequence

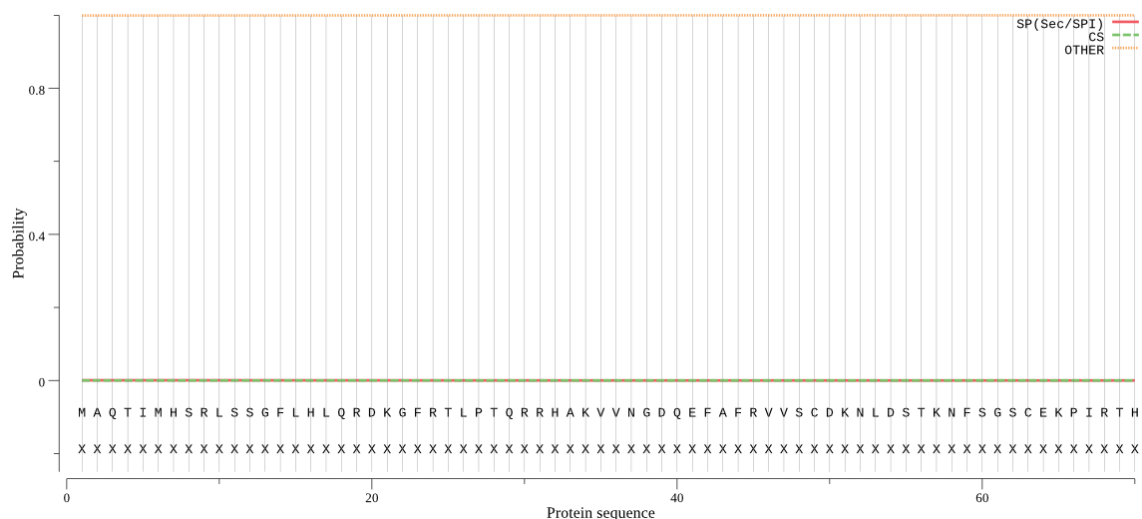


Figure S13. SignalP-5.0 results for PT1 from *Pastinaca sativa*.

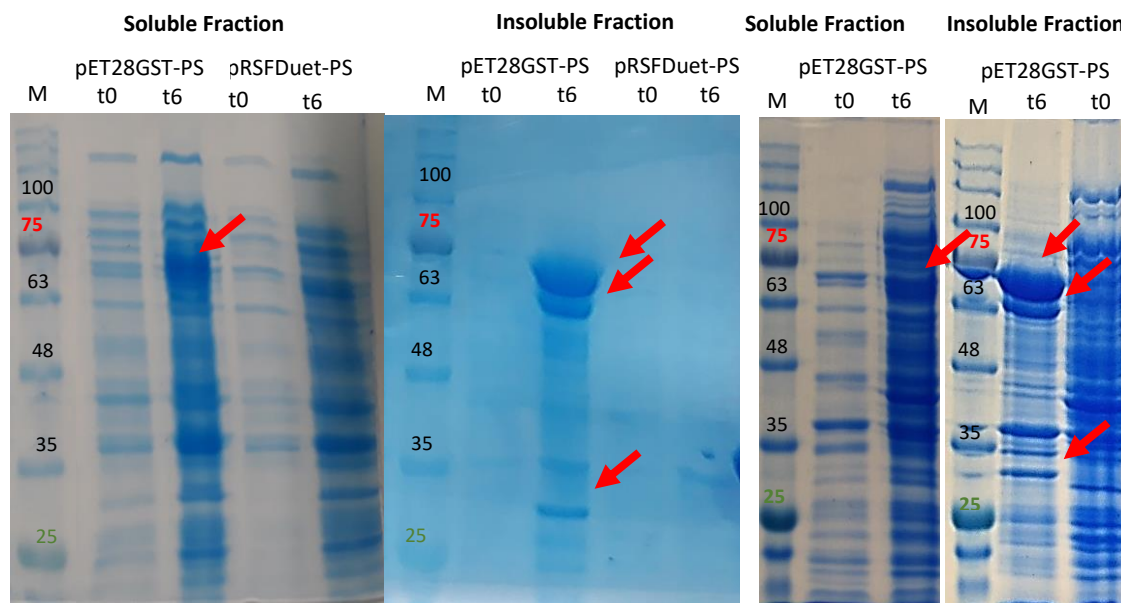


Figure S14. Protein SDS gels showing psoralen synthase (PS) expression using different plasmids (pRSFDuet-PS and pET28GST-PS) at time zero (t0) of induction and after 6 h (t6) of induction. PS is expected around 55.9 kDa and GST+PS at 82.08 kDa. Arrows indicate where it is possible to observe the bands of interest. M: marker (NZYColour Protein Marker II – NZYTech).

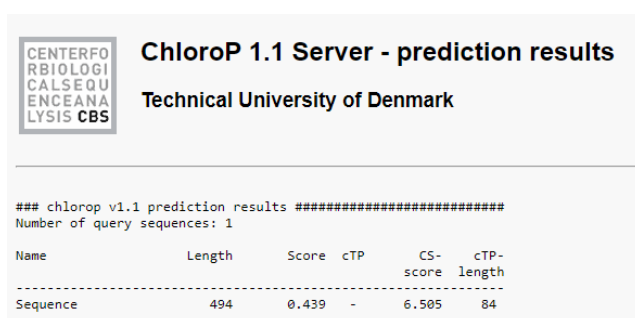


Figure S15. ChloroP results for PS from *Ammi majus*.

iPSORT Prediction

Predicted as: *having a mitochondrial targeting peptide*

Sequence (Type: plant)

```
1 MKMLE QNPQY LYFFS LFLVT IFLYK MLTLK KTPLK NLPPS PPQYP IIGNL
51 HQIGP DPQAS LRDLA QKYGK LMFLK FGTVP VLVVS SADAA REALK THDLV
101 FADRP YSSVA NKIFY NGKDM VFARY TEYHR QVKSI CVTQL LSNKR VNSFH
151 YVREE EVDLL VQNLE NSHSK VANLT ELLIE VTGNV VCRVS VSGSD KVDSE
201 KILIL EIMDM LGYSR SIEDF FPLLG WVDWL TGLRG KVAEA AKGVD TFLEG
251 VLKEH LSTTG SKYND FVSIL LEIQE ADAGS SMDNE CIKSL INDMML GAGTE
301 TISTA LEWTL AALIK NPDAM FKLQN EVREI GKGKS KISEA DLVKM NYLQA
351 VMKES MRLVF TAPLL VPREA RQDIK FMGYD ISSGT QVLIN AWATA RDPLL
401 WDKPE EFRPE RFLNS PIDYK GFHYE FLPGF AGRRG CPGIQ FAMCI NELVV
451 ANLVH KFNFE LPDGK RLEDL DMTAA SGITL RKKSP LLVVA RPHV
```

Values used for reasoning

Node	Answer	View	Substring	Value(s)	Plot
1. Signal peptide?	No	Average Hydropathy (KYTJ820101)	[6,25]	0.685 (>= 0.9225? No)	show
2. Mitochondrial or chloroplast ?	Yes	Average Negative Charge (FAUJ880112)	[1,30]	0.0333333 (< 0.083? Yes)	show
		Indexing: AI1 Pattern: 22121222 (ins/del <= 2)	[1,30]	MKMLEQNPQYLYFFSLFLV-TIFLYKMLTLK 202202022222222222-21222022220 22121222	--
3. Mitochondrial?	Yes	Average Isoelectric Point (ZIMJ680104)	[1,15]	5.82467 (>= 6.21? No)	show
		Indexing: AI2 Pattern: 100100110 (ins/del <= 3)	[1,15]	MKMLEQ--NPQYLYFFS 0100-00--020000000 100100110	--

Figure S16. iPSORT results for PS from *Ammi majus*.

TMHMM result

```
# PS Length: 494
# PS Number of predicted TMHs: 1
# PS Exp number of AAs in TMHs: 21.929769999999999
# PS Exp number, first 60 AAs: 19.06224
# PS Total prob of N-in: 0.80603
# PS POSSIBLE N-term signal sequence
PS TMHMM2.0 inside 1 11
PS TMHMM2.0 TMhelix 12 29
PS TMHMM2.0 outside 30 494
```

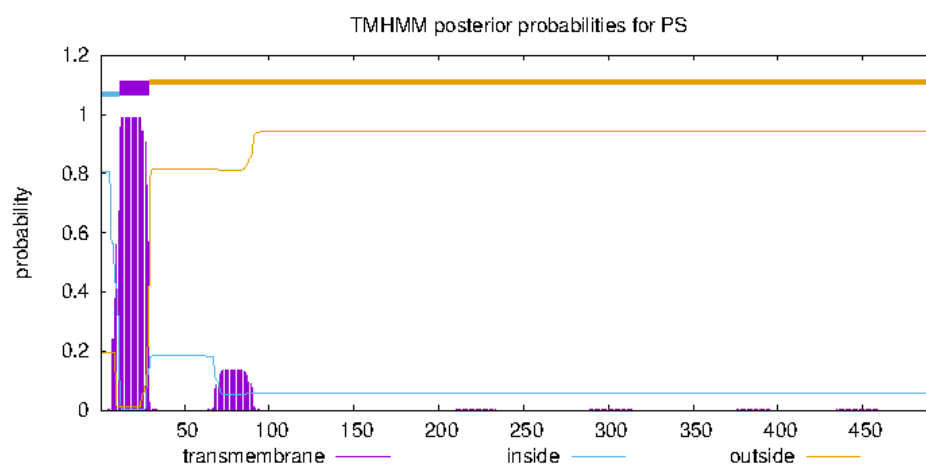


Figure S17. TMHMM-2.0 results for PS from *Ammi majus*.

LOCtree 3 Protein Subcellular Localization Prediction System

Domain: Eukaryota

Details	Protein ID	Score	Expected Accuracy	Localization Class	Gene Ontology Terms	Annotation Type
Details	PS	25	84%	endoplasmic reticulum	endoplasmic reticulum GO:0005783(IEA);	PSI-BLAST

Figure S18. LocTree3 results for PS from *Ammi majus*.

BaCelLo

Balanced Subcellular Localization Predictor

Home

Prediction

Datasets

Proteomes

Info

Results

Predictor chosen: **Plants**

Name of the sequence:	Localization:	Localization Steps:
PS	Chloroplast	(Intracellular -> Organelle -> Chloroplast)

Figure S19. BaCelLo results for PS from *Ammi majus*.

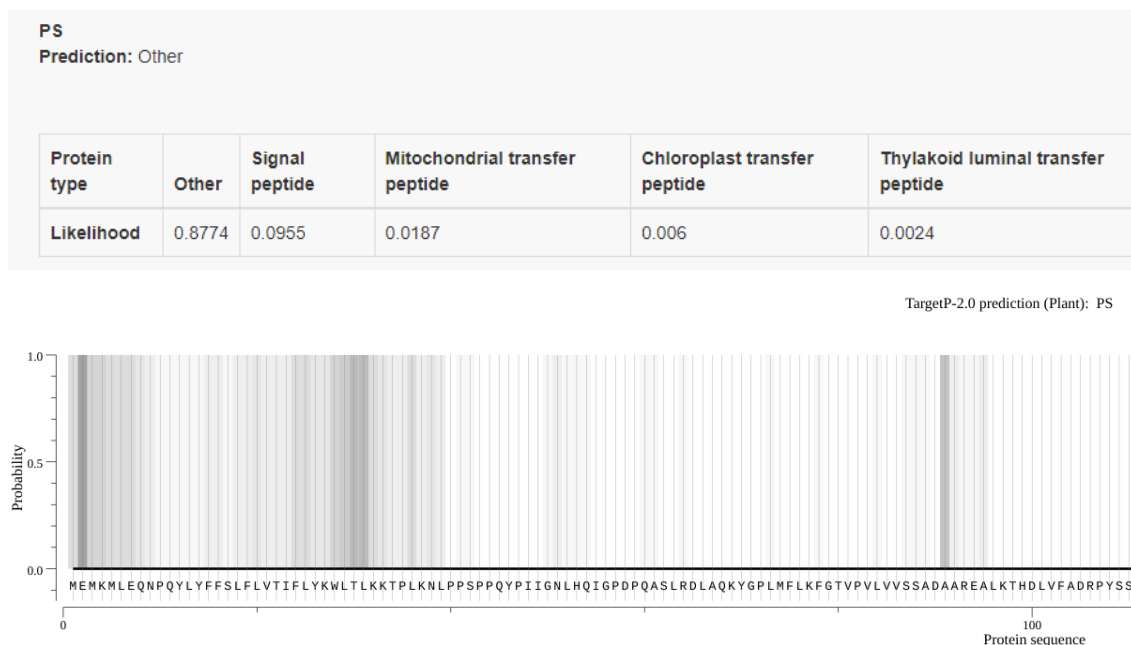


Figure S20. TargetP-2.0 results for PS from *Ammi majus*.

Prediction: Other

Protein type	Signal Peptide (Sec/SPI)	Other
Likelihood	0.006	0.994

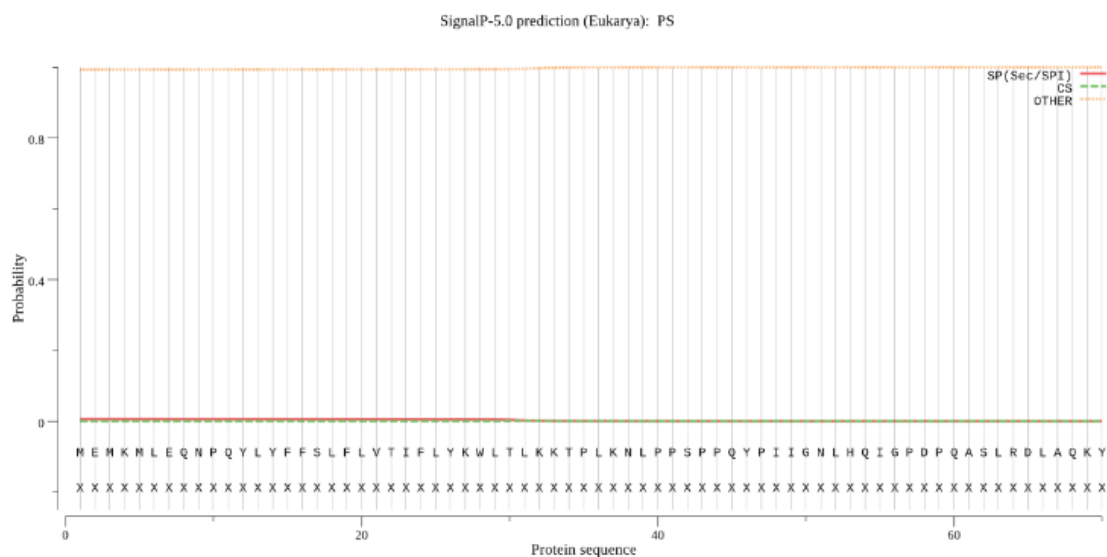


Figure S21. SignalP-5.0 results for PS from *Ammi majus*.

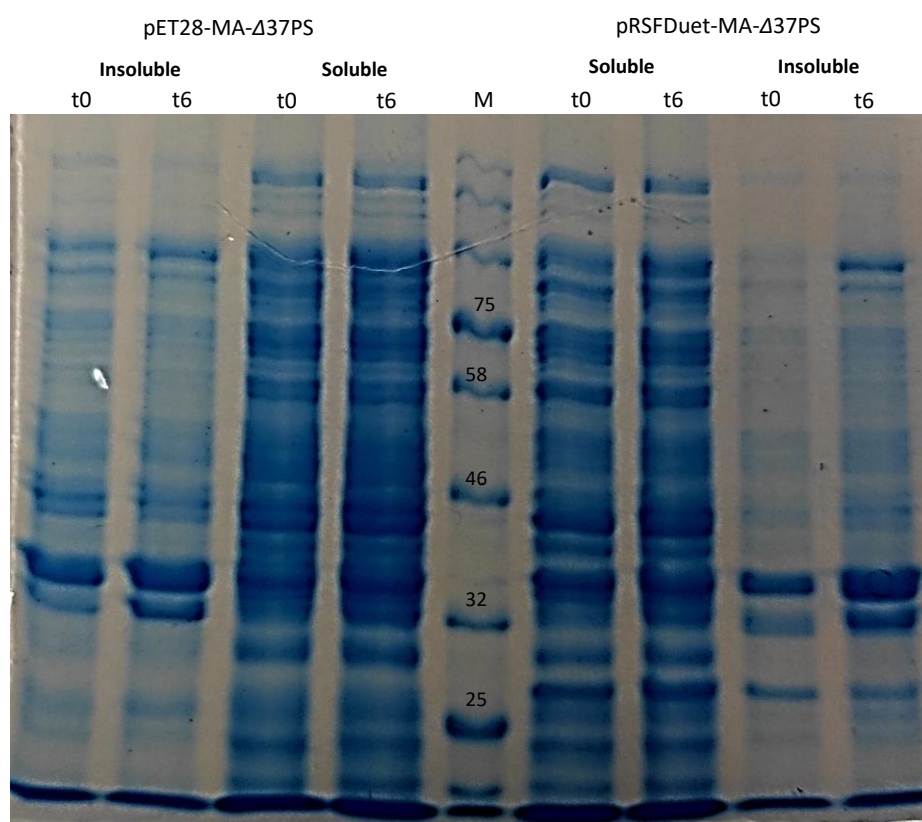


Figure S22. Protein SDS gels showing $\Delta 37$ aa psoralen synthase (PS) expression with MA N-terminal using different plasmids (pRSFDuet-MA- $\Delta 37$ PS and pET28-MA- $\Delta 37$ PS) at time zero (t0) of induction and after 6 h (t6) of induction. PS is expected around 52–54 kDa. M: marker (Blue Prestained Protein Standard, Broad Range - NEB).

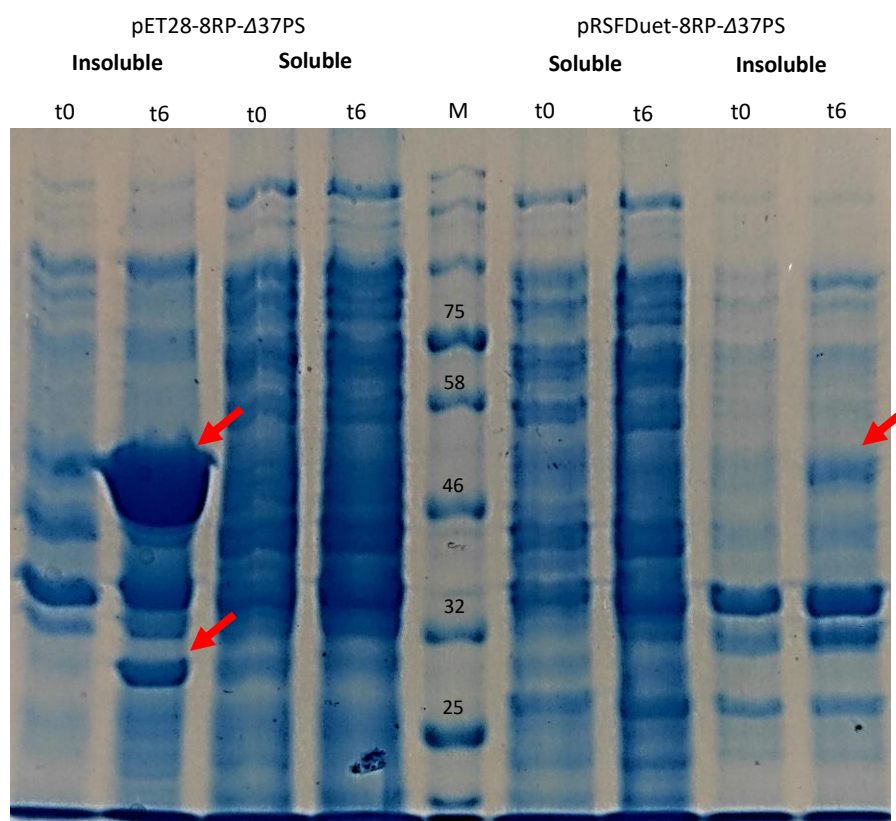


Figure S23. Protein SDS gels showing $\Delta 37$ aa psoralen synthase (PS) expression with 8RP N-terminal using different plasmids (pRSFDuet-8RP- $\Delta 37$ PS and pET28-8RP- $\Delta 37$ PS) at time zero (t0) of induction and after 6 h (t6) of induction. PS is expected around 53-55 kDa. Arrows indicate where it is possible to observe the bands of interest. M: marker (Blue Prestained Protein Standard, Broad Range - NEB).

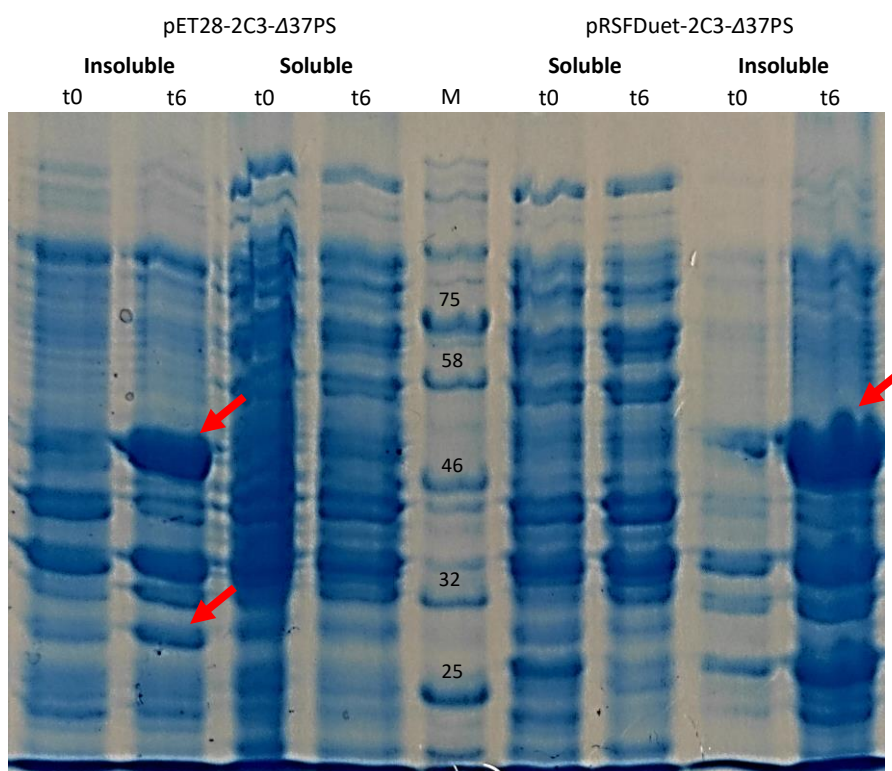


Figure S24. Protein SDS gels showing $\Delta 37$ aa psoralen synthase (PS) expression with 2C3 N-terminal using different plasmids (pRSFDuet-2C3- $\Delta 37$ PS and pET28-2C3- $\Delta 37$ PS) at time zero (t0) of induction and after 6 h (t6) of induction. PS is expected around 53-55 kDa. Arrows indicate where it is possible to observe the bands of interest. M: marker (Blue Prestained Protein Standard, Broad Range - NEB).

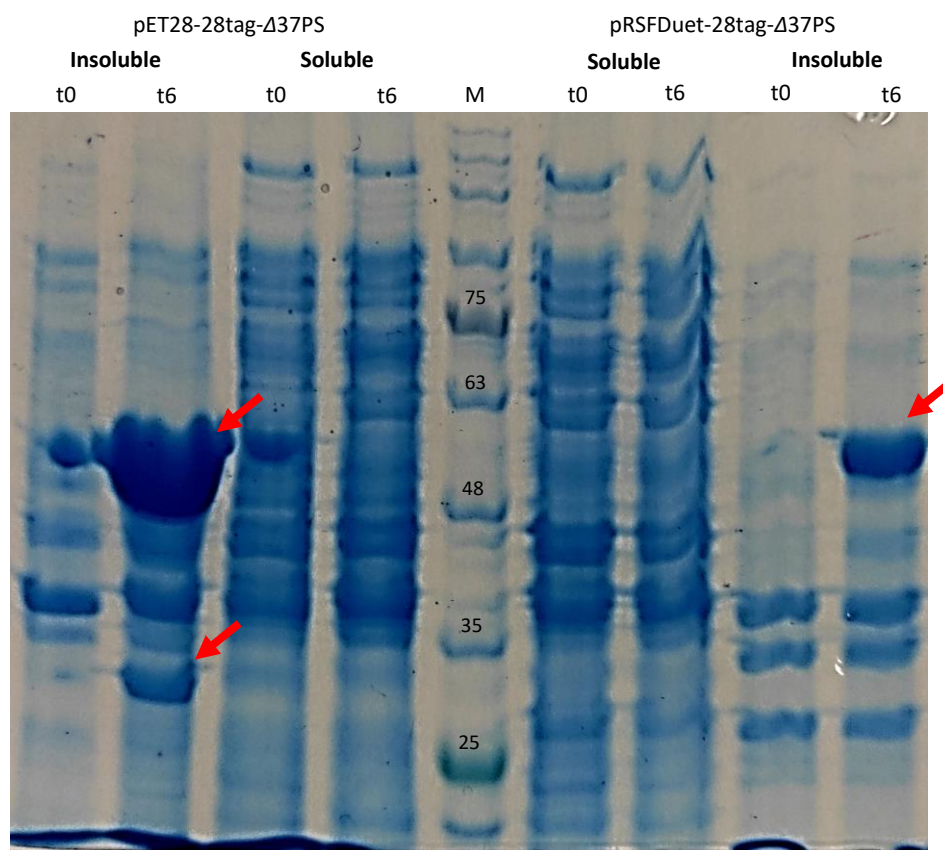


Figure S25. Protein SDS gels showing $\Delta 37$ aa psoralen synthase (PS) expression with 28tag N-terminal using different plasmids (pRSFDuet-28tag- $\Delta 37$ PS and pET28-28tag- $\Delta 37$ PS) at time zero (t0) of induction and after 6 h (t6) of induction. PS is expected around 55-57 kDa. Arrows indicate where it is possible to observe the bands of interest. M: marker (NZYColour Protein Marker II – NZYTech).

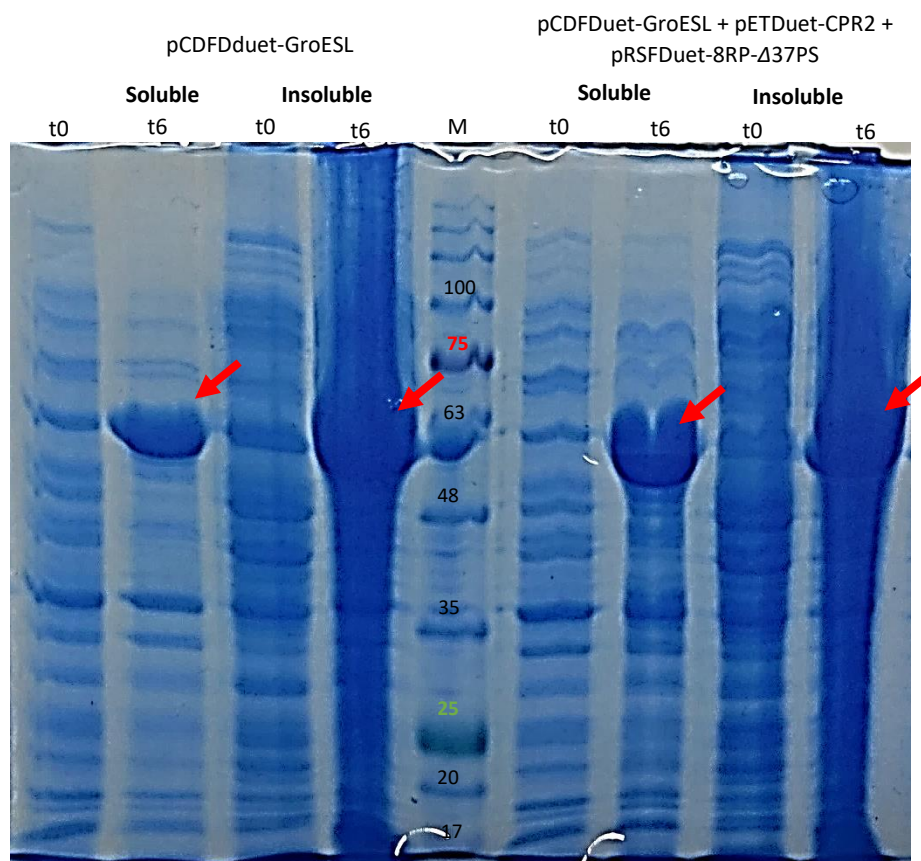


Figure S26. Protein SDS gels showing GroESL expression alone and when combined with cytochrome P450 reductase (CPR2) and 8RP- Δ 37PS at time zero (t0) of induction and after 6 h (t6) of induction. GroEL, GroES, CPR2 and 8RP- Δ 37PS are expected at round 57, 10, 79 and 55 kDa, respectively. Arrows indicate where it is possible to observe the bands of interest. M: marker (NZYColour Protein Marker II – NZYTech).

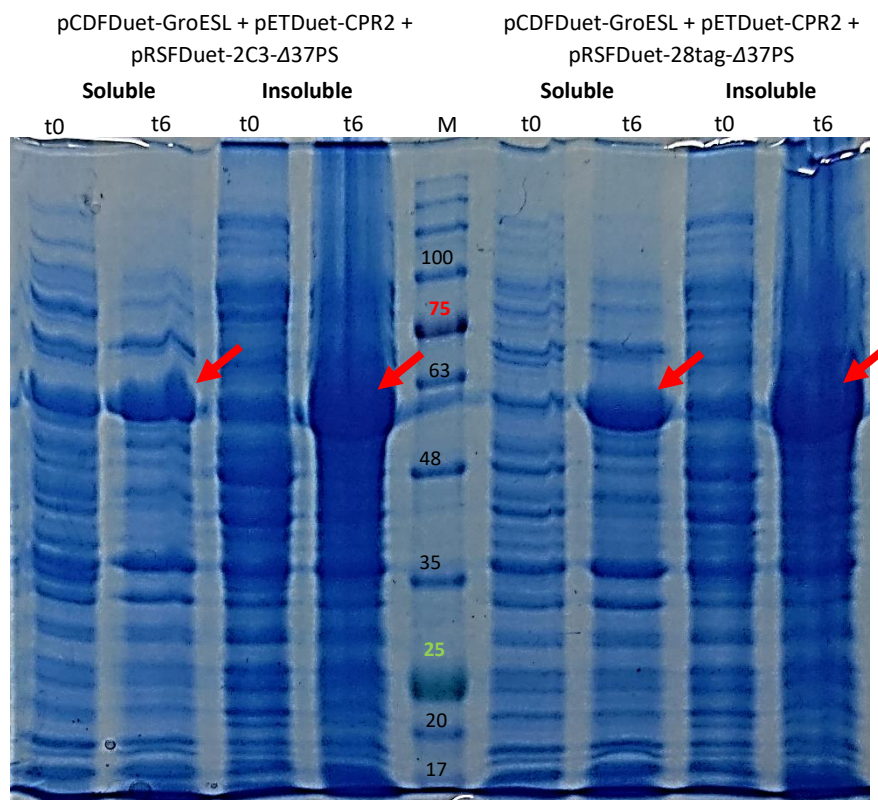


Figure S27. Protein SDS gels showing GroESL expression combined with cytochrome P450 reductase (CPR2) and 2C3-Δ37PS or 28tag-Δ37PS at time zero (t0) of induction and after 6 h (t6) of induction. GroEL, GroES, CPR2, 2C3-Δ37PS and 28tag-Δ37PS are expected at round 57, 10, 79, 55 and 56 kDa, respectively. Arrows indicate where it is possible to observe the bands of interest. M: marker (NZYColour Protein Marker II – NZYTech).

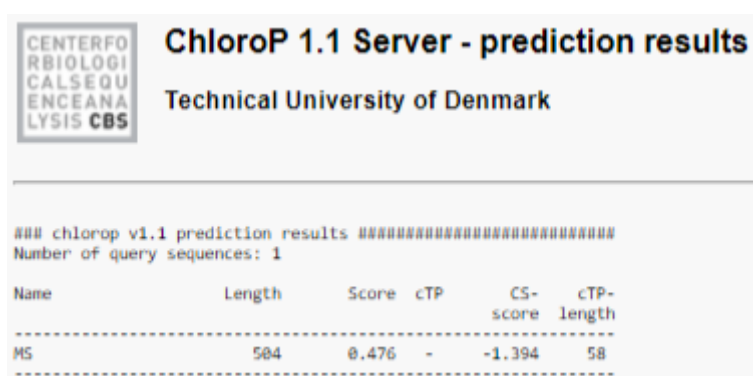


Figure S28. ChloroP results for MS from *Ficus carica*.

LOCTREE 3
Protein Subcellular Localization Prediction System

Domain: Eukaryota

Details	Protein ID	Score	Expected Accuracy	Localization Class	Gene Ontology Terms	Annotation Type
Details	MS	42	89%	endoplasmic reticulum membrane	endoplasmic reticulum membrane GO:0005789(IEA); integral to membrane GO:0016021(IEA);	PSI-BLAST

Figure S29. LocTree3 results for MS from *Ficus carica*.

Home

Prediction

Datasets

Proteomes

Info

BaCellLo

Balanced Subcellular Localization Predictor

Results

Predictor chosen: **Plants**

Name of the sequence:	Localization:	Localization Steps:
MS	Secretory	(Secretory)

Figure S30. BaCelLo results for MS from *Ficus carica*.

iPSORT Prediction

Predicted as: having a signal peptide

Sequence (Type: plant)

1	MSMDI	FTSL	YLALI	LFSL	QVFRS	FAPPK	HKRLP	PGPKP	RPIIG	SLLLE
51	GDQPH	RSLAR	LSESY	GPFMH	LKLQG	VTTVV	ISSTT	MAKEV	LQANS	QVSS
101	RTITD	ASRAH	RHSDF	SNMVL	PVSP	WRNLR	KISNS	HLSS	KALDG	NMELR
151	NKKVK	ELLND	VKHSV	QAGEA	VEIAS	LSFRA	TLNLL	STTF	SMDMA	DDTNS
201	VTLKE	LGKHE	SHMME	ELKAP	NLADY	FPFLQ	KIDQP	GIRRR	NTVTF	RKLIN
251	LFGR	IDQRL	KVREA	SGSLK	DDDLT	DTLIN	MMVVD	QEKKE	QDLQK	TIIEH
301	FLDL	FSAGT	ETTST	TLEWA	MAELV	KAPEI	MSKAR	AELDQ	VIGKG	NQVKE
351	SDVSR	LPVLQ	AIVKE	TFRIH	PTAPL	LIPRK	ADSDI	EISDY	IIPDK	AQVTV
401	NWNAI	GRDSS	TWENP	DKFIP	ERFLD	IDIDV	GGRFD	KLIPF	GAGRR	ICPGF
451	PLAMR	HLMML	LGSL	HSFWD	KLEDG	VRPDA	NWIDE	KFGLT	LQMAQ	PLRAI
501	PVPTK	H								

Values used for reasoning					
Node	Answer	View	Substring	Value(s)	Plot
1. Signal peptide?	Yes	Average Hydropathy (KYTJ820101)	[6,25]	1.605 (>= 0.9225? Yes)	show
2. Mitochondrial or chloroplast ?	No	Average Negative Charge (FAUJ80112)	[1,30]	0.0333333 (< 0.083? Yes)	show
		Indexing: AI1 Pattern: 22121222 (ins/del <= 2)	[1,30]	HWDIPTSLLYLA-LILFFLQVFRSFAFK 222012222222-212222222122220 22121222	--
3. Mitochondrial?	No	Average Isoelectric Point (ZIMJ680104)	[1,15]	5.62467 (>= 6.21? No)	show
		Indexing: AI2 Pattern: 100100110 (ins/del <= 3)	[1,15]	HWDIPTSLLYLAL 000020000000002 NOMATCH	--

Figure S31. iPSORT results for MS from *Ficus carica*.

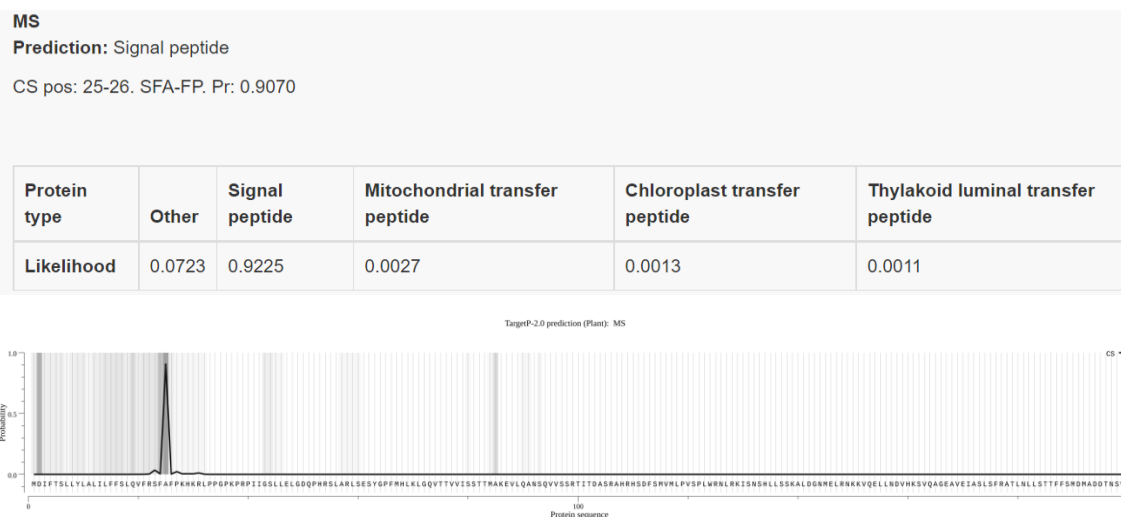


Figure S32. TargetP-2.0 results for MS from *Ficus carica*.

Prediction: Signal peptide (Sec/SPI)
 Cleavage site between pos. 25 and 26: SFA-FP. Probability: 0.535'

Protein type	Signal Peptide (Sec/SPI)	Other
Likelihood	0.6097	0.3903

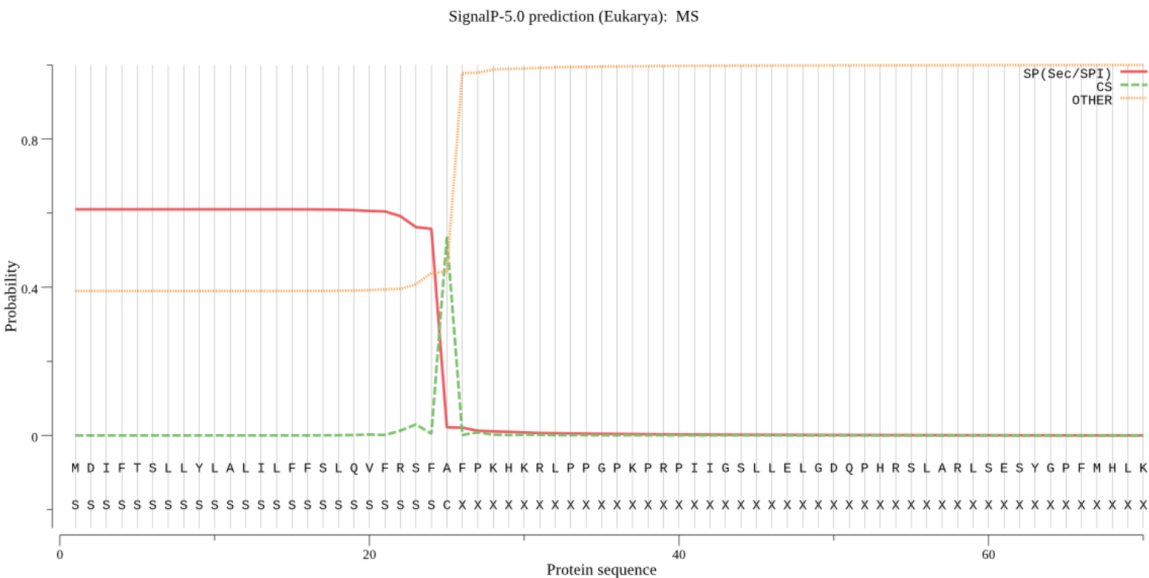


Figure S33. SignalP-5.0 results for MS from *Ficus carica*.

TMHMM result

```
# MS Length: 504
# MS Number of predicted TMHs: 1
# MS Exp number of AAs in TMHs: 19.092809999999999
# MS Exp number, first 60 AAs: 19.04846
# MS Total prob of N-in: 0.95858
# MS POSSIBLE N-term signal sequence
MS   TMHMM2.0   inside    1    6
MS   TMHMM2.0   TMhelix   7   26
MS   TMHMM2.0   outside   27  504
```

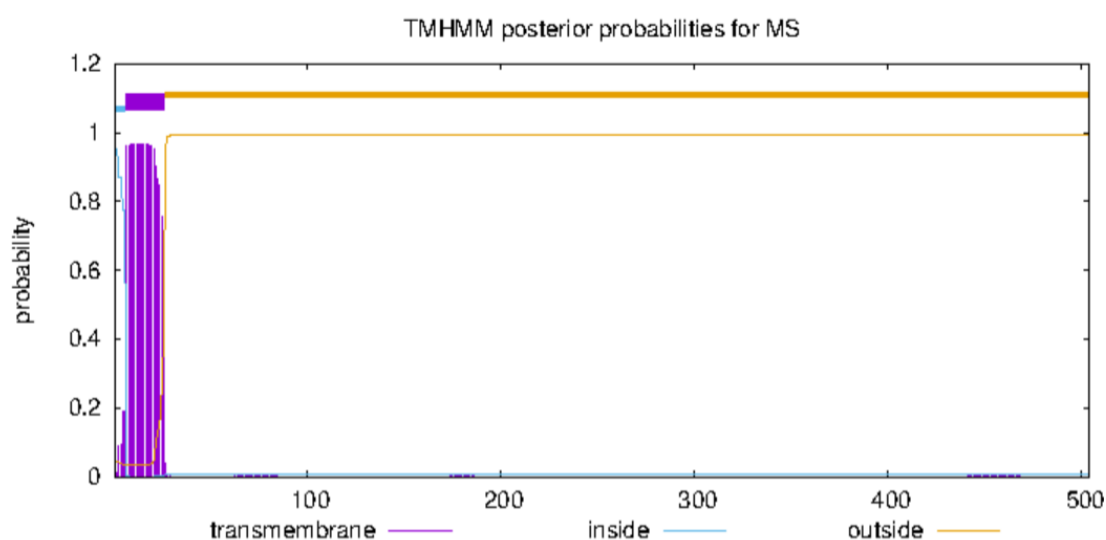


Figure S34. TMHMM-2.0 results for MS from *Ficus carica*.