

Volume 73 (2017)

Supporting information for article:

Structure of Rubisco from *Arabidopsis thaliana* in complex with 2-carboxyarabinitol-1,5-bisphosphate

Karin Valegård, Dirk Hasse, Inger Andersson and Laura H. Gunn

## **Table S1** The location and nature of amino acid differences between the A.*thaliana* Rubisco SSu isoforms

Residue positions that differ between the isoforms are only included for the mature SSu peptide. See Fig. 2 and Fig. S1 for descriptions of SSu and LSu secondary structure features, respectively.

	Amino acid		Location				
Residue #	RbcS3	RbcS1A	In SSu	In holoenzyme structure	Proximal to	Nature of change	
2	Lys	Gln	N-terminus	Solvent exposed, packs against LSu	LSu helix α8	Charged/uncharged residue	
					LSu helix αK	Potential ionic influence on LSu Glu454 ir RbcS3 SSus	
22	Thr/Ser	Thr	N-terminal to helix αA	Solvent exposed		Minor difference in side chain	
24	Val	Ser	Helix αA	Solvent exposed	SSu helix αA is near LSu helix α8	Hydrophobic/polar	
34	Leu	lle	Helix αA	Solvent exposed		Minor difference in side chain	
58	Thr	Ser	Apex of βA-βB loop	Solvent channel	LSu helix α3 in two different LSus	Minor difference in side chain	
96	Gly	Asn	Loop αB-βC	Solvent exposed	SSu C-tail	Glycine / polar side chain	
					Start of SSu βA-βB loop		
124	Asp/Glu	Gly	C-terminus	Not resolved in structure		Charged /uncharged residue	
125	Ala	-	C-terminus	Not resolved in structure		Absent in RbcS1A	

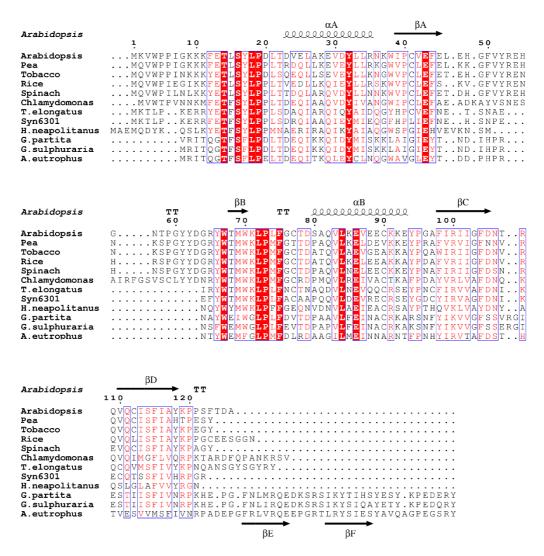
Arabidopsis				тт	βA <b>ττ</b> 22
Arabidopsis Pea Tobacco Rice Spinach Chlamydomonas T.elongatus Syn6301 H.neapolitanus G.partita G.sulphuraria A.eutrophus	MSPQTET MSPQTET MSPQTET MSPQTET MSPQTET MVPQTET MAYTQSK	KAKVGFKAGV KASVGFKAGV KASVGFKAGV KASVEFKAGV SQKVGYQAGV SQKVGYQAGV SAAGYKAGV IKNSRYESGV IKNSRYESGV	KDYKL.TYYT KEYKL.TYYT KDYKL.TYYT KDYKL.TYYT KDYRL.TYYT KDYRL.TYYT KDYKL.TYYT KEYRQ.TYWM IPYAKMGYWN IPYAKMGYWN	P D YQ TKD TD I L P E YQ TKD TD I L P E YE TKD TD I L P E YE T LD TD I L P D Y TV VRD TD I L P D Y T VRD TD I L P D Y T P KD TD I L P E YT P LD SD I L P D YQ V KD TD V L P D YQ V KD TD V L	AAFRVTPOPGVPPE AAFRVSPOPGVPPE AAFRVTPOPGVPFE AAFRVFPOPGVPFE AAFRFSPOPGVPAD ACFKITPOPGVDRE ALFRVTPOPGVDPI ALFRVTPOPGVDPI
Arabidopsis	αB 222222222 60	ηΑ <u>00000</u> 70	αC βΕ	TT	$\begin{array}{ccc} \beta C & \eta B \\ & & 2 \\ 1 \\ 0 \\ 0 \\ \end{array} $
Arabidopsis Pea Tobacco Rice Spinach Chlamydomonas T.elongatus Syn6301 H.neapolitanus G.partita G.sulphuraria A.eutrophus	EAGAAVAAESSTGTW EAGAAVAAESSTGTW EAGAAVAAESSTGTW EAGAAVAAESSTGTW EAGAAVAAESSTGTW EAGAAVAAESSTGTW EAAAAVAAESSTGTW EAAAAVAAESSTGTW EAAAAVAAESSTGTW EAAAAVAAESSTATW	TTVWTDGLTS TTVWTDGLTS TTVWTDGLTS TTVWTDGLTS TTVWTDGLTS TTVWTDGLTS TTVWTDLLTD TTVWTDLLTD TTVWTDLLTD TTVWTDLLTD	LDRYKGRCYH LDRYKGRCYER LDRYKGRCYH LDRYKGRCYH LDRYKGRCYH LDRYKGRCYH UDRYKGRCYH MDRYKGKCYH MDRYKGRAYR MDLYRAKAYK ADLYRAKAYK	$\begin{split} \overrightarrow{I} & \overrightarrow{E} \overrightarrow{P} \overrightarrow{V} \overrightarrow{P} \overrightarrow{G} \overrightarrow{E} \overrightarrow{E} \overrightarrow{T} \overrightarrow{O} \overrightarrow{F} \\ \overrightarrow{I} & \overrightarrow{E} \overrightarrow{P} \overrightarrow{V} \overrightarrow{P} \overrightarrow{G} \overrightarrow{E} \overrightarrow{D} \overrightarrow{N} \overrightarrow{O} \overrightarrow{F} \\ \overrightarrow{I} & \overrightarrow{E} \overrightarrow{V} \overrightarrow{V} \overrightarrow{G} \overrightarrow{E} \overrightarrow{D} \overrightarrow{N} \overrightarrow{O} \overrightarrow{Y} \\ \overrightarrow{I} & \overrightarrow{E} \overrightarrow{P} \overrightarrow{V} \overrightarrow{P} \overrightarrow{G} \overrightarrow{E} \overrightarrow{D} \overrightarrow{N} \overrightarrow{O} \overrightarrow{Y} \\ \overrightarrow{I} & \overrightarrow{E} \overrightarrow{P} \overrightarrow{V} \overrightarrow{P} \overrightarrow{G} \overrightarrow{E} \overrightarrow{D} \overrightarrow{N} \overrightarrow{O} \overrightarrow{Y} \\ \overrightarrow{I} & \overrightarrow{E} \overrightarrow{P} \overrightarrow{V} \overrightarrow{P} \overrightarrow{G} \overrightarrow{E} \overrightarrow{E} \overrightarrow{D} \overrightarrow{O} \overrightarrow{P} \\ \overrightarrow{I} & \overrightarrow{E} \overrightarrow{P} \overrightarrow{V} \overrightarrow{P} \overrightarrow{G} \overrightarrow{E} \overrightarrow{E} \overrightarrow{D} \overrightarrow{O} \overrightarrow{E} \\ \overrightarrow{I} & \overrightarrow{E} \overrightarrow{D} \overrightarrow{V} \overrightarrow{O} \overrightarrow{V} \overrightarrow{P} \overrightarrow{N} \overrightarrow{P} \overrightarrow{O} \overrightarrow{V} \\ \overrightarrow{V} \overrightarrow{D} \overrightarrow{V} \overrightarrow{P} \overrightarrow{N} \overrightarrow{P} \overrightarrow{N} \overrightarrow{P} \overrightarrow{V} \overrightarrow{O} \overrightarrow{V} \overrightarrow{P} \overrightarrow{N} \overrightarrow{P} \overrightarrow{V} \overrightarrow{O} \overrightarrow{V} \overrightarrow{P} \overrightarrow{N} \overrightarrow{P} \overrightarrow{V} \overrightarrow{O} \overrightarrow{V} \overrightarrow{P} \overrightarrow{N} \overrightarrow{P} \overrightarrow{V} \overrightarrow{V} \overrightarrow{O} \overrightarrow{V} \overrightarrow{P} \overrightarrow{N} \overrightarrow{P} \overrightarrow{V} \overrightarrow{V} \overrightarrow{V} \overrightarrow{V} \overrightarrow{V} \overrightarrow{V} \overrightarrow{V} V$	I AYV <b>AY</b> PLD <b>LFEEG</b> I AYV <b>AY</b> X I D <b>LFEEG</b> I AY I AY PLD <b>LFEEG</b> FAFIAY PLD <b>LFEEG</b> FAYIAY FLD <b>LFEEG</b> FAYIAY ELD <b>LFEEG</b>
Arabidopsis	αD ηC 202020202 202 120	βD 130	αE 2020	222	$\begin{array}{ccc} \chi F & & \beta I \\ \underline{0} & \underline{0} & \underline{0} \\ 160 & & 170 \end{array}$
Arabidopsis Pea Tobacco Rice Spinach Chlamydomonas T.elongatus Syn6301 H.neapolitanus G.partita G.sulphuraria A.eutrophus	SVTNMFTSIVGNVFG SVTNMFTSIVGNVFG SVTNMFTSIVGNVFG SVTNMFTSIVGNVFG SVTNMFTSIVGNVFG SVTNMLTSIVGNVFG SVTNMLTSIVGNVFG SVTNILTSIVGNVFG SVVNVFTSLVGNVFG SIANLTASIIGNVFG SIANLTASI	FKALAALRLE FKALRALRLE FKALRALRLE FKALRALRLE FKALRALRLE FKALRALRLE FKALRALRLE FKAVRCLRLE FKAVRCLRLE FKAVKALRLE FKAVKALRLE	DIRTPRAYTK DIRTPYAYVK DIRTPYAYVK DIRTPYAYVK DIRTPYYSK DIRTPYAYVK DIRTPYAYVK DIRTPYAYVK DIRTPYAYVK DIRTPYALVK DWRTPLAYVK DWRTPLAYVK DWRTPTAYIK	$ \mathbf{T} F \bigcirc \mathbf{G} P P H \mathbf{G} I \bigcirc \mathbf{V} \\ \mathbf{T} F \bigcirc \mathbf{G} P P H \mathbf{G} I \bigcirc \mathbf{V} \\ \mathbf{T} F \bigcirc \mathbf{G} P P H \mathbf{G} I \bigcirc \mathbf{V} \\ \mathbf{T} F \bigcirc \mathbf{G} P P H \mathbf{G} I \bigcirc \mathbf{V} \\ \mathbf{T} F \bigcirc \mathbf{G} P P H \mathbf{G} I \bigcirc \mathbf{V} \\ \mathbf{T} F \bigcirc \mathbf{G} P P H \mathbf{G} I \bigcirc \mathbf{V} \\ \mathbf{T} F \bigcirc \mathbf{G} P P H \mathbf{G} I \bigcirc \mathbf{V} \\ \mathbf{T} F \bigcirc \mathbf{G} P P H \mathbf{G} I \bigcirc \mathbf{V} \\ \mathbf{T} F \bigcirc \mathbf{G} P P H \mathbf{G} I \bigcirc \mathbf{V} \\ \mathbf{T} F \bigcirc \mathbf{G} P P H \mathbf{G} I \bigcirc \mathbf{V} \\ \mathbf{T} F \bigcirc \mathbf{G} P P H \mathbf{G} I \bigcirc \mathbf{V} \\ \mathbf{T} F \bigcirc \mathbf{G} P P H \mathbf{G} I \bigcirc \mathbf{V} \\ \mathbf{T} F \bigcirc \mathbf{G} P H \mathbf{G} I \bigcirc \mathbf{V} I \\ \mathbf{T} F \bigcirc \mathbf{G} P H \mathbf{G} I \bigcirc \mathbf{V} I \\ \mathbf{T} F \bigcirc \mathbf{G} P H \mathbf{G} I \bigcirc \mathbf{V} I \\ \mathbf{T} F \bigcirc \mathbf{G} P H \mathbf{G} I \bigcirc \mathbf{V} I \\ \mathbf{T} F \bigcirc \mathbf{G} P H \mathbf{G} I \bigcirc \mathbf{V} I \\ \mathbf{T} F \bigcirc \mathbf{G} P H \mathbf{G} I \bigcirc \mathbf{V} I \\ \mathbf{T} F \bigcirc \mathbf{G} P H \mathbf{G} I \\ \mathbf{G} \cap \mathbf{V} I \\ \mathbf{G} F \\ \mathbf{G} \\ \mathbf{G} F \\ \mathbf{G} \\ \mathbf{G}$	ERDKLNKYGRPLLG ERDKLNKYGRPLLG ERDKLNKYGRPLLG ERDKLNKYGRPLLG ERDKLNKYGRPLLG ERDKLNKYGRPLLG ERDKLNKYGRPLLG ERDLLNKYGRPLLG ERDLLNKYGRPLLG ERERLDKFGRPLLG
Arabidopsis	180	αl 2020200 <b>190</b>	β2 200 <b>TT</b>		α2 00000000000000 220 230
Arabidopsis Pea Tobacco Rice Spinach Chlamydomonas T.elongatus Syn6301 H.neapolitanus G.partita G.sulphuraria A.eutrophus	CTIKPKLGLSAKNYG CTIKPKLGLSAKNYG CTIKPKLGLSAKNYG CTIKPKLGLSAKNYG CTIKPKLGLSAKNYG CTIKPKLGLSAKNYG CTIKPKLGLSAKNYG CTIKPKLGLSAKNYG CTIKPKLGLSGKNYG ATTKPKLGLSGKNYG	RAVYECLRGG RAVYECLRGG RAVYECLRGG RAVYECLRGG RAVYECLRGG RAVYECLRGG RAVYECLRGG RAVYECLRGG RVYYEALKGG	LDFTKDDENV LDFTXDDENV LDFTXDDENV LDFTXDDENV LDFTXDDENV LDFTXDDENI LDFTKDDENI LDFTKDDENI LDFVXDDENI LDFVKDDENI	N S Q P F M R W R D R N S Q P F M R W R D R N S Q P F M R W R D R N S Q P F M R W R D R N S Q P F Q R W R D R N S Q P F Q R W R D R N S Q P F Q R W R D R N S Q P F M R W R D R N S Q P F M R W R B R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R N S Q P F M R R N S Q P F M R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N S Q P F M R R N R N S Q P F M R R N R N R N R N R N R N R N R N R N	FLFCAEALYKSQAE FLFCAEALYKAQAE FLFCAEALYKAQAE FLFCAEALYKAQAE FLFVAEALYKAQAE FLFVADAIHKAQAE FLFVADAIHKAQAE FLFVADAIHKSQAE FLFVQDATETAEAQ YLFTMEAVNKASAA
Arabidopsis	Ω β3 240	α3 200000000 250	<u>β4</u> 260	→2022 2222	α4 0000000000
Arabidopsis Pea Tobacco Rice Spinach Chlamydomonas T.elongatus Syn6301 H.neapolitanus G.partita G.sulphuraria A.eutrophus	TGEIKGHYLNATAGT TGEIKGHYLNATAGT TGEIKGHYLNATAGT TGEIKGHYLNATAGT	CEEMIKRAVF CEEMIKRAVF CEEMIKRAVF CEEMIKRAVF	ARELGVPIVM ARELGVPIVM ARELGVPIVM ARELGVPIVM	HDYLTGGFTAN HDYLTGGFTAN HDYLTGGFTAN HDYLTGGFTAN	ISLSHYCRDNGLLL ITLSHYCRDNGLLL ISLAHYCRDNGLLL ISLAHYCRDNGLLL

1

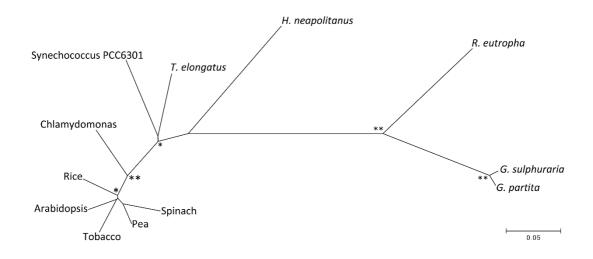
Arabidopsis Pea Tobacco Rice Spinach Chlamydomonas T.elongatus Syn6301 H.neapolitanus G.partita G.sulphuraria A.eutrophus	HIHRAMHAVI HIHRAMHAVI HIHRAMHAVI HIHRAMHAVI HIHRAMHAVI HIHRAMHAVI HIHRAMHAVI HIHRAGNSTY HLHRAGNSTY	310 RQKNHGMHFRV RQKNHGMHFRV RQKNHGIHFRV DRQKNHGIHFRV DRQKNHGIHFRV DRQKNHGIHFRV DRQKNHGIHFRV DRQKNHGIHFRV SRQKNHGMNFRV	LAKALRLSG LAKALRMSG LAKALRMSG LAKALRLSG LAKALRMSG LAKCLRMSG LAKCLRLSG LTKILRLSG I CKWMRMAG	DHIHAGTVVGKL DHIHAGTVVGKL DHIHAGTVVGKL DHIHSGTVVGKL DHIHSGTVVGKL DHIHGTVVGKL DHIHGTVVGKL DHLHGTVVGKL DHLHAGTVVGKL DHIHAGTVVGKL	α6 QQQQQQQQQQQ 340 SORESTLGFVDLLRD EGEREITLGFVDLLRD EGEREDITLGFVDLLRD EGEREVTLGFVDLLRD EGEREVTLGFVDLLRD EGEREVTLGFVDLLRE EGDKASTLGFVDLRE EGDKASTLGFVDLRE EGDFIITRGFYKTLLL EGDPIITRGFYKTLLL EGDPLTVQGYYNVCRD
Arabidopsis Pea Tobacco Rice Spinach Chlamydomonas T.elongatus Syn6301 H.neapolitanus G.partita G.sulphuraria A.eutrophus	DYIKKDRSRG DFVEQDRSRG DYTEKDRSRG DYTEKDRSRG NYIEQDRSRG DHIEADRSRG SFIPEDRSRG PKLERNLQEG PKLERNLQEG	IYFTQDWVSLPG IYFTQDWVSLPG IYFTQDWVSMPG IYFTQDWXSMPG IYFTQDWASMPG VFFTQDWASMPG IFFDQDWGSMPG IFFDQDWGSLF LFFDMEWASLRK LFFDMEWASLRK	VIPVASGGI VLPVASGGI VIPVASGGI VLPVASGGI VMPVASGGI VMPVASGGI VLPVASGGI VMPVASGGI VMPVASGGI	IVWHMPALTEIFG IVWHMPALTEIFG IVWHMPALTEIFG IVWHMPALVEIFG IVWHMPALVEIFG IVWHMPALVEIFG IVWHMPALVNIFG IAGQMHQLIHYLG	β8 αI Δ00 410 DDSVLQFGGGTLGHPW DDSVLQFGGGTLGHPW DDSVLQFGGGTLGHPW DDSVLQFGGGTLGHPW DDAVLQFGGGTLGHPW DDAVLQFGGGTLGHPW DDSVLQFGGGTLGHPW DDSVLQFGGGTLGHPW DDSVLQFGGGTIGHPD DDVVLQFGGGTIGHPD DDVVLQFGGGTIGHPQ
Arabidopsis Pea Tobacco Rice Spinach Chlamydomonas T.elongatus Syn6301 H.neapolitanus G.partita G.sulphuraria A.eutrophus	$\begin{array}{c} 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0$	VALEAC VQARNE VALEAC VQARNE VALEAC VQARNE VALEAC VQARNE VALEAC TQARNE VALEAC VQARNE VALEAC VQARNE VALEAC VEARNQ VALEAM ILARNE	GRDLAREGN/ GRDLAREGN GRDLAREGN GRDLAREGN GRDLAREGGI GRDLYREGGI GRDLYREGGI GRDLYREGG RDLEKEGKE NRDYLTEGPE	200000 200   450   2011REACKWSPEI   201REACKWSPEI   201REACKWSACGAL	xx       460     470       AAACEVWKEITFNFPT       AAACEVWKEIKFEFPA       AAACUWKEIKFEFPA       AAACUWKEIKFEFPA       AAACUWKEIKFEFPA       AAALDUWKEIKFEFPA       AAALUWKIKFEFPA       AAALUWEIKFEFPA       AAALUWKIKFEFPA       AAALUWKIKFEFPA       AAALUWKIKFEFPA       AAALUWKEIKFEFPA       AAALUWKEIKFEFPA       AAALUWKEIKFEFPA       AAALUWKEIKFEFPA       AAALUWKIKFEFPA       AAALUWKIKEIKFEFPA       AA
Arabidopsis Pea Tobacco Rice Spinach Chlamydomonas T.elongatus Syn6301 H.neapolitanus G.partita G.sulphuraria A.eutrophus	IDKLDGQE MDTL VDVLDK VDKLDS MDTV QDTI VDKLL VDKLDTQNR. TDTSDFVETP TDTSDFVETP TDTSDFAPTA	  IANI IANI			

**Figure S1** Structure-based sequence alignment of Rubisco LSu sequences. See Fig. 2 figure legend for a description of residue colours. Residue numbering along the top refers to *A*. *thaliana* Rubisco. *Symbols above* blocks of sequences annotate the Rubisco LSu secondary structure from coordinates 5IU0:  $\alpha$ ,  $\alpha$ -helix;  $\eta$ , 310- helix;  $\beta$ ,  $\beta$ -strand; TT, tight  $\beta$ -turns. *Symbols below* blocks of sequences indicate the location of residues that contribute to the active site (blues triangles), the catalytic lysine residue (blue star). The Rubisco LSu structural features are labelled according to convention (Knight *et al.*, 1990), where secondary

structure elements are named  $\beta$ A,  $\beta$ B...,  $\alpha$ A,  $\alpha$ B... etc, except for the 8 C-terminal  $\beta\alpha$  barrel units, which are numbered  $\beta$ 1,  $\beta$ 2...,  $\alpha$ 1,  $\alpha$ 2... etc. The sequence alignment was created using the accession numbers 4HHH\_A (Pea), 4RUB\_A (Tobacco), 1WDD\_A (Rice), 8RUC\_A (Spinach), 1GK8\_A (Chlamydomonas), 3ZXW\_A (*T. elongatus*), 1RBL\_A (*Synechococcus* sp. 6301), 1SVD\_A (*H. neapolitanus*), 1BWV\_A (*G. partita*), 4F0K\_A (*G. sulphuraria*) and 1BXN\_A (*A. eutrophus*).



**Figure S2** Structure-based sequence alignment of Rubisco SSu sequences. See Fig. 2 figure legend for a description of symbols and residue colours. *Symbols above* blocks of sequences annotate the Rubisco SSu secondary structure elements (αA, αB, βA, βB, βC and βD; Knight *et al.*, 1990) from coordinates 5IU0. *Symbols below* blocks of sequences indicate the location of the βE and βF sheets (from co-ordinates 1BWV) unique to form IC and ID Rubiscos. The sequence alignment was created using the accession numbers 4HHH\_S (Pea), 4RUB\_S (Tobacco), 1WDD\_S (Rice), 8RUC\_I (Spinach), 1GK8\_I (Chlamydomonas), 3ZXW\_B (*T. elongatus*), 1RBL\_I (*Synechococcus* sp. 6301), 1SVD\_M (*H. neapolitanus*), 1BWV\_S (*G. partita*), 4F0K\_B (*G. sulphuraria*) and 1BXN\_I (*A. eutrophus*).



**Figure S3** Evolutionary relationships of Rubisco LSus. The evolutionary history was inferred using the Minimum Evolution method (Rzhetsky & Nei, 1992). The optimal tree with the sum of branch length = 0.98484718 is shown. Replicate trees in which the associated taxa clustered together in  $\geq 95\%$  (\*\* = 100%, \* = 99%) of the bootstrap iterations (1000 replicates) are shown at branch points (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method (Nei & Kumar, 2000) and are in the units of the number of amino acid differences per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm (Nei & Kumar, 2000) at a search level of 1. The Neighbor-joining algorithm (Saitou & Nei, 1987) was used to generate the initial tree. The analysis involved 12 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 455 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura *et al.*, 2013).

1B 2B 3B 1A 5IU0	MASSMLSSAA VVTSPAQATM VAPFTGLKSS ASFPVTRKAN NDITSITSNG FTAT. KA T M.A NAA	
1B 2B 3B 1A 5IU0	GRVSCMKVWP PIGKKKFETL SYLPDLTDVE LAKEVDYLLR NKWIPCVEFE	[100] [100] [100] [100] [100]
1B 2B 3B 1A 5IU0	LEHGFVYREH GNTPGYYDGR YWTMWKLPLF GCTDSAQVLK EVEECKKEYP	[150] [150] [150] [150] [150]
1B 2B 3B 1A 5IU0	GAFIRIIGFD NTRQVQCISF IAYKPPSFTD A [181] E . [181] E . [181] NG - [181] G - [181]	

**Figure S4** Sequence alignment of the *A. thaliana* Rubisco SSu precursors. The *A. thaliana* Rubisco structure presented in this study (pdb 5IU0) contains the RbcS1B isoform. Residues are numbered relative to the precursor peptide sequence. Identity is represented by dots, and gaps by dashes, and the transit peptide is shaded grey. See Fig. 2 figure legend for accession numbers.

## **Supplementary references**

Diederichs, K. & Karplus, P. A. (1997). Nat. Struct. Biol. 4, 269-275.

Felsenstein, J. (1985). Evol. Int. J. Org. Evol. 39, 783-791.

Nei, M. & Kumar, S. (2000). Molecular Evolution and Phylogenetics Oxford University Press.

Rzhetsky, A. & Nei, M. (1992). Mol. Biol. Evol. 9, 945.

Saitou, N. & Nei, M. (1987). Mol. Biol. Evol. 4, 406-425.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). *Mol. Biol. Evol.* **30**, 2725–2729.