Supporting Information

Identification of the allosteric site for neutral amino acids in the maize C4-isozyme of phosphoenolpyruvate carboxylase: The critical role of Ser100

Lilian González-Segura[1], Carlos Mújica-Jiménez[1], Javier-Andrés Juárez-Díaz[2], Rodrigo Güémez-Toro[1], León P. Martínez-Castilla[1],
and Rosario A. Muñoz-Clares[1*]

[1]Departamento de Bioquímica, Facultad de Química, [2]Departamento de Biología Comparada, Facultad de Ciencias. Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510, Ciudad de México, México.

Running title: Allosteric-site for neutral amino acids in *Zm*PEPC-C4

*To whom correspondence should be addressed: clares@unam.mx

Materials include Table S1, Table S2, Figure S1, Figure S2, Figure S3, Figure S4, Figure S5, Figure S6

Table S1. PEPC-C4 isozymes used in the multiple sequence analysis

[a]Amino acid sequences were retrieved from GeneBank with the exception of those from *Miscanthus sinensis, Panicum virgatum*, *Setaria italic, Setaria viridis* and *Sorghum bicolor* that were found in Phytozome 12.1.5; only those sequences having a serine at position equivalent to 780 of *Zea mays* PEPC-C4 (*Zm*PEPC-C4, CAA33317) were considered. [b]Most sequences are fragments lacking part of the N-terminal and/or C-terminal regions, only those marked with * are complete; those marked with † are the result of adding two consecutive sequence fragments reported in separate papers. [c]Numbers in parenthesis indicate the bibliographic reference of the paper where the sequence was reported, when any, given below the table. [d]Residues numbering is that of *Zm*PEPC-C4 (CAA33317).

PLANT FAMILY	C4-PLANT SPECIES	ACCESION NUMBER[a]	RESIDUES[b] (Ref.)[c]	POSITION 100[d]	POSITION 937[d]
MONOCOTS					
Cyperaceae					
	Bulbostylis barbata	CAR63753	427 (1)	—	—
	Cyperus capitatus	CAR63763	446 (1)	—	—
	Cyperus distichus	CAR63822	456 (1)	—	—
	Cyperus esculentus	ARK19505	965*	Lys	Asp
	Cyperus iria	CAR63823	400 (1)	—	—
	Cyperus longus	CAR63765	456 (1)	—	—
	Cyperus papyrus	CAR63766	453 (1)	—	—
	Cyperus pedunculatus	CAR63806	400 (1)	—	—
	Cyperus rotundus	CAR63768	400 (1)	—	—
	Cyperus sanguinolentus	CAR63805	456 (1)	—	—
	Cyperus ustulatus	CAR63770	456 (1)	—	—
	Cyperus ustulatus	CAR63771	435 (1)	—	—
	Eleocharis baldwinii	CAR63773	400 (1)	—	—
	Eleocharis vivipara	CAR63788	456 (1)	—	—
	Eleocharis vivipara	BAC19851	968* (2)	Lys	Asp
	Fimbristylis dichotoma	CAR63791	262 (1)	—	—
	Fimbristylis dichotoma	CAR63792	273 (1)	—	—
	Fimbristylis ferruginea	CAR63794	400 (1)	—	—
	Fimbristylis littoralis	CAR63796	400 (1)	—	—
	Rhynchospora globosa	CAR63809	429 (1)	—	—
	Rhynchospora globosa	CAR63810	350 (1)	—	—
	Rhynchospora rubra	CAR63812	400 (1)	—	—
Poaceae					
	Aeluropus littoralis	AJR16760	229	—	—
	Alloteropsis angusta	CCA60833	502 (3)	—	Gly
	Alloteropsis cimicina	CCA60994	846 (3)	—	Glu
	Alloteropsis semialata subs semialata	CCA60991	835 (3)	—	Gly
	Andropogon gerardii	CAM84081	439 (4)	—	—
	Anthaenantia lanata	CAM84072	439 (4)	—	—
	Apluda mutica	AHA39157	303 (5)	—	—
	Aristida adscensionis	CAM84082	438 (4)	—	—
	Aristida rhiniochloa	CAM84083	438 (4)	—	—
	Arundinella deppeana	CAM84086	439 (4)	—	—
	Bothriochloa ischaemum	CAM84087	439 (4)	—	—
	Bothriochloa saccharoides	CAM84049	439 (4)	—	—

PLANT FAMILY	C4-PLANT SPECIES	ACCESION NUMBER[a]	RESIDUES[b] (Ref.)[c]	POSITION 100[d]	POSITION 937[d]
	Capillipedium parviflorum	CAM84091	439 (4)	—	—
	Cenchrus americanus	ADV92634	959* (6)	Ser	Gly
	Centropodia forskalii	CAM84094	439 (4)	—	—
	Chloris gayana	AAG42288	955*	Ser	Asp
	Chrysopogon serrulatus	AHA39105	347 (5)	—	—
	Chrysopogon zizanioides	CAM84148	439 (4)	—	—
	Coix lacryma-jobi	CAM84097	439 (4)	—	—
	Cynodon dactylon	CAM84101	439 (4)	—	—
	Cymbopogon citratus	CAM84102	439 (4)	—	—
	Dactyloctenium aegyptium	CAM84047	439 (4)	—	—
	Digitaria ciliaris	CAM84051	439 (4)	—	—
	Digitaria didactyla	CAM84050	439 (4)	—	—
	Digitaria sanguinalis	CAM83971	627 (4)	—	Asp
	Echinochloa crus-galli	AAX98688	964*	Gly	Gly
	Echinochloa esculenta	CAM84052	439 (4)	—	—
	Eleusine coracana	AEG78556	278 (7)	—	—
	Eleusine floccifolia	AEF58974	260 (7)	—	—
	Eleusine indica	CAM84054	439 (4)	—	—
	Eleusine intermedia	AEF58978	146 (7)	—	—
	Eleusine jaegeri	AEF58979	260 (7)	—	—
	Eleusine kigeziensis	AEG78571	278 (7)	—	—
	Eleusine kigeziensis	AEG78572	278 (7)	—	—
	Eleusine multiflora	AEF58990	259 (7)	—	—
	Eleusine tristachya	AEF58992	259 (7)	—	—
	Enteropogon prieurii	CAM84056	439 (4)	—	—
	Eragrostis capensis	CAM84058	439 (4)	—	—
	Eragrostis minor	CAM84059	439 (4)	—	—
	Eriochloa nana	CCB84861	424 (3)	—	—
	Eulalia aurea	CAM84060	439 (4)	—	—
	Hyparrhenia hirta	CAM84064	439 (4)	—	—
	Imperata cylindrica	CAM84068	439 (4)	—	—
	Lepturus repens	CAM84073	439 (4)	—	—
	Megathyrsus maximus	CCA60995	936 (3)	Ser	Glu
	Megathyrsus maximus	CBV65831	687 (8)	—	—
	Melinis minutiflora	CAM84075	439 (4)	—	—
	Microstegium sp. PC-2007	CAM84104	439 (4)	—	—
	Miscanthus sinensis	Misin18G159600	961*	Ser	Gly
	Neurachne munroi	CCK33010	588 (3)	—	—
	Panicum capillare	CAM84110	439 (4)	—	—
	Panicum coloratum	CAM84116	439 (4)	—	—
	Panicum fluviicola	CBL93742	801 (8)	—	Gly
	Panicum hallii	PAN23802	961*	Gly	Gly
	Panicum laetum	CBT21623	961* (8)	Gly	Gly
	Panicum miliaceum	CBV65830	750 (8)	—	—
	Panicum pansum	CBT21624	688 (8)	—	—
	Panicum phragmitoides	CBT21625	801 (8)	—	Gly
	Panicum schinzii	CBY89214	717 (8)	—	—
	Panicum turgidum	CBR26851	795 (8)	—	Gly
	Panicum virgatum	Pavir.J09522	961*	Gly	Gly
	Paraneurachne muelleri	CCK33005	588 (3)	—	—
	Paspalum conjugatum	CAM84121	439 (4)	—	—

PLANT FAMILY	C4-PLANT SPECIES	ACCESION NUMBER[a]	RESIDUES[b] (Ref.)[c]	POSITION 100[d]	POSITION 937[d]
	Paspalum dilatatum	CAM84122	439 (4)	—	—
	Paspalum paniculatum	CCB84879	884 (3)	Gly	—
	Paspalum quadrifarium	CAM84120	439 (4)	—	—
	Saccharum officinarum	CAC08829	961* (9)	Ser	Gly
	Saccharum spontaneum	CAC85930	961* (9)	Ser	Gly
	Setaria adhaerens	CCB84858	427 (3)	—	—
	Setaria italica	Si005789m	964*	Ser	Gly
	Setaria megaphylla	CCB84863	439 (3)	—	—
	Setaria palmifolia	CAM84130	439 (4)	—	—
	Setaria plicata	CAM84131	439 (4)	—	—
	Setaria viridis	Sevir.4G143500	964*	Ser	Gly
	Sorghum bicolor	Sb10g021330	1028*	Ser	Gly
	Sorghum halepense	AHA39146	347 (5)	—	—
	Sorghum interjectum	AHA39132	347 (5)	—	—
	Sorghum intrans	AHA39142	347 (5)	—	—
	Sorghum laxiflorum	AHA39139	346 (5)	—	—
	Sorghum matarankense	AHA39134	347 (5)	—	—
	Sorghum sorghoides	AHA39136	347 (5)	—	—
	Sorghum stipoideum	AHA39113	347 (5)	—	—
	Sorghum sudanense	AHA39145	347 (5)	—	—
	Sorghum timorense	AHA39111	347 (5)	—	—
	Sporobolus anglicus	CAM84132	439 (4)	—	—
	Sporobolus festivus	CAM84134	439 (4)	—	—
	Sporobolus schoenoides	CAM84099	439 (4)	—	—
	Sporobolus sp. Hodkinson s.n.	CAM84090	439 (4)	—	—
	Stenotaphrum dimidiatum	CAM84136	439 (4)	—	—
	Stipagrostis pennata	CAX65714	658 (10)	—	—
	Stipagrostis pennata Lausanne	CAX65713	503 (10)	—	—
	Tetrapogon cenchriformis	CAM84140	439 (4)	—	—
	Tragus racemosus	CAM84144	439 (4)	—	—
	Tripogonella minima	CAM84145	439 (4)	—	—
	Tristachya leucothrix	CAM84146	439 (4)	—	—
	Urochloa villosa	CAM84088	439 (4)	—	—
	Zea mays	CAA33317	970*(11)	Ser	Gly
	Zoysia japonica	CAM84149	439 (4)	—	—
	Zuloagaea bulbosa	CBR26850	688 (8)	—	—
EUDICOTS					
Aizoaceae					
	Trianthema portulacastrum	AIF35268	400 (12)	—	Glu
Amaranthaceae					
	Alternanthera pungens	AAY28729	966* (13)	Lys	Glu
	Amaranthus hypochondriacus	AAB18633	964* (14)	Lys	Asp
	Bienertia cycloptera	ALH25077/ AHN09924	879† (15,16)	Lys	Glu
	Bienertia sinuspersici	ABG20459	968* (17)	Lys	Glu

Table S1 (*continued*)

PLANT FAMILY	C4-PLANT SPECIES	ACCESION NUMBER[a]	RESIDUES[b] (Ref.)[c]	POSITION 100[d]	POSITION 937[d]
	Bienertia sinuspersici	AHN09926	510 (15)	—	Glu
	Gomphrena globosa	AIF35261	204 (12)	—	Glu
	Haloxylon persicum	AIF35262	519 (12)	—	Glu
	Suaeda acuminata	ALH25076/ AHN09928	879† (15,16)	Lys	Glu
	Suaeda aralocaspica	ABG20460	851 (17)	—	Glu
	Suaeda eltonica	ABG20461	830 (17)	—	Glu
	Suaeda eltonica	ALH25062/ AHN09938	879† (15,16)	Lys	Glu
	Tidestromia lanuginose	AIF35263	517 (12)	—	Glu
	Tidestromia valdesiana	AIF35265	519 (12)	—	Glu
Asteraceae					
	Flaveria bidentis	BAZ95841	966*	Lys	Glu
	Flaveria trinervia	CAA43601	966* (18)	Lys	Glu
Molluginaceae					
	Hypertelis cerviana	CBM40391	521 (12)	—	Glu

REFERENCES:

1. Besnard, G., Muasya, A. M., Russier, F., Roalson, E. H., Salamin, N. and Christin, P. A. (2009) Phylogenomics of C4 photosynthesis in sedges (Cyperaceae): Multiple appearances and genetic convergence. *Mol. Biol. Evol.* **26,** 1909–1919

2. Agarie,S., Kai,M., Takatsuji,H. and Ueno, O. (2002) Environmental and hormonal regulation of gene expression of C4 photosynthetic enzymes in the amphibious sedge *Eleocharis vivipara. Plant Sci.* **163,** 571–580

3. Christin, P. A., Edwards, E. J., Besnard, G., Boxall, S. F., Gregory, R., Kellogg, E. A., Hartwell, J. and Osborne, C. P (2012) Adaptive evolution of C4 photosynthesis through recurrent lateral gene transfer. *Curr. Biol.* **22,** 445–449

4. Christin, P. A., Salamin, N., Savolainen, V., Duvall, M. R. and Besnard, G. (2007) C4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr. Biol.* **17,** 1241–1247

5. Liu, Q., Liu, H., Wen, J. and Peterson, P. M. (2014) Infrageneric phylogeny and temporal divergence of *Sorghum* (Andropogoneae, Poaceae) based on low-copy nuclear and plastid sequences. *PLoS ONE* **9,** E104933

6. Singh, J., Reddy, G. M., Agarwal, A., Chandrasekhar, K., Sopory, S. K., Reddy, M. K. and Kaul, T. (2012) Molecular and structural analysis of C4-specific PEPC isoform from *Pennisetum glaucum* plays a role in stress adaptation. *Gene* **500,** 224–231

7. Liu, Q., Triplett, J. K., Wen, J. and Peterson, P. M. (2011) Allotetraploid origin and divergence in *Eleusine* (Chloridoideae, Poaceae): evidence from low-copy nuclear gene phylogenies and a plastid gene chronogram. *Ann. Bot.* **108,** 1287–1298

8. Zimmermann, T., Bocksberger, G., Brüggemann, W., Berberich, T. (2013) Phylogenetic relationship and molecular taxonomy of African grasses of the genus *Panicum* inferred from four chloroplast DNA-barcodes and nuclear gene sequences. *J. Plant Res.* **126,** 363–371

9. Besnard, G., Offmann, B., Robert, C., Rouch, C. and Cadet, F. (2002) Assessment of the C(4) phosphoenolpyruvate carboxylase gene diversity in grasses (Poaceae). *Theor. Appl. Genet.* **105,** 404–412

10. Christin, P. A. and Besnard, G. (2009) Two independent C4 origins in Aristidoideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. *Am. J. Bot.* **96,** 2234–2239

11. Hudspeth, R. L. and Grula, J. W. (1989) Structure and expression of the maize gene encoding the phosphoenolpyruvate carboxylase isozyme involved in C4 photosynthesis. *Plant Mol. Biol.* **12,** 579–589

12. Christin, P. A., Arakaki, M., Osborne, C. P., Brautigam, A., Sage, R. F., Hibberd, J. M., Kelly, S., Covshoff, S., Wong, G. K., Hancock, L. and Edwards, E. J. (2014) Shared origins of a key enzyme during the evolution of C4 and CAM metabolism. *J. Exp. Bot.* **65,** 3609–3621

13. Gowik, U., Engelmann, S., Blasing, O. E., Raghavendra, A. S. and Westhoff, P. (2006) Evolution of C4 phosphoenolpyruvate carboxylase in the genus *Alternanthera*: gene families and the enzymatic characteristics of the C4 isozyme and its orthologues in C3 and C3/C4 *Alternantheras. Planta* **223,** 359–368

14. Rydzik, E. and Berry, J. O. (1996) The electronic plant gene register. *Plant Physiol.* **110,** 713–714

15. Rosnow, J. J., Edwards, G. E. and Roalson, E. H. (2014) Positive selection of Kranz and non-Kranz C4 phosphoenolpyruvate carboxylase amino acids in Suaedoideae (Chenopodiaceae). *J. Exp. Bot.* **65,** 3595–3607

16. Rosnow, J. J., Evans, M. C., Kapralov, M. V., Cousins, A. B., Edwards, G. E., Roalson, E. H. (2015) Kranz and single-cell forms of C4 plants in the subfamily Suaedoideae show kinetic C4 convergence for PEPC and Rubisco with divergent amino acid substitutions. *J. Exp. Bot.* **66,** 7347–7358

17. Lara, M. V., Chuong, S. D., Akhani, H., Andreo, C. S. and Edwards, G. E. (2006) Species having C4 single-cell-type photosynthesis in the Chenopodiaceae family evolved a photosynthetic phospho*enol*pyruvate carboxylase like that of Kranz-type C4 species. *Plant Physiol.* **142,** 673–684

18. Poetsch, W., Hermans, J. and Westhoff, P. (1991) Multiple cDNAs of phosphoenolpyruvate carboxylase in the C4 dicot *Flaveria trinervia. FEBS Lett.* **292,** 133–136

**Table S2. CAM-PEPC and C3-PEPC isozymes that have Ser780**

[a]Every sequence included in this table has serine at position equivalent to 780 of *Zea mays* PEPC-C4 (*Zm*PEPC-C4, CAA33317) despite of being from a CAM or C3 plant as indicated within parenthesis. [b]Amino acid sequences were retrieved from GeneBank except those from *Helianthus annuus* that were found in Phytozome 12.1.5. [c]Most sequences are fragments lacking part of the N-terminal and/or C-terminal regions; only the sequence marked with * is complete. [d]Numbers in parenthesis indicate the bibliographic reference of the paper where the sequence was reported, when any, given below the table.

PLANT FAMILY	PLANT SPECIES[a] (Photosynthetic metabolism)	ACCESION NUMBER[b]	NUMBER OF RESIDUES[c] (Ref.)[d]
MONOCOTS			
Cyperaceae			
	Rhynchospora colorata (C3)	CAR63802	456 (1)
EUDICOTS			
Anacampserotaceae			
	Anacampseros albissima (CAM)	AIF35194	404 (2)
	Anacampseros kurtzii (C3)	AIF35195	197 (2)
	Anacampseros retusa (CAM)	AIF35196	415 (2)
	Grahamia bracteata (C3)	AIF35197	196 (2)
Asteraceae			
	Cynara cardunculus (CAM)	KVH88808	849* (3)
	Helianthus annuus (C3)	HanXRQChr17g0560511	269
	Helianthus annuus (C3)	HanXRQChr07g0190711	251
Basellaceae			
	Anredera baselloides (CAM)	AIF35179	518 (2)
	Anredera cordifolia (CAM)	AIF35180	189 (2)
	Anredera ramosa (CAM)	AIF35182	186 (2)
	Ullucus tuberosus (CAM)	AIF35183	389 (2)
Cactaceae			
	Echinocactus grusonii (CAM)	AIF35207	518 (2)
	Echinocereus pectinatus (CAM)	AIF35404	504 (2)
	Ferocactus leucacanthus (CAM)	AIF35209	196 (2)
	Hylocereus undatus (CAM)	AHF21553	958* (2)
	Leuenbergeria aureiflora (CAM)	AIF35199	393 (2)
	Leuenbergeria bleo (CAM)	AIF35200	196 (2)
	Leuenbergeria guamacho (CAM)	AIF35203	393 (2)
	Leuenbergeria portulacifolia (CAM)	AIF35204	196 (2)
	Mammillaria plumosa (CAM)	AIF35208	196 (2)
	Opuntia cochenillifera (CAM)	AIF35403	518 (2)
	Pereskia aculeata (CAM)	AIF35198	518 (2)
	Pereskia diaz-romeroana (CAM)	AIF35201	393 (2)
	Pereskia grandifolia (CAM)	AIF35202	196 (2)
	Pereskiopsis gatesii (CAM)	AIF35205	195 (2)
	Tephrocactus articulatus (CAM)	AIF35206	196 (2)
	Weingartia kargliana (CAM)	AIF35210	196 (2)
Didiereceae			
	Alluaudiopsis marnieriana (CAM)	AIF35185	195 (2)
	Alluaudia procera (CAM)	AIF35184	195 (2)
	Didierea madagascariensis (CAM)	AIF35187	293 (2)
	Portulacaria longipedunculata (CAM)	AIF35230	527 (2)

PLANT FAMILY	PLANT SPECIES[a] (Photosynthetic metabolism)	ACCESION NUMBER[b]	NUMBER OF RESIDUES[c] (Ref.)[d]
Euphorbiaceae			
	Ricinus communis (C3)	EEF27881	852
Halophytaceae			
	Halophytum ameghinoi (CAM)	AIF35178	518 (2)
Talinaceae			
	Talinella pachypoda (CAM)	AIF35193	196 (2)
	Talinum fruticosum (CAM)	AIF35192	196 (4)
	Talinum paniculatum (CAM)	AIF35191	196 (2)

REFERENCES:

1. Bruhl, J. J. and Wilson K. L. (2007) Towards a comprehensive survey of C3 and C4 photosynthetic pathways in Cyperaceae. *Aliso* **23,** 99–148

2. Christin, P. A., Arakaki, M., Osborne, C. P., Brautigam, A., Sage, R. F., Hibberd, J. M., Kelly, S., Covshoff, S., Wong, G. K., Hancock, L. and Edwards, E. J. (2014) Shared origins of a key enzyme during the evolution of C4 and CAM metabolism. *J. Exp. Bot*. **65,** 3609–3621

3. Jaradat, A. A. (2010) Genetic resources of energy crops: Biological systems to combat climate change. *Aust. J. Crop Sci.* **4,** 309–323

4. http://tropical.theferns.info/viewtropical.php?id=Talinum+fruticosum
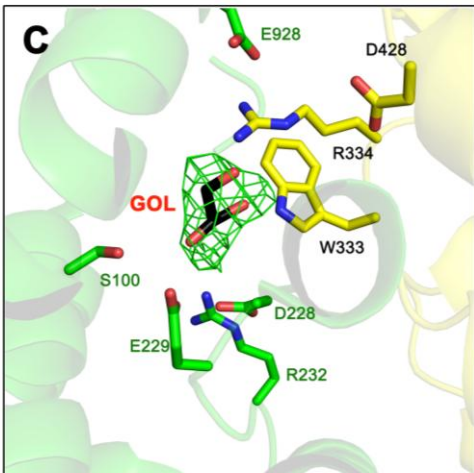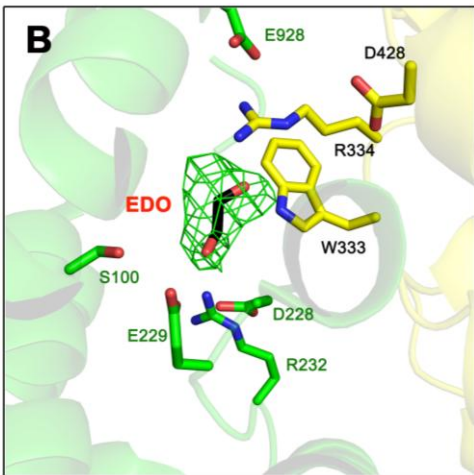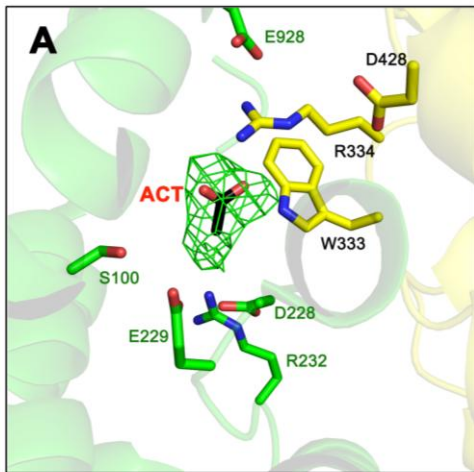
Supplementary Figure S1. Omit electronic density map of the ligand bound in the allosteric-site for neutral amino acids of *Zm*PEPC-C4. The simulated annealing omit electron density map (F_o-F_c) contoured at 3.0 σ level is shown as a green mesh. Acetate (ACT, panel *A*), ethylene glycol (EDO, panel *B*) or glycerol (GOL, panel *C*) were modeled into this omit electron density map. As can be seen, none of these molecules gave a good fit. The side chains of amino acid residues are shown as sticks with oxygen atoms in red, nitrogen atoms in blue, and carbon atoms in green (subunit A) or yellow (subunit B). ACT, EDO and GOL are shown as sticks colored similarly but with black carbons. The simulated annealing omit map was calculated using Phenix (1) following simulated annealing refinement at 3,000 K. Images were generated using PyMOL.
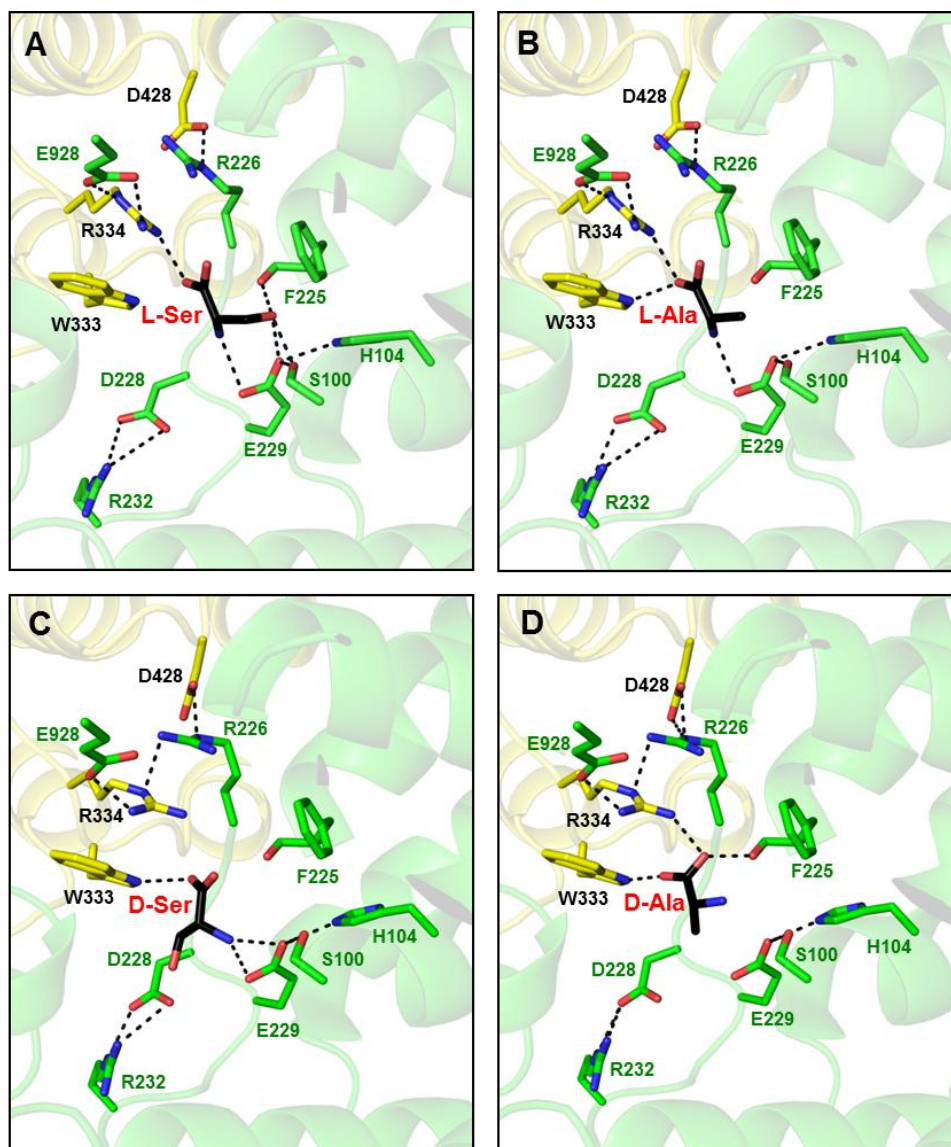
Supplementary Figure S2. Energy-minimized models of D/L-alanine and D/L-serine molecules docked into the allosteric-site for neutral amino acids of the *Zm*PEPC-C4 5VYJ structure. The glycine molecule in this crystal structure was manually removed from the electronic density map before rigidly docking D/L-alanine and D/L-serine molecules, which α-carboxyl and α-amino groups make the same interactions as the corresponding groups of the bound glycine. Docking simulations were performed using the PyMOL building mode and then subjected to a 1000-step energy-minimization process using the Amber force field parameters in the UCSF Chimera (2). The side chain of amino acid residues and the bound amino acids are shown as sticks with oxygen atoms in red, nitrogen atoms in blue, and carbon atoms in green (subunit A), yellow (subunit B) or pink (D/L-alanine and D/L-serine ligands). Hydrogen bonds (cut-off 3 Å) are depicted as black dashed lines. The four amino acids fit well inside the neutral amino acid allosteric site and made several interactions with protein residues, consistent with their proven ability to activate the enzyme. Images were generated using PyMOL.
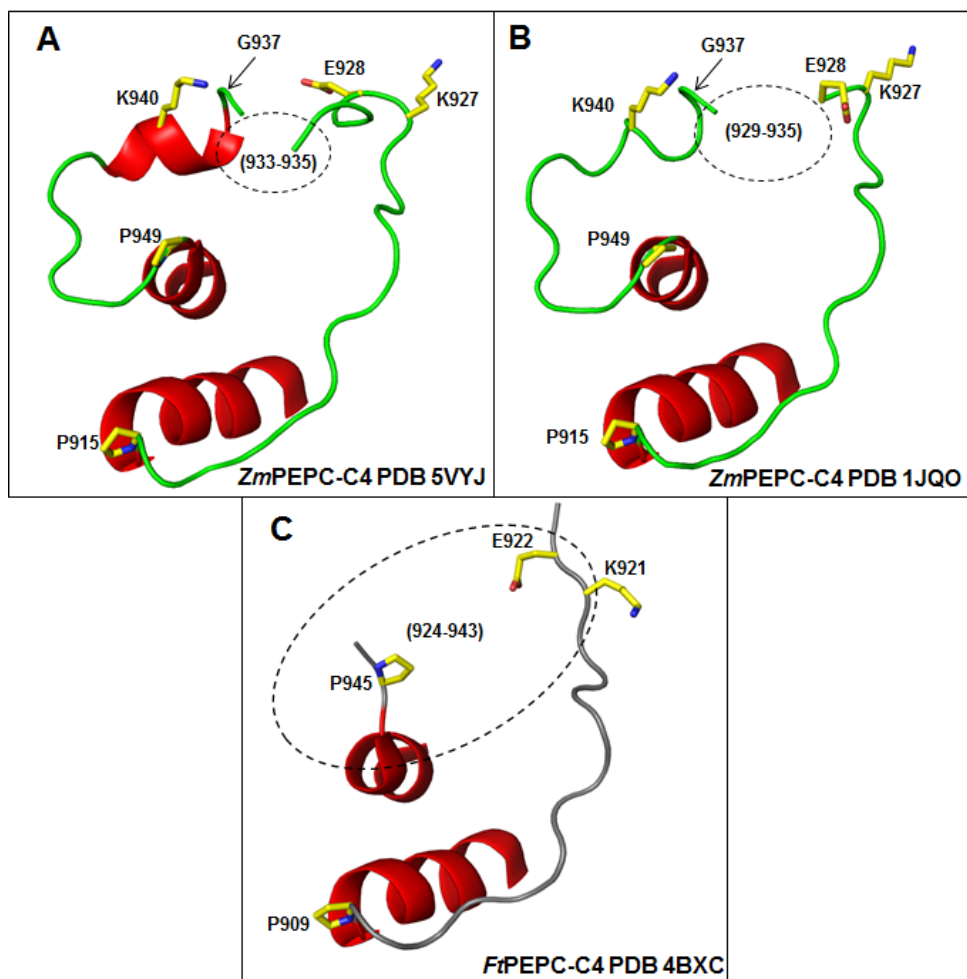
Supplementary Figure S3. Exposed loop at the carboxy-terminal region of PEPC-C4 isoenzymes. Cartoon representation of the exposed loop extending from Pro915/Pro909 to Pro949/945 (*Zm*PEPC-C4/*Ft*PEPC-C4 numbering); these two proline residues are absolutely conserved. The side chains of the two proline flanking the loop as well as those of Lys927/921 and Glu928, when observed in the crystal structure, are shown as sticks with carbon atoms in yellow, oxygen in red and nitrogen in blue. A black arrow marks the position of Gly937 in the two *Zm*PEPC-C4 structures. (*A*) In the 5VYJ *Zm*PEPC-C4 crystal structure, which has a glycine molecule bound at its allosteric site, three residues (Asn933, Lys934 and Pro935) do not exhibit electronic density, and five residues (Leu938, Val939, Lys940, Leu941 and Asn942) form a short α-helix within the loop. (*B*) In the 1JQO *Zm*PEPC-C4 crystal structure, which has an empty allosteric site for neutral amino acids, seven residues (Phe929, Ala930, Asp931, Glu932, Asn933, Lys934 and Pro935) do not exhibit electronic density. (*C*) In the 4BXC *Ft*PEPC-C4 crystal structure no electronic density was found associated with the side chains of residues 924 to 943, suggesting a high flexibility of this region. The dotted ellipses represent the disordered stretches of residues with no electronic density. Images were generated using PyMOL.
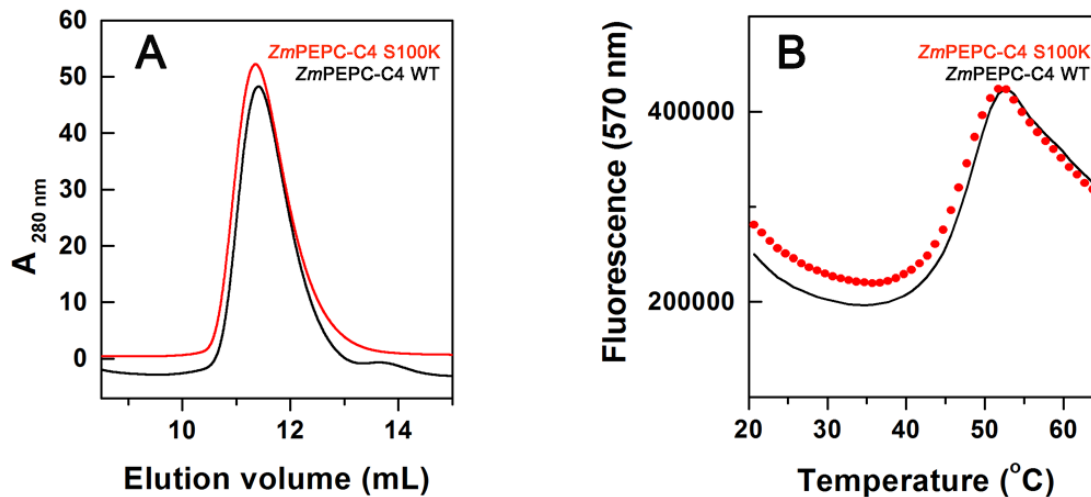
Supplementary Figure S4. Oligomeric state and thermal stability of wild-type and S100K *Zm*PEPC-C4 enzymes. (A) Size-exclusion chromatography elution profile obtained using a Superdex 200 10/300 (GE Healthcare Life Sciences) column in an ÄKTA Pure 25 (GE Healthcare Life Sciences). The proteins (at 0.8 mg/mL) were in 50 mM Hepes-KOH, pH 7.4, buffer containing 50 mM NaCl, 10% (v/v) glycerol and 2 mM 2-mercaptoethanol. The column was eluted with the same buffer at 0.5 mL/min rate. (B) Thermal denaturation curves obtained using a StepOnePlus Real Time PCR system. The proteins (at 2.2 mg/mL) were in the same buffer as for the size-exclusion chromatography experiments. Heat-induced transitions were determined by following the changes in SYPRO Orange (Sigma-Aldrich) fluorescence intensity. The temperature range was 20–90 $^{\circ}$C and the scan rate 1.0 $^{\circ}$C/min. A representative curve of three different experiments is shown in the figure. Transitions were evaluated with a Boltzmann function using the Protein Thermal Shift™ Software v1.0. The estimated apparent melting temperatures (appT_m) were 48.8 $^{\circ}$C for the wild-type and 47.75 $^{\circ}$C for the S100K mutant.
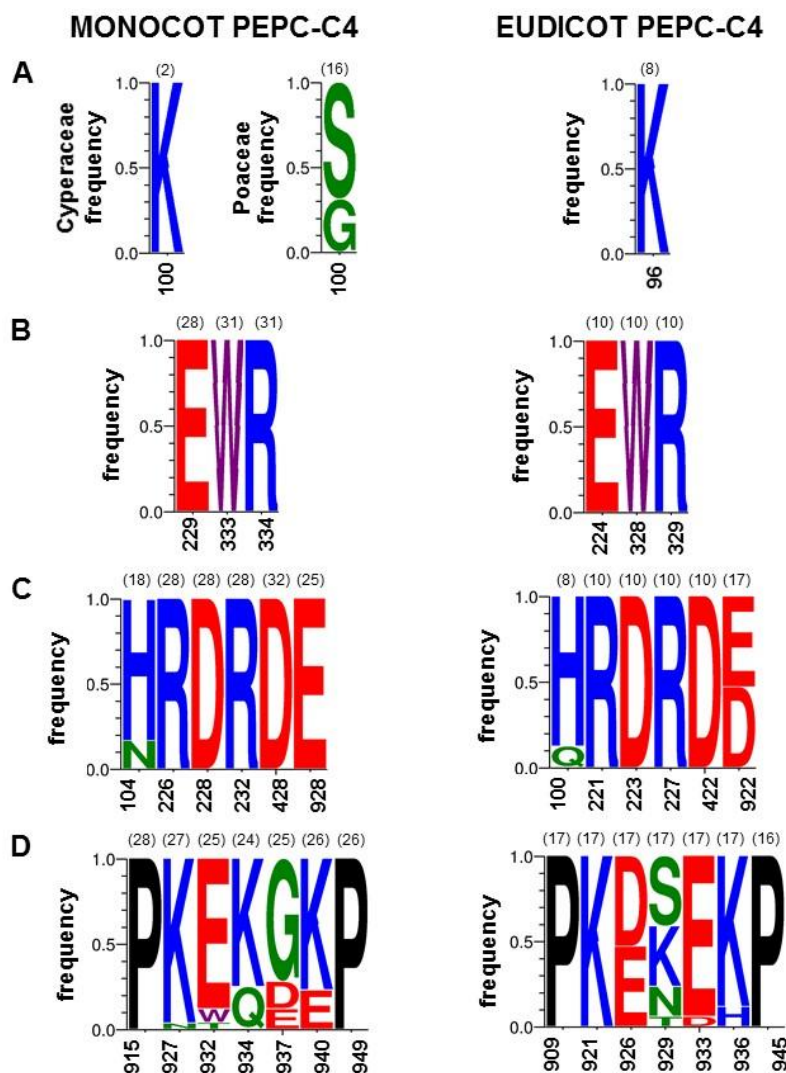
Supplementary Figure S5. Sequence logos of PEPC-C4 residues located at or near the region equivalent to that of the allosteric site for neutral amino acids of *Zm*PEPC-C4. The sequences of PEPC-C4 isozymes were selected considering as C4 those PEPCs that have a serine at the position equivalent to 780 in *Zm*PEPC-C4 (CAA33317) (3) and that, in addition, belong to a C4 plant. Residues numbering is that of *Zm*PEPC-C4 for monocot sequences and that of *Ft*PEPC-C4 (CAA43601) for eudicot sequences. (*A*) Conservation of residue at position 100/96. (*B*) Conservation of residues directly involved in binding the neutral amino acid. (*C*) Conservation of residues involved in the web of hydrogen bonds that stabilize the conformation of the neutral amino acid allosteric site. (*D*) Conservation of residues that have been related to the allosteric site for neutral amino acids by previous site-directed mutagenesis studies, as well as the two proline residues that flank the loop to which they belong. The amino acids color scheme was according to their chemical properties: polar (G, S, T, Q, N), green; aromatic (W), purple; basic (K, R, H), blue; acidic (D, E), red; and hydrophobic (P), black. Given that most of the retrieved sequences were partial, within parenthesis above each position is given the number of sequences analyzed.
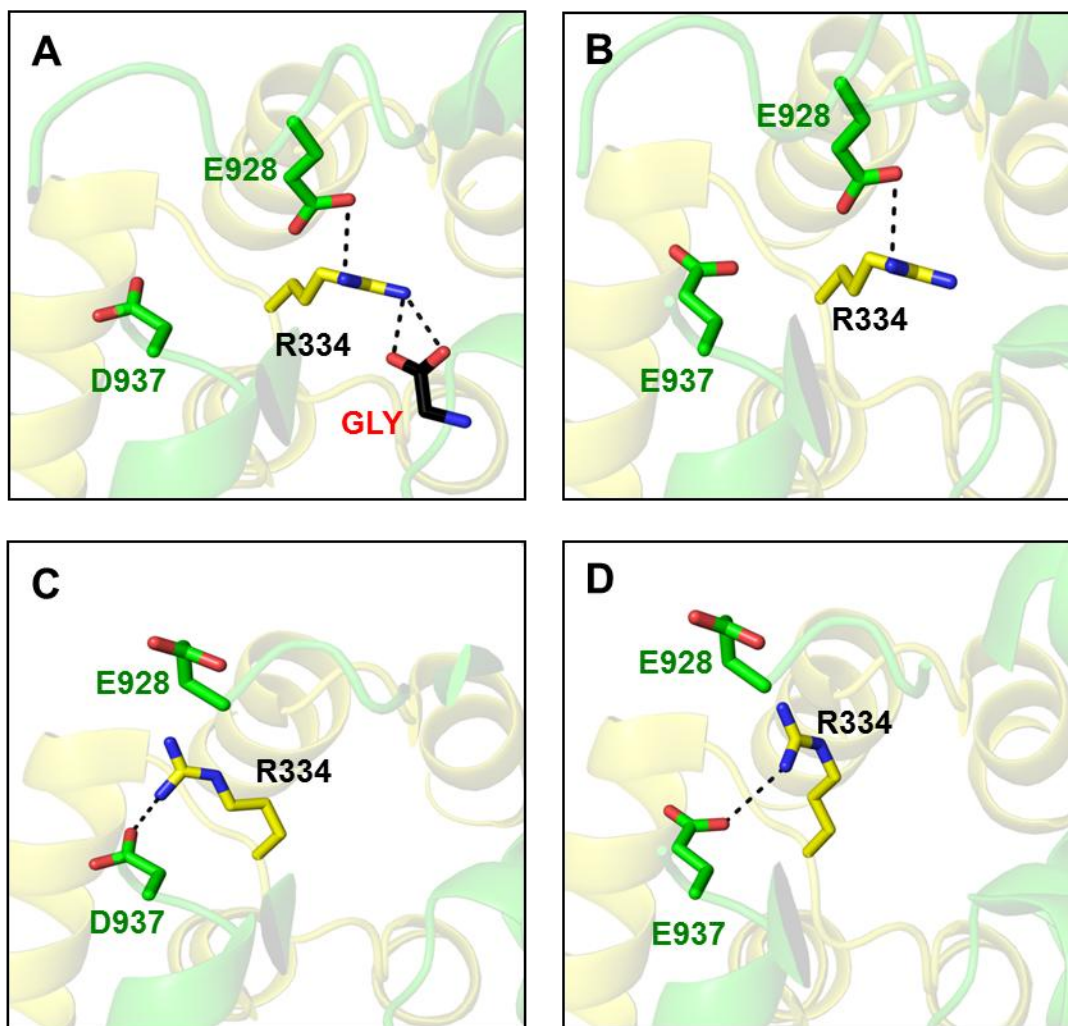
Supplementary Figure S6. *In silico* **change of Gly937 of *Zm*PEPC-C4 for aspartate or glutamate.** Cartoon representation of the region near Gly937 showing the relevant amino acid residues and the activator glycine molecule as sticks. *In silico* mutations of Gly937 for aspartate or glutamate were generated using the 5VYJ crystal structure reported here and the standard rotamer library of Coot (4). Hydrogen bonds (cut-off 3 Å) are depicted as black dashed lines. Coloring is as in Supplementary Fig. S1. When glycine is bound to its allosteric site, the model shows that Glu/Asp937 are exposed to the solvent, not interacting with any other protein residue, and Arg334 is in the "inside" conformation interacting with the activator glycine molecule (panels *A* and *B*). But Asp/Glu937 could form a hydrogen bond with Arg334 of the opposing subunit if the arginine residue is modeled in the exposed "outside" conformation observed in the *Zm*PEPC-C4 1JQO crystal structure, where there is no ligand bound in the allosteric-site for neutral amino acids (panels *C* and *D*). Images were generated using PyMOL.

REFERENCES

1. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954
2. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004) UCSF Chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612
3. Bläsing, O. E., Westhoff, P. and Svensson, P. (2000) Evolution of C4 phosphoenolpyruvate carboxylase in *Flaveria*, a conserved serine in the carboxyterminal part of the enzyme is a major determinant for C4-specific characteristics. *J. Biol. Chem.* **275,** 27917–27923
4. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132