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

Lilian González-Segura¹, Carlos Mújica-Jiménez¹, Javier Andrés Juárez-Díaz², Rodrigo Güémez-Toro¹, León P. Martínez-Castilla¹, and Rosario A. Muñoz-Clares¹*

¹Departamento de Bioquímica, Facultad de Química, ²Departamento 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 ZmPEPC-C4

To whom correspondence should be addressed: Rosario A. Muñoz-Clares, Departamento de Bioquímica, Facultad de Química, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510, Ciudad de México, México. Telephone number: +52 55 5622 3718; e-mail: clares@unam.mx

Keywords: Zea mays L.; monocot and eudicot PEPC-C4 isozymes; allosteric regulation; enzyme kinetics; protein structure; crystal structure; site-directed mutagenesis; amino acid residues conservation.
ABSTRACT

The photosynthetic phosphoenolpyruvate carboxylase isozymes from C4 plants (PEPC-C4) play a critical role in their atmospheric CO₂ assimilation and productivity. They are allosterically activated by phosphorylated trioses or hexoses, such as D-glucose-6-phosphate, and inhibited by L-malate or L-aspartate. Additionally, PEPC-C4 isozymes from grasses are activated by glycine, serine or alanine, but the allosteric site for these compounds remained unknown. Here we report a new crystal structure of the isozyme from Zea mays (ZmPEPC-C4) with glycine bound at the monomer-monomer interfaces of the two dimers of the tetramer, making interactions with residues of both monomers. This binding site is close to, but different from, the one proposed to bind glucose-6-phosphate. Docking experiments indicated that D/L-serine or D/L-alanine could also bind to this site, which does not exist in the PEPC-C4 isozyme from the eudicot plant Flaveria, mainly because of a lysyl residue at the equivalent position of Ser100 in ZmPEPC-C4. Accordingly, the ZmPEPC-C4 S100K mutant is not activated by glycine, serine or alanine. Amino acid sequence alignments showed that PEPC-C4 isozymes from the monocot family Poaceae have either serine or glycine at this position while those from Cyperaceae and eudicots families have lysine. The size and charge of the residue equivalent to Ser100 is not only crucial for the activation of PEPC-C4s by neutral amino acids but also affect their affinity for the substrate phosphoenolpyruvate and their allosteric regulation by glucose-6-phosphate and malate, accounting for the reported kinetics differences between PEPC-C4 isozymes from monocot and eudicot plants.

INTRODUCTION

During photosynthesis, in order to fix and reduce carbon and to synthesize simple sugars, the ATP and NADPH generated in the light reactions are used in the carbon-fixation reactions. In many plant species, known as C3-type, these reactions occur exclusively in the chloroplast by means of the C3-pathway, also known as the Calvin-Benson cycle, in which the starting and ending compound is ribulose 1,5-bisphosphate (RuBP). In this pathway, the carboxylation step is catalyzed by RuBP carboxylase/oxygenase (RUBISCO). The majority of angiosperms are of the C3-type, but there is a group of plants—called the C4-type—which include agronomically important crops as maize (Zea mays), sorghum (Sorghum bicolor), sugarcane (Saccharum officinarum), and common millet (Panicum miliaceum) as well as most of the weeds—in which the Calvin-Benson cycle is not the only pathway used in the carbon-fixation reactions. In the C4-plants, two metabolic pathways are involved in CO₂ assimilation: the C4- and the C3-pathways (1). The C4-pathway provides a CO₂-concentrating mechanism that ensures high CO₂ levels in the environment of the C3-pathway enzymes, greatly favoring the carboxylating activity of RUBISCO over its oxygenating activity. This results in an important reduction in photorespiration—a process that in C3-plants causes significant losses in photosynthetic efficiency—as well as in a better use of water and nitrogen. As a consequence, C4-plants exhibit an increased productivity, especially in sunny, dry and hot environments.

The initial and highly regulated step in the C4-pathway takes place in the cytosol of mesophyll cells and consists in the irreversible carboxylation of 2-phosphoenolpyruvate (PEP) by the C4-photosynthetic isozyme of phosphoenolpyruvate carboxylase (orthophosphate: oxalacetate carboxy-lyase (phosphorylating), EC 4.1.1.31) yielding oxalacetate and Pi. The carboxylating reactions of PEPC-C4 and RUBISCO take place in different cell types in the leaves of those plants that have the two-cells C4-pathway (also called Kranz-type because of the anatomy of their leaves), which is by far the most spread, whereas these reactions take place in different intracellular compartments in those plants that have the single-cell C4-pathway (2,3). In addition to transport of some of the pathway
intermediates between cells or intracellular compartments, in the two kinds of C4-pathways the three essential steps are the initial fixation of CO₂ by a PEPC-C4 isozyme to form a C4 acid, the decarboxylation of the C4 acid to release CO₂ in the RUBISCO environment, and the regeneration of the primary CO₂ acceptor PEP.

Due to the increased need of food to satisfy the requirements of the growing human population, much effort has been, and still is, devoted to the improvement of the photosynthetic performance and therefore the yield of C3 plants—particularly those of agronomically and economically important crops such as wheat (Triticum aestivum), rice (Oryza sativa), barley (Hordeum vulgare), soybean (Glycine max) or potato (Solanum tuberosum). This goal has been attempted by introducing C4-pathway genes, particularly the one coding for the PEPC-C4 isozyme from maize, using conventional transgenic techniques, so that they could have the CO₂ concentrating mechanism of C4 plants.

The importance of the reaction catalyzed by PEPC-C4 isozymes in the photosynthetic metabolism of C4 plants is underscored by their allosteric regulation (reviewed in reviewed in 4-6), which is particularly complex in the case of isozymes from some monocot C4 plants, such as that from maize (ZmPEPC-C4) (Fig. 1). It has been known for long that at physiological pH, ZmPEPC-C4 is activated by phosphorylated trioses and hexoses, particularly D-glucose-6-phosphate (G6P) (7-13), by its substrate PEP in its free form, i.e. not complexed with Mg²⁺ ions (13,14), as well as by the neutral amino acids glycine, serine and alanine (8-11,15-19). In addition, it is allosterically inhibited by L-malate (20). Whereas the activation by G6P and the inhibition by L-malate or L-aspartate are common to most PEPC enzymes biochemically characterized to date, the activation by neutral amino acids has been only found in PEPC-C4 isozymes from monocot grasess (Poaceae family) (8,9,11,15, 16,19). As exemplified by ZmPEPC-C4, the activation by neutral amino acids of the PEPC-C4 isozymes from grasses appears to be crucial for their activity under conditions close to those assumed to be physiological, given that malate inhibition is significantly reverted by these compounds but not by G6P (18). The structural basis of these kinetic differences between monocot and eudicot PEPC-C4 isozymes are not known at present.

The location in the three-dimensional structure of different PEPC enzymes of the allosteric sites for carboxylic acids and for phosphorylated sugars was predicted by site-directed mutagenesis and confirmed by X-ray crystallography (reviewed in 6). In order to understand the mechanism underlying allosteric regulation it is of pivotal importance to identify the protein residues that interact with the allosteric effector. The allosteric site for malate was unequivocally identified in the three-dimensional structure of the PEPC-C4 isoenzyme from Flaveria trinerva (FtPEPC-C4), which was crystallized bound to an aspartate molecule (21; PDB code 3ZGE). This allosteric site appears to be highly conserved in PEPC enzymes, since that of FtPEPC-C4 is very similar to the one previously found in the PEPC enzyme from Escherichia coli (22-24; PDB codes 1QB4, 1FIY, 1JQN). The allosteric site for G6P of ZmPEPC-C4 was proposed to be located between the two monomers of a dimeric unit of the native tetrameric PEPC structure because a sulfate anion was found bound at this place (24). This suggested location was in agreement with previous site-directed mutagenesis studies on residues assumed to be involved in binding G6P (25,26). Moreover, this is a conserved region in other plant,PEPCs as shown in the crystal structures of the PEPC-C3 isozymes from Flaveria pringlei (21, PDB code 3ZGB) and Arabidopsis thaliana (PDB code 5FDN). However, Schlieper et al. (27) have recently questioned the relevance of this site for activation. With regard to the allosteric site for neutral amino acids, its existence has been predicted on the basis of biochemical data, which indicate that it is distinct from the
allosteric site for G6P (11,13,17,18,28), but its position remained unknown. Site-directed mutagenesis experiments on ZmPEPC-C4 suggested that it might be formed by residues of the monomer-monomer interface of the dimeric units of the native tetramer, contiguously to the proposed G6P allosteric site (26), and by residues in a loop at the carboxy-terminal region of the protein (29).

With the aim of unequivocally locating and characterizing the allosteric site for neutral amino acids, and to find out the structural differences between monocot and eudicot PEPC-C4 isozymes accounting for their kinetic difference, specially in regard to their sensitivity to neutral amino acids activation, we obtained a new crystal structure of the ZmPEPC-C4 isozyme which showed, for the first time, a glycine molecule bound to its allosteric site at the predicted monomer-monomer interface, close to the proposed G6P-allosteric site. In addition, we found that a serine residue (Ser100) which was not previously identified by site-directed mutagenesis and does not directly interact with the bound glycine, is critical for allowing the binding of the neutral amino acids, while a lysine at this position, as found in the isozyme from Flaveria trinervia (FpPEPC-C4) would impede their binding. Accordingly, the ZmPEPC-C4 S100K mutant was insensitive to activation by neutral amino acids. Multiple alignment studies of the known PEPC-C4 amino acid sequences indicated that a serine or glycine is present at this position in PEPC-C4 isozymes from the Poaceae family, but it is a lysine in the isozymes from the monocot sedges (Cyperaceae family) or from eudicots families, thus explaining the reported lack of activation by neutral amino acids of eudicot PEPC-C4 isozymes. In addition, Ser100 appears to play an important role in determining other kinetic properties of ZmPEPC-C4, and presumably of every PEPC-C4 from grasses, different from those of the eudicot PEPC-C4 isozymes.

RESULTS

Crystallographic evidence of the allosteric-site for neutral amino acids in ZmPEPC-C4

The structure of the recombinant ZmPEPC-C4 in complex with glycine was determined at 3.3 Å resolution. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB, www.rcsb.org) with the accession code 5VYJ. All data collection and refinement statistics of the structure are summarized in Table 1.

This new crystal structure consists of a αβ-barrel with eight central β-strands surrounded by at least 44 α-helices (Fig. 2A), very similar to the other known PEPC three-dimensional structures—those of the enzyme from E. coli (22-24; PDB codes 1QB4, 1FY1 and 1JQN), the previously reported ZmPEPC-C4 crystal structure (24; PDB code 1JQO), the FpPEPC-C4 isozyme (21,27; PDB codes 4BXH, 4BXC and 3ZGE), and the C3 isozymes from F. pringlei (FpPEPC-C3; 21; PDB code 3ZGB) and Arabidopsis thaliana (PDB code 5FDN). The asymmetric unit of the crystal reported here contained the four subunits that constitute the biological unit of the enzyme, which can be considered a dimer of dimers, thus allowing the observation of the dimer-dimer interface for the first time. The asymmetric units of the previously reported PEPC crystals were either a dimer (1JQO) or two monomers belonging to different dimeric units of the tetramer (1QB4, 4JQN, 4BXH, 4BXC, 3ZGE, 3ZGB, and 5FDN). Also for the first time, in each monomer of the 5VYJ crystal a clear electron density was observed in crevices that open to the solvent at the monomer-monomer interface of the dimeric units of the tetramer, adjacent to the G6P allosteric site. A glycine molecule with occupation of one accounted for the extra electronic density, as indicated by a simulated annealing omit map Fe-Fe at 3.0 σ (Figs. 2B and 2C). Modeling other low molecular weight compounds present in the crystallization medium—glycerol, acetate or ethylene glycol—gave no satisfactory results (Supplementary Fig. S1).
The glycine molecule makes hydrogen bonds through its \(\alpha\)-carboxyl group with the side chains of Arg334 and Trp333 of one subunit, and through its \(\alpha\)-amino group with the side-chain carboxyl group of Glu229 of the neighboring subunit. Glu229 in turn is hydrogen bonded to the side chains of Ser100 and His104 of its own subunit. In addition to its binding to the \(\alpha\)-carboxyl of the bound glycine, Arg334 participates in a web of interactions with side chains of residues of the opposing subunit that contribute to the stabilization of the monomer-monomer interface. Of particular relevance appears to be the interaction with Glu928 and with Arg226, which is in turn bound to the side-chain carboxyl of Asp428 of the same subunit than Arg334. In the ZmPEPC-C4 5VYJ crystal structure, Arg334 and Glu928 close the allosteric site, acting as a lid that should be open to allow the entrance of the amino acid activator. The “open” conformation of this allosteric site can actually be observed in the previously reported 1JQO ZmPEPC-C4 structure, as shown in Fig. 2D, in which no activator molecule is bound. In the empty allosteric site for neutral amino acids of the 1JQO crystal there is free access of the amino acid activator due to the “outside” conformations of both Arg334 and Glu928. The high B-factors values of the side-chains of these two residues in the 1JQO structure (96 for Arg334 and 132 for Glu928, much higher than the average B-factors values) compared with those in the 5VYJ crystal (22 and 20, respectively, similar to the average B-factors value), indicate their mobility in the absence of the activator glycine, and the flexibility of this region needed to allow the entry and binding of the activator in the allosteric site. Indeed, the conformation of the residues surrounding the \(\alpha\)-amino group of glycine is similar in both structures, except that of Arg334, which side chain moves about 5 Å from a position exposed to the solvent in the 1JQO structure towards the \(\alpha\)-carboxyl group of glycine in the 5VYJ structure reported here (Fig. 2E). This movement, most likely induced by the binding of glycine, makes room for the side chain of Glu928, which carboxyl group also moves about 11 Å from its position in the protein surface towards the guanidinium group of Arg334 of the neighboring subunit, as shown in Fig. 2E. Also, in the 5VYJ structure, the imidazol ring of His53 rotates about 25 degrees from its position in the 1JQO structure, interacting with Glu928 of its own subunit (not shown for clarity of the figure). Therefore, the binding of glycine significantly contributes to the stabilization of the monomer-monomer interface of the ZmPEPC-C4 isozyme, not only by bridging the two monomers through the hydrogen bonds made by its \(\alpha\)-amino and \(\alpha\)-carboxyl groups, but also by promoting the formation of several hydrogen bonds not present when the activator is not bound. This may account, at least in part, for the activation effect of glycine.

The 5VYJ crystal reported here also show how the allosteric site for neutral amino acids and the proposed allosteric site for G6P are connected in the ZmPEPC-C4 isozyme, thus explaining the synergistic effects of both kinds of allosteric activators observed before (11,12,17,18). This connection is established by Asp228, which side-chain carboxyl group makes hydrogen bonds with the side chain of Arg232 (Fig. 2C), a conserved residue in PEPC-C4 isozymes (see below). This agrees with the previous observation that the affinity for glycine decreased when Asp228 was changed for an asparagine and that the affinity for both glycine and G6P decreased when Arg232 was changed for a glutamine (26). Therefore, the interactions of Asp228 with Arg232 appear to play an important role in the activation by both kinds of activators, neutral amino acids and G6P.

To explore whether L-serine or L-alanine—the other two neutral amino acids that have been reported as activators of ZmPEPC-C4 (11,15,18)—could also be bound in the same site as Gly, as well as to learn the possible mode of their binding, we constructed in silico energy-minimized models with L-alanine or L-serine docked in this region. We also docked D-serine and D-alanine inside the Gly-binding...
site given that they were also found to be activators of the ZmPEPC-C4 isozyme (11). In the models, shown in supplementary Fig. S2, the side chains of L/D-alanine or L/D-serine fit well into this site without producing any steric clash, and their α-carboxyl and α-amino groups interact with protein residues in the same manner as do the glycine groups in the crystal structure reported here. The results of these simulations support that the site where glycine was observed bound in the 5VYJ crystal is indeed the allosteric site for neutral amino acids.

**Comparison of ZmPEPC-C4-Gly and FtPEPC-C4 crystal structures**

To find out the structural differences between the PEPC-C4 isozymes that are allosterically activated or non-activated by neutral amino acids, we compared the region of the allosteric site for neutral amino acids of the 5VYJ crystal structure with the equivalent region in the reported crystal structures of FtPEPC-C4. In spite that every residue that is interacting with glycine in the 5VYJ ZmPEPC-C4 crystal structure is conserved in the FtPEPC-C4 isozyme—although some of them have a different conformation—we observed a critical change in a nearby residue which does not make a direct interaction with the bound glycine: Ser100 in the maize enzyme is changed for a lysine (Lys96) in the Flaveria enzyme. Interestingly, the side-chain amino group of Lys96 occupies a position similar to that of the α-amino group of the bound glycine in the ZmPEPC-C4 5VYJ crystal (Fig. 3A), but it is hydrogen bonded to the carboxyl group of Asp223 (equivalent to Asp228 in the maize enzyme) instead of being bonded to the carboxyl group of Glu224 (equivalent to Glu229 in the maize enzyme). Moreover, the side chain carboxyl of Asp223 has a position comparable to that of the α-carboxyl of the glycine bound in the maize enzyme and forms hydrogen bonds with the same tryptophanyl (Trp328 in FtPEPC-C4 equivalent to Trp333 in ZmPEPC-C4) and arginyl (Arg329 in FtPEPC-C4 equivalent to Arg334 in ZmPEPC-C4) residues as the activator molecule in the maize enzyme. Consequently, in the Flaveria PEPC-C4 crystals, the neutral amino acid allosteric-site does not exist, since the side chains of Lys96 and Asp223 totally fill up the cavity where glycine binds in the ZmPEPC-C4 5VYJ crystal (Fig. 3B). The different position of Asp223 in the FtPEPC-C4 crystals with respect to that of the equivalent Asp228 in the two known ZmPEPC-C4 structures is due to a -95° change in the psi angle of this aspartate, resulting in the movement of its side chain towards the region equivalent to the neutral amino acids allosteric site. Also, in the FtPEPC-C4 crystals Glu224 (equivalent to Glu229 in the maize enzyme) has a different conformation than in the ZmPEPC-C4 crystals; its side-chain is outside the Gly-binding site, interacting with Asn362 of the opposing subunit (Fig. 3C), while in the maize crystals the asparagine at the equivalent position (Asn368) is exposed to the solvent. Interestingly, the interactions between the activator molecule and protein residues that stabilize the dimer-dimer interface in the region of the neutral amino acid allosteric site of the maize enzyme are made between protein residues in the Flaveria enzyme. Finally, the side chain of Glu922 of FtPEPC-C4 (equivalent to Glu928 of ZmPEPC-C4, a residue that interacts with Arg334 in the 5VYJ crystal) is only observed in two of the FtPEPC-C4 crystals where it is exposed to the solvent (Supplementary Fig. S3).

**Biochemical evidence supporting the identification of the allosteric site for neutral amino acids in the ZmPEPC-C4-Gly crystal structure**

In order to prove the critical role of the residue at position 100 of ZmPEPC-C4 in its allosteric activation by neutral amino acids, we substituted Ser100 for a lysine by site-directed mutagenesis and compared the kinetics of activation by neutral amino acids of the S100K mutant with those of the wild-type ZmPEPC-C4 enzyme (Fig. 4). The assays were conducted at pH and concentrations of substrates believed
to be close to those prevailing in vivo under illumination conditions (pH 7.4, bicarbonate 0.1 mM, PEP 3 mM, free Mg\(^{2+}\) 0.4 mM, L-malate 20 mM) as previously discussed by Tovar-Méndez et al (18). As shown in Figures 4A and 4B, glycine greatly increases the wild-type recombinant enzyme activity even in the presence of the high L-malate concentrations characteristic of the day period, in a similar manner to that observed in the enzyme purified from maize leaves (18). In contrast, glycine was not able to activate the S100K mutant, neither in the absence nor in the presence of the inhibitor, thus confirming the importance of the size and charge of the residue at position 100 for the amino acid activator to bind to its allosteric site. Neither was the S100K mutant enzyme activated by L-serine and L-alanine (Figs. 4C and D). The kinetic parameters estimated from the neutral amino acid saturation data are given in Table 2.

The ZmPEPC-C4 S100K mutant also showed other significant kinetic differences with the ZmPEPC-C4 wild type (Fig. 5 and Table 3). These differences mainly consist in a higher affinity for the substrate and activator (Figs. 5A and B) and a lower affinity for the inhibitor (Fig. 5C). Interestingly, G6P counteracted the inhibition by L-malate of the S100K mutant (Fig. 5D), while it did not in the wild-type enzyme, as previously reported (18). Thus, the residue at position 100 of ZmPEPC-C4 appears to be determinant not only of its activation by neutral amino acids, but also of its general kinetic behavior. Interestingly, the global kinetic properties of the S100K mutant resemble more those of the PEPC-C4 isozymes from eudicot plants (19), than to the wild-type ZmPEPC-C4. These kinetic differences couldn’t be attributed to changes in the association state, folding or stability of the mutant enzyme. Both the mutant and wild-type recombinant enzymes were obtained in the soluble fraction of the E. coli cells extracts with a similar yield, indicating that both recombinant proteins were properly folded. Also, the native structure of both proteins is tetrameric (Supplementary Fig. S4A) and both exhibit a similar thermal stability in thermal-shift experiments with apparent transition temperatures (\(T_{m}\)) of 48.8 °C for the wild-type and 47.75 °C for the S100K mutant (Supplementary Fig. S4B).

Residue conservation at the allosteric site for neutral amino acid in plant PEPC-C4 isozymes

Since PEPC-C4 isozymes from monocots are activated by neutral amino acids while isozymes from eudicots are not, we considered of interest to investigate the degree of conservation at the amino acid positions of the allosteric site for neutral amino acids found in the ZmPEPC-C4 crystal structure (PDB 5VYJ). Using the proposed criterion to identify PEPC-C4 isozymes, that is, the presence of serine at position equivalent to 774 of the FtPEPC-C4 isozyme (30) equivalent to position 780 of the ZmPEPC-C4 isozyme, a total of 172 non-redundant protein sequences (allelic forms excluded) were initially identified as PEPC-C4 in the RefSeq protein database of the National Center for Biotechnology Information (NCBI), and two in the Phytozone v12.1.5 database. Of these, only 138 sequences are from C4 plants (118 from monocots and 20 from eudicots) and thus we consider them to be true PEPC-C4 isozymes (Supplementary Table S1); the rest were from eudicot C3 plants (3 sequences) or CAM plants (31 sequences) (Supplementary Table S2). The latter finding is consistent with the proposed shared origin of some C4 and CAM PEPC isozymes (31). Another PEPC sequence that possesses a Ser at position 780 is the one encoded by the gene EEF27881 (GeneBank code) from Ricinus communis, which is a bacterial–like PEPC, similar to those previously found in Arabidopsis and rice (32) and therefore was not included in our analysis. The only so far known sequence for a PEPC-C4 isozyme from a C4 plant from the Hydrocharitaceae family (that from Hydrilla verticillata) do not have a serine at this position. Hence, the presence of Ser780 is not a sufficient criterion to distinguish
PEPC-C4 from other PEPC isozymes, as previously pointed out by others (33).

Although our sequence analysis was restricted due to the fact that most of the reported PEPC-C4 sequences were partial, lacking the amino- and/or carboxyl-terminal regions (only 13 sequences from monocots and 5 sequences from eudicots were complete) and in the case of eudicots only sequences from 4 of the 14 families known to contain at least one C4 member have been reported to date, after multiple amino acid sequences alignments, were prepared sequence logos of selected residues of the monocot and eudicot sequences that show them to evaluate their degree of conservation. In the 16 PEPC-C4 sequences from grasses that show the residue at position 100 there is serine, as in ZmPEPC-C4, or glycine (11 and 5, respectively), which is a conservative change from the point of view of size and charge. The Hydrilla verticillata PEPC-C4 sequence also has a serine at position equivalent to Ser100 of ZmPEPC-C4, but in Hydrilla enzyme the possible activation by neutral amino acids has not been tested. The two PEPC-C4 sequences reported from sedges that include this position have lysine, similar to the only 8 PEPC-C4 sequences from different eudicot families that include the residue at this position (Supplementary Fig. S5A). The PEPC-C4 isozymes from sedges have not been biochemically characterized yet, so we do not know whether they are activated by neutral amino acids or not, although on the basis of the presence of lysine at position 100 we would predict that they are not. In regard to the PEPC-C4 isozymes from eudicots, the ones studied at this respect (Amaranthus retroflexus, Amaranthus tricolor, Amaranthus hypochondriacus, Portulaca oleracea) are insensitive to neutral amino acids (15,19). The evolutionary conservation of these two different kinds of residues—small and neutral or large and positively charged—at this critical position strongly supports the importance of this trait. We also found that the residues directly involved in the binding of the activator glycine (Glu229, Trp333 and Arg334, ZmPEPC-C4 numbering) are absolutely conserved in both monocot and eudicot PEPC-C4 sequences (Supplementary Fig. S5B). In regard to the residues involved in the web of hydrogen bonds that stabilize the conformation of the neutral amino acid allosteric site, His104, Arg226, Asp428, Glu928 (residues that interact with those that bind the activator molecule) Asp228 and Arg232 (the residue that link the neutral amino acid and the proposed G6P allosteric sites, as mentioned above) also are totally conserved, or conservatively changed in the case of His104 (changed to asparagine in some monocots or to glutamine in some eudicots sequences) and of Glu928, which is aspartate in some of the eudicot sequences (Supplementary Fig. S5C).

The conservation of other ZmPEPC-C4 residues that have been related to the allosteric site for neutral amino acids by previous site-directed mutagenesis studies (Lys927, Glu932, Lys934, Gly937, Lys940) (29), as well as the two Pro residues (Pro915 and Pro949) that flank the highly flexible loop in which these residues are included, was also investigated (Supplementary Fig. S5D). The polar nature of the charged residues is highly conserved and the two prolines are absolutely conserved, suggesting that this exposed loop exist in every PEPC-C4 isozyme and that the flexibility of the loop flanked by them is important for their activity and/or allosteric regulation. The most conspicuous change is that of Gly937, which is the most frequent residue in the PEPC-C4 monocot sequences, although some of them have aspartate or glutamate, and is glutamate and less often aspartate in the eudicot PEPC-C4 sequences.

DISCUSSION

Identification of the allosteric site for neutral amino acids

In order to understand the mechanism underlying allosteric regulation it is of pivotal importance to identify the protein residues that interact with the allosteric effectors. The crystal structure of ZmPEPC-C4 in complex with glycine reported here reveals the position in the
three-dimensional structure and the structural details of the allosteric site for neutral amino acids, which remained unknown. The comparison of ZmPEPC-C4 5VYJ crystal structure with that of the FlPEPC-C4 isozyme already known indicated that the main structural reason for the lack of activation by amino acids in PEPC-C4 isozymes from eudicot plants is the presence of a lysine at position 100 of the maize enzyme instead of the serine or alanine residue that the isozymes from monocots have. The crystallographic data show that there is no room for the binding of the neutral amino acids in the Flaveria enzyme because of the large and positively charged side-chain of this lysine and the interactions it makes with a nearby aspartate residue conserved in both monocot and eudicot PEPC-C4 isozymes. In fact, the cavity between dimers that forms the allosteric site for neutral amino acids in the ZmPEPC-C4 does not exist in the FlPEPC4 enzyme. Moreover, in spite of the scarce number of complete sequences reported to date, the conservation of serine at position 100 in the PEPC-C4 sequences from grasses and the presence of lysine at the equivalent position in those from sedges and eudicot strongly support the critical role played by this residue in determining the different susceptibility to activation by neutral amino acids of these two kinds of PEPC-C4 isozymes, which was confirmed by the lack of activation by these compounds of the ZmPEPC-C4 S100K mutant. Since other kinetic properties were also significantly affected in this mutant, we propose that at this respect there are two different kinds of PEPC-C4 isozymes: the Ser/Gly100-type or the Lys100-type. The available PEPC-C4 sequences indicate that of the three monocot families that include species with C4 metabolism (Cyperaceae, Hydrocharitaceae and Poaceae), the isozymes from plants of the Poaceae family (grasses), which are economically the most important ones, and probably the only PEPC-C4 from the Hydrocharitaceae family, are of the Ser/Gly100-type type, while those from plants of the Cyperaceae family (sedges) are of the Lys100-type. Likewise PEPC-C4 sequences from eudicot plants appears to be of the Lys100-type according to the absolute conservation of this residue in the known sequences that show this position.

Ser100 was not identified as an important residue for glycine activation in the previously reported site-directed mutagenesis studies, which concluded that Arg226, Asp228, Glu229, and Arg232 are the important ones (26). The ZmPEPC-C4 5VYJ crystal structure reported here with a glycine molecule bound shows that Glu229 actually participates in the binding of the activator molecule, and that the other three residues are involved in maintaining the proper conformation of the allosteric site by means of a web of hydrogen bonds (Fig. 2C). Later, it was suggested that Lys927 and Gly937 form the glycine binding site of the maize enzyme and that other charged residues in the loop to which they belong are also important for binding of the activator (29). However, the ZmPEPC-C4 5VYJ crystal structure reported here shows that neither Lys927 nor Gly937 participate in the binding of the activator. The residues mentioned in the referred work belong to an exposed loop that extends from Pro915 to Pro940 in ZmPEPC-C4 and from Pro909 to Pro945 in FlPEPC-C4. This loop is so flexible that a great part of it could not be observed in the reported crystal structures (Supplementary Fig. S3A). Lys927 is exposed to the solvent in every plant PEPC structure so far reported and Lys940 is also exposed to the solvent in the maize crystal structures but is not observed in the Flaveria ones. The residues equivalent to Glu932 and Lys934 are not observed in any of the reported PEPC-C4 crystal structures, while the residue at position equivalent to Gly937 is not observed in the FlPEPC-C4 crystal structures. Gly937 was proposed to be critical for the ability of monocot PEPC-C4 isozymes to be activated by neutral amino acids in contrast to the eudicot isozymes, because it was assumed that the presence of glycine at this position is a distinct feature of monocot PEPC-C4 isozymes that distinguish them from the eudicots ones where there is an aspartate or
glutamate at this position (29). However, our sequence analysis indicated that not all monocot sequences have Gly937, some of them, even of the Poaceae family, have aspartate or glutamate (Supplementary Table S1). This finding suggests that the residue at position 937 is not as critical for neutral amino acid activation as is the residue at position 100. Nevertheless, to find out the possible structural reason for the reported effect that the change of Gly937 of the maize enzyme for an aspartate had on the response to glycine (29), we performed the in silico mutations G937D and G937E (Supplementary Fig. S6). Using the ZmPEPC-C4 5VYJ crystal structure with glycine bound, we found that either an aspartate or glutamate at position equivalent to Gly937 would be exposed to the solvent, not interacting with other protein residue in any of their possible rotamers (Supplementary Fig. S6A and B). But when we simulated the possible exposed conformations of Arg334 and Glu928 that could occur when the allosteric site for neutral amino acid is empty as indicated by the ZmPEPC-C4 1QJO structure, we found rotamers of Arg334 and of Asp/Glu937 that can interact (Supplementary Fig. S6C and D). This interaction would stabilize the exposed conformation of Arg334, opposing to its movement towards the bound activator molecule to form the critical hydrogen bond with the α-carboxyl group, as well as the other hydrogen bonds other protein residues that stabilize the activator-bound conformation of the allosteric site. Therefore, it appears that the presence of aspartate/glutamate at position 937 in PEPC-C4 isozymes from monocot plants could decrease the affinity to neutral amino acids by somehow limiting the movement of Arg334 into the allosteric site, where this residue importantly contributes to the binding of the carboxyl group of the activator. Additionally, the substitution of charged residues for neutral ones in the Pro915 to Pro949 loop could affect the dynamics of this loop, which in turn may negatively affect the allosteric transition induced by the binding of neutral amino acids. This dynamic effect could explain the decreased sensitivity to glycine caused by the amino acid substitutions in this loop.

**Implications for the understanding of the kinetics and allosteric regulation of the PEPC-C4 isozymes**

In the absence of the inhibitor L-malate, neutral amino acids activate ZmPEPC-C4 roughly to the same extent than G6P (8,10,11,18), but when this enzyme is assayed in the presence of malate, glycine or serine counteract the inhibition of the ZmPEPC-C4 enzyme by L-malate in a much higher degree than G6P (17,34). This is particularly true when the assays are carried out at conditions close to those prevailing during the day—i.e., high L-malate concentrations (35,36) and low CO₂ concentration (37)—, when G6P is unable to significantly revert the inhibition caused by L-malate while glycine or L-serine offset this inhibition, producing an enzyme almost as active as that in the absence of the inhibitor (18,19 and results in the present work shown in Fig. 4B). At first sight, this result contradicts previous reports of G6P offsetting malate inhibition of ZmPEPC-C4 isozymes (10,17) and other PEPC-C4 isozymes such as that from *Hydrida* (38) and those from eudicot plants reported by Rosnow (39). In the case of the eudicot isozymes, the reports are consistent with the results obtained with the ZmPEPC-C4 S100K mutant presented here. In the case of the monocot PEPC-C4 isozymes it has to be taken into account that these previous results were obtained at non-physiologically high Mg²⁺ ions and bicarbonate concentrations and under those experimental conditions, the concentration of the preferred substrate, the complex Mg-PEP (13), when the total PEP concentration is kept at 3 mM is almost saturating, which greatly increases the affinity for the activator G6P and greatly diminished that for the inhibitor L-malate. But under conditions near to those estimated to prevail during the day, when the concentration of Mg-PEP is well below saturation, the activation by glycine and/or L-serine, not by G6P, appears to be crucial for
achieving appreciable levels of ZmPEPC-C4 activity and, therefore, a significant rate of CO₂ assimilation (18). Given this, we believe that it is of pivotal importance that PEPC-C4 isozymes are in vitro tested at near physiological concentrations of their substrates in order to correctly evaluate the role played by their allosteric effectors.

If the activation by neutral amino acids is so critical, the question arises: ¿how do contend with L-malate inhibition those C4 plants whose PEPC-C4 isozyme is not activated by neutral amino acids? Noteworthy, the kinetic properties of the ZmPEPC-C4 S100K mutant suggest that it would be more active than the wild type enzyme at low PEP concentrations in the absence of the activator regardless of the absence or presence of L-malate. Also of particular relevance at this respect is that G6P is able to revert in a significant extent the inhibition by malate of the S100K mutant, contrary to the wild-type enzyme. These kinetic properties of the S100K mutant are, at least qualitatively, more similar to those of the PEPC-C4 isozyme from the eudicot plant Amaranthus hypochondriacus (AhPEPC-C4) than to those of the wild type ZmPEPC-C4. Our findings suggest that these kinetic differences between the Ser/Gly100-type and the Lys100-type PEPC4 isozymes could explain why the PEPC-C4 isozymes from eudicot plants, and likely also those from sedges although they have not been kinetically characterized to date, do not require the activation by neutral amino acids to efficiently work under physiological conditions.

CONCLUSION

The crystallographic and site-directed mutagenesis data reported here unequivocally identify the location of the allosteric site for neutral amino acids in the ZmPEPC-C4 three-dimensional structure and pointed out the importance of the residue at position 100 of ZmPEPC-C4 for neutral amino acid activation and other kinetic properties, which was confirmed by site-directed mutagenesis experiments. Residues conservation analysis supports our proposal that the nature of the residue at this position plays a similar critical role on the catalysis and allosteric properties of every PEPC-C4 isozyme. We believe that our findings are of great biotechnological interest given the current efforts for the molecular engineering of grass C3 plants, such as rice, wheat and barley in order to improve their productivity.

EXPERIMENTAL PROCEDURES

DNA constructs and site-directed mutagenesis

The ppc-C4 gene coding for ZmPEPC-C4 with the GeneBank accession number CAD60555 cloned into plasmid pTM94, was a kind gift from Prof. Izui (Kyoto University). The protein encoded by this gene differs from the one encoded by the gene with GeneBank accession number CAA33317 in only two amino acid residues: Pro482 and Asp 509 in CAD60555 are Ser482 and Glu509, respectively, in CAA33317. These are polymorphic changes that seem not to affect either the structure or function of the PEPC-C4 isozyme. The gene was transferred from the plasmid pTM94 to the pET32a(+) (Novagen) expression vector by ligating the NcoI-HindIII digested insert into the previously linearized plasmid with the same restriction enzymes. Since this gene contains an internal NcoI restriction site, a partial NcoI digestion was used to isolate the full-length gene prior to its insertion into pET32a(+). The plasmid was purified from transformed TOP-10 E. coli cells selected by growing the cells at 100 μg/ml ampicillin, and analyzed by restriction analysis, PCR and sequencing of the insert. The ZmPEPC-C4 sequence within pET32a(+) did not contain any mutations and was in frame to be expressed as a fusion protein with a thrombin-cleavable double tag of thioredoxin and 6X-histidine towards the N-terminus of the ZmPEPC-C4 protein. Once the construct was verified, it was used to transform BL21-CodonPlus(DE3)-RIL E. coli cells (Novagen) to overexpress the gene.

To generate the ZmPEPC-C4 S100K mutant, the forward and reverse primers 5ʼ...
GCCATCCTCGGGCAGTCCATCCTCG
AC 3’ and 5’
GTGCAGGATGGACTTCGCCACGAGGAT
GGC 3’, respectively, were used to substitute serine at position 100 by lysine. The ZmPEPC-
C4 gene cloned into the pET32a(+) plasmid was used as the template for the site-directed
mutagenesis by PCR, using the QuickChange II site-directed mutagenesis kit (Agilent
Technologies), following manufacturer’s instructions. The DNA was sequenced to
confirm that the desired mutation was present and that no other nucleotide changes occurred.
Finally, the mutant protein was overexpressed in BL21-CodonPlus(DE3)-RIL E. coli cells
(Novagen).

**ZmPEPC-C4 expression, purification and assay**

The overexpression of both the recombinant wild-type and S100K ZmPEPC-
C4 proteins was achieved by inducing cells grown in Luria-Bertani (LB) broth added with
100 mg mL⁻¹ of ampicillin at an OD₆₀₀ of 0.6 by the addition of 0.1 mM isopropyl β-D-
thiogalactoside (IPTG) followed by an incubation of 7 h at 25 ºC. Cells were harvested by
centrifugation at 6,000 g for 10 min. The pellet was suspended in 10 mL of 50 mM
Hepes-KOH buffer, pH 7.5, containing 2 mM 2-mercaptoethanol, 50 mM KCl, 1 mM
phenylmethanesulfonyl fluoride (PMSF) and 10% (v:v) glycerol (Buffer A), and sonicated
for 20 min. Cell debris was removed by centrifugation at 15,000 rpm for 20 min. The
ZmPEPC-C4 protein was purified using the Protino Ni-TED resin (Macherey-Nagel, Düren, Germany) from which it was eluted
with imidazole 150 mM in Buffer A. Imidazole
was then removed by centrifugal concentration, using Amicon Ultra 30 (Millipore). The concentrated enzyme was digested for 2 h at 25
ºC with enterokinase (New England Biolabs),
and then applied to a MonoQ chromatography
column and eluted with a gradient from 0 to
400 mM of potassium phosphate, pH 7.4.
Protein concentrations were determined
spectrophotometrically by A280, using the
extinction coefficient of 111,730 M⁻¹ cm⁻¹
predicted from amino acid sequence with ExPASy ProtParam (40).

ZmPEPC-C4 activity was assayed spectrophotometrically at 30 °C in a coupled
enzymatic assay with malate dehydrogenase by
monitoring NADH oxidation at 340 nm (ε₅₅₀ =
6,220 M⁻¹ cm⁻¹), as described (28). The
standard assay medium consisted of 100 mM
HEPES-KOH buffer, pH 7.3, containing 0.1
mM NaHCO₃, 0.2 mM NADH, 3 mM total
PEP, 0.4 mM free Mg²⁺ and five units of
malate dehydrogenase. Enzyme activity in the
presence of allosteric effectors was assayed using
the standard assay with the indicated
effector concentrations. Each assay was
performed in duplicate and with at least two
different enzyme preparations. One unit is
defined as the amount of enzyme needed to
catalyze the formation of 1 μmol of oxalacetate
per min under our experimental conditions.

**Crystallization, X-ray data collection, structure solution and refinement**

Crystals of ZmPEPC-C4 with glycine
bound were grown at 18 ºC by the hanging-
drop vapor diffusion method mixing equal
volumes of protein and reservoir solution,
which consisted of 100 mM TRIS-HCl pH 8.5,
200 mM sodium acetate trihydrate, 100 mM
potassium/sodium tartrate-4-hydrate, and 15%
w/v) PEG 4000. Protein concentration was 10
mg mL⁻¹ in 10 mM HEPES-KOH buffer, pH 7.5, 1 mM DTT, 10 mM MgCl₂ and 100 mM
glycine. The crystals were cryoprotected in
reservoir solution plus 20% (v/v) glycerol and
cryo-cooled in N₂. X-ray diffraction data of the
ZmPEPC-C4-Gly were collected at 100 K at the
Advanced Photo Source, GM/CA, beamline 23
ID-B located at the Argonne National
Laboratory, Chicago IL.

The data were integrated using XDS
(41) and scaled and truncated with programs
from the CCP4 suite (42). Because of
anisotropy, the dataset was submitted to the
Diffraction Anisotropy Server (http://services.mbi.ucla.edu/anisoscale) (43) to
perform ellipsoidal truncation and anisotropic
scaling. The ellipsoidal truncated dataset, with the upper resolution limit cut at 3.7 Å, 3.3 Å, and 2.7 Å along a*, b*, and c* respectively, was used in later stages of the model refinement. The initial phases were obtained by molecular replacement with the program Phaser (44) using the coordinates of the reported ZmPEPC-C4 structure (PDB code 1JQO) as a starting model. Alternating cycles of automatic and manual refinement were carried out with the standard protocols of Phenix (45) monitoring the $R_{\text{work}}$ and $R_{\text{free}}$ split during the whole process. Phenix was also used for atomic positions, group atomic displacement parameters (ADP), and translation, libration and screw-rotation displacement (TLS). The simulated annealing omit map was calculated with Phenix following simulated annealing refinement at 3,000 K. The program Coot (46) was used to analyze the electron density maps ($2F_o-F_c$ and $F_o-F_c$). Structural alignments were performed with Coot and PyMOL (http://www.pymol.org).

**Kinetic data analysis**

Kinetic data were analyzed by nonlinear regression calculations. Initial velocity data obtained varying the concentration of activator at constant concentration of substrates were fitted to Eq. 1:

$$v_a = \left\{ v_a \max \times [A]^b / (A_{0.5}^b + [A]^b) \right\} + v_0 \quad \text{(Eq. 1)}$$

where $v_a$ and $v_0$ are the experimentally determined initial velocities in the presence and absence of activator, respectively; $v_a \max$ is the estimated maximum activity at saturating activator concentrations; $[A]$ is the activator concentration; $A_{0.5}$ is the concentration of activator that gives half-maximum activation at fixed concentrations of substrate; and $h$ is the Hill number indicative of the degree of cooperativity in the binding of the activator.

Kinetic data obtained at varied concentration of substrate PEP were fitted to a Hill equation (Eq. 2):

$$v_0 = V_{\max} \times [S]^b / (S_{0.5}^b + [S]^b) \quad \text{(Eq. 2)}$$

where $v_0$ is the experimentally determined initial velocity; $V_{\max}$ the estimated maximum velocity; $[S]$ the concentration of the substrate; $S_{0.5}$ the concentration of substrate that gives half-maximum velocity; and $h$ the Hill number indicative of the degree of cooperativity in the binding of the substrate.

When the concentration of L-malate was varied at constant concentration of substrates, the experimental data were fitted to Eq. 3:

$$v_i = \left\{ v_0 \times I_{S_{0.5}}^b / (I_{S_{0.5}}^b + [I]^b) \right\} + v_{i \min} \quad \text{(Eq. 3)}$$

where $v_i$ and $v_0$ are the experimentally determined initial velocities in the presence and absence of inhibitor, respectively, $v_{i \min}$ is the estimated residual velocity at saturation of the inhibitor, $[I]$ is the inhibitor concentration, $I_{S_{0.5}}$ is the concentration of inhibitor that gives half-maximum inhibition at fixed concentrations of substrates, and $h$ the Hill number indicative of the degree of cooperativity in the binding of the inhibitor.

**Retrieval and sequence analysis of PEPC-C4 orthologs**

Plant genomes typically contain several sequences coding for members of the PEPC family. Some of them were considered alleles given their high similarity at the nucleotide and amino acid sequences or because they mapped to a single locus, when mapping information was available, but other were considered true isozymes because they are located in different loci. To clarify this, we searched all available PEPC orthologs from plants, performing a blastp search on the RefSeq collection of the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov) protein database using as query the amino acid sequence of ZmPEPC-C4 (GenBank accession CA33317 that corresponds to the PEPC isozyme involved in C4 photosynthesis in maize (47) coded by a *ppc-C4* gene (48). In order not to exclude any candidate plant PEPC-C4, the E-value threshold was set at 10; the scoring matrix used was BLOSUM62, with gap opening and gap extension costs of 11 and 1 respectively. All retrieved sequences belong to
Viridiplantae. To include sequences for which we would have mapping, allelism and paralogy information, we also performed blastp searches on version 11.1.5 of the Phytozome database (https://phytozome.jgi.doe.gov/), which contains data for 63 wholly sequenced Viridiplantae genomes. Only hits with an E-value higher than 10e-3 were excluded. The resulting sequences were aligned with ClustalX2 (49), and the alignment was refined by hand, in some cases with the help of structural alignments using homology models, and stripped of redundant sequences.

For the construction of the sequence logos that show the conservation of residues at positions related to the neutral amino acid allosteric site, plant PEPC-C4 sequences were chosen based on the criterion of having serine at position equivalent to ZmPEPC-C4 780 (30). Only sequences belonging to a C4 plant were considered; PEPC sequences from C3 or CAM plants that have this residue were excluded. Sequence logos were constructed with the WebLogo3 server (http://weblogo.berkeley.edu/; 50,51).
ACKNOWLEDGEMENTS
We are indebted to Prof. Izui (Kyoto University) for providing us the ZmPEPC-C4 cDNA. We are also very grateful to M.S. F. Barreda-Huerta and Dr. I. Bustos-Jaimes for subcloning the pepc-C4 gene, to B.Sc. S.P. Rojas-Trejo and Dr. H. Serrano-Posada for preliminary crystallization attempts of the ZmPEPC-C4 isozyme, and to Dr. E. Rudiño-Piñera and Dr. A. Zárate-Romero for helpful discussions. We thank the staff of beamline 23 ID-B(GM/CAT) of the Advanced Photon Source at Argonne National Laboratory for their use and assistance with X-ray beamline and advice during collection. We also thank the lecturers of the “2013 CCP4/APS Crystallographic School” at Argonne National Laboratory for helpful discussions of data processing and refinement strategies and Dr. Dominika Borek for assistance with structure solution.

CONFLICT OF INTEREST
The authors declare that they have no conflict of interest with the contents of this article.
REFERENCES

1. Hatch, M. D. (1987) C4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochem. Biophys. Acta* **895**, 81–106
2. Bowes, G., Rao, S. K., Estavillo, G. M. and Reiskind, J. B. (2002) C4 mechanisms in aquatic angiosperms: comparisons with terrestrial C4 systems. *Funct. Plant Biol.* **29**, 379–392
3. Edwards, G. E., Franceschi, V. R. and Voznesenskaya, E. V. (2004) Single cell C4 photosynthesis versus the dual-cell (Kranz) paradigm. *Annu. Rev. Plant Biol.* **55**, 173–196
4. Andreo, C. S., González, D. H., Iglesias, A. A. (1987) Higher plant phosphoenolpyruvate carboxylase: Structure and regulation. *FEBS Lett.* **213**, 1–8
5. Chollet, R., Vidal, J. and O’Leary, M. H. (1996) Phosphoenolpyruvate carboxylase: A ubiquitous, highly regulated enzyme in plants. *Annu. Rev. Plant Physiol. Plant. Mol. Biol.* **47**, 273–298
6. Izui, K., Matsumura, H., Furumoto, T. and Kai, Y. (2004) Phosphoenolpyruvate carboxylase: A new era of structural biology. *Annu. Rev. Plant Biol.* **55**, 69–84
7. Wong, K. F. and Davies, D. D. (1973) Regulation of phosphoenolpyruvate carboxylase of *Zea mays* by metabolites. *Biochem. J.* **131**, 451–458
8. Uedan, K. and Sugiyama, T. (1976) Purification and characterization of phosphoenolpyruvate carboxylase from maize leaves. *Plant Physiol.* **57**, 906–910
9. Stiborová, M. and Lebllová, S. (1985) Activation of maize (*Zea mays* L.) phosphoenolpyruvate carboxylase by glucose-6-phosphate and glycine: effects of pH and Mg2+. *Photosynthetica* **19**, 177–182
10. Doncaster, H. D. and Leegood, R. C. (1987) Regulation of phosphoenolpyruvate carboxylase activity in maize leaves. *Plant Physiol.* **84**, 82–87
11. Bandarian, V., Poehner, W. J. and Grover, S. D. (1992) Metabolite activation of crassulacean acid metabolism and C4 phosphoenolpyruvate carboxylase. *Plant Physiol.* **100**, 1411–1416
12. Mújica-Jiménez, C., Castellanos-Martínez, A. and Muñoz-Clares, R. A. (1998) Studies of the allosteric properties of maize leaf phosphoenolpyruvate carboxylase with the phosphoenolpyruvate analog phosphomycin as activator. *Biochim. Biophys. Acta* **1386**, 132–144
13. Tovar-Méndez, A., Rodríguez-Sotres, R., López-Valentín, D. M. and Muñoz-Clares, R. A. (1998) Re-examination of the roles of PEP and Mg2+ in the reaction catalysed by the phosphorylated and non-phosphorylated forms of phosphoenolpyruvate carboxylase from leaves of *Zea mays*. *Biochem. J.* **332**, 633–642
14. Rodríguez-Sotres, R. and Muñoz-Clares, R. A. (1990) Kinetic evidence of the existence of a regulatory phosphoenolpyruvate binding site in maize leaf phosphoenolpyruvate carboxylase. *Arch. Biochem. Biophys.* **276**, 180–190
15. Nishikido, T. and Takanashi, H. (1973) Glycine activation of PEP carboxylase
from monocotyledoneous C₄ plants. *Biochim. Biophys. Res. Commun.* **53**, 126–133

16. Garson, L. and Gray, V. (1991) Regulation of phosphoenolpyruvate carboxylase from maize leaves by nitrate and alanine. *Biochem. Int.* **23**, 299–305

17. Gillinta, J. and Grover, S. D. (1995) Kinetic interactions of glycine with substrates and effectors of phosphoenolpyruvate carboxylase from maize leaves. *Photosynth. Res.* **45**, 121–126

18. Tovar-Méndez, A., Mújica-Jiménez, C., Muñoz-Clares, R. A. (2000) Physiological implications of the kinetics of maize leaf phosphoenolpyruvate carboxylase. *Plant Physiol.* **123**, 149–160

19. Güemez-Toro, R., Mújica-Jiménez, C. and Muñoz-Clares, R. A. (2012) Allosteric regulation of the photosynthetic C₄ isoenzyme of phosphoenolpyruvate carboxylase: a comparative study between enzymes from monocot and eudicots. *J. Mex. Chem. Soc.* **56**, 58–66

20. González, D. H., Iglesias, A. A. and Andreo, C. S. (1984) On the regulation of phosphoenolpyruvate carboxylase activity from maize leaves by L-malate. Effect of pH. *J. Plant Physiol.* **116**, 425–434

21. Paulus, J. K., Schlieper, D. and Groth, G. (2013) Geater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. *Nat. Commun.* **4**, 1518

22. Matsumura, H., Terada, M., Shirakata, S., Inoue, T., Yoshinaga, T., Izui, K. and Kai, Y. (1999) Plausible phosphoenolpyruvate binding site revealed by 2.6 Å structure of Mn²⁺-bound phosphoenolpyruvate carboxylase from *Escherichia coli*. *FEBS Lett.* **458**, 93–96

23. Kai, Y., Matsumura, H., Inoue, T., Terada, K., Nagara, Y., Yoshinaga, T., Kihara, A., Tsumura, K. and Izui, K. (1999) Three-dimensional structure of phosphoenolpyruvate carboxylase: A proposed mechanism for allosteric inhibition. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 823–828

24. Matsumura, H., Xie, Y., Shirakata, S., Inoue, T., Yoshinaga, T., Ueno, Y., Izui, K. and Kai, Y. (2002) Crystal structures of C₄ form maize and quaternary complex of *E. coli* phosphoenolpyruvate carboxylases. *Structure* **10**, 1721–1730

25. Takahashi-Terada, A., Kotera, M., Ohshima, K., Furumoto, T., Matsumura, H., Kai, Y. and Izui, K. (2005) Maize phosphoenolpyruvate carboxylase. Mutations at the putative binding site for glucose 6-phosphate caused desensitization and abolished responsiveness to regulatory phosphorylation. *J. Biol. Chem.* **280**, 11798–11806

26. Yuan, J., Sayegh, J., Mendez, J., Sward, L., Sanchez, N., Sanchez, S., Waldrop, G. and Grover, S. (2006) The regulatory role of residues 226-232 in phosphoenolpyruvate carboxylase from maize. *Photosynth. Res.* **88**, 73–81

27. Schlieper, D., Förster, K., Paulus, J. K. and Groth, G. (2014) Resolving the activation site of positive regulators in the plant phosphoenolpyruvate carboxylase. *Mol. Plant* **7**, 437–440
28. Tovar-Méndez, A., Mújica-Jiménez, C. and Muñoz-Clares, R. A. (1997) Desensitization to glucose 6-phosphate of phosphoenolpyruvate carboxylase from maize leaves by pyridoxal 5’phosphate. Biochim. Biophys. Acta 1337, 207–216

29. Gonzalez, L., Sánchez, S., Home, J., Kanzaki, G. and Grover, S. (2007) Mutational analysis of a proposed glycine binding site in PEP carboxylase. FASEB J 21, 806.3

30. 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

31. Christin, P. A., Arakaki, M., Osborne, C. P., Bräutigam, 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 of C4 and CAM metabolism. J. Exp. Bot. 65, 3609–36021

32. Sánchez, R. and Cejudo, J. (2003) Identification and expression analysis of a gene encoding a bacterial-type phosphoenolpyruvate carboxylase from Arabidopsis and rice. Plant Physiol. 132, 949–957

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

34. Gao, Y. and Woo, K. C. (1996) Regulation of phosphoenolpyruvate carboxylase in Zea mays by protein phosphorylation and metabolites and their roles in photosynthesis. Aust. J. Plant Physiol. 23, 25–32

35. Leegood, R. C. (1985) The intercellular compartmentation of metabolites in leaves of Zea mays L. Planta 164, 163–171

36. Stitt, M. and Heldt, H. W. (1985) Generation and maintenance of concentration gradients between the mesophyll and bundle sheath in maize leaves. Biochim. Biophys. Acta 80, 400–414

37. Heineke, D., Riens, B., Grosse, H., Hoferichter, P., Peter, U., Flügge, U. I. and Heldt, H. W. (1991) Redox transfer across the inner chloroplast envelope membrane. Plant Physiol. 95, 1131–1137

38. Rao, S. K., Reiskind, J. B., Bowes, G. (2008) Kinetic analyses of recombinant isoforms of phosphoenolpyruvate carboxylase from Hydrilla verticillata leaves and the impact of substituting a C4-signature serine. Plant Sci. 174, 475–483

39. 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

40. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. and Bairoch, A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31, 3784–3788
41. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
42. Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
43. Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. and Eisenberg, D. (2006) Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8060–8065
44. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674
45. 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
46. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
47. Hudspeth, R. L. and Grula, J. W. (1989) Structure and expression of the maize gene encoding the phosphoenolpyruvate carboxylase isoenzyme involved in C₄ photosynthesis. *Plant Mol. Biol.* **12**, 579–589
48. 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.* **14**, 1241–1247
49. Larkin, M. A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948
50. Crooks, G. E., Hon, G., Chandonia, J. M. and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190
51. Schneider, T. D. and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100
52. 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
FOOTNOTES
This work was supported by DGAPA-UNAM (PAPIIT grant IN216911) and CONACYT (grant 252123) to R.A.M.-C. RGT was the recipient of a CONACYT doctoral scholarship.
Table 1. Crystallographic data collection and refinement statistics of ZmPEPC-C4 in complex with Gly

| **Data collection**\(^a\) |  |
|--------------------------|------------------|
| X-ray source             | 23-ID B, APS     |
| Wavelength (Å)           | 0.9793           |
| Space group              | \(P2_12_12_1\)   |
| Cell dimensions          |  |
| \(a, b, c\) (Å)          | 156.95, 167.24, 242.90 |
| \(α, β, γ\) (º)          | 90, 90, 90       |
| \(R_{merge}\) (%)        | 12.2 (43.7)      |
| \(I/σ(I)\)               | 11.3 (2.8)       |
| Completeness (%)         | 93.28 (62.19)    |
| Unique reflections       | 90116 (5966)     |
| Redundancy               | 3.9 (2.4)        |

| **Refinement**           |  |
| Resolution range (Å)     | 57.1-3.3\(^\dagger\) (3.4-3.3) |
| \(R_{work}/R_{free}\) (%)| 21.0/25.3         |
| No. of atoms             |  |
| Protein                  | 29265             |
| Glycine                  | 25                |
| Acetate                  | 52                |
| Average \(B\)-factors (Å\(^2\)) |  |
| Protein                  | 27.50             |
| Glycine                  | 12.75             |
| Acetate                  | 22.77             |
| rms deviations from ideal stereochemistry |  |
| Bond lengths (Å)         | 0.007             |
| Bond angles (º)          | 0.98              |
| Ramachandran Plot        |  |
| Most favored regions (%) | 94.30             |
| Allowed regions (%)      | 5.02              |

| **PDB code**             | 5VYJ              |

\(^a\)Values in parentheses correspond to the highest resolution shell. \(^\dagger\)Diffracted anisotropically to a resolution of 3.7 Å, 3.3 Å, and 2.7 Å along \(a^*\), \(b^*\), and \(c^*\), respectively. Data were anisotropically truncated and scaled with the Diffraction Anisotropy Server (http://services.mbi.ucla.edu/anisoscale).
Table 2. Apparent kinetic parameters of ZmPEPC-C4 wild-type and mutant S100K enzymes

| Parameter                     | Wild-type | S100K |
|-------------------------------|-----------|-------|
| Saturation by Gly             |           |       |
| $v_0$                         | 4.5       | 8.5   |
| $v_{\text{max}}$              | 24.3 ± 0.3| NE    |
| $A_{0.5}$                     | 2.0 ± 0.1 | NE    |
| Saturation by Gly + 20 mM malate |           |       |
| $v_0$                         | 0.02      | 2.9   |
| $v_{\text{max}}$              | 11.5 ± 0.2| NE    |
| $A_{0.5}$                     | 16.3 ± 0.6| NE    |
| Saturation by L-Ala           |           |       |
| $v_0$                         | 4.8       | 8.3   |
| $v_{\text{max}}$              | 23.4 ± 0.3| NE    |
| $A_{0.5}$                     | 5.9 ± 0.3 | NE    |
| Saturation by L-Ser           |           |       |
| $v_0$                         | 5.8       | 10.1  |
| $v_{\text{max}}$              | 16.4 ± 0.6| NE    |
| $A_{0.5}$                     | 6.2 ± 0.9 | NE    |

$^a$Values ± S.D. were estimated by the best fit to Eq. 1, fixing the $h$ value to 1 and $v_0$ to the experimentally determined values. Velocities ($v_0$ and $v_{\text{max}}$) are given as units/mg protein and $A_{0.5}$ as mM. Assays were carried out at pH 7.3, 0.4 mM free Mg$^{2+}$, 0.1 mM bicarbonate and 3 mM total PEP. NE, values not estimated due to the absence of activation.
Table 3. Apparent kinetic parameters of ZmPEPC-C4 wild-type and mutant S100K enzymes

| Parameter                        | Wild-type | S100K  |
|----------------------------------|-----------|--------|
| Saturation by total PEP         | \( V_{\text{max}} \) | 19.9 ± 0.8 | 13.4 ± 0.4 |
|                                  | \( S_{0.5} \)  | 17.5 ± 1.3 | 4.6 ± 0.5  |
|                                  | \( h \)     | 1.6 ± 0.1  | 1.8 ± 0.3  |
| Saturation by G6P\(^b\)         | \( v_0 \)   | 5.4       | 12.6       |
|                                  | \( v_{a\text{ max}} \) | 22.7 ± 0.2 | 13.1 ± 0.8 |
|                                  | \( A_{0.5} \) | 3.4 ± 0.1  | 0.9 ± 0.1  |
|                                  | \( h \)     | 1.4 ± 0.0  | 1.8 ± 0.5  |
| Saturation by G6P + 20 mM malate\(^b\) | \( v_0 \)   | 0.01      | 2.2        |
|                                  | \( v_{a\text{ max}} \) | NE       | 13.9 ± 0.7 |
|                                  | \( A_{0.5} \) | NE       | 5.9 ± 0.7  |
|                                  | \( h \)     | NE       | 1.0 ± 0.0  |
| Saturation by malate\(^b\)      | \( v_0 \)   | 6.9       | 14.4       |
|                                  | \( v_{1\text{ min}} \) | 2.21     | 7.6        |
|                                  | \( I_{50} \) | 0.23 ± 0.0 | 1.0 ± 0.01 |
|                                  | \( h \)     | 1.3 ± 0.1  | 4          |

\(^a\)Values ± S.D. were estimated by the best fit to Eq. 2 for saturation by total PEP, to Eq. 1 for saturation by G6P in the absence and presence of malate, or to Eq. 3 for saturation by malate; \( v_0 \) was fixed to the experimentally determined values. Velocities (\( v_0, V_{\text{max}}, v_{a\text{ max}} \) and \( v_{1\text{ min}} \)) are given as units/mg protein and \( S_{0.5}, A_{0.5} \) and \( I_{50} \) as mM. Assays were carried out at pH 7.3, 0.4 mM free Mg\(^{2+}\) and 0.1 mM bicarbonate. \(^b\)The apparent kinetic parameters for the allosteric effectors were determined at fixed 3 mM total PEP. NE, values not estimated due to the low enzyme activity under these conditions.
FIGURE LEGENDS

Figure 1. Schematic representation of the allosteric regulation of ZmPEPC-C4.
In C4-plants, the first carboxylation reaction that incorporates the atmospheric CO₂ in an organic compound is catalyzed by PEPC-C4 isozymes located in the mesophyll cells. The product of this reaction, oxaloacetate, is then reduced to malate, as in maize shown here, or converted to aspartate in other plants; both compounds are important allosteric inhibitors of most of PEPC enzymes. During the illumination period L-malate accumulated up to high concentrations in the mesophyll cells, from where passively diffuses to the bundle-sheath cells, where it is decarboxylated to yield CO₂—which is then incorporated in the Calvin-Benson cycle by RUBISCO—and pyruvate, which moves back to the mesophyll cells regenerating phosphoenolpyruvate (PEP). The allosteric activators trioses-P and hexoses-P are mainly produced in the bundle-sheath cells from the 3-phosphoglycerate (PGA) formed in the RUBISCO-catalyzed reaction, but also may be produced in mesophyll cells given that PGA may diffuse from the bundle-sheath cells. Neutral amino acids (glycine and serine) are intermediates of the photorespiration pathway; they may also be transported to the mesophyll cell, where they activate PEPC-C4 isozymes from monocot grasses thus counteracting malate inhibition.

Figure 2. The allosteric-site for neutral amino acids of ZmPEPC-C4. (A) Cartoon and surface representations of the fold and secondary structure elements of the tetrameric ZmPEPC-C4 in complex with glycine. Red arrows indicate the position of the allosteric-site for neutral amino acids with a glycine bound, of the active site with an acetate bound, of the putative allosteric-site for G6P with also an acetate bound, and of the empty allosteric site for L-malate. (B) Transverse section of a dimeric unit showing the glycine molecule bound inside its allosteric-site. Residues are colored green or yellow depending on the monomer. (C) Stereoview showing the residues and interactions of the allosteric-site for neutral amino acids. The simulated annealing omit map (F₀-Fₑ) of the bound glycine molecule is shown contoured at 3.0 σ level (green mesh). (D) Surface representation of the monomer/monomer interface showing the crevice of the neutral amino acid allosteric site as seen in the previously reported 1JQO ZmPEPC-C4 crystal structure, where this site is empty and open. The red arrow marks the position of the glycine activator in the 5VYJ ZmPEPC-C4 crystal structure reported here. (E) Differences in the conformation of critical residues of the allosteric-site for amino acids between the 5VYJ ZmPEPC-C4 in complex with glycine (green carbon atoms) and the 1JQO ZmPEPC-C4 structure (magenta carbon atoms). Hydrogen bonds are depicted as black dashed lines (panel C), or green and magenta (panel E); cutoff is 3.0 Å. In panels B, C and E the side chains of relevant protein residues are shown as sticks with oxygen atoms in red, nitrogen in blue, and carbons in green, yellow or magenta. Glycine and acetate molecules are depicted as spheres colored similarly but with black carbons. Panels A, C and E were generated using
PyMOL (http://www.pymol.org). Panels B and D were generated using the UCSF Chimera package (52).

Figure 3. Region of FtPEPC-C4 equivalent to the allosteric-site for neutral amino acids in ZmPEPC-C4. (A) Cartoon representation showing that Lys96 (equivalent to Ser100 of ZmPEPC-C4) interacts with Asp223 (equivalent to Asp228 of ZmPEPC-C4), which in turn is hydrogen bonded to Arg329 (equivalent to Arg334 of ZmPEPC-C4) and Trp328 (equivalent to Trp333 of ZmPEPC-C4). Amino acid side-chains are depicted as sticks with carbon atoms in orange or gray (depending on the monomer of the dimeric unit), oxygen atoms in red and nitrogen atoms in blue. (B) Surface representation of the monomer/monomer interface of the FtPEPC-C4 dimer, showing that the allosteric site for neutral amino acids does not exist because of the position of the above mentioned residues does not leave a cavity between the two monomers. The side chains of relevant protein residues are shown as sticks with oxygen atoms in red, nitrogen in blue, and carbons in black. Of the three reported crystal structures of FtPEPC-C4, the one shown here is that with the PDB code 3ZGE. (C) Differences in the conformation of critical residues of the allosteric-site for amino acids between the 5VYJ ZmPEPC-C4 (green carbon atoms) and FtPEPC-C4 3ZGE (orange carbon atoms) crystal structures. Hydrogen bonds are depicted as dashed lines: black in panel A, green (maize isozyme) or orange (Flaveria isozyme) in panel C; cutoff is 3.0 Å. Panel A and C were generated using PyMOL (http://www.pymol.org/) and panel B using UCSF Chimera package (52).

Figure 4. Loss of response to neutral amino acids of the ZmPEPC-C4 S100K mutant. Saturation by Gly in the absence (A) or presence of 20 mM L-malate (B) of the wild-type and S100K mutant enzymes. (C) and (D) Saturation by L-Ala and L-Ser, respectively, of the wild-type and S100K mutant enzymes. Initial velocity data are depicted as black circle symbols (wild-type) or red circle symbols (S100K mutant). Activity assays were carried out at pH 7.4 and 30 °C, 3 mM total PEP, 0.4 mM free Mg$^{2+}$, and 0.1 mM bicarbonate; 20 mM L-malate was included in the assay medium when indicated. The results are mean values from two independent experiments. The points in the figures are the experimentally determined values, whereas the curves are calculated from fits of these data to Eq. 1 given in the Experimental Procedures section.

Figure 5. Kinetics of the ZmPEPC-C4 S100K mutant. (A) Saturation by total PEP. (B) Saturation by G6P. (C) Saturation by malate. (D) Saturation by G6P in the presence of 20 mM L-malate. Initial velocity data are depicted as black circle symbols (wild-type) or red square symbols (S100K mutant). Experimental conditions were as those given in the legend of Fig. 4. The initial velocity data were fitted to Eq. 1, 2 or 3, as appropriate.
Figure 1

MESOPHYLL CELL

PHOTORESPIRATION

BUNDLE-SHEATH CELL

Atmospheric CO₂

PEP

PEPC-C4

Malate

C₄ CYCLE

CALVIN CYCLE

PGA

Trioses-P

Hexoses-P

Glycine

Serine

CO₂

O₂

CO₂

RUBISCO
Figure 3
Figure 4

A. Specific Activity (U/mg prot.) vs. [Gly] (mM) for ZmPEPC-C4 WT and ZmPEPC-C4 S100K.

B. Specific Activity (U/mg prot.) vs. [Gly] (mM) for ZmPEPC-C4 WT and ZmPEPC-C4 S100K with +Malate 20 mM.

C. Specific Activity (U/mg prot.) vs. [L-Ser] (mM) for ZmPEPC-C4 WT and ZmPEPC-C4 S100K.

D. Specific Activity (U/mg prot.) vs. [L-Ala] (mM) for ZmPEPC-C4 WT and ZmPEPC-C4 S100K.
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, Carlos Mújica-Jiménez, Javier Andrés Juárez-Díaz, Rodrigo Güemes-Toro, León P. Martínez-Castilla and Rosario A. Muñoz-Clares

J. Biol. Chem. published online May 9, 2018

Access the most updated version of this article at doi: 10.1074/jbc.RA118.002884

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts