A novel mechanism of allosteric regulation of archaeal phosphoenolpyruvate carboxylase: a combined approach to structure-based alignment and model assessment

Hiroyoshi Matsumura1,2, Katsura Izui3 and Kenji Mizuguchi1,4,5

1Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK. 2Department of Applied Chemistry, Graduate School of Engineering, Osaka University, Suita, Osaka 565-0871, Japan. 3Department of Biotechnological Science, Kinki University, Kinokawa, Wakayama 649-6493, Japan and 4Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 OWA, UK

To whom correspondence should be addressed.
E-mail: kenji@cryst.bioc.cam.ac.uk

Phosphoenolpyruvate carboxylase (PEPC) catalyzes the irreversible carboxylation of phosphoenolpyruvate (PEP) and plays a crucial role in fixing atmospheric CO2 in C4 and CAM plants. The enzyme is widespread in plants and bacteria and mostly regulated allosterically by both positive and negative effectors. Archaeal PEPCs (A-PEPCs) have unique characteristics in allosteric regulation and molecular mass, distinct from their bacterial and eukaryotic homologues, and their amino acid sequences have become available only recently. In this paper, we generated a structure-based alignment of archaeal, bacterial and eukaryotic PEPCs and built comparative models using a combination of fold recognition, sequence and structural analysis tools. Our comparative modeling analysis identified A-PEPC-specific strong interactions between the two loops involved in both allostery and catalysis, which explained why A-PEPC is not influenced by any allosteric activators. We also found that the side-chain located three residues before the C-terminus appears to play a key role in determining the sensitivity to allosteric inhibitors. In addition to these unique features, we revealed how archaeal, bacterial and eukaryotic PEPCs would share a common catalytic mechanism and adopt a similar mode of tetramer formation, despite their divergent sequences. Our novel observations will help design more efficient molecules for ecological and industrial use.

Keywords: allosteric regulation/carboxylase/CO2 fixation/homology recognition/sequence alignment

Introduction

Phosphoenolpyruvate carboxylase (PEPC) (EC 4.1.1.31) catalyzes the irreversible β-carboxylation of phosphoenolpyruvate (PEP) by bicarbonate in the presence of Mg2+, to yield oxaloacetate (OAA) and inorganic phosphate (Figure 1). The enzyme has been isolated in higher plants, algae and many kinds of bacteria (O’Leary, 1982; Chollet et al., 1996; Kai et al., 2003; Izui et al., 2004). Although the enzyme’s primary role is to supply OAA, a C4-dicarboxylic acid source, to the tricarboxylic acid cycle in nonphotosynthetic tissues, PEPC also plays an important role in catalyzing the first committed step in the fixation of atmospheric CO2 in C4 plants. In plants and photosynthetic bacteria, ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyzes the carboxylation of 5-ribulose 1,5-bisphosphate with CO2, the primary event in CO2 fixation but Rubisco also catalyzes the oxygenation of the substrate (the process called photorespiration). Both reactions occur simultaneously and in competition, which decreases the efficiency of CO2 assimilation (Ogren, 1984). In C4 plants, such as maize and sugarcane, PEPC, with its high affinity for relatively inert bicarbonate (one of the dissolved forms of CO2), produces C4 compounds, which are then decarboxylated to supply Rubisco with a high concentration of CO2 (Hatch, 1987). This CO2 enrichment allows C4 plants to suppress Rubisco oxygenation more effectively and to sustain higher rates of photosynthesis than C3 plants.

Understanding the structure–function relationship of PEPC is of both academic and industrial interest. Owing to its high affinity for inert bicarbonate, PEPC has the potential to convert abundant atmospheric CO2 into more useful compounds. PEPC genes have been introduced into C3 plants, such as Arabidopsis (Chen et al., 2004), rice (Ku et al., 1999), potato (Rademacher et al., 2002) and tobacco (Kogami et al., 1994; Hausler et al., 2002) to improve the efficiency of CO2 fixation in C3 photosynthesis (Furbank and Taylor, 1995). High-level expression of the maize PEPC gene was observed in transgenic rice plants and these transgenic plants showed reduced O2 inhibition of photosynthesis (Ku et al., 1999). However, the expressed PEPC gene in transgenic plants does not achieve maximal activity in vivo, presumably because of the presence of an allosteric inhibitor or dephosphorylation (see below) (Matsuoka et al., 2001). The ability to modify allosteric properties of PEPC will be an important step forward towards resolving this issue.

All known plant and bacterial PEPCs are tetrameric enzymes with molecular weights ranging from 400 to 440 kDa. The relatively high level of sequence conservation (Toh et al., 1994) suggests that the enzymes from various organisms share essentially a common catalytic mechanism. Biochemical and structural studies (Chollet et al., 1996; Kai et al., 1999; Matsumura et al., 1999; Matsumura et al., 2002) proposed a catalytic mechanism (Figure 1), in which CO2 is generated as an intermediate to form a C–C bond (Step 5 of Figure 1). The activities of most PEPCs are regulated by a variety of allosteric effectors; in bacteria, most PEPCs are activated by acetyl-CoA and inhibited by L-aspartate or L-malate, whereas in plants most enzymes are activated by glucose 6-phosphate (G6P) and inhibited by L-malate. In addition, the activities of some PEPCs in C4 plants are controlled by reversible phosphorylation mediated by a specific protein kinase (Hartwell et al., 1999; Tsuchida et al., 2001). Crystal structures of Escherichia coli PEPC (Ec-PEPC) (Kai et al., 1999) and maize PEPC (Zm-PEPC) (Matsumura et al., 2002) revealed an elaborated TIM-barrel (β2α)-fold,
with the active site on the C-terminal end of the barrel. The Ec-PEPC structure, complexed with the allosteric inhibitor L-aspartate and a substrate analogue, is considered to represent an inactive state of the enzyme, while the Zm-PEPC structure, complexed with the allosteric activator analogue sulfate, is considered to represent an active state. The structural studies identified two residues, His11 and Arg243, critical for both catalysis and allosteric regulation; His11 is considered to deprotonate the carboxy group of carboxyphosphate to generate CO$_2$ (Terada and Izui, 1991) (Steps 4 and 5 of Figure 1) and Arg243 is considered to stabilize substrates and an intermediate (Yano et al., 1995). (In this paper, all residue numbers refer to alignment positions in Figure 2 unless stated otherwise.) Both residues move more than 10 Å between the active and inactive forms of the enzyme (Matsumura et al., 2002). In addition to His11 and Arg243, a mutagenesis study (Kai et al., 1999) identified the positively charged side chain of Arg364 (703 in E.coli numbering) to be responsible for bicarbonate-binding. Arg364 lies on a highly flexible loop in the crystal structure and is likely to transfer bicarbonate into the active site. The crystal structures revealed a hydrophobic pocket in the active site, in which the intermediate CO$_2$ would be accommodated (Rose et al., 1969; Izui et al., 1983; Matsumura et al., 1999). Although the structural studies provided a better understanding of both catalytic and allosteric mechanisms, these results are yet to deliver more efficient PEPC-introduced transgenic plants. The lack of three-dimensional (3D) structural information about non-allosteric PEPC has hampered such developments.

Until recently, the genes of archaeal PEPC (A-PEPC) have not been identified although two enzymes from Sulforbus acidetifus (Sac) (Sako et al., 1997) and Methanothermus socialis (Mso) (Sako et al., 1996) have been characterized. Recently, A-PEPC genes from Sulforbus solfataricus (Sso) (Ettema et al., 2004) and Methanothermobacter thermautotrophicus (Mth) (Patel et al., 2004) have been identified, and A-PEPC genes have been reported to be encoded in most of the sequenced archaeal genomes (Ettema et al., 2004; Patel et al., 2004). The reported multiple sequence alignments suggested A-PEPC to be homologous (i.e. evolutionarily related) to bacterial and eukaryotic PEPC (BE-PEPC) (Ettema et al., 2004; Patel et al., 2004). However, A-PEPC has several unique features distinguishable from BE-PEPC (Table 1). The activities of A-PEPC are almost unaffected by any allosteric activators including acetyl-CoA and G6P. The activities of Sso- and Sac-PEPCs are inhibited by L-malate or L-aspartate but those of Mso- and Mth-PEPCs are unaffected despite sharing highly similar sequences. Although A-PEPC, such as BE-PEPC, is likely to form a homotetramer (Sako et al., 1996; Sako et al., 1997; Patel et al., 2004), its molecular weight of ~250 kDa is only half that of BE-PEPC.

Owing to the low sequence similarity between A-PEPC and BE-PEPC, the published alignments are either incomplete (covering only the highly conserved parts) or not without ambiguity, and no attempt has been made to utilize the
Fig. 2. Structure-based multiple sequence alignment of archaeal and eukaryotic PEPC. The upper line indicates the positions of the core secondary structures expressed as a helix (green), strand (blue), and other regions (white). The second and third lines indicate the secondary structure prediction of Mth-PEPC with the PredictProtein server (Rost et al., 2004) (H, helix; E, strand; blank, other) and the predicted disorder region with GlobPlot2.1 (Linding et al., 2003) (D, disorder region; blank, other). The fourth and the fifth lines indicate the positions of the motifs and key residues involved in catalysis (in blue), hydrophobic pocket (in yellow), Mg$^{2+}$ binding (in red) and the unmodeled loop (**). Numbers in ‘<’ are the residue number of the first amino acid in the alignment. Numbers in ‘( )’ between conserved blocks are insert lengths. The next five lines correspond to the crystal structures of Ec-PEPC (PDB entry 1FIY) and Zm-PEPC (PDB entry 1JQO, chain A), three comparative models of Sso-, Sac- and Mth-PPECs, with their structural features annotated with JOY (Mizuguchi et al., 1998). Thirteen homologues of A-PEPC are also included. The formatting convention of JOY is as follows: red, α-helices; blue, β-strands; upper case letters, solvent inaccessible; lower case letters, solvent accessible; boldface, hydrogen bonds to mainchain amides; underlining, hydrogen bonds to mainchain carbonyls; and italic, positive mainchain torsion angles. The sources of sequences are as follows (abbreviation, organism, Uniprot accession number): Ec, E.coli, P00864; Zm, Zea mays, P04711; Sso, S.solfataricus P2, Q97WG4; Sac, S.acidocaldarius, Q4JCJ1; Mth, M.thermautotrophicum, O27026; Mka, M.kandleri AV19, Q8TYV1; Afu, A.fulgidus DSM 4303, Q96YS2; Hnr, Halobacterium sp. NRC-1, Q9HN43, Pfu, P.furiosus DSM 3638, Q8TZL5, Pae, P.aerophilum str. IM2, Q8ZT64; Cpe, C.perfringens str. 13, Q8XLE8; Pab, P.abyssi GE5, Q9V2Q9; Pho, P.horikoshii OT3, O57764; Sto, S.tokodaii str. 7, Q96YS2; Mba, M.barkeri, Q469A3; Mma, M.mazei Go1, Q8PS70.

### Table 1: Sequence Alignment of PEPCs

<table>
<thead>
<tr>
<th>Accession</th>
<th>Organism</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>P00864</td>
<td>E.coli</td>
<td>S51079.1</td>
</tr>
<tr>
<td>P04711</td>
<td>Zea mays</td>
<td>K95461.1</td>
</tr>
<tr>
<td>Q97WG4</td>
<td>S.solfataricus P2</td>
<td>P11750</td>
</tr>
<tr>
<td>Q4JCJ1</td>
<td>S.acidocaldarius</td>
<td>Q97641</td>
</tr>
<tr>
<td>O27026</td>
<td>M.thermautotrophicum</td>
<td>O27026</td>
</tr>
<tr>
<td>Q8TYV1</td>
<td>M.kandleri AV19</td>
<td>Q8TYV1</td>
</tr>
<tr>
<td>Q96YS2</td>
<td>Halobacterium sp. NRC-1</td>
<td>Q96YS2</td>
</tr>
<tr>
<td>Q8TZL5</td>
<td>P.furiosus DSM 3638</td>
<td>Q8TZL5</td>
</tr>
<tr>
<td>Q9V2Q9</td>
<td>P.abyssi GE5</td>
<td>Q9V2Q9</td>
</tr>
<tr>
<td>O57764</td>
<td>P.horikoshii OT3</td>
<td>O57764</td>
</tr>
<tr>
<td>Q96YS2</td>
<td>S.tokodaii str. 7</td>
<td>Q96YS2</td>
</tr>
<tr>
<td>Q469A3</td>
<td>M.barkeri</td>
<td>Q469A3</td>
</tr>
<tr>
<td>Q8PS70</td>
<td>M.mazei Go1</td>
<td>Q8PS70</td>
</tr>
</tbody>
</table>

### Notes
- The sequence alignment was performed using ClustalW2 with default parameters.
- The secondary structure predictions were generated using PredictProtein and GlobPlot2.1.
- The crystal structures were analyzed using JOY.

Mechanisms of allosteric regulation

Mechanisms of allosteric regulation
information about the 3D structure in two distinct states (active and inactive). In this paper, we have generated a multiple sequence alignment of archaeal, bacterial and eukaryote PEPCs using a combination of fold recognition, sequence and structural analysis tools. Based on this new alignment, we have built multiple structural models for A-PEPC to address the following questions: (i) how does A-PEPC maintain its unique allosteric features? (ii) how does A-PEPC exhibit its catalytic activity, given its low sequence similarity to BE-PEPC and reduced size? (iii) how does A-PEPC form a tetramer? Our modeling study provides a novel insight into the unique enzymatic features of A-PEPC, which will help design more efficient molecules for ecological and industrial use.

Materials and methods

Structure-based alignment and comparative modeling of A-PEPC

We chose Sso-, Sac- and Mth-PEPC as a representative set of A-PEPC for comparative modeling analysis. (No sequence is available for Mso-PEPC.) First, homologues of A-PEPC were collected by running WU-Blast2 (Altschul and Gish, 1996) against UniProt (Apweiler et al., 2004) with default parameters, using Sso-, Sac- and Mth-PEPCs as query. The following 14 hits in archaeal species were collected: Sso, Sulfolobus solfataricus P2, Q97WG4; Sac, Sulfolobus acidocaldarius, Q4JC1; Mth, Methanobacterium thermotrophicum, O27026; Mka, Methanopyrus kandleri AV19, Q8TYV1; Afu, Archaeoglobus fulgidus DSM 4304, O28786; Hnr, Halobacterium sp. NRC-1, Q9HN43; Pfu, Pyrococcus furiosus DSM 3638, Q8TZL5; Pae, Pyrobaculum aerophilum str. IM2, Q8ZT64; Cpe, Clostridium perfringens str. 13, Q8XLE8; Pab, Pyrococcus abyssi GE5, Q9V2Q9; Pho, Pyrococcus horkhosii OT3, O57764; Sto, Sulfolobus tokodaii str. 7, Q96YS2; Mba, Methanosarcina barkeri, Q469A3; Mma, Methanosarcina mazei G01, Q8PS70.

Second, a structural profile for BE-PEPC was prepared. Following the domain definitions in the CATH database (Orengo et al., 1997), only the conserved core was extracted from the crystal structures of Ec-PEPC (PDB entry 1FIY) (Kai et al., 1999) and Zm-PEPC (PDB entry 1JQO) (Matsumura et al., 2002) (for more details see Results and discussion section). The core consisted of residues 128–141, 197–290 and 390–883 for the PDB entry 1FIY and residues 167–180, 236–330 and 450–970 for chain A of the PDB entry 1JQO, respectively. The two processed structures were superimposed with STAMP (Russell and Barton, 1992) and a structure-based alignment was generated with COMPARER (Sali and Blundell, 1990). This alignment was then converted into a structural profile, a position-specific score matrix used by the sequence–structure homology recognition program FUGUE (Shi et al., 2001).

Finally, the A-PEPC sequences were added to the structural profile of BE-PEPC. This was achieved either by aligning the A-PEPC sequences alone first with MAFFT (Katoh et al., 2002) and then combining them with the structural profile with FUGUE, or by adding the A-PEPC sequences one by one with FUGUE, or a combination of both. The resulting alignment was formatted with JOY (Mizuguchi et al., 1998) and visually inspected.

Based on this alignment, we built a series of models for the 3D structure of A-PEPCs excluding residues on one long insertion corresponding to residues 154–176 for Sso-, 166–188 for Sac- and 152–174 for Mth-PEPCs (Table 2). The unmodeled region is marked with **** in key residue line in Figure 2. For each template–target pair, multiple models were built with MODELLER 8.1 (Fiser and Sali, 2003) and the best model selected using Verify3D (Eisenberg et al., 1997) and Prosa2003 (Sippl, 1993).

Both the alignment and models were examined based on the output from Verify3D, Prosa2003 and JOY, as well as the positions of the secondary structures predicted by Predict-Protein server (Rost et al., 2004) and PSI-PRED (McGuffin et al., 2000) and of the disordered regions predicted by GLOBPLOT2.1 (Linding et al., 2003) and DISOPRED (Jones and Ward, 2003). The alignment was adjusted manually assuming that the TIM-barrel scaffold should be highly conserved (for more details see Results and discussion section). If necessary, the structural profile was updated using different core structures and the whole process was repeated.

Results and discussion

Structure-based alignment revealed general features of A-PEPC

The alignment between A-PEPC and BE-PEPC proved to be far from straightforward, due to the low sequence similarity and the large differences in sequence length. Different
alignment programs produced slightly different results depending on the parameters used and the representative sequences chosen. We have noticed that the CATH database (Orengo et al., 1997) defines two domains for BE-PEPC, the TIM-barrel fold for the catalytic core called ‘PEP-binding domain’ (CATH code 3.20.20.60) and an additional all-\alpha domain (1.20.1100.10), which is discontiguous and consists of four segments. In addition, no domain code is assigned for one region (residues 273–389 of 1FIY) forming a four-helical bundle. Our initial alignments suggested that the TIM-barrel fold domain was highly conserved, but the other regions were not. Especially, the N-terminal two segments of all-\alpha domain and the four-helical bundle were missing in A-PEPC, although the C-terminal segments were moderately conserved between BE-PEPC and A-PEPC. Therefore, we decided to remove the N-terminal two segments of the all-\alpha domain and the four-helical bundle from the two representative structures, Ec-PEPC (PDB entry 1FIY) and Zm-PEPC (PDB entry 1JQO). These processed structures were superimposed and a structure-based alignment generated (Figure 2). An iterative method, detailed in Materials and methods section, was employed to add the sequences of A-PEPC to this structure-based alignment, based on the assumption that A-PEPC should adopt the complete TIM-barrel fold. More specifically, the alignment was built to optimize the following criteria: (i) none of the eight core \(\beta\)-strands and \(\alpha\)-helices of the TIM-barrel fold is deleted. (ii) Their positions should match the predicted secondary structures. (iii) The known catalytic residues should be conserved. We noted that \(\beta_2\) was especially difficult to align because of its low sequence similarity to BE-PEPC. Therefore, we performed many rounds of alignment and model building (Figure 3) and validated the results with multiple sources of information.

The final alignment shown in Figure 2 revealed that structurally important residues were conserved throughout BE-PEPC and A-PEPC. Many of the buried hydrophobic residues in BE-PEPC are highly conserved in A-PEPC (e.g. at 158, 159, 328, 346 and 371). In addition, there are well-conserved residues that are hydrogen-bonded to main chain or other side chain groups in the structures of BE-PEPC.

Table 2. Characteristics and qualities of comparative models for A-PEPCs

<table>
<thead>
<tr>
<th>Template</th>
<th>Sources</th>
<th>Sso-PEPC</th>
<th>Sac-PEPC</th>
<th>Mtb-PEPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction between loop(\gamma1)-loop(\gamma7)</td>
<td>Zm (active)</td>
<td>Ec (inactive)</td>
<td>Zm (active)</td>
<td>Ec (inactive)</td>
</tr>
<tr>
<td>O</td>
<td>X</td>
<td>O</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>Interaction between R243-L-Asp</td>
<td>NA</td>
<td>O</td>
<td>NA</td>
<td>O</td>
</tr>
<tr>
<td>Prosa score</td>
<td>−3.36</td>
<td>−3.24</td>
<td>−3.63</td>
<td>−3.67</td>
</tr>
<tr>
<td>Verify3D scores for loop(\gamma7)</td>
<td>0.48</td>
<td>0.28</td>
<td>0.51</td>
<td>0.50</td>
</tr>
</tbody>
</table>

\(^a\)O, interaction observed; X, no interaction observed.

\(^b\)O, interaction observed; X, no interaction observed; NA, L-aspartate is not included in the models.
(Asp53, Gln269 and Arg370). Highly conserved motifs are found (e.g. loops at 7–14 connecting β1 and α1 and at 237–245 connecting β7 and α7). The acidic residues involved in Mg²⁺ binding (at 161 and 199) are also completely conserved. The alignment was further validated by evaluating the quality of the comparative models (see below).

One major difference between our new alignment and the previously published ones is in the region between β1 and α4. In the alignment reported previously (Ettema et al., 2004), the number of amino acid residues between β1 and β2 is only 16, which is too short to provide a loop–helix–loop segment and does not agree with the predicted secondary structures. In our alignment, the predicted positions of secondary structures match those in the known structures while maintaining a long enough loop (35 amino acid residues) between β1 and β2.

As a consequence, our alignment highlighted a shift (an apparent misalignment) in β2, α2, β3 and α4 in the previously reported alignment. Because the active site lies on the C-terminal end of the TIM-barrel, it is crucial to align reliably all the loops connecting the β-strands.

**Strong interactions between two loops stabilize the active conformation of A-PEPC**

Two mobile loops in BE-PEPC have been shown to be essential for its catalytic reaction and allosteric regulation (Matsumura et al., 2002). In the inactive state of Ec-PEPC, the side chain of Arg243 (587 in *E.coli* numbering) is bound to an allosteric inhibitor L-aspartate and kept away from the active site (see Figure 4A). In the active state of Zm-PEPC, the loop containing this Arg (at positions 236–245 in

---

**Fig. 4.** (A) A schematic representation of the conformational changes of two loops (loopβ1α1 and loopβ7α7) in the active (in red) and inactive (in blue) states of BE-PEPC. Cofactors (L-aspartate and G6P), and the side chains of H11, D13, S239 and R243 are shown. The side chain of R243 is bound to an allosteric inhibitor L-aspartate and kept away from the active site in the inactive form, while it moves towards the active site upon the release of L-aspartate in the active form. The side chain of H11 is kept away from the active site in the inactive state, while in the active state, it moves towards the active site due to the interactions of several Arg residues on this loop with the allosteric activator G6P. Arg183, Arg184, Arg231 and Arg372 are shown in maize numbering. (B) The corresponding two loops in the active-state model of A-PEPC. Interactions between the two loops are shown as orange dots. The side chains of H11, D13, S239 and R243 are shown. The red dots indicate that the residues involved in G6P binding are missing in A-PEPC. (C and D) The active site structures of Zm-PEPC [in green, (C)] and Sso-PEPC [in magenta, (D)], viewed from the same orientation as in Figure 3. The side chains of Arg243, Asp or Thr at position 243, and Ser239 are shown as sticks. The interactions between the two loops are represented by orange dots.
Mechanism of allosteric regulation

Figure 2; referred to as loopβ7z7 hereafter) moves towards the active site and Arg243 is displaced by 15 Å. Similarly in the active state of Zm-PEPC, His11 on another loop (at positions 8–14 in Figure 2; referred to as loopβ1x1 hereafter) moves 10 Å toward the active site, owing to the interactions of several Arg residues on this loop with the bound sulfate molecule, an analogue of the allosteric activator G6P. The side chains of Arg243 and His11 are likely to be located near the active site when it exhibits catalytic activity.

Although the amino acid sequences of loopβ7z7 and loopβ1x1 are relatively well-conserved in all PEPCs (Figure 2), A-PEPCs are not reported to be activated by any activators. This difference between BE-PEPC and A-PEPC fits the observation that four amino acid residues involved in allosteric activator binding (Arg183, Arg184, Arg231 and Arg372 in maize numbering) are missing in A-PEPCs (Figure 4B). However, the question still remains as to what conformations these two loops would adopt in A-PEPC.

Do they adopt structures similar to those in the active state and therefore, A-PEPC does not require allosteric activation or does A-PEPC have a completely different regulation mechanism?

To answer this question, we built comparative models for Sso-, Sac- and Mth-PEPCs. For each A-PEPC, two models were built based on the coordinates of Ec-PEPC (referred to as the E-model) and Zm-PEPC (referred to as the Z-model), respectively. The overall quality of the Z-models (i.e. active state-like) is comparable to that of the corresponding E-models (i.e. inactive state-like) according to the model evaluation programs Verify3D (Eisenberg et al., 1997) and Prosa2003 (Sippl, 1993). The Verify3D scores were positive for most regions and Prosa2003 produced an acceptable range of model scores (Table 2). However, the sequence-environment compatibility scores from Verify3D were generally higher for the loopβ7z7 in the Z-models than those in the E-models (see Table 2), suggesting that this loop in A-PEPCs may adopt a conformation more similar to that in the active state. (Loopβ1x1 is close to the N-terminus and cannot be evaluated with Verify3D.)

We conclude that the modeled structures of the two loops (loopβ1x1 and loopβ7z7) are sufficiently reliable for the following reasons. First, the two loops are at the C-terminal end of the β-barrel, which is the most highly conserved region of eukaryotic and A-PEPCs (Toh et al., 1994) (Figure 2). Indeed, the sequence conservation in these regions is sufficiently high and they can be considered as part of the ‘structurally conserved regions’, as termed normally in comparative modeling. (We use the word ‘loop’ simply to refer to regions between standard secondary structures and it should not be confused with ‘structurally variable regions’). The TQHPD and TAHPD motifs in loopβ1x1 and the GXGXXFGXG and GRRGXXGRGG motifs in loopβ7z7 are completely conserved in archaeal and eukaryotic PEPCs, respectively. Second, the two loops are surrounded by other parts of the molecule, which severely restricts the available conformations and interactions.

In all the Z-models of A-PEPC, the two loops appear to interact via a salt-bridge between the side chain of Asp13 on loopβ1x1 and the side chain of Arg243 on loopβ7z7 and a hydrogen-bond between the side chains of Asp13 and Ser239 (Figure 4D). No salt-bridges between the two loops are observed in BE-PEPC (Figure 4C). Ser239 is substituted with Gly in all BE-PEPCs thus no hydrogen-bonds can be formed. Although Arg243 is essential for catalysis and is completely conserved in all A- and BE-PEPCs, Asp13 is completely conserved in A-PEPCs but not in BE-PEPCs (see Figure 2), indicating that the salt-bridge between the side chains of Asp13 and Arg243 is unique in A-PEPCs. In A-PEPCs, the amino acid residue at position 239 is also highly conserved as Ser or Thr, whereas in BE-PEPCs, the residue at 239 is completely conserved as Gly. These observations suggest that the two loops of A-PEPCs are already stabilized and adopt an active conformation without any activators. This hypothesis is in agreement with the biochemical data (Ettema et al., 2004; Patel et al., 2004) that the enzymatic activity of A-PEPC is insensitive to any allosteric activators. In BE-PEPC, the position 13 is mostly occupied by Thr, which is generally expected to interact with Arg less strongly than Asp. The lack of strong interactions between the two loops would make them more flexible, which may be an important factor for the allosteric regulation of BE-PEPC. Mutating Thr13 to Asp should test this hypothesis.

Some A-PEPCs are susceptible to allosteric inhibition due to a unique C-terminal interaction

The crystal structure of Ec-PEPC complexed with the allosteric inhibitor L-aspartate shows its binding pocket composed of Arg243, Lys422, Arg447 and Asn480 (Kai et al., 1999). Of these four amino acid residues, only the catalytically essential Arg243 is conserved in Mso- and Mth-PEPCs. This seems to be the reason why Mso- and Mth-PEPC are insensitive to allosteric inhibitors (Sako et al., 1997; Ettema et al., 2004). However, two other A-PEPCs, Sso- and Sac-PEPCs also conserve only Arg243, but these are inhibited by allosteric inhibitors, such as l-aspartate or l-malate. Although Mso and Mth belong to a group (Euryarchaeote) phylogenetically distinct from that of Sso and Sac (Crenarchaeote) (Woese et al., 1990), they all share relatively high sequence similarity. The previous report proposed different types of allosteric regulation for BE-PEPC and A-PEPC (Ettema et al., 2004) but did not explain the difference between Sso/Sac and Mth/Mso. To investigate this issue, we built comparative models for Sso-, Sac- and Mth-PEPCs in a complex form including l-aspartate, based on the coordinates of Ec-PEPC complexed with l-aspartate (Table 2).

In the Sso- and Sac-PEPC models, l-aspartate is bound to the side chain of Arg243 (see Figure 5A, middle and Figure 5B). In addition, Arg418 interacts with l-aspartate, substituting the role of Asn480 in Ec-PEPC (881 in E.coli numbering) and Arg478 and Phe272 form a pocket (instead of Arg880 and Met616 in Ec-PEPC), which can accommodate l-aspartate in a manner similar to that observed in Ec-PEPC. Therefore, these models suggest that l-aspartate may inhibit both Ec-PEPC and Sso- and Sac-PEPCs in a similar manner.

On the other hand, no interaction was observed between Arg243 and l-aspartate in the Mth-PEPC model. The large side chain of Phe480 (conserved in Mth and Mso) appears to hinder preferable conformations of Arg418, which in turn blocks the interaction between Arg243 and l-aspartate (Figure 5A, right). Attempts to remodel the side chain conformations excluded the possibility that the side chain of Phe480 flips out of this position, due to the presence of the surrounding conserved amino acid residues, such as Gln269.
The relatively small side chains at 480 in Sso- and Sac-PEPCs (Ala and Ser, respectively) do not seem to affect the conformation of the spatially nearby side chain of Arg418. These modeling results agree with available biochemical data on A-PEPCs (Sako et al., 1996; Sako et al., 1997; Ettema et al., 2004; Patel et al., 2004). The third residue from the C-terminus should be a critical residue on allosteric inhibition because the residue at this position directly participates in binding of L-aspartate in BE-PEPC. In addition, the previous biochemical study showed that a C-terminal tetrapeptide was essential for catalytic activity (Dong et al., 1999). However, no site-directed mutagenesis has been performed on this site and this is an obvious target for further experimental studies.

**A flexible bicarbonate binding loop is conserved in A-PEPC**

The crystal structures of Zm-PEPC and Ec-PEPC revealed a loop (at positions 362–370) in multiple conformations (Kai et al., 1999; Matsumura et al., 2002). This loop contains three Arg residues highly conserved within BE-PEPCs (at 364, 365 and 370; Figure 2). A site-directed mutagenesis study suggested that the positively charged side chain of Arg364 (703 in E.coli numbering) is associated with bicarbonate-binding (Kai et al., 1999) and is also essential for the catalytic function. We hypothesized that if A-PEPC and BE-PEPC share a common reaction mechanism, this loop in A-PEPC should play a similar role. This prediction is consistent with the results from the disorder prediction programs GLOBPLOT2.1 (Linding et al., 2003) and DISOPRED (Jones and Ward, 2003), which strongly suggested that the loop at positions 362–370 in A-PEPC is likely to be natively unstructured (Figure 2). Indeed, most of the predicted disordered regions are on the C-terminal end of the β-barrel, suggesting the involvement of these loops in ligand-binding and catalysis. Between BE- and A-PEPC, two Arg residues at positions 364 and 370 are strictly conserved and Pro and Gly are also partially conserved in this region. The importance of this loop in the enzymatic function of PEPC has not been reported before.

**A-PEPC conserves a hydrophobic pocket for carboxylation**

Studies on the reaction mechanism of PEPC suggested that CO2 would be generated and the carboxylation reaction would occur in a hydrophobic pocket around the methylene group of PEP (Rose et al., 1969; Izu et al., 1983). Crystal structures of Zm-PEPC and Ec-PEPC revealed a hydrophobic pocket comprising Trp49, Leu159, Met194, Gly196, Gly235 and Thr267, consistent with its suggested role in the reaction.
mechanism. Therefore, this hydrophobic pocket should also be conserved in A-PEPC if it shares a common catalytic mechanism. In the alignment in Figure 2, A-PEPC has hydrophobic residues in the equivalent positions of these residues but they are not completely conserved. We investigated the corresponding pocket with respect to its shape, hydrophobicity and size. As shown in Figure 6, the shape and hydrophobicity of the pockets of BE-PEPC and A-PEPC are very similar to each other. The volume of the hydrophobic pocket of Zm-PEPC and Sac-PEPC was also very similar, which was calculated by CASTp (Binkowski et al., 2003) to be 126.6 and 130.1 Å³, respectively. The volume of CO₂ is estimated to be ~30 Å³ and therefore, the pocket is large enough to allow the conformational changes of carboxyphosphate during Steps 3 and 4 of Figure 1 and to accommodate CO₂. These observations suggest that the reaction mechanism of carboxylation of A-PEPC is very similar to that of BE-PEPC.

Common features in tetramer formation
All known BE-PEPCs are tetrameric enzymes with molecular weights ranging from 400 to 440 kDa. The crystal structures show that each protomer has two main interfaces, one more extensive than the other, thus creating a tetramer consisting of a dimer of dimers (the two dimers being the lower and upper pairs, respectively in Figure 7A).

Experimental evidence suggests that A-PEPCs also forms a homotetramer with molecular weights of ~250 kDa (Sako et al., 1996; Sako et al., 1997; Patel et al., 2004). However, A-PEPC partially lacks the subunit–subunit interface in BE-PEPC (see Figures 3 and 7). Using the A-PEPC models, we calculated optimal docking areas (ODAs) (Fernandez-Recio et al., 2005), a set of continuous surface patches with the optimal protein–protein docking desolvation energy. A previous analysis correctly located known protein–protein interfaces in 80% of the cases (Fernandez-Recio et al., 2005).

As a positive control, we also calculated ODAs for a protomer of Zm-PEPC.

ODAs with the lowest significant energy values (<~10 kcal/mol) are shown in gray in Figure 7B. Zm-PEPC has one significant patch (on the right-hand side of the image), which agrees with the stronger intersubunit interface (pale circle) within a dimer. This interface is maintained by both hydrophobic and electrostatic interactions. No significant ODA was detected for the second, weaker interface (at the top of the image), presumably because this interface is essentially maintained by electrostatic interactions alone.

Sso-, Sac- and Mth-PEPCs have larger ODAs even with a more stringent cutoff value (<~20 kcal/mol in Figure 7C) but this is probably because of the inaccuracies in the models (e.g. the sub-optimal packing of side chains and missing loops). Ignoring the size of the ODAs, two major patches can be still observed. The larger one (on the right-hand side of the image) corresponds to a region spatially equivalent to the stronger interface in BE-PEPC and thus, the molecule may form a dimer in a similar way to that of BE-PEPC (pale circle) although the interface residues would be different. It should be emphasized, however, that the putative interfaces of A-PEPC, while showing reasonable model quality, include probably the least accurately modeled surface due to the regions eliminated from the alignment. The second, smaller patch (at the top of the image) is equivalent to the weaker interface in Zm-PEPC. Therefore, A-PEPC may form a tetramer in a manner similar to that of BE-PEPC despite the lack of conservation of the interface residues.

Most allosteric enzymes are multisubunit proteins and in PEPC, structural rearrangements across the subunits could be involved in a cooperative allosteric transition. However, our recent structural studies on full-length maize PEPC in the active state show only the local conformational changes near the active site (as we discussed above) and no rearrangement across the subunits. (The crystal structure of full-length maize

Fig. 6. The amino acid residues forming a hydrophobic pocket in the active site, displayed as Connolly surfaces (Connolly, 1983) and sticks, in Zm-PEPC (A) and Sac-PEPC (B) as a representative of A-PEPC. This figure is prepared using PyMol (The PyMOL Molecular Graphics System, http://www.pymol.org). The structures are viewed from the same orientation as in Figure 3. (The colour version of Figure 6 is available in Supplementary Data.)
PEPC will be described elsewhere.) This result suggests that the local structural changes near the active site dominate the allosteric transition for BE-PEPCs, and probably A-PEPCs as well.

Conclusions

Our comparative analysis of A-PEPCs revealed both common and unique features of their allostery, catalysis and oligomer formation. We showed how A- and BE-PEPC would adopt a common catalytic mechanism and similar modes of tetramer formation despite the low sequence similarity between them. We identified strong interactions between the two loops (loopβ1x1 and loopβ7x7) involved in both allostery and catalysis, a unique feature in A-PEPC which explained why A-PEPC is not influenced by any allosteric activators. We also found that the side chain located three residues before the C-terminus appears to play a key role in determining the sensitivity to allosteric inhibitors. Although, we cannot rule out the possibility that A-PEPC could be inhibited by a completely different mechanism, our observations led to several specific hypotheses that can be tested by direct experimentation. For instance, site-directed mutagenesis to change Thr13 to Asp is predicted to decrease the sensitivity of BE-PEPC to its allosteric inhibitors. Transgenic plants carrying this modified PEPC gene may overcome the problem of allosteric inhibition and achieve more efficient CO₂ assimilation. Such a combination of computational and experimental studies should help design more efficient molecules for ecological and industrial use.

Fig. 7. (A) A schematic representation of the tetramer of BE-PEPC. Each protomer has two main interfaces, one more extensive than the other, thus creating a tetramer consisting of a dimer of dimers. (B and C) Optimal desolvation areas (ODAs) calculated on the surface of Zm-PEPC (B) and Sac-PEPC (C). Residues are colored according to their corresponding ODA values (gray for ODA values <−10.0 and <−20.0 kcal/mol for Zm- and Sac-PEPCs; black otherwise). The two circles show the stronger and weaker intersubunit interface, respectively. Zm-PEPC and Sso-PEPC are viewed from the same orientation as in Figure 3. (The colour version of Figure 7 is available in Supplementary Data.)

Acknowledgements

H.M. is a recipient of Japanese Society for the Promotion of Science postdoctoral fellowship for research abroad (2004). The authors thank Dr Juan Fernandez-Recio for his kind assistance with ODA calculation.

References

Mechanism of allosteric regulation


Received April 17, 2006; revised May 23, 2006; accepted May 30, 2006

Edited by Amnon Horovitz