A NOVEL TWO-DOMAIN ARCHITECTURE WITHIN THE AMINO ACID KINASE ENZYME FAMILY REVEALED BY THE CRYSTAL STRUCTURE OF *Escherichia coli* GLUTAMATE 5-KINASE

Clara Marco-Marín[1], Fernando Gil-Ortiz,[1] Isabel Pérez-Arellano,[2] Javier Cervera,[2] Ignacio Fita[3] and Vicente Rubio[1,*]

[1]*Instituto de Biomedicina de Valencia (IBV-CSIC) and Center for Biomedical Research on Rare Diseases (CIBERER-ISCIII), Jaume Roig 11, Valencia-46010, Spain*

[2]*Centro de Investigación Príncipe Felipe (FVIB-CSIC), Avda. Autopista del Saler 16, Valencia-46013, Spain*

[3]*Instituto de Biología Molecular de Barcelona (IBMB-CSIC). Institute for Research in Biomedicine. Parc Científic, Josep Samitier 1-5, 08028-Barcelona, Spain.*

Present address: F. Gil-Ortiz, Centro de Investigación Príncipe Felipe (FVIB-CSIC), Avda. Autopista del Saler 16, Valencia-46013, Spain

* *Corresponding author*:

Vicente Rubio

Instituto de Biomedicina de Valencia

Jaume Roig 11, Valencia-46010, Spain

E-mail: rubio@ibv.csic.es Tel. +34 963 391 772 Fax. +34 963 690 800

Short title: Structure of γ-glutamyl kinase of *Escherichia coli*

Summary. Glutamate 5-kinase (G5K) makes the highly unstable product glutamyl-5-phosphate (G5P) in the initial, controlling step of proline/ornithine synthesis, being feed-back inhibited by proline or ornithine, and causing, when defective, clinical hyperammonaemia. We have determined two crystal structures of G5K from *Escherichia coli*, at 2.9- and 2.5-Å-resolution, complexed with glutamate and sulphate, or with G5P, sulphate and the proline analog 5-oxoproline. *E. coli* G5K presents a novel tetrameric (dimer of dimers) architecture. Each subunit contains a 257-residue AAK domain, typical of acylphosphate-forming enzymes, with the characteristic $\alpha_3\beta_8\alpha_4$ sandwich topology. This domain is responsible for catalysis and proline inhibition and has a crater on the β sheet C-edge that hosts the active centre and the bound 5-oxoproline. Each subunit also contains a 93-residue C-terminal PUA domain, typical of RNA-modifying enzymes, which presents the characteristic $\beta_5\beta_4$ sandwich fold and three α helices. The AAK and PUA domains of one subunit associate non-canonically in the dimer with the same domains of the other subunit, leaving a negatively charged hole between them that hosts two Mg ions in one crystal, in line with the G5K requirement of free Mg. The tetramer, formed by two dimers interacting exclusively through their AAK domains, is flat and elongated, and has in each face, pericentrically, two exposed active centres in alternate subunits. This would permit the close apposition of two active centres of bacterial glutamate-5-phosphate reductase (the next enzyme in the proline/ornithine-synthesising route), supporting the postulated channelling of G5P. The structures clarify substrate binding and catalysis, justify the high glutamate specificity, explain the effects of known point mutations, and support the binding of proline near glutamate. Proline binding may trigger the movement of a loop that encircles glutamate, and which participates in a hydrogen bond network connecting active centres which is possibly involved in the cooperativity for glutamate.

Keywords: Δ^1-Pyrroline 5-carboxylate synthetase; proline synthesis; ornithine synthesis; amino acid kinase family; PUA domain.

Abbreviations used: AAK, amino acid kinase; CK, carbamate kinase; G5K, L-glutamate 5-kinase; G5P, glutamate 5-phosphate; G5PR, glutamyl-5-phosphate reductase; MAD, multiple anomalous dispersion; MIR, multiple isomorphous replacement; NAGK, N-acetyl-L-glutamate kinase; PUA, a domain named after pseudouridine synthases and archaeosine-specific transglycosylases; P5CS, Δ^1-pyrroline 5-carboxylate synthetase; rmsd, root mean square deviation; SAD, single anomalous dispersion; SIR, single isomorphous replacement.

INTRODUCTION

Proline is not only a protein building block, but it also fulfils other important functions. It protects cells from osmotic stress, it is a scavenger of free radicals, a transient form of nitrogen storage, a source of reducing power, and is involved in pH regulation.[1-5] Microorganisms and plants make proline from glutamate in three steps catalysed by glutamate 5-kinase (G5K), glutamyl 5-phosphate reductase (G5PR) and pyrroline 5-carboxylate reductase (Figure 1).[6-8] Mammals also use G5K and G5PR to synthesize ornithine (Figure 1).[9-10] This latter function is crucial for proper ammonia detoxification in humans, as highlighted in Δ^1-pyrroline 5-carboxylate synthetase deficiency (OMIM 138250; http://www.ncbi.nlm.nih.gov), an inborn error in which a G5K-inactivating mutation causes clinical hyperammonaemia, with hypoornithininemia and hypoprolinemia.[11] G5K plays a key role in proline or ornithine biosynthesis since it catalyses the first step of these biosynthetic routes and since it is the subject of feed-back inhibition by the final products, these being proline in microorganisms and plants,[6,7] and ornithine in animals.[10]

In animals and plants G5K is the N-terminal moiety of a bifunctional polypeptide called Δ^1-pyrroline 5-carboxylate synthetase (P5CS). The C-terminal moiety of this polypeptide is the enzyme G5PR.[7,10] The bifunctionality serves the purpose of channelling between G5K and G5PR of the G5K product glutamyl 5-phosphate (G5P). G5P is highly unstable,[12] given the rapid intramolecular reaction between its amino group and its γ-acylphosphate group, yielding 5-oxoproline (also called pyroglutamate) (Figure 1).[13] Bacteria have monofunctional G5K and G5PR,[14-16] but the channelling of G5P between these two enzymes is also believed to be necessary in prokaryotes.[6,13] However, the possibility of G5P surviving long enough to be used by a monofunctional reductase, at least when the reductase is highly abundant, cannot be

completely ruled out. This is in view of the fact that the *Corynebacterium glutamicum* and *Escherichia coli asd* genes, which encode the enzyme aspartate β-semialdehyde dehydrogenase (an enzyme that catalyses the same reaction as G5PR using aspartyl-4-phosphate as substrate), when expressed at high levels, suppressed the mutational loss of G5PR in *E. coli*.[17] Nevertheless, compelling evidence for the association of G5K and G5PR is provided by the observation that G5PR has to be included in the assay of *Escherichia coli* G5K (in the absence of the G5PR substrate NADPH) to efficiently convert G5P to its hydroxamate.[13]

G5K contains a domain of ~260-residues in both prokaryotes and eukaryotes that exhibits substantial sequence identity over its entire length with carbamate kinase (CK) and N-acetyl-L-glutamate kinase (NAGK).[18] These two enzymes catalyse the same type of reaction as G5K, the synthesis of a carboxylic-phosphoric anhydride, using ATP as the phosphoryl group donor. Both enzymes belong to the same structural family called the amino acid kinase (AAK) family (Pfam: PF00696; http://www.sanger.ac.uk/Software/Pfam), a family characterised by a structure conforming to an $\alpha_3\beta_8\alpha_4$ sandwich fold with a typical constant topology.[18] Similarly to plant and animal G5Ks, most bacterial G5Ks including the archetypical *E. coli* enzyme[19] have an AAK domain linked C-terminally to another domain. In the case of the bacterial enzymes, this other domain has only ~110 residues and appears to be a PUA domain on the basis of its sequence.[20,21] PUA domains, named after pseudouridine synthases and archaeosine-specific transglycosylases (Pfam: PF01472),[20] are ~80-residue domains that exhibit a characteristic β-sandwich fold and one or two α helices.[22-23] They are found in RNA-modifying enzymes. Therefore, their identification in bacterial G5Ks was surprising, although it was speculated[20] that they might be involved in the reported regulation of gene expression effected by the *proB* gene

(encoding G5K) of *Bacillus subtilis*.[23,24] Although the actual function of the PUA domain of bacterial G5Ks is unknown, this domain is not required for substrate binding, catalysis or regulation, since neither the natural nor the artificial lack of this domain in either *Streptococcus thermophilus*[15] G5K or the PUA domain-deleted *E. coli* G5K,[21] respectively, prevents catalysis or proline inhibition in these G5Ks. Furthermore, the PUA domain-deleted *E. coli* enzyme appears on gel filtration to be tetrameric, similarly to the wild-type enzyme,[21] and thus this domain is not essential for tetramer formation. Nevertheless, the deletion of this domain increased and decreased the proline and Mg requirements for the inhibition and the activity of *E. coli* G5K, respectively, and also abolished proline-triggered aggregation of the enzyme tetramers into higher oligomers.[21] This indicates that the PUA domain might play some role in modulating the enzymatic properties of bacterial G5Ks.

It is essential to determine the structure of a bacterial G5K having a PUA domain to confirm that this domain really conforms to the canonical domain fold. The enzyme structure should be crucial for understanding substrate binding, catalysis and regulation in G5Ks. In addition, it should clarify the respective interactions of the AAK and PUA domains in the building of an apparently tetrameric architecture that has no precedent within the AAK family. The structure might shed light on the roles and the potential of both domains for mediating interactions with G5PR and possibly for mediating gene regulation. We herein answer some of these questions, by determining two crystal structures for *E. coli* G5K, at 2.9 and 2.5 Å resolution, one having bound glutamate and sulphate, and the other having bound G5P, sulphate and the proline analog 5-oxoproline. Our structures reveal a novel architecture within the AAK and PUA domain families which appears to be well adapted for G5P channelling. These structures largely clarify substrate binding and catalysis, and account for the reported

effects of some mutations.[26] Furthermore, they provide clues as to the mechanism of allosteric regulation by proline, pointing to the binding of the effector near the substrate glutamate.

RESULTS AND DISCUSSION

Structure determination

E. coli G5K posed huge challenges to crystallisation and phasing. Tetragonal and monoclinic crystals (Table 1) diffracting at 2.5 Å or 2.9 Å resolution and belonging to the space groups $P4_12_12$ or $P2_1$, respectively, either grew after several months and could not be reproduced (tetragonal crystals) or grew slowly and their reproducibility was poor (monoclinic crystals). Phasing attempts failed using MAD, SAD, MIR SIR and selenomethionine or heavy atom derivatives including Br or I derivatives, or using molecular replacement with *E. coli* NAGK[18] or *Pyrococcus furiosus* CK[27] models (23% and 19% sequence identity, respectively, with residues 1-257 of *E. coli* G5K). Phases were finally obtained for the tetragonal crystal data by molecular replacement, using as a model the structure (deposited in the PDB by a structural genomics consortium as file 2AKO) of a putative G5K from *Campylobacter jejuni* which has no PUA domain. This protein exhibits ~33% sequence identity with residues 1-257 of *E. coli* G5K. After introducing a model of the archetypical PUA domain of the archaeosine tRNA-guanine transglycosylase from *Pyrococcus horikoshii*,[22] a dimer was obtained in the asymmetric unit. Each subunit is composed of one AAK and one PUA domain. These domains interact with the same domains of the other subunit to form the dimer. Application of the $P4_12_12$ crystallographic symmetry produced a tetramer, in agreement with the results of gel filtration data indicating that the enzyme is tetrameric.[21] In this tetramer, two dimers interact by less extensive interfaces which are provided exclusively by the AAK domains. A large mass of non-protein electron density was found at the active centre of each subunit. This mass was interpreted as one molecule of each, L-glutamyl-5-phosphate, 5-oxoproline and sulphate. The model for both subunits lacked 41 or 43 residues beginning at residue 172, because of the absence of a defined electron density

for these residues. However, the model for the complete polypeptide (residues 3-367; residues 1-2 were not visible) was obtained at 2.9 Å resolution in two subunits (D and E) of the monoclinic crystal. The solution for this monoclinic crystal, obtained by molecular replacement using the dimer determined with the tetragonal crystal as a model, consisted of two tetrameric molecules (each one exhibiting non-crystallographic point group *222* symmetry) formed by subunits A-D and E-H, respectively (Figure 2 (a) to (c)). Non-protein densities in this monoclinic crystal corresponded to: glutamate and one sulphate in the active centre; two Mg ions in a negatively charged hole found between the domains at the dimer interface (Figure 2(d)); one bound glutamate in the PUA domain (Figure 2 (e)). The two tetramers in the monoclinic crystal and the tetramer generated in the tetragonal crystal, as well as all the enzyme subunits, are essentially identical. The root mean square deviation (rmsd) values for the superimpositions of the C^{α} atoms are 0.64-1.74 Å (mean 1.19 Å), when comparing complete tetramers, and 0.14-0.57 Å (mean 0.47 Å), when comparing individual subunits (residues 172-213, missing in the subunits from the tetragonal crystal, were excluded in all the superimpositions).

A novel subunit architecture made up of the canonical AAK and PUA domains.

The *E. coli* G5K subunit is composed of an N-terminal catalytic domain (with the modelled polypeptide chain spanning from Asp3 to Ala257) and a C-terminal domain (residues from Ala275 to Arg367) connected by a hinge that includes the short and irregular αI helix. The catalytic domain exhibits a typical AAK fold (Figures 3 (a), (b) and (c)).[18] The superimposition of this domain with *C. jejuni* G5K yields an rmsd value of 1.5 Å for 216 equivalent Cα atoms. The C-terminal domain exhibits the topology of the archetypical PUA domain from the *P. horikoshii* archaeosine tRNA-

guanine transglycosylase[22] with an rmsd value of 1.4 Å for the superimposition of the C^{α} atoms from 71 equivalent residues. The *E. coli* G5K AAK domain contains a mainly parallel eight-stranded β-sheet sandwiched by two layers of three and four α-helices. The α_3 helix layer looks towards the other dimer, whereas the α_4 layer looks towards the C-terminal domain of the same subunit. The AAK domain can be further divided into the N-terminal lobe (residues 1-163) and the C-terminal lobe (residues 164-257). This domain exhibits two prominent bulges, which are also found in all other AAK family enzymes and emerge from the two lobes at the opposite ends of the β sheet C-edge. G5K differs from most other enzymes of the AAK family[18,27,28] in the fact that it has no lid covering the active centre. The loops that connect helices B and C and strand β3 and helix D are not long, they do not cover the site for the phosphorylatable substrate as in NAGK,[18] and are oriented away from this site. Thus, the active centre is very wide and open and is surrounded by the two bulges, resembling a large crater (Figure 2 (b)).

The *E. coli* G5K PUA domain is nucleated by a characteristic β sandwich that is its hallmark,[22,23] and which is composed of mainly antiparallel five-stranded and four-stranded sheets running at right angles. This domain also has two helices (J and K) found in other PUA domains,[22,29] and one extra helix (L) located at the end of helix K and belonging to a short insertion spanning from residue His342 to Tyr354. At the free edge of the β sandwich a crevice is formed between helix J, strand β10 and the turn leading to β11, where one glutamate molecule was unexpectedly found in subunits A, D, E, F and G of the monoclinic crystal, but not in the tetragonal crystal (Figure 2 (e)). This glutamate molecule appears to interact with its site almost exclusively through hydrophobic interactions, whereas polar atoms remain mostly exposed towards the solvent. On the opposite side of the β sandwich edge, helices K and L participate in the interactions with the PUA domain of a neighbouring subunit. Therefore, both faces of

the β sandwich are exposed, and the PUA domain exhibits its five-stranded sheet on the same side of the subunit as the active centre of the AAK domain.

Finally, helix I belonging to the linker between the AAK and PUA domains is embedded in a four-helix bundle with helices G, E and a helical turn preceding helix D. The face of helix I that looks towards the PUA domain concentrates most of the contacts with the latter domain, in particular with the β16 end from the five-stranded sheet (Figure 3 (b)).

The first tetramer within the AAK family.

Both G5K crystal structures correspond to a molecular tetramer with 222 symmetry and a shape approaching a flattened ellipsoid of radii 34, 50 and 76 Å (Figure 2 (b) and (c)). The intersubunit interactions are very extensive (2150 $Å^2$ of buried surface per subunit, Figure 4(a)) across the twofold axis corresponding to the longest radius (the R axis; the three molecular symmetry axes are named, by convention and historical reasons, P, Q and R[30]). The interactions are much less extensive across the Q axis (430 $Å^2$ buried surface per subunit, Figure 4(b)), whereas they are non-existent across the P axis, around which there is a central hole of ~5×8 Å. The G5K tetramer can be defined as the dimer across the P or Q axis of the dimer formed across the R axis (Figure 2 (b) and (c)). Although there are few interactions across the Q axis, the tetramer appears to be highly stable, not dissociating even in dilute solution (shown by gel filtration).[21] In the R dimer, the AAK and the PUA domains interact only with the AAK and the PUA domain of the neighbouring subunit, respectively, and the narrowings corresponding in each subunit to the interdomain boundary merge into a central, negatively charged hole of ~8 Å diameter (Figure 2(d)). Two masses of density found in the monoclinic crystal within this hole have been interpreted as Mg ions. These

ions may play a structural role, since *E. coli* G5K needs free Mg for activity.[21] In agreement with the involvement of the PUA domain in forming this hole, the deletion of the PUA domain altered the Mg requirement. [21] The centres of mass of the four domains of each dimer are approximately coplanar, but the dimer end corresponding to the AAK domains is thicker and wider (~100×50×40 $Å^3$) than the opposite end (~60×25×35 $Å^3$), which corresponds to the interacting PUA domains. The flat and large surface of intersubunit contact within the dimer, is split by the connection between domains into large (1663 $Å^2$) and small (468 $Å^2$) portions contributed, respectively, by the AAK and the PUA domains (Figure 4(a)). The large portion corresponds to the surface of dimerization of the other AAK family enzymes CK, NAGK and UMP kinase,[18,27,28] consisting of the β3 end of the αβα sandwich of the AAK domain. A cross-grid is formed between the elements of this end of the sandwich and the same elements of the other subunit (Figure 5). In this cross-grid, the long C helices cross each other with different degrees of rotation, depending on the enzyme. The rotation around an axis that perpendicularly penetrates the intersubunit interface at this C-helices crossover point is of ~110º for NAGK and CK,[18,27] ~190º for UMP kinase[28] and ~260º for *E. coli* G5K. In addition, the crossover point is shifted from the third turn of helix C in NAGK and CK to the fifth turn in G5K. The orientation of these interacting surfaces largely determines the dimer shape.

G5K is the first known structure in which two PUA domains interact. However, the surface used for this interaction (the edge of the β sandwich corresponding to helix K and the β11-β12 connection, Figure 4(a)) is essentially the same utilised in archaeosine tRNA-guanine transglycosylase for the interaction of the PUA domain with the catalytic domain.[22] In known structures of PUA domain-RNA complexes[23,29,31] the 4-stranded sheet is the face of the β sandwich which contacts the RNA. In G5K, this

face is exposed, flat and smooth, presenting the conserved (in microbial G5Ks having a PUA domain) positive residues His360 and Arg361 (Figure 6). In the neighbouring AAK domain on the same face of the subunit, there is a nearby patch of exposed positive residues that are conserved only in PUA domain-containing G5Ks (Arg111, Arg118 and, in the interdomain linker, Arg265). These two positive patches might be involved in potential interactions with RNA, as in other PUA domain-containing enzymes, in which the PUA domain four-stranded face and a neighbouring region in the catalytic domain are involved in the RNA interactions[23,29,31]. A model (not shown) for such putative complex of G5K, built based on the tRNA complex of archaeosine tRNA-guanine transglycosylase, has the tRNA 3´-end at the site of the PUA domain-bound glutamate. This raises possibilities of control by aminoacyl tRNAs that deserve further analysis, particularly in the context of the involvement of *proB* of *B. subtilis* in the regulation of the expression of other genes.[24,25]

The interface between R dimers in the tetramer consists, for each pair of interacting subunits, in a 4-helix bundle formed by helices A and C of the 3-helix layer of the AAK domain. Contacts are mainly hydrophobic (mediated by Arg25, Ala26, Val29, Ile93 and Tyr94) although H bonds (NηArg25:OγSer92 and NαAla26:OαIle93) and the ion pair Glu30:Arg33 are also involved (Figure 4(b)). In the tetramer, the angle formed by the planes of both dimers when viewed along the R axis is only ~20º. Thus, the tetramer is elongated and planar, with ~140 Å maximum dimension, and has two active centres exposed on each face, on alternate subunits, in the central, wider and thicker part of the molecule surrounding the central hole and corresponding to the AAK domains (Figures 2 (b) and (c)). The leading role of the AAK domains in mediating the intersubunit contacts would account for the persistence of the tetramer upon PUA domain deletion.[21] In turn, the abolition by the deletion of the PUA domain[21] of proline-

triggered G5K aggregation into higher oligomers is possibly explained by the contacts in the $P2_1$ crystal. Here the tetramers approximately run parallel and are shifted along their longer axis by a full subunit (Figure 2 (a)), interacting by AAK-PUA domain contacts that would be lost upon PUA domain deletion.

The active centre

One sulphate and one glutamate were found in the large crater of the AAK domain in all the subunits of the monoclinic crystal (except subunit H, where the density for glutamate was poorly defined) (Figure 6(a)). The tetragonal crystal also contained the sulphate, but the density corresponding to the glutamate was larger, extending towards the sulphate and also, with discontinuity, in the region near the Cα of the glutamate (Figure 6 (b) and (c)). This large density is accounted for by a glutamyl 5-phosphate (G5P) and by a 5-oxoproline. The G5P would have been produced by the enzyme from the glutamate present and the 0.2 mM ATP found to contaminate the ADP in the crystallization drop. 5-Oxoproline might be generated by spontaneous intramolecular cyclisation of G5P. Glutamate and G5P bind in an open pit formed in the N-lobe of the AAK domain, between β4, the β2-αB junction and the two ends of the large β4-αE loop. Judging from the ADP complex of *C. jejuni* G5K (PDB code 2AKO) and from the AMPPNP-NAGK complex,[18] the sulphate approximately occupies the site that corresponds to the β-phosphate of the nucleotide (Figure 7(a)). It interacts with residues that are equivalent to those interacting with the β-phosphate in these other enzymes (Ser14, Thr169 and Lys217, in *E. coli* G5K; His11, Ser162 and Lys210 in *C. jejuni* G5K; Gly11, Ser180 and Lys217 in *E. coli* NAGK).[18] Lys217 was proposed in NAGK to help stabilise the negative charge developing in the β-phosphate of the nucleotide.[18,32] According to this proposal, the mutations Lys217Ala and Lys217Arg

triggered a >100-fold decrease in the activity of *E. coli* G5K.[26] The Thr169Ala mutation, but not the Thr169Ser mutation,[26] caused a ~10- and a ~20-fold increase and decrease, respectively, in the apparent K_m and V_{max} for ATP. This strongly suggests that the hydrogen bond between Thr169 and the β-phosphate is important for ATP binding and catalysis.

The glutamate found in the monoclinic crystals binds extended, with one of its molecular faces being totally exposed (Figure 7(a)). The correspondence of this conformation with that of the diequatorial form of L-*cis*-cycloglutamate, and the lack of steric constrains to accommodate the three extra carbons of a cycloglutamate, seem to justify the much better K_m of G5K for the conformationally more rigid cycloglutamate than for glutamate.[13] The extended conformation and the interactions mediated by the α and γ carboxylates and by the α-amino group of the L-glutamate justify the high specificity of G5K for this substrate.[13,21] The α and γ-carboxylates (Figures 6(a) and 7(a)) sit near the N-termini of helices B and E, respectively, making hydrogen bonds with main chain N-atoms of the first turn of these helices (either Gly51, Ala52 or Ile53, depending on the subunit, or, in helix E, Asn149) as does the γ-carboxylate with the γO of Ser50. The α-amino group forms one salt bridge with the γ-COO^- of Asp137, anchoring glutamate to the floor of the binding site. The mutation to glutamate of the corresponding aspartate (Asp147) in the G5K from tomato caused a decrease of ~5-fold in G5K activity,[33] which is to be expected if the mutation caused a drastic decrease in the affinity for glutamate. The α-amino and α-carboxylate of glutamate also make hydrogen bonds (in some subunits) with the δ-O and δ-N groups from Asn134, respectively. Asn134 is part of a hydrogen bond chain that links the two catalytic centres in the R dimer and which also involves the conserved residues Asn80, Gln100 and Asn149 (Figure 7(b)). Preliminary mutational studies (Carmona and Cervera, in

preparation) support the involvement of this hydrogen bond chain in modulating cooperativity between subunits. Asn134 and Asn149 conform most of the floor of the glutamate binding pocket, with the Asn149 side chain running antiparallel to the glutamate chain.

Active centre residues change little in the tetragonal crystal, relative to the monoclinic crystal. The binding of the sulphate and of the α-amino and γ-carboxyl groups of G5P is essentially identical in both crystals, and involves the same residues (Figure 6 (c)). In the tetragonal crystal, however, the G5P α-COO^- interacts with the ξ-N of Lys145, moving away from the carboxylate of the 5-oxoproline. Therefore, the glutamate and the glutamate moiety of G5P appear to be bound in different ways, possibly depending on either the presence or the absence of 5-oxoproline (or proline). The phosphate of G5P interacts with the ξ-N of the invariant Lys10, similarly to what happens with the γ-phosphate of ATP and Lys8 in *E. coli* NAGK.[18,32] Site-directed mutagenesis studies in both G5K and NAGK[26,34] revealed the importance of this lysine for catalysis. 5-Oxoproline sits flat on the side chain of Ile53 with its amidic O group hydrogen-bonded to the γ-hydroxyl of Thr13, and with its carboxylate, which occupies approximately the same site as the glutamate α-carboxylate in the monoclinic crystal, bound to both the amino group and the side-chain of Asn134 (Figures 6(b) and (c)).

The G5K structure highlights a key role for Asp150 in the organisation of the active centre of the enzyme. Although Asp150 does not interact with the ligands, it makes bonds with the ξN atoms of both Lys10 and Lys217, and appears to orient these two substrate-binding and catalytic lysine residues (Figures 6 (a) and (b)). In agreement with this proposed role, the Asp150Ala or Asp150Asn mutations inactivated the enzyme,[26] similarly to what was observed with the mutation to Glu of the corresponding

E. coli NAGK residue, Asp162 (the equivalent residue, that also interacts with both catalytic lysines).[34]

Taking together our data on the complexes containing sulphate and glutamate or G5P, and the deposited structure of *C. jejuni* G5K containing bound ADP, it is clear that the G5K active centre (Figure 7(a)) replicates the basic traits of the active centre of NAGK, the best studied enzyme of the AAK family.[18,32] Thus, the nucleotide and the phosphoryl acceptor bind on the C- and N-lobes of these enzymes, respectively, over the main sheet C-edge, with the polyphosphate chain of the nucleotide looking towards the N-domain and the adenine ring looking in the opposite direction. The phosphoryl group that is transferred, represented by the phosphoryl group of G5P, sits at the junction between both lobes. Catalysis of the transfer involves two invariant lysines (Lys10/Lys8 and Lys217/Lys217 in G5K/NAGK) that are connected and mutually oriented towards the substrates by a key organising aspartate (Asp150/Asp162 for G5K/NAGK). In both enzymes the β1-αA and β2-αB junctions include small residues that provide interactions with the ATP β phosphate (represented here by the sulphate), the transferred phosphoryl and the acceptor carboxylate. This carboxylate virtually sits identically in the two complexes studied herein and in NAGK. Instead, the other end of the glutamate molecule, in the two conformations observed in this study, differs importantly from the way of binding to NAGK.[18,32] While the basic catalytic machinery is therefore fixed, the details of the binding of the specific substrate that is to be phosphorylated vary, as also observed previously with UMP and aspartate among other enzymes of the AAK family.[28,35]

G5K regulation and potential bases for channelling

Although we were unable to obtain G5K crystals containing proline, the binding in the active centre of the tetragonal crystal, next to the glutamate site, of the proline analog 5-oxoproline fits the recent conclusion of kinetic and mutagenesis studies which suggests that proline and glutamate bind at overlapping sites.[26] Since oxoproline is a poor inhibitor of bacterial G5K,[36] the binding of proline must trigger effects that are not triggered by oxoproline. The positive charge on the proline imino group that is not present in oxoproline may trigger conformational changes mediated by Asp148, a residue whose mutation to asparagine triggered a dramatic increase in the $I_{0.5}^{Pro}$,[26] and which belongs to the mobile β4-αE loop. The existence of very important conformational changes in this loop, including Asp148, is supported by the comparison of the *E. coli* and *C. jejuni* G5K structures (Figure 8). Many mutations found to hamper proline inhibition concentrate in the β4-αE loop (or in the equivalent region in other G5Ks),[26,33] which includes two residues involved in the hydrogen bond network interlinking the glutamate sites of two subunits, and may be involved in the cooperativity for glutamate (Figure 7(b)). In this respect, proline importantly increases the cooperativity for glutamate.[26] The tendency of the enzyme to aggregate in the presence of proline[21] may be the result of a change in the extent to which this external loop is exposed, triggered by proline binding. Interestingly, the longer form of mammalian G5K produced by alternative splicing, which has two extra residues in the region corresponding to the β4-αE loop, is not feed-back inhibited by ornithine, whereas the shorter form is inhibited by ornithine.[10] Thus, ornithine and proline respective inhibition of the mammalian and the bacterial/plant enzymes may involve in both cases changes in the β4-αE loop, and thus may take place in both cases in the same way. In any case, the vast majority of the reported mutations that abolish proline inhibition map in the N-lobe of the AAK domain,[26,33] around the site for glutamate,

again supporting the nearness of the glutamate and inhibitor sites. UMP kinase represents another example among the enzymes of the AAK family in which the feed-back inhibitor binds overlapping with the substrate that is phosphorylated, although with this enzyme the overlap is complete.[37] G5K and UMP kinase would represent one way of inhibition diametrally different to the feed-back inhibition of NAGK by arginine, since in the latter enzyme arginine binds at a remote site relative to the active centre.[38] None of the traits of the arginine site of NAGK is found in G5K (not illustrated) excluding the possibility that G5K would resemble NAGK in this respect.

A novel feature of G5K is its tetrameric organisation, which differs widely from the dimeric or hexameric architectures described for other AAK enzymes,[18,27,28,35,38] This organisation appears to be highly stable despite the modest surface of interaction between the two dimers forming the tetramer, and may be an adaptation to allow G5P channelling. The structure of the monoclinic G5K crystal containing bound G5P reveals that this product can remain intact in the active centre of G5K. It should be possible to transfer G5P to G5PR without cyclisation if the active centres of G5K and G5PR are apposed intimately and the thiol group of the active centre cysteine of G5PR reacts with G5K-bound G5P. This appears to be feasible in the bienzymatic complex modelled in Figure 9. A mechanism of half of the sites reactivity would appear possible in this complex of both bacterial enzymes in which each enzyme would interact with the other by alternatively using one of its molecular faces. This proposed model agrees with the fact that G5K and G5PR belong to a single polypeptide (P5CS) in higher eukaryotes. Thus, a model of the bienzymatic P5CS can be generated from the above proposed bacterial complex. This P5CS model assumes a tetrameric organisation nucleated by the four AAK domains organised as found here in *E. coli* G5K. A pair of G5PR domains mutually related as in the recently determined structure of the human G5PR domain

dimer of P5CS (PDB file 2H5G) would be oriented to each of the sides of the tetramer of G5K domains. In the human enzyme, an Arg84Gln mutation was reported to cause pyrroline-5-carboxylate synthetase deficiency.[11] Although this residue is not constant in all G5Ks (Gly18 is the corresponding residue in *E. coli* G5K), the residue belongs to the β1-αA junction which is centrally located at the site of phosphoryl group transfer. Thus, it is conceivable that structural changes triggered by mutations at this position may inactivate the enzyme. In any case, it would be important to determine the crystal structure of human P5CS to experimentally document the deleterious effect of this mutation, and to experimentally determine the architecture of the G5K-G5PR complex. Efforts in this direction are presently under way in our laboratory.

MATERIALS AND METHODS

Enzyme crystallization and data collection.

E. coli (DH5α strain, from Clontech) *proB* cloning into pET-22b to yield pGKE, and the overexpression and purification of G5K have been reported.[39] The sequence of the enzyme presents a single amino acid change of a conservative nature (I129V), relative to the deposited sequence for *E. coli* G5K (file P0A7B5, Swissprot Database; http://www.expasy.org/uniprot). Bipyramidal-shaped crystals of approximately 0.3 mm maximal dimension were grown as reported,[39] in about four to five months, at 294 K, using the vapour-diffusion approach, in hanging drops prepared by mixing 1.5 μl of reservoir solution (1.6 M $MgSO_4$, 0.1 M KCl, 0.1 M MES pH 6.5) and 1.5 μl of enzyme solution (10 mg ml^{-1} G5K in 50 mM Tris-HCl pH 7.2, 20 mM KCl, and 1 mM dithiothreitol) containing 160 mM Na L-glutamate, 30 mM $MgCl_2$, and 6 mM ADP. In addition to these crystals, small plates of 0.1 mm maximal dimension were prepared in the same way, using a reservoir solution containing 1.45 M $MgSO_4$, 20 mM $CaCl_2$, 0.1 M MES pH 6.5. Cryoprotection of the crystals was accomplished by immersion in 10 % (v/v) glycerol-supplemented crystallisation solution. Complete X-ray diffraction datasets were collected at 100 K from single crystals at the European Synchrotron Radiation Facility (ESRF), Grenoble, France (Table 1), attaining 2.5 Å and 2.9 Å resolution with the bipyramidal crystals and the plates, respectively. Data were processed with MOSFLM, SCALA and TRUNCATE.[40] Space group for the bipyramidal crystals was $P4_12_12$ with unit cell parameters a= b= 101.1 Å, c= 178.6 Å. The space group for the plates was $P2_1$ with unit cell parameters a= 96.3 Å, b= 124.1 Å, c= 144.9 Å, α= γ= 90º, β= 94.0º. Packing density calculations for the observed cell sizes and for a mass of 38.9 kDa for the G5K polypeptide, agree with the presence of two and

eight molecules of the enzyme polypeptide in the asymmetric unit of the $P4_12_12$ and $P2_1$ crystals, respectively, for corresponding solvent contents of 58% and 55 %.

Structure determination and refinement.

Statistics for data collection and refinement are summarised in Table 1. Phasing was successful by molecular replacement with MOLREP[41] with diffraction data between 50 and 4 Å resolution for the $P4_12_12$ crystal, using a polyalanine search model derived from the structure of the *C. jejuni* putative G5K dimer (PDB file 2AKO), a protein that exhibits 33 % sequence identity with the N-terminal 257 amino acids of the 367-residue polypeptide chain of *E. coli* G5K. The solution corresponded to two subunits in the asymmetric unit forming one molecular dimer. Rigid body refinement and simulated annealing with program CNS[42] were applied to the initial model, followed by several rounds of model building with COOT[43] and O,[44] alternating with positional refinement with REFMAC5.[45] All the diffraction data (50-2.5 Å) were used throughout the refinement process, except the 5 % randomly selected data for calculating R_{free}. After several rounds of model building and refinement some of the elements of the PUA domain were incorporated into the model and these guided the fitting of the entire PUA domain, taken from the structure of archaeosine tRNA transglycosylase (PDB entry 1IQ8)[22], using MOLREP.[41] TLS refinement was performed assuming one rigid body per subunit. Water molecules were assigned with REFMAC5,[45] and were visually confirmed. One molecule of glutamyl phosphate and one of 5-oxoproline were included in each one of the subunits, and were refined using libraries generated with REFMAC5. After several rounds of refinement with REFMAC5 alternating with model building with O,[44] R_{factor}/R_{free} converged to final values of 19.3/24.3. The final model for the two subunits, obtained at a resolution of 2.5 Å, encompassed residues 3 to 367. However,

residues 172-213 and 172-211 in subunits A and B, respectively, could not be traced because of the lack of electron density. The model presented satisfactory stereochemistry (checked with Procheck).[46]

To determine the phases for the $P2_1$ crystal, a partially refined model of the complete G5K dimer found in the $P4_12_12$ crystal was used for molecular replacement. The correct solution yielded four dimers in the asymmetric unit, organised as two tetramers, one of them composed of subunits A to D and the other of subunits E to H. Model building and refinement was performed essentially as described for the $P4_12_12$ crystal, applying non-crystallographic symmetry restraints that were progressively relaxed, mainly for amino acid side-chains, in the final rounds of refinement. R_{factor}/R_{free} converged to final values of 19.7/24.7. The final model, at 2.9 Å resolution, for the eight subunits in the asymmetric unit, encompassed the entire polypeptide chain from residue 3 in subunits E and D, but lacked the following residues in the other subunits: 202-211 (subunit A); 202-213 and 367 (B); 203-211 (G and C); 176-185 and 196-213 (H); 203-213 (F). One glutamate and one sulphate were found in each active centre, except in subunit H where the glutamate was missing. A density corresponding to a glutamate molecule was found associated to the PUA domain in five subunits,.

Other methods

Buried surface areas were calculated with NACCESS (http://wolf.bms.umist.ac.uk/naccess) using a probe of radius 1.4 Å. Superposition of structures was carried out with program LSQKAB.[47] The complex between G5K and G5PR was modelled by superimposing the P axis of the G5K tetramer with one of the 2-fold symmetry axes of the G5PR tetramer of *Thermotoga maritima* (PDB 1O20), approaching the active centres whereas steric clashes were minimized. A virtually

identical model was obtained for the complex using the *E. coli* G5PR structure, modelled[48] from the *T. maritima* structure on the basis of the sequence alignment of both enzymes, exhibiting 46% sequence identity. The closed conformation of the G5PR subunit was generated by superposing the two domains of the *T. maritima* G5PR subunit on the corresponding domains of the NAD-complexed subunit of class 3 aldehyde dehydrogenase (from rat, PDB 1AD3), which is homologous to G5PR. ATP was determined luminometrically with luciferase[49] using a commercial kit (ATP Bioluminiscence CLS kit, from Roche Diagnostics). Figures were generated using MOLSCRIPT,[50] BOBSCRIPT,[51] RASTER3D,[52] PYMOL (DeLano, http://www.pymol.org), MSMS[53] and DINO (Phillippsen, http://cobra.mih.unibas.ch/dino).

Protein Data Bank accession codes

Coordinates and structure factors are deposited in the Protein Data Bank (PDB) with accession codes 2J5T and 2J5V.

Acknowledgements. This work was supported by grants BFU2004-05159, BFU2004-04472 from the Spanish Ministry of Education and Science, and PI052838 from the Spanish Ministry of Health. Data collection at beamline ID14.4 and BM16 of ESRF were under the auspices of the EU and the Spanish Ministry of Science and Technology, respectively. We thank ESRF personnel for their help.

REFERENCES

1. Csonka, L. N. & Hanson, A. D. (1991). Prokaryotic osmoregulation: genetics and physiology. *Annu. Rev. Microbiol.* **45**, 569-606.

2. Delauney, A. J. & Verma, D. P. S. (1990). Proline biosynthesis and osmoregulation in plants. *Plant J.* **4**, 215-223.

3. Smirnoff, N. & Cumbes, Q. J. (1989). Hydroxyl radical scavenging activity of compatible solutes. *Phytochemistry*, **28**, 1057-1060.

4. Kohl, D. H., Schubert, K. R., Carter, M. B., Hagedorn, C. H. & Shearer, G. (1988). Proline metabolism in N_2-fixing root nodules: energy transfer and regulation of purine synthesis. *Proc. Natl. Acad. Sci. U S A*, **85**, 2036-2040.

5. Venekamp, J. H., Lampe, J. E. M. & Koot, J. T. M. (1989). Organic acids as sources of drought-induced proline synthesis in field bean plants, *Vicia faba* L. *J. Plant Physiol.* **133**, 654-659.

6. Leisinger T. (1996). Biosynthesis of proline in: *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology (Neidhardt, F.C., Ed.), pp. 434-441, ASM Press, Washington, DC.

7. Hu, C. A., Delauney, A.J. & Verma D.P. (1992). A bifunctional enzyme (Δ^1-pyrroline-5-carboxylate synthetase) catalyzes the first two steps in proline biosynthesis in plants. *Proc. Natl. Acad. Sci. U S A.*, **89**, 9354-9358.

8. Aral, B. & Kamoun, P. (1997). The proline biosynthesis in living organisms. *Amino Acids,* **13**, 189-217.

9. Alonso, E. & Rubio, V. (1989). Participation of ornithine aminotransferase in the synthesis and catabolism of ornithine in mice. Studies using gabaculine and arginine deprivation. *Biochem. J.* **259**, 131-138.

10. Hu, C. A. Lin, W. W., Obie, C. & Valle, D. (1999). Molecular enzymology of mammalian Δ^1-pyrroline-5-carboxylate synthase. Alternative splice donor utilization generates isoforms with different sensitivity to ornithine inhibition. *J. Biol. Chem.* **274,** 6754–6762.

11. Baumgartner, M. R., Hu, C. A., Almashanu, S., Steel, G., Obie, C., Aral, B., *et al.* (2000). Hyperammonemia with reduced ornithine, citrulline, arginine and proline: a new inborn error caused by a mutation in the gene encoding Δ^1-pyrroline-5-carboxylate synthase. *Hum. Mol. Genet.* **9,** 2853-2858.

12. Katchalsky, A. & Paecht, M. (1954). Phosphate anhydrides of amino acids. *J. Am. Chem. Soc.* **76**, 6042-6044.

13. Seddon, A. P., Zhao, K. Y. & Meister, A. (1989). Activation of glutamate by γ-glutamate kinase: formation of γ-cis-cycloglutamyl phosphate, an analog of γ-glutamyl phosphate. *J. Biol. Chem.* **264**, 11326-11335.

14. Deutch, A. H., Rushlow, K. E. & Smith, C. J. (1984). Analysis of the *Escherichia coli proBA* locus by DNA and protein sequencing. *Nucleic Acids Res.* **12,** 6337-6355.

15. Massarelli, I., Forlani, G., Ricca, E. & De Felice, M. (2000). Enhanced and feedback-resistant γ-glutamyl kinase activity of an *Escherichia coli* transformant carrying a mutated *proB* gene of *Streptococcus thermophilus*. *FEMS Microbiol. Lett.* **182,** 143-147.

16. Page, R., Nelson, M. S., von Delft, F., Elsliger, M. A., Canaves, J. M., Brinen, L. S., *et al.* (2004). Crystal structure of γ-glutamyl phosphate reductase (TM0293) from *Thermotoga maritima* at 2.0 Å resolution. *Proteins*, **54**, 157-161.

17. Serebrijski, I., Wojcik, F., Reyes, O. & Leblon, G. (1995). Multicopy suppression by *asd* gene and osmotic stress-dependent complementation by heterologous *proA* in *proA* mutants. *J. Bacteriol.* **177**, 7255-7260.

18. Ramón-Maiques, S., Marina, A., Gil-Ortiz, F., Fita, I. & Rubio, V. (2002). Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the amino acid kinase enzyme family, during catalysis. *Structure,* **10,** 329-342.

19. Smith, C. J., Deutch, A. H. & Rushlow, K. E. (1984) Purification and characteristics of a γ-glutamyl kinase involved in *Escherichia coli* proline biosynthesis. *J. Bacteriol*. **157**, 545-551.

20. Aravind, L. & Koonin, E. V. (1999). Novel predicted RNA-binding domains associated with the translation machinery. *J. Mol. Evol.* **48**, 291-302.

21. Perez-Arellano, I., Rubio, V. & Cervera, J. (2005). Dissection of *Escherichia coli* glutamate 5-kinase: functional impact of the deletion of the PUA domain. *FEBS Lett*. **579,** 6903-6908.

22. Ishitani, R., Nureki, O., Fukai, S., Kijimoto, T., Nameki, N., Watanabe M., *et al.* (2002). Crystal structure of archaeosine tRNA-guanine transglycosylase. *J. Mol. Biol.* **318**, 665-677.

23. Pan, H., Agarwalla, S., Moustakas, D. T, Finer-Moore J., & Stroud R. M. (2003). Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit, *Proc. Natl. Acad. Sci. USA,* **100**, 12648-12653.

24. Ogura, M., Kawata-Mukai, M., Itaya, M., Takio, K. & Tanaka, T. (1994). Multiple copies of the *proB* gene enhance degS-dependent extracellular protease production in *Bacillus subtilis*. *J. Bacteriol*. **176**, 5673-5680.

25. Ogura, M. & Tanaka, T. (1996). Transcription of *Bacillus subtilis degR* is sigma D dependent and suppressed by multicopy *proB* through sigma D. *J. Bacteriol.* **178**, 216-222.

26. Perez-Arellano, I., Rubio, V. & Cervera, J. (2006). Mapping active site residues in glutamate-5-kinase. The substrate glutamate and the feed-back inhibitor proline bind at overlapping sites. *FEBS Lett*. **580**, 6247-6253

27. Ramón-Maiques, S., Marina, A., Uriarte, M., Fita, I. & Rubio, V. (2000). The 1.5 Å resolution crystal structure of the carbamate kinase-like carbamoyl phosphate synthetase from the hyperthermophilic archaeon *Pyrococcus furiosus*, bound to ADP, confirms that this thermostable enzyme is a carbamate kinase, and provides insight into substrate binding and stability in carbamate kinases. *J. Mol. Biol.* **299,** 463-476.

28. Marco-Marin, C., Gil-Ortiz, F. & Rubio V. (2005). The crystal structure of *Pyrococcus furiosus* UMP kinase provides insight into catalysis and regulation in microbial pyrimidine nucleotide biosynthesis. *J. Mol. Biol.* **352**, 438-54.

29. Li, L. & Ye, K. (2006). Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature*, **443**, 302-307.

30. Rossmann, M. G., Adams, M. J., Buehner, M., Ford, G. C., Hackert, M. L., Liljas, A. *et al.* (1973). Letter: Molecular symmetry axes and subunit interfaces in certain dehydrogenases. *J. Mol. Biol.* **76**, 533-537.

31. Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S. & Yokoyama, S. (2003). Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. *Cell*, **113**, 383-394.

32. Gil-Ortiz, F., Ramon-Maiques, S., Fita, I. & Rubio, V. (2003). The course of phosphorus in the reaction of N-acetyl-L-glutamate kinase, determined from the structures of crystalline complexes, including a complex with an AlF_4^- transition state mimic. *J. Mol. Biol.* **331**, 231-244.

33. Fujita, T., Maggio, A., Garcia-Rios, M., Stauffacher, C., Bressan R. A. & Csonka L. N. (2003). Identification of regions of the tomato γ-glutamyl kinase that are involved in allosteric regulation by proline. *J. Biol. Chem.* **278**, 14203-14210.

34. Marco-Marín, C., Ramón-Maiques, S., Tavárez, S. & Rubio, V. (2003) Site-directed mutagenesis of *Escherichia coli* acetylglutamate kinase and aspartokinase III probes the catalytic and substrate-binding mechanisms of these amino acid kinase family enzymes and allows three-dimensional modelling of aspartokinase. *J. Mol. Biol*. **334**, 459-476.

35. Kotaka, M., Ren, J., Lockyer, M., Hawkins, A. R. & Stammers, D.K. (2006). Structures of R- and T-state *Escherichia coli* aspartokinase III: Mechanisms of the allosteric transition and inhibition by lysine. *J. Biol. Chem*. **281**, 31544-31552.

36. Krishna, R.V. & Leisinger, T. (1979). Biosynthesis of proline in *Pseudomonas aeruginosa*. Partial purification and characterization of γ-glutamyl kinase. *Biochem. J.* **181**, 215-222.

37. Briozzo, P., Evrin, C., Meyer, P., Assairi, L., Joly, N., Barzu, O. & Gilles, A. M. (2005). Structure of *Escherichia coli* UMP kinase differs from that of other nucleoside monophosphate kinases and sheds new light on enzyme regulation. *J. Biol. Chem*. **280**, 25533-25540.

38. Ramón-Maiques, S., Fernandez-Murga, M. L., Gil-Ortiz, F., Vagin, A., Fita, I. & Rubio, V. (2006). Structural bases of feed-back control of arginine biosynthesis, revealed by the structures of two hexameric N-acetylglutamate kinases, from *Thermotoga maritima* and *Pseudomonas aeruginosa*. *J. Mol. Biol*. **356**, 695-713.

39. Perez-Arellano, I., Gil-Ortiz, F., Cervera, J. & Rubio, V. (2004). Glutamate-5-kinase from *Escherichia coli*: gene cloning, overexpression, purification and crystallization of the recombinant enzyme and preliminary X-ray studies. *Acta Crystallogr. D Biol. Crystallogr*. **60**, 2091-2094.

40. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760-763.

41. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30,** 1022-1025.

42. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 905-921.

43. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* **60**, 2126-2132.

44. Jones, T. A., Zou, J.-Y., Cowan S., & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. Sect. A*, **47**, 110-119.

45. Murshudov, G. N., Vagin, A.A.& Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst. D Biol. Crystallogr.* ***53***, 240-255.

46. Laskowski, R. A. , MacArthur, M. W., Moss D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.

47. Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallog. Sect. A*, **32**, 922-923.

48. Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815.

49. McElroy, W. D., Seliger, H. H. & White, E. H. (1969). Mechanism of bioluminescence, chemiluminescence and enzyme function in the oxidation of firefly luciferin. *Photochem. Photobiol.* **10**, 153-170.

50. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.

51. Esnouf, R. M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 938-940.

52. Merritt, E. A., & Murphy, M. E. P. (1994). Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 869-873.

53. Sanner, M. F., Olson, A. J., & Spehner, J. C. (1996). Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, **38**, 305-320.

FIGURE LEGENDS

Figure 1. Pathway of proline synthesis in microorganisms and plants and of ornithine synthesis in mammals. Enzymes are enclosed in grey boxes. Feed-back inhibition of microbial and plant G5Ks by proline and of animal G5Ks by ornithine is indicated with broken arrows. The dotted arrow indicates the spontaneous cyclisation of G5P to 5-oxoproline that is an abortive side-reaction.

Figure 2. The structure of *E. coli* glutamate 5-kinase. (a) Backbone representation of two views of the asymmetric unit of the $P2_1$ crystal, formed by two tetramers coloured blue and red. The unit cell is shown as a black lined box. (b) and (c) Surface representation of two perpendicular views of the glutamate 5-kinase tetramer, with the substrates in space-filling representation. Different colours are used for each subunit with different intensities for the amino acid kinase (AAK, darker) and PUA (lighter) domains. The molecular P, Q and R axes are identified, and are represented with broken lines. (d) Magnesium binding site represented with the electrostatic potential of the solvent accessible surface. Positively and negatively charged regions are coloured blue and red, respectively, with the intensity of the colour being proportional to the local potential. The $2F_{obs}$-F_{calc} omit electron density map contoured at the 1 σ level, is also shown (in green) with the bound magnesium ions as the purple spheres. (e) Semitransparent surface representation of the PUA domain showing the bound glutamate (ball and stick representation) and its corresponding $2F_{obs}$-F_{cal} omit electron density map contoured at the 1 σ level.

Figure 3. Structure of the *E. coli* G5K subunit (a) Stereo view of the superposition of the C^{α} traces of G5K from *E. coli* (green) and *C. jejuni* (pink). The absence of the PUA

domain is evident in the *C. jejuni* enzyme. (b) The *E. coli* G5K subunit structure with the catalytic AAK (green) and the PUA (orange) domains. Secondary elements and the link between domains (purple) are also indicated. Two molecules of glutamate, a sulphate ion and a magnesium ion bound to the G5K subunit are represented with balls and sticks. (c) Correspondence between the amino acid sequence (single-letter code) and the secondary structure. Superposed cylinders denote α-helices, and arrows denote β-strands. T colour code is as in (b). Residues having decreased accessibility upon the binding of glutamyl 5-phosphate, sulphate, 5-oxoproline or glutamate in the amino acid kinase domain, are indicated with cyan, violet, brown or blue triangles respectively. Residues having decreased accessibility upon the binding of glutamate in the PUA domain are indicated with yellow triangles. Residues that form the negatively charged hole where magnesium binds are shown with red triangles. Open and closed circles indicate residues that exhibit decreased accessibility upon homodimer and tetramer formation, respectively.

Figure 4. Oligomerisation surfaces of one subunit across the (a) R and (b) Q dimers (see in the text). Contacting non-polar (reddish) and polar (light green) atoms are indicated together with the corresponding structural elements.

Figure 5. Comparison between the dimers of *E. coli* NAGK, *P. furiosus* UMP kinase (UMPK) and *E. coli* G5K. Left panel, ribbon representation of the homodimers of NAGK, UMPK and G5K, to highlight the drastically different dimer architectures. Subunits on the left (lighter) have the same orientation, in the three enzymes. Right panel, elements of the two subunits involved in dimer formation, to show their different relative orientations in the three enzymes.

Figure 6. Substrate binding. (a) Stereoview of the glutamate 5-kinase (G5K) active centre, in the monoclinic crystal. The bound glutamate and sulphate are shown in ball and stick representation. Amino acid side chains are shown in thinner trace. The ($2F_{obs}$-F_{cal}) omit electron density map, contoured at the 1 σ around the bound substrates is also shown (green). Hydrogen bonds and ion pairs are shown as broken black lines. (b) Similarly for the tetragonal crystal, with the corresponding bound molecules of glutamyl 5-phosphate, 5-oxoproline and sulphate (c) Stereo view of the superimposition of the substrate binding region and of the bound ligands in the monoclinic (colored) and tetragonal crystals of *E. coli* G5K. The only significant differences are observed between the glutamate and glutamylphosphate. Oxoproline is present only in the tetragonal crystal.

Figure 7. The active centre. (a) Semitransparent surface representation of a detailed view of the G5K binding site, showing bound glutamate in stick representation as well as an ADP molecule modelled from the structure of *C. jejuni* G5K in the presence of this nucleotide. Important side chains are represented in bonds and identified. (b) Stereo view showing the hydrogen bond network between the glutamate binding sites of the two subunits of the R dimer. Each subunit is shown in a different colour.

Figure 8. Movement of the β4-αE loop. Stereo view of the superposition of the *E. coli* (green) and *C. jejuni* (violet) loops and neighboring elements including (in stick representation) bound 5-oxoproline. The side chain of the important residue for proline binding, Asp148 (Asp141 of *C. jejuni* G5K), is also represented, to illustrate its radical change of orientation from looking outwards to looking inwardly to the site.

Figure 9. Possible interaction between the glutamate 5-kinase (G5K) and the glutamyl 5-phoshate reductase (G5PR)**.** Semitransparent surface representation of two perpendicular views of the G5K tetramer, with the substrates in space-filling representation. A dimer of the G5PR from *T. maritima* is shown (ribbon representation) with one subunit (orange) presenting the open conformation as observed in the crystal structure of this enzyme in the absence of substrates (PDB 1O20), and with the other subunit (yellow) presenting a closed conformation modelled from class 3 aldehyde dehydrogenase complexed to NAD (PDB 1AD3). The catalytic and NADPH binding domains of G5PR are identified.

Table 1. Data collection and refinement statistics

	Tetragonal crystals	Monoclinic crystals
A. Data collection		
ESRF Beamline/Wavelenght (Å)	BM16/0.979	ID14.4/0.979
Space group	$P4_12_12$	$P2_1$
Unit cell (Å or º)	a = b = 101.1, c = 178.6	a = 96.30, b = 124.11, c = 144.93; γ = 94.0
Resolution range (Å)	48.8-2.5 (2.64-2.50)	50-2.9 (3.06-2.90)
Reflections, total/unique	367,219/ 61,566 (29,004/ 8,941)	339,415/ 74,908 (49,569/ 10,882)
Completeness (%)	99.5 (99.5)	99.3 (99.6)
I/σ	6.8 (1.7)	9.7 (2.9)
R_{sym} (%)[a]	6.8 (41.6)	16.5 (49.4)
B. *Refinement statistics*		
Resolution range (Å)	49 - 2.5	50 - 2.9
R-factor/ R_{free} (%)	19.3/24.3	19.7/24.7
Molecules and atoms refined	Polypeptide chains: 2	Polypeptide chains : 8
	Protein atoms: 4,872	Protein atoms: 21,344
	Glutamyl-5-phosphate molecules: 2	Glutamate molecules : 12
	5-Oxoproline molecules:2	Sulphate ions: 8
	Sulphate ions: 2	Magnesium ions: 8
	Magnesium ions: 1	Chloride ions: 20
	Water molecules: 166	Water molecules: 103[c]

Rmsd from ideal		
Bond lengths (Å)	0.011	0.009
Bond angles (°)	1.35	1.16
Average B-factor (Å^2)		
Protein	36.9	19.0
Water	35.9	5.7[b]
Glutamate		28.4
Glutamyl 5-phosphate	42.0	
5-Oxoproline	38.6	
Sulphate	55.7	54.4
Magnesium	19.2	39.9
Chloride		29.2
Ramachandran plot (%)[c]		
Most favoured	94.1	90.3
Additionally allowed	5.9	9.4
Generously allowed	0	0.3
Disallowed	0	0

Values in parentheses are data for the highest resolution shell

[a] $R_{sym} = \Sigma \left| I - \langle I \rangle \right| / \Sigma I$, where I is the observed intensity and <I> is the average intensity of multiple observations of symmetry-related reflections.

[b] These waters correspond to well-defined densities. However, at the 2.9-Å resolution of this structure some of the waters may actually correspond to ions, as suggested by the low B-factors.

[c] Calculated using PROCHECK[46]

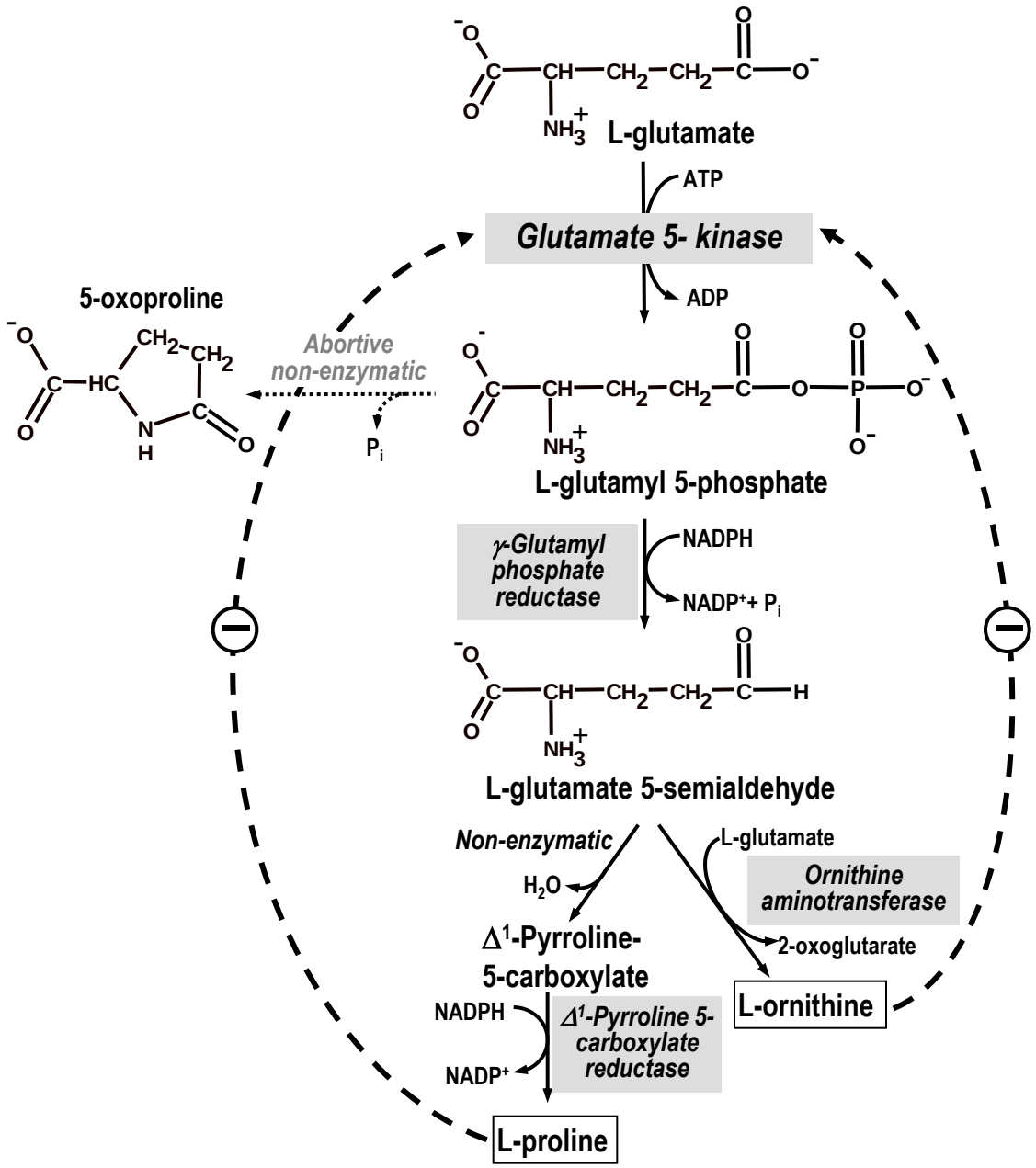

L-glutamate
ATP
Glutamate 5- kinase
ADP
5-oxoproline
Abortive non-enzymatic
P_i
L-glutamyl 5-phosphate
γ-Glutamyl phosphate reductase
NADPH
$NADP^+ + P_i$
L-glutamate 5-semialdehyde
Non-enzymatic
H_2O
L-glutamate
Ornithine aminotransferase
2-oxoglutarate
Δ^1-Pyrroline-5-carboxylate
L-ornithine
NADPH
Δ^1-Pyrroline 5-carboxylate reductase
$NADP^+$
L-proline

Figure 1

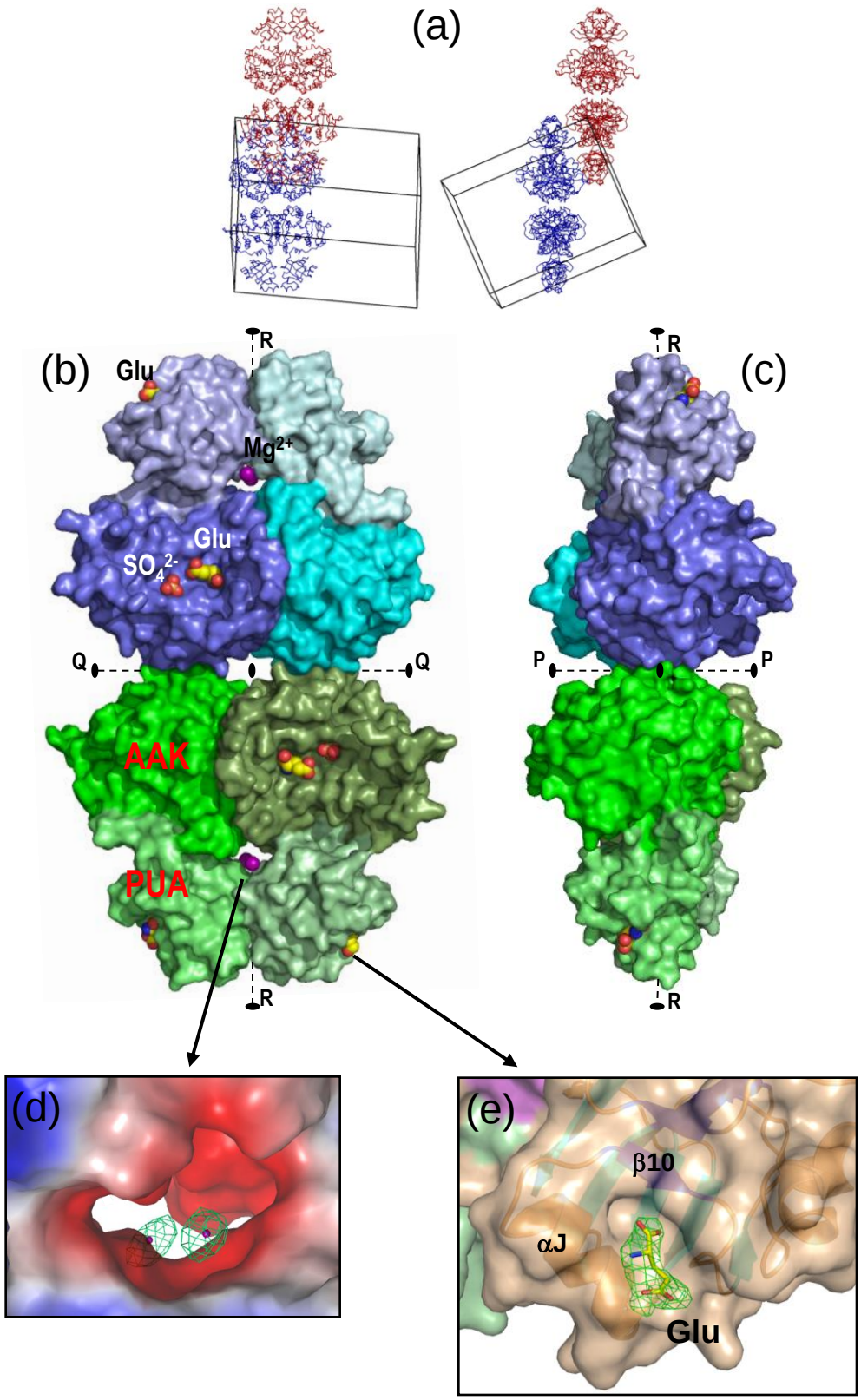


Figure 2

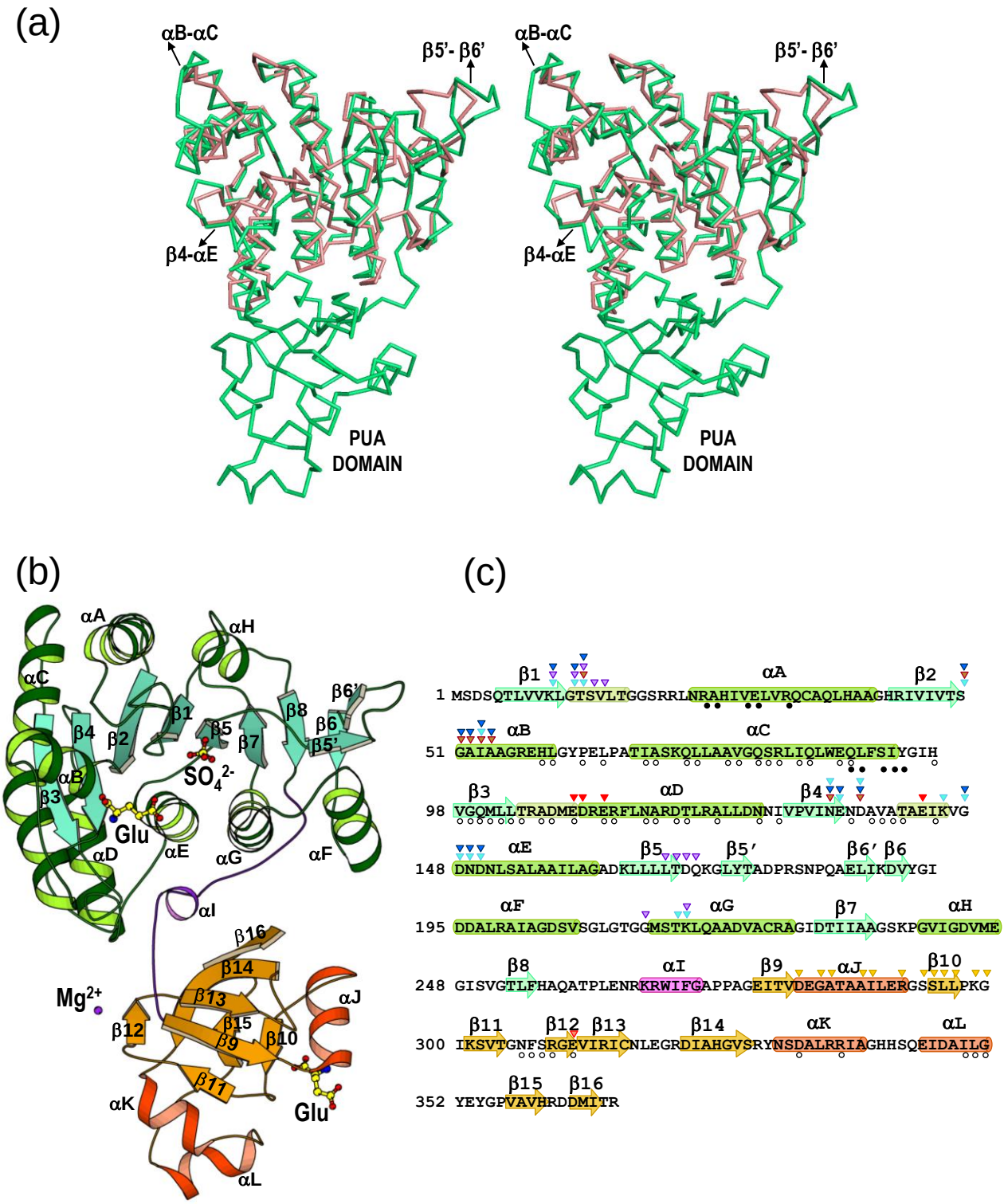

(a)
αB-αC
β5'- β6'
β4-αE
PUA
DOMAIN
αB-αC
β5'- β6'
β4-αE
PUA
DOMAIN
(b)
αA
αH
αC
β6'
β1
β8
β6
β4
β2
β5
β7
β5'
αB
β3
SO4 2-
Glu
αD
αE
αG
αF
αI
β16
β14
Mg2+
β13
αJ
β12
β15
β9
β10
β11
αK
Glu
αL
(c)
β1
αA
β2
1 MSDSQTLVVKLGTSVLTGGSRRLNRAHIVELVRQCAQLHAAGHRIVIVTS
αB
αC
51 GAIAAGREHLGYPELPATIASKQLLAAVGQSRLIQLWEQLFSIYGIH
β3
αD
β4
98 VGQMLLTRADMEDRERFLNARDTLRALLDNNIVPVINENDAVATAEIKVG
αE
β5
β5′
β6′ β6
148 DNDNLSALAAILAGADKLLLLTDQKGLYTADPRSNPQAELIKDVYGI
αF
αG
β7
αH
195 DDALRAIAGDSVSGLGTGGMSTKLQAADVACRAGIDTIIAAGSKPGVIGDVME
β8
αI
β9
αJ
β10
248 GISVGTLFHAQATPLENRKRWIFGAPPAGEITVDEGATAAILERGSSLLPKG
β11
β12
β13
β14
αK
αL
300 IKSVTGNFSRGEVIRICNLEGRDIAHGVSRYNSDALRRIAGHHSQEIDAILG
β15
β16
352 YEYGPVAVHRDDMITR

Figure 3

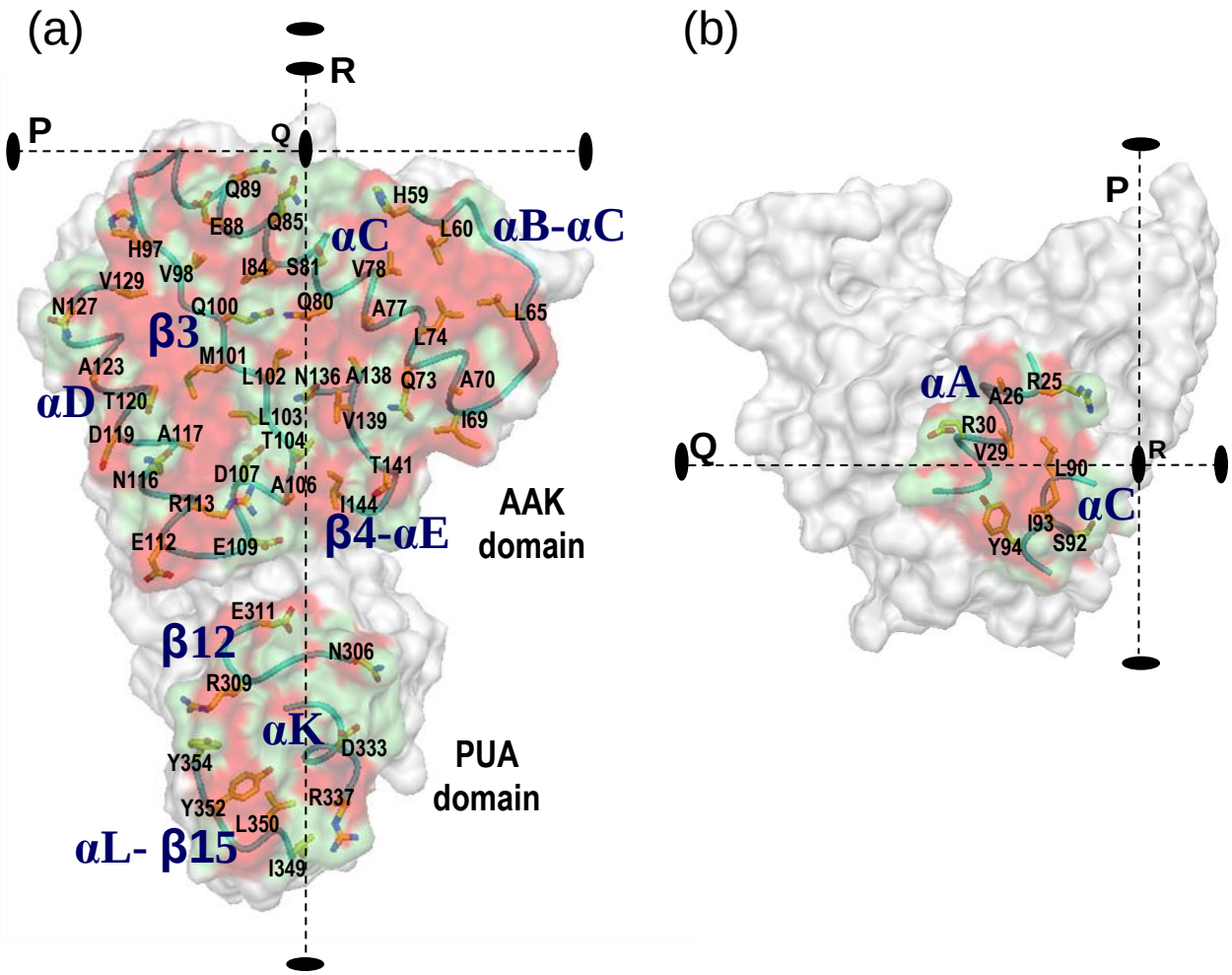
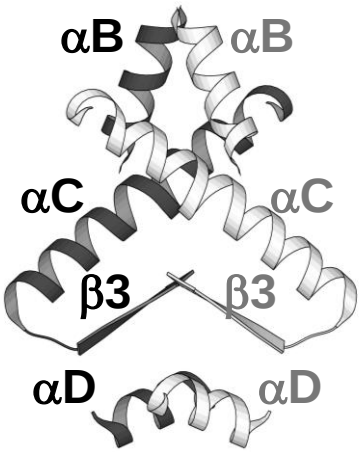

(a)
R
P
Q
Q89
H59
E88
Q85
αC
L60
αB-αC
H97
I84
S81
V78
V129
V98
N127
Q100
Q80
A77
L65
β3
L74
M101
A123
L102
N136
A138
Q73
A70
αD
T120
L103
V139
I69
D119
A117
T104
T141
N116
D107
A106
R113
I144
AAK
domain
β4-αE
E112
E109
E311
β12
N306
R309
αK
D333
PUA
domain
Y354
R337
Y352
L350
αL- β15
I349
(b)
P
αA
R25
A26
R30
V29
Q
R
L90
αC
I93
Y94
S92

Figure 4

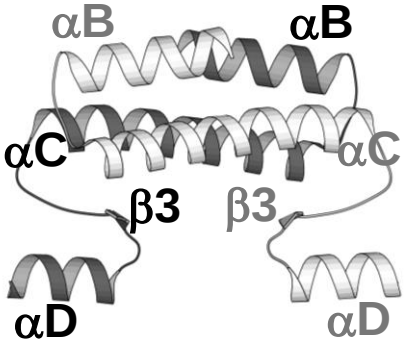
(a)

NAGK



(b)

UMPK



(c)

G5K

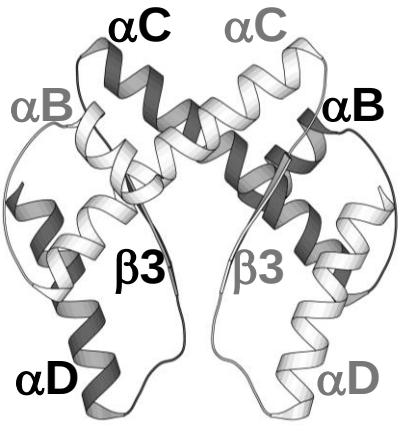


Figure 5

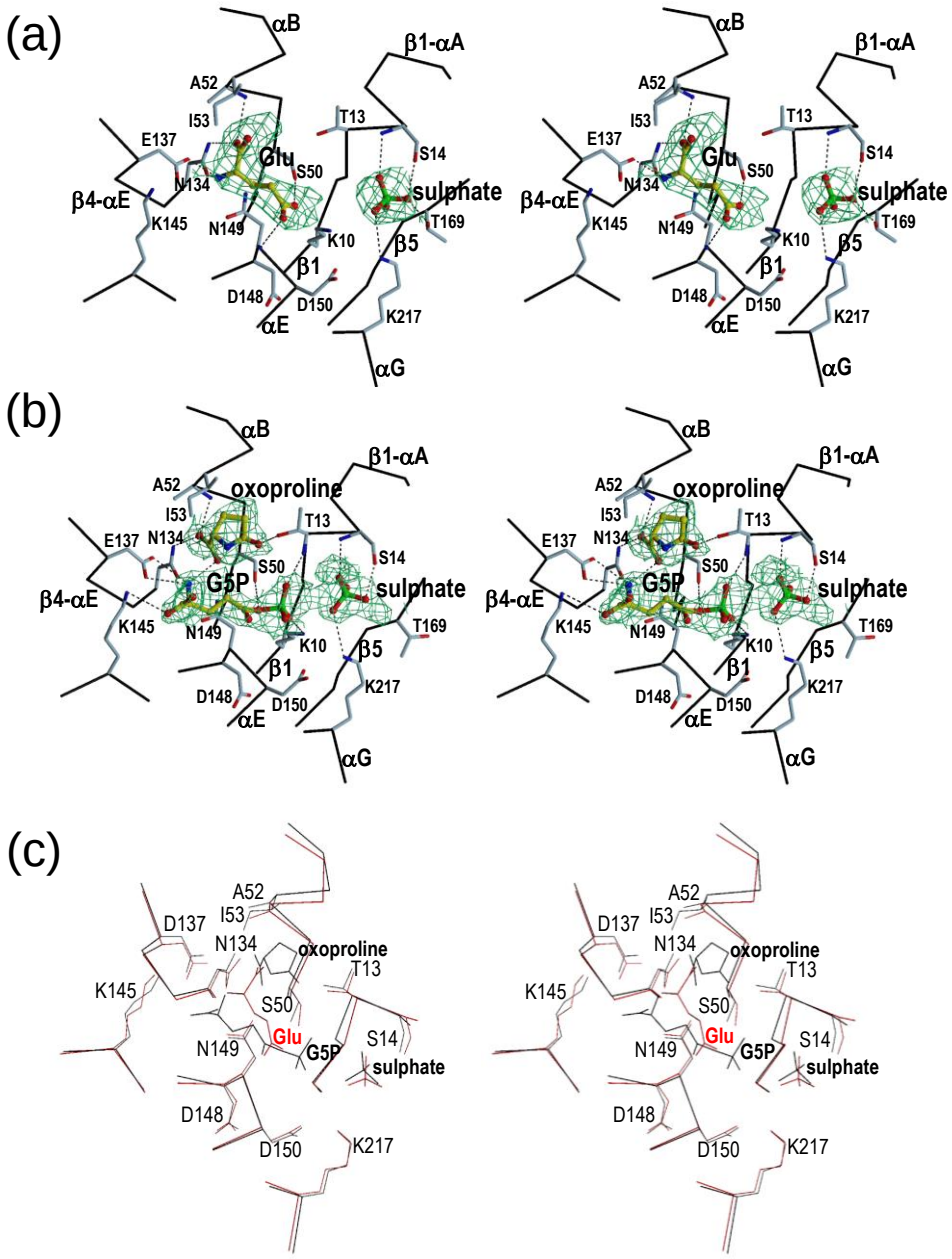
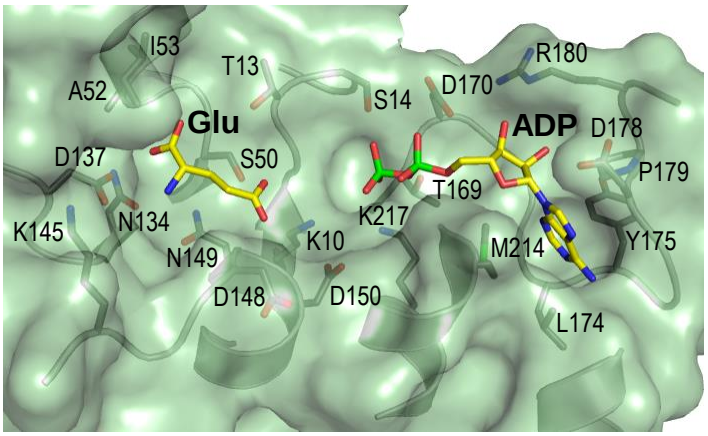

(a)
αB
β1-αA
A52
I53
T13
E137
Glu
S50
S14
sulphate
N134
β4-αE
K145
N149
T169
K10
β5
β1
D148
D150
K217
αE
αG
(b)
oxoproline
G5P
(c)
D137

Figure 6

(a)


I53
T13
R180
A52
S14
D170
Glu
ADP
D178
D137
S50
P179
T169
K217
K145
N134
K10
M214
Y175
N149
D148
D150
L174

(b)

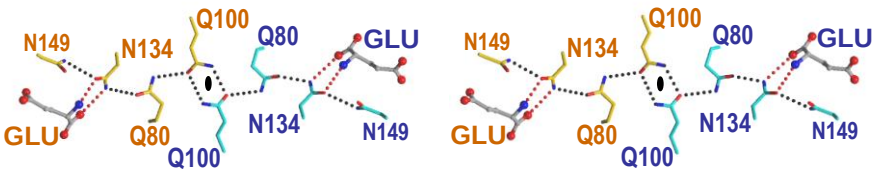

N149
N134
Q100
Q80
GLU
GLU
Q80
Q100
N134
N149
N149
N134
Q100
Q80
GLU
GLU
Q80
Q100
N134
N149

Figure 7

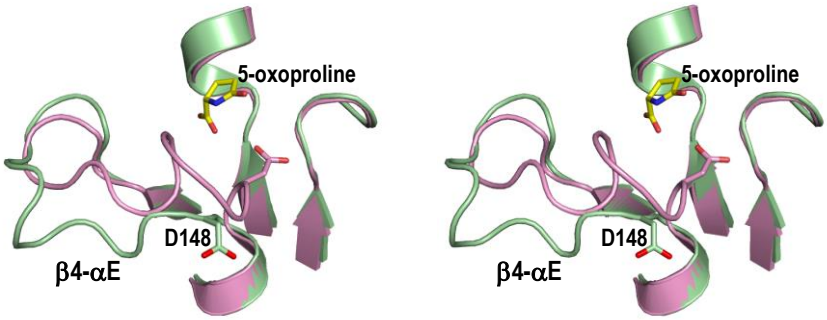


Figure 8

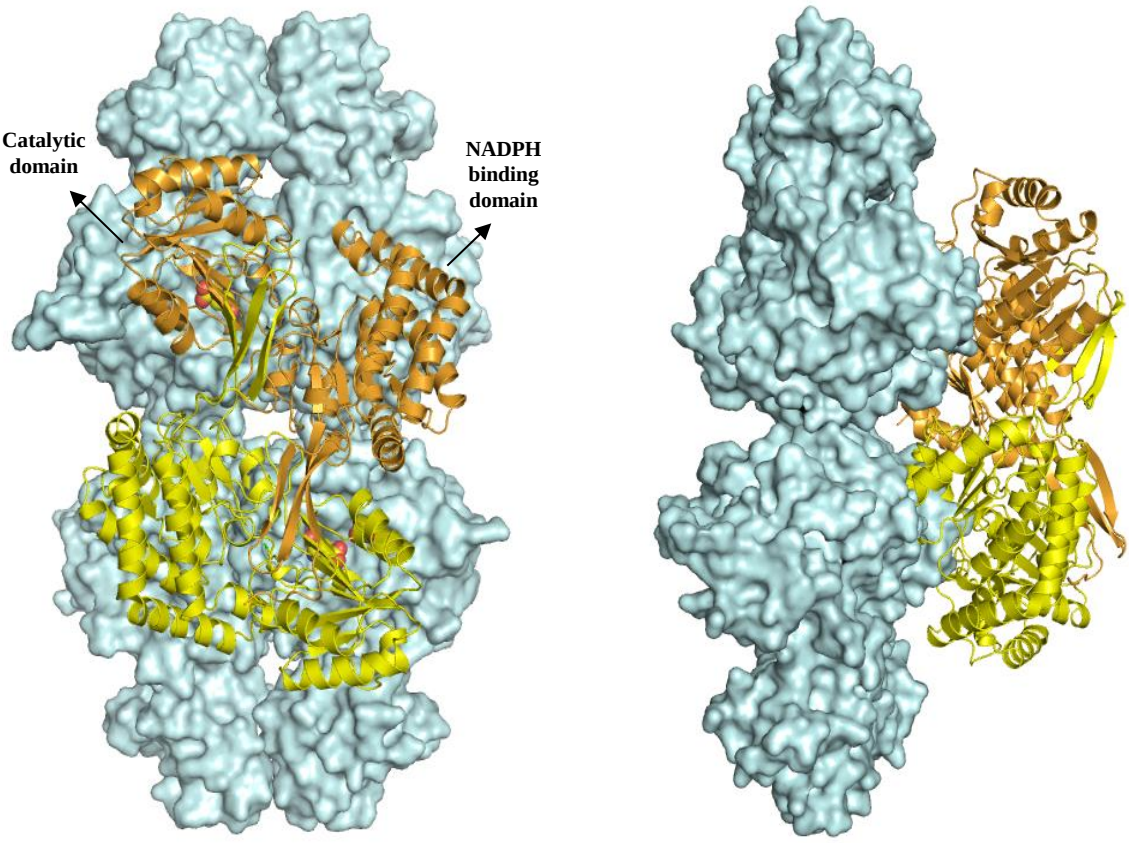


Figure 9