SUPPLEMENTARY MATERIAL

Enzyme closure and nucleotide binding structurally lock guanylate kinase

Olivier Delalande, Sophie Sacquin-Mora and Marc Baaden

Supplementary Text: Full simulation and analysis methods

Simulations

Low resolution Brownian dynamics.

All low resolution Brownian dynamics (BD) simulations were run with the ProPHet (Probing Protein Heterogeneity) program (1,2). In this approach, the protein is represented using an elastic network model (ENM). In contrast to most common coarse-grain models which use a single pseudoatom per residue (3), we use a more detailed representation (4) that involves 2-3 pseudoatoms for each residue and enables different amino acids to be distinguished. Pseudoatoms closer than the cutoff parameter R_c=9 Å are joined by Gaussian springs and all such springs have identical spring constants of γ=0.42 N m^{-1} (0.6 kcal mol^{-1} Å^{-2}). The springs are taken to be relaxed (with lengths $d_{ij} = d_{ij}^0$) for the experimentally observed conformation of the protein, in this case the crystallographic conformation of GK_{MT} with PDB-id 1S4Q.

Mechanical properties are obtained from 50,000 BD steps at 300 K. The simulations are analyzed in terms of the fluctuations on the mean distance between each pseudoatom belonging to a given amino acid residue and the pseudoatoms belonging to the remaining residues of the protein. The inverse of these fluctuations yields an effective force constant k_i describing the ease of moving a pseudoatom with respect to the overall protein structure.

$$k_i = \frac{3k_B T}{\left\langle \left(d_i - \langle d_i \rangle\right)^2 \right\rangle},$$

where $\langle\rangle$ denotes an average taken over the whole simulation and $d_i = \langle d_{ij} \rangle_{j*}$ is the average distance from particle i to the other particles j in the protein (the sum over j* implies the exclusion of the pseudoatoms belonging to residue i). Distances between the C_α pseudoatom of residue i and the C_α pseudoatoms of the adjacent residues *i-1* and *i+1* are excluded since they are virtually constant. The force constant for each residue k is the average of the force constants for all its constituent pseudo atoms i. We will use the term “rigidity profile” to describe the ordered set of force constants for all the residues of the protein.

In ProPHet we can use the fluctuations of the interresidue distances d_{ij} to compute force constants similar to those obtained by Eyal and Bahar (5) with a one-point-per-residue ENM. The resulting directional force constants (DFC) are simply calculated as:

$$k_{ij} = \frac{3kT}{\left\langle \left(d_{ij} - \langle d_{ij} \rangle^2\right)^2 \right\rangle},$$

and can be used to construct a complete map of the mechanical resistance of a protein in response to all possible pulling directions.

Molecular Dynamics Simulations.

All molecular dynamics simulations were run with the Yasara program and the Yamber3 forcefield (6). The simulations described in Supplementary Table 1 were run in the NVT ensemble at a temperature of 298 K, after solvation of GK within rectangular explicit solvent boxes using the TIP3P water model under periodic boundary conditions and after initial equilibration. A double integration timestep of 1fs for bonded and 2fs for non-bonded interactions was used. The simulation cell was neutralized using Yasara's default salt

concentration of 0.9% NaCl. Production trajectories of 20 nanoseconds were collected at 2 ps intervals for each of the six states representing reaction intermediates.

Forcefield parameterization. All MD simulations set up with the Yasara program used the implemented AutoSMILES procedure to parameterize bound substrate, product or co-factor nucleotides.

pKa calculations were performed with the Yasara software (www.yasara.org, 7) at neutral pH, where maximal GK activity is observed (8). No significant protonation state differences were apparent between the *apo*/open, *apo*/closed and GMP-bound/closed states shown in Figure 2. On the basis of these pKa calculations, we used a consensus protonation state where all histidines are neutral, including the His93 and His97 residues which are part of or close to the catalytic site. All other ionisable residues are in their standard protonation states.

Quality evaluation of the GK_{MT} closed model. *M. Muscus* GK (1LVG, closed structure) has a high sequence identity towards *M. Tuberculosis* GK (1S4Q, open structure), lying between 37% (ALIGN) and 43% (BLASTP) depending on the sequence alignment and gap definitions. Homology models produced from a structural pattern sharing over 20% sequence identity are commonly considered as highly reliable. Figure 3 shows that all residues defining the catalytic site are fully conserved (*) or highly conserved (:) after replacement by strictly equivalent amino acids (S/T, E/D or D/N, with respect to the C=O group). Less conserved (.) residues (A31, V32) are bearing only backbone H-bond contacts with substrates/products, but they belong to the P-loop that is considered as an extremely conserved motif in all guanylate kinases. Superimposition of the homology model to the experimental structures (Figure S1A) confirms the good quality of our closed model designed for *M. Tuberculosis*. More precise evaluation of the homology model quality has been performed using Procheck (9) and MolProbity (10) as well as following a protocol that we have previously described (11). Confronting homology models to Ramachandran criteria, Procheck recorded 92.9% of residues in the most favoured regions (without any disallowed ones). On the same structure, MolProbity recorded 95.5% of Ramachandran favoured residues (without any outliers); checking the 1LVG closed GK form used as a structural pattern led to 96.8% of the residues in the favoured Φ/Ψ regions. RMSD measured after 2ns MD trajectory for the closed state models (3 to 6) are ranging from 1.2Å (State4) to 1.5Å (State3 and 5), which is an excellent proof of model reliability considering the 2.1Å resolution of the 1LVG structure used as pattern for the homology-based construction (see ref. 11 for a detailed discussion). Looking at the conservation in the spatial positioning of the catalytic residues (Figures S1B and S1C), and taking into account that the homology model is the only structure to bear all substrates and cofactors, thus necessarily modulating the position of linking amino acids, the spatial definition of the catalytic site is remarkably conserved. In Figure S2, a comparison between the homology model and the experimental structures is shown, indicating a great conservation of the highly positive ATP-binding pocket (principally defined by the P-loop motif) in its topography and in the spatial distribution of hydrophobic potential and electrostatic potential.

Analysis

The tools included in the Yasara and Gromacs3.3 (12) software packages were used for post-processing and analysis of the MD trajectories. The VMD software (13) was used for visualization and graphical representations. The DynDom (14) web server was used for sub-domain determination in the final models. Molecular Hydrophobic Potentials (MHP) have been computed with the Platinum server (15) and electrostatics with the APBS program (16). Charts were prepared with the XmGrace program.

References

1. Sacquin-Mora, S., and R. Lavery. 2006. Investigating the local flexibility of functional residues in hemoproteins. *Biophys. J.* 90:2706-2717.
2. Sacquin-Mora, S., E. Laforet, and R. Lavery. 2007. Locating the active sites of enzymes using mechanical properties. *Proteins* 67:350-359.
3. Tozzini, V. 2005. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* 15:144-150.
4. Zaccharias, M. 2003. Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.* 12:1271-1282.
5. Eyal, E., and I. Bahar. 2008. Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. *Biophys. J.* 94:3424-3435.
6. Krieger, E., T. Darden, …, and G. Vriend. 2004. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins* 57:678-83.
7. Krieger, E., J. E. Nielsen, …, and G. Vriend. 2006. Fast empirical pK(a) prediction by Ewald summation. *J. Mol. Graph. Model.* 25:481-6.
8. Oeschger, M. P. 1978. Guanylate kinase from escherichia coli B. *Methods Enzymol.* 51:473-482.
9. Laskowski, R. A., M. W. MacArthur, …, and J. M. Thornton. 1993. PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Cryst.* 26:283-291.
10. Davis, I. W., A. Leaver-Fay, …, and D. C. Richardson. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35:W375-83.
11. Law, R. J., C. Capener, …, and M. S .P. Sansom. 2005. Membrane protein structure quality in molecular dynamics simulation. *J Mol Graph Model.* 24:157-165.
12. van der Spoel, D., E. Lindahl, …, and H. J. C. Berendsen. 2005. GROMACS: Fast, Flexible and Free. *J. Comput. Chem.* 26:1701-1718.
13. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD - Visual Molecular Dynamics. *J. Molec. Graphics* 14 :33-38.
14. Lee, R. A., M. Razaz, and S. Hayward. 2003. The DynDom database of protein domain motions. *Bioinformatics* 19:1290-1291.
15. Efremov, R. G., A. O. Chugunov, ..., and E. Jacoby. 2007. Molecular lipophilicity in protein modeling and drug design. *Curr. Med. Chem.* 14:393-415.
16. Baker, N. A., D. Sept, …, and J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98:10037-10041.
17. Hible, G., L. Renault, …, and J. Cherfils. 2005. Calorimetric and crystallographic naylsis of the oligomeric structure of escherichia coli GMP kinase. *J. Mol. Biol.* 352:1044-1059.
18. Hible, G., P. Christova, …, and J. Cherfils. 2006. Unique GMP-binding site in mycobacterium tuberculosis guanosine monophosphate kinase. *Proteins* 62:489-500.

Table S1: Summary of all-atom molecular dynamics simulations.

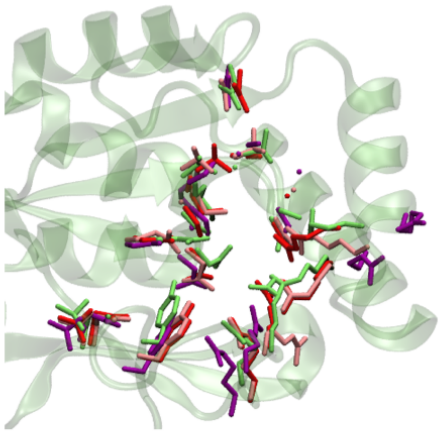
State	Initial structure	Substrates, products or co-factors	Conformer	Acronym	Simulation duration	Nb water	Nb ions	Cell size
1	1S4Q	None	Open	O_{APO}	20 ns	19427	108	93 x 83 x 78 Å
2	1ZNX	GMP	Semi-open	O_{GMP}	20 ns	19438	108	93 x 83 x 78 Å
3	1LVG-based homology model	GMP	Closed	C_{GMP}	20 ns	8103	44	71 x 67 x 57 Å
4	*same as state 3*	GMP : ATP : Mg^{2+}	Closed	$C_{GMP, ATP}$	20 ns	8070	45	71 x 67 x 57 Å
5	*same as state 3*	GDP : ADP : Mg^{2+}	Closed	$C_{GDP,ADP}$	20 ns	8124	46	71 x 67 x 57 Å
6	*same as state 3*	None	Closed	C_{APO}	20 ns	8119	44	71 x 67 x 57 Å

Table S2: RMS deviations (RMSD, in Å) calculated between models describing each step of the proposed enzymatic pathway (see Figure 2). The number of each state is given in parenthesis, as well as the simulation acronym.

	(1) O_{APO}	(3) C_{GMP}	(4) $C_{GMP, ATP}$	(5) $C_{GDP, ADP}$
(2) O_{GMP}	1.5	> 5	> 5	> 5
(4) $C_{GMP, ATP}$	> 5	1.4	-	1.4
(5) $C_{GDP, ADP}$	> 5	1.8	1.4	-
(6) C_{APO}	> 5	1.4	1.4	1.6

A

B



C

Backbone RMSD for catalytic residues

	$C_{GMP, ATP}$	1LVG	2QOR	2AN9
$C_{GMP, ATP}$	0	1.10	1.23	2.15
1LVG	1.10	0	1.03	2.07
2QOR	1.23	1.03	0	1.73
2AN9	2.15	2.07	1.73	0

Figure S1: (A) Backbone superimposition of the homology model (green) to the experimental structures of 1LVG (red), 2QOR (pink) and 2AN9 (purple). (B) three-dimensional fit and (C) backbone rmsd (in Å) for catalytic residues of these closed structures (see Figure 3 for definition and sequence alignment).

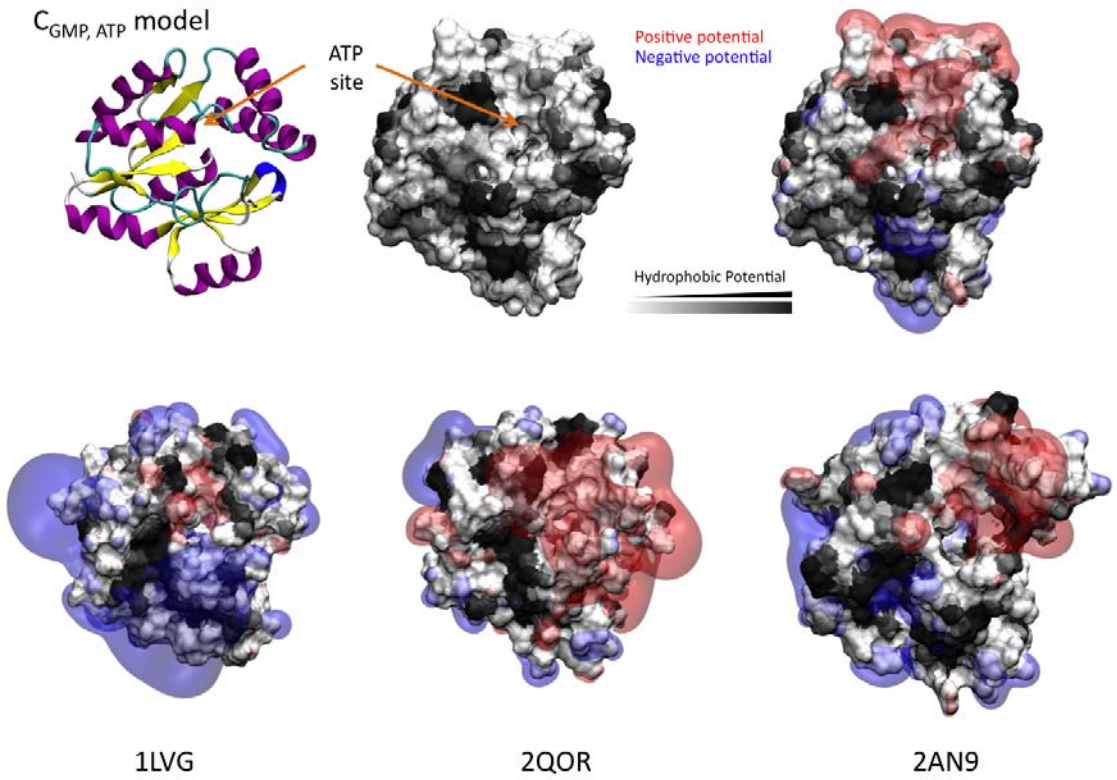


Figure S2: Surface properties of the ATP binding site in closed GK structures. Molecular Hydrophobic Potentials (MHPs) have been computed with the Platinum server (15) and electrostatics with the APBS program (16).

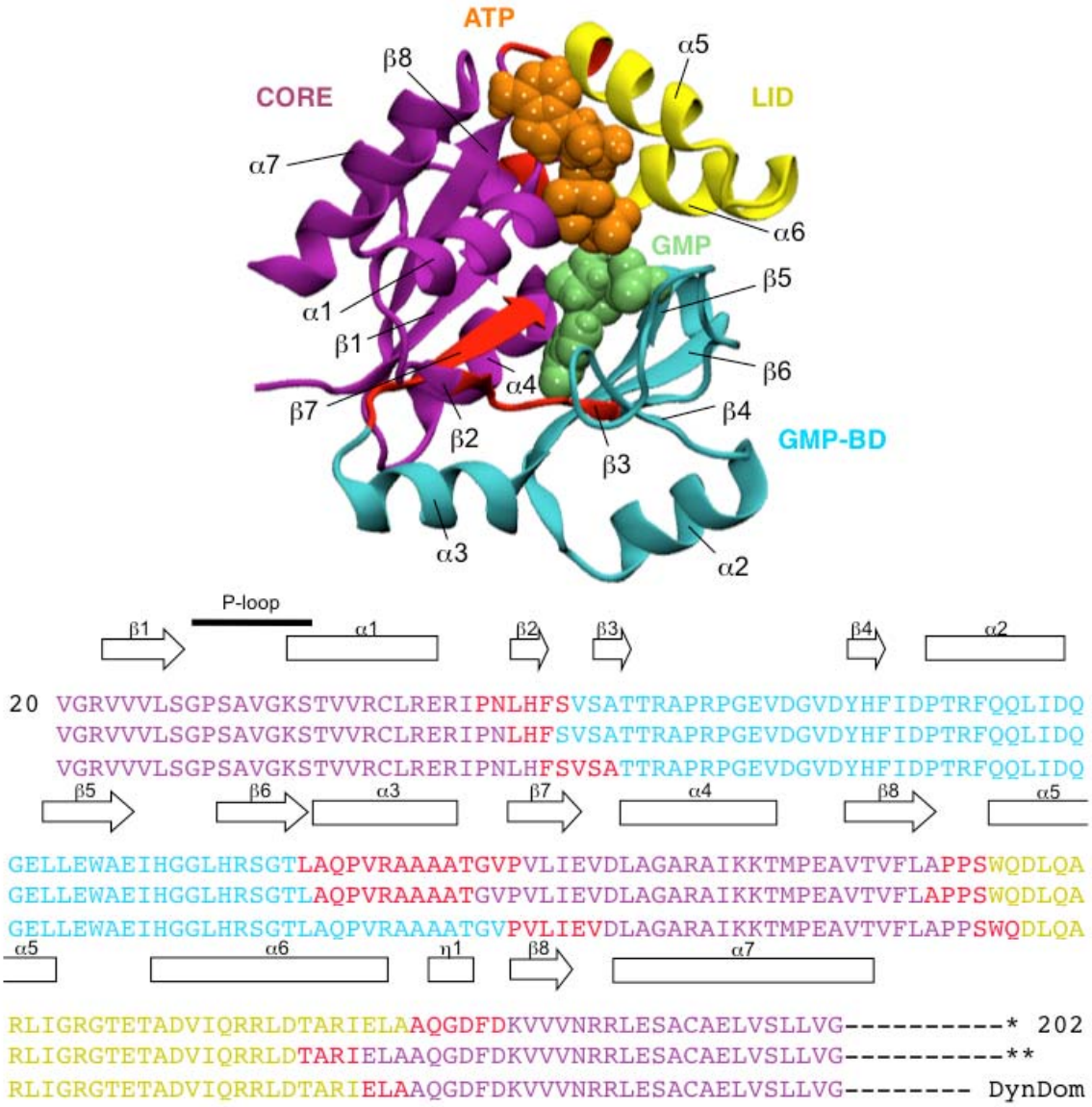


Figure S3: Extended Figure 1 from the main text. On top, secondary structure elements and the three structural domains of GKMT as defined by the DynDom server (http://fizz.cmp.uea.ac.uk/dyndom/dyndomMain.do) are reported on a cartoon representation with hinge regions in red and each domain in a distinct colour: CORE (purple), GMP-BD (cyan) and LID (yellow). Below, two domain definitions from previous experimental studies (*17, **18) are reported on the GK_{MT} primary sequence. The DynDom domain definition obtained from analyzing open and closed GK_{MT} models is shown on the third line.

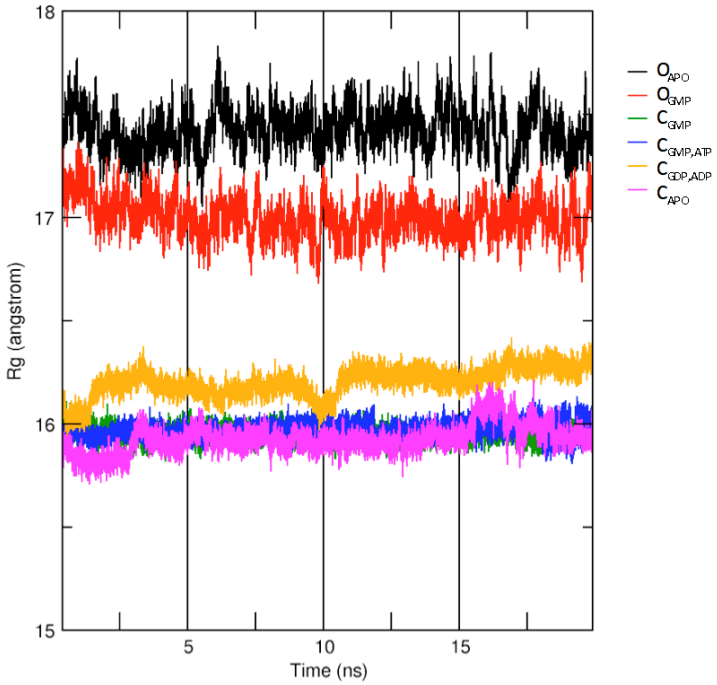


Figure S4: Radius of gyration (in Å) for the six states describing the enzymatic pathway all along the MD trajectory shown in Figure 2.

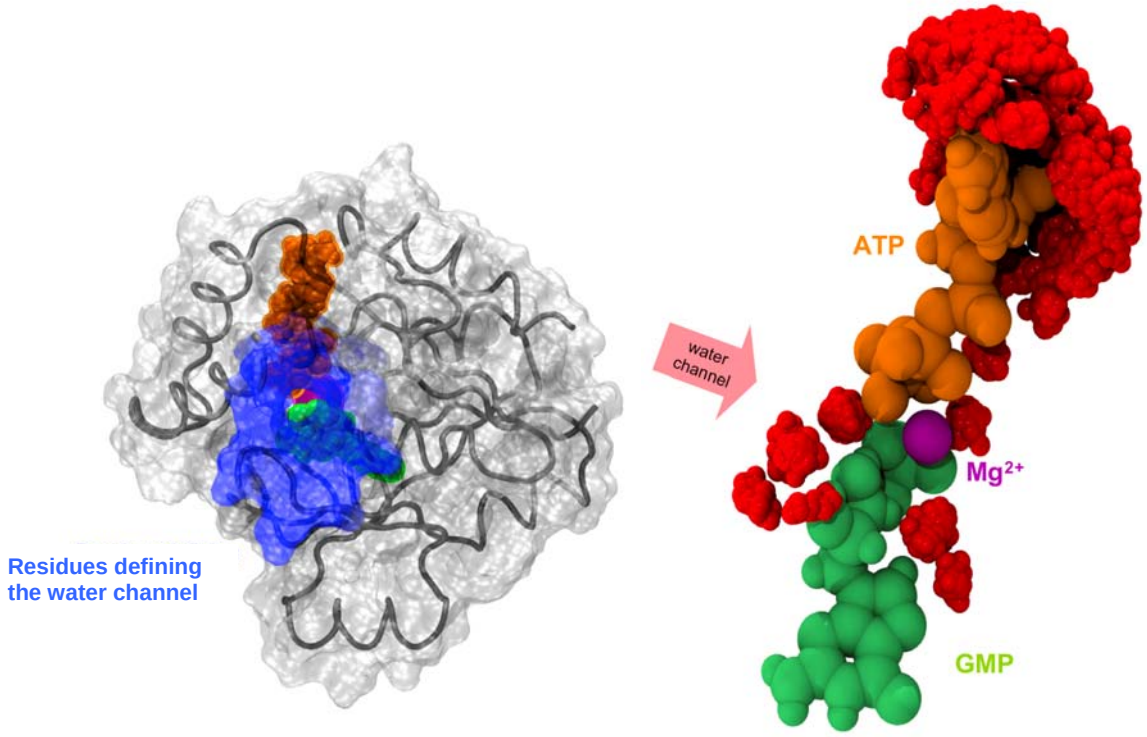


Figure S5: On the left, back view of the substrate binding pocket highlighting GK residues (blue transparent mesh) that define walls for a water channel. The first ring, closest to the substrates, is composed of G28, P29, S30, K34, I92, H93, D121, L122, A170 and E173. The second ring comprises the A90, E91, G94, A123, T169 and I172 residues. The backbone is shown as grey tube and substrates are represented in spacefilling representation (GMP in green, ATP in orange and the magnesium ion in purple). On the right, active site hydration observed during the last 10ns of the $C_{GMP,\ ATP}$ simulation. GMP, ATP and the magnesium ion are shown in green, orange and purple, respectively. Oxygen atoms of water molecules closer than 2.4Å with respect to the substrates are shown as red spheres (cluster built on the overall trajectory). ATP is clearly more solvent accessible than the GMP substrate. The water molecules localized inside the protein are driven through a hypothetical water channel formed at the interface of the three structural domains GMP-BD, LID and CORE. Condensed water oxygen clusters indicate stabilization of water molecules in the surroundings of bridging phosphates (d<4.5Å), in agreement with two GK structures available in the protein databank (PDB-ids 1LVG and 1EX7). Some of these clusters might be involved in the phosphate transfer reaction.

OPEN GK•GMP

CLOSED GK•GMP

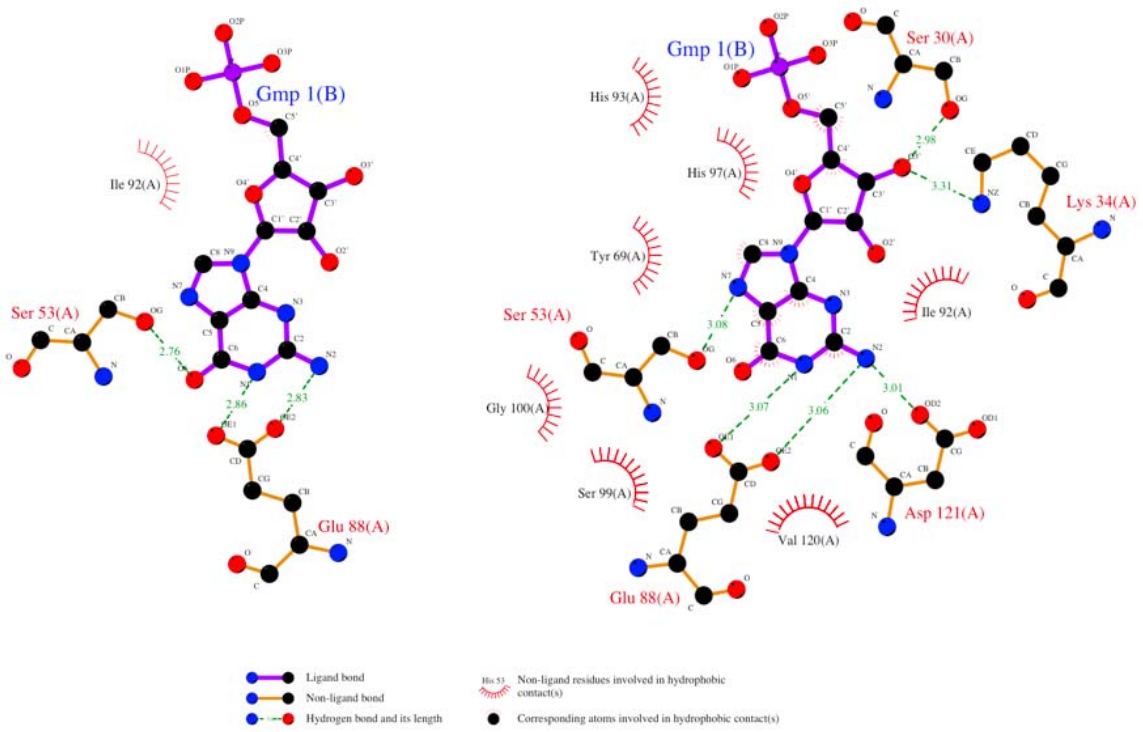


Figure S6A: Residues of the GK_{MT} active site involved in hydrogen-bonding or steric contacts with the GMP substrate for the O_{GMP} model (left) and the C_{GMP} model (right). The conformations were derived from average structures over the trajectories.

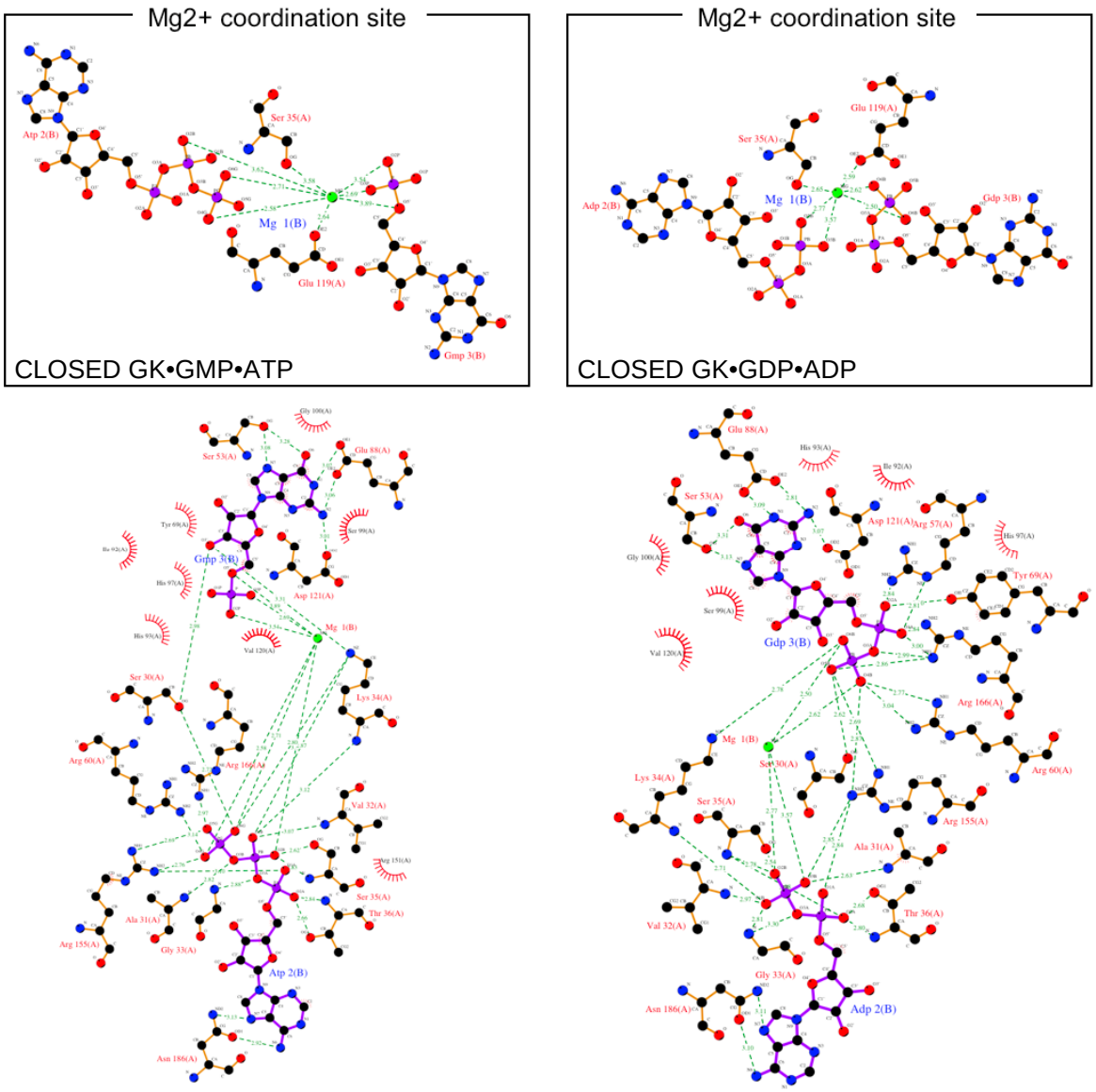


Figure S6B: Residues of the GK_{MT} active site involved in H-bonding or steric contacts with substrates or products, respectively, for the $C_{GMP,\ ATP}$ model (left) and the $C_{GDP,\ ADP}$ model (right). Conformations were generated from average structures over 20ns trajectories. Coordinating residues and substrate groups for the single Mg^{2+} ion are presented in inlays on top.

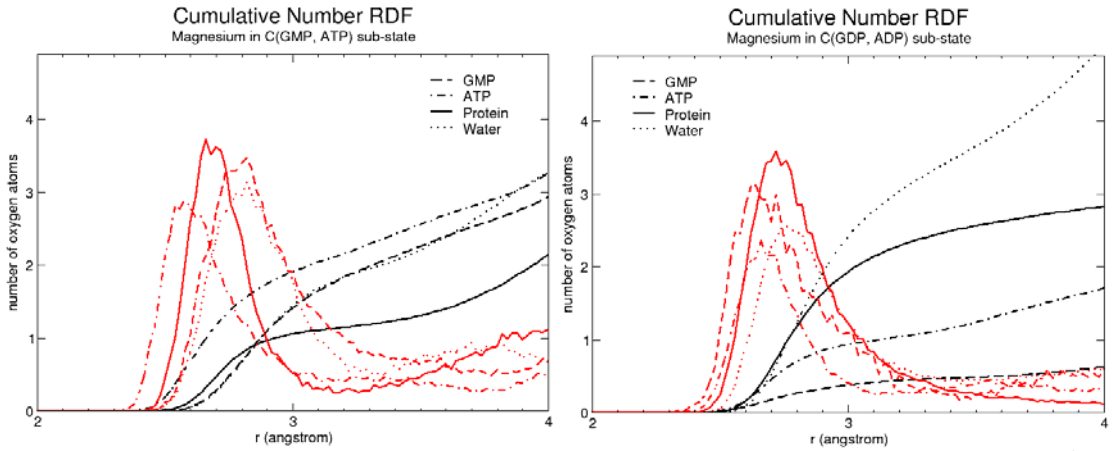


Figure S7: Radial distribution function (in red) of oxygen atoms around the Mg^{2+} ion in both sub-states $C_{GMP,\ ATP}$ (left panel) and $C_{GDP,\ ADP}$ (right panel). Black lines correspond to the cumulative number of oxygen atoms belonging to GMP/GDP (dashed lines), ATP/ADP (dot-dash lines), protein (solid lines) and water (dotted lines) molecules for a given distance from magnesium (x-axis in Å).

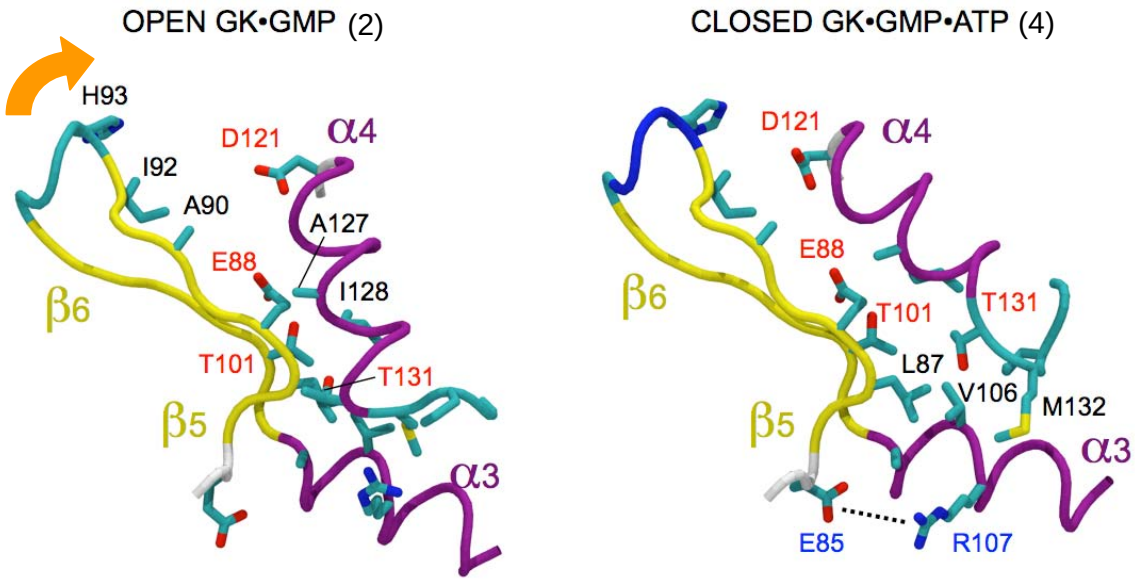


Figure S8: Comparison of the CORE/GMP-BD interface in the open (left, state 2) and closed (right, state 4) forms of GK_{MT}. In the O_{GMP} model (left), domain contact is stabilized between the N-terminal part of the α3 helix and C-terminal α4 helix by a large hydrophobic patch (L87, E88, A103, V106, R107, A127, I128, T131, M132). The interface is newly defined between the β5 strand and the N-terminal part of the α4 helix in the $C_{GMP,\ ATP}$ model (right) after an important modification of this region. The E88-A127 hydrophobic interaction is maintained in both structures whereas the very specific E88-T101 H-bonding contact disappears after closure of the GK enzyme. A new salt bridge contact is created between the E85 and R107 residues in the $C_{GMP,\ ATP}$ form.

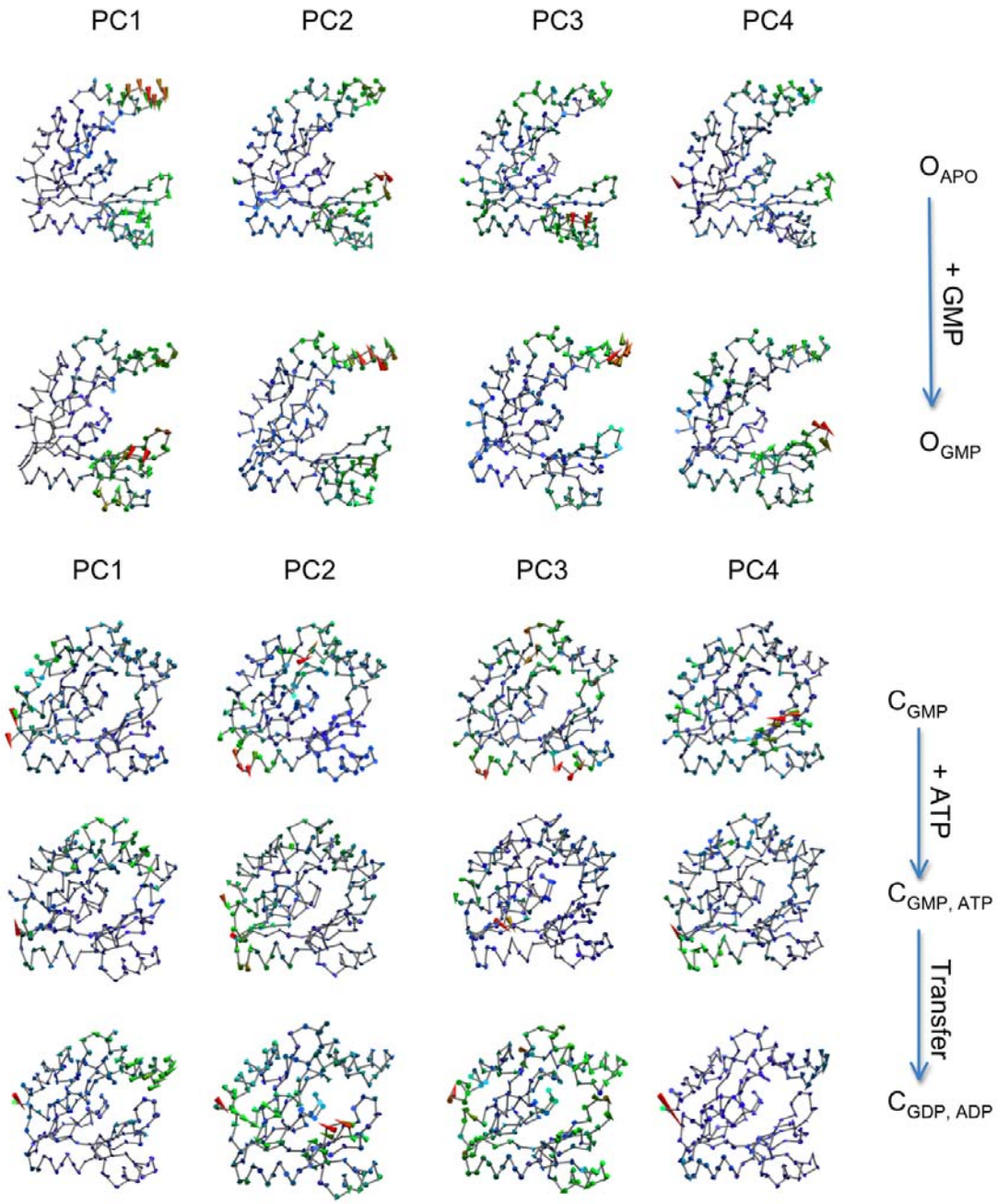


Figure S9: Principal components analysis for all trajectories of the five sub-states strictly describing the enzymatic pathway (sub-states 1 to 5 in Figure 2).

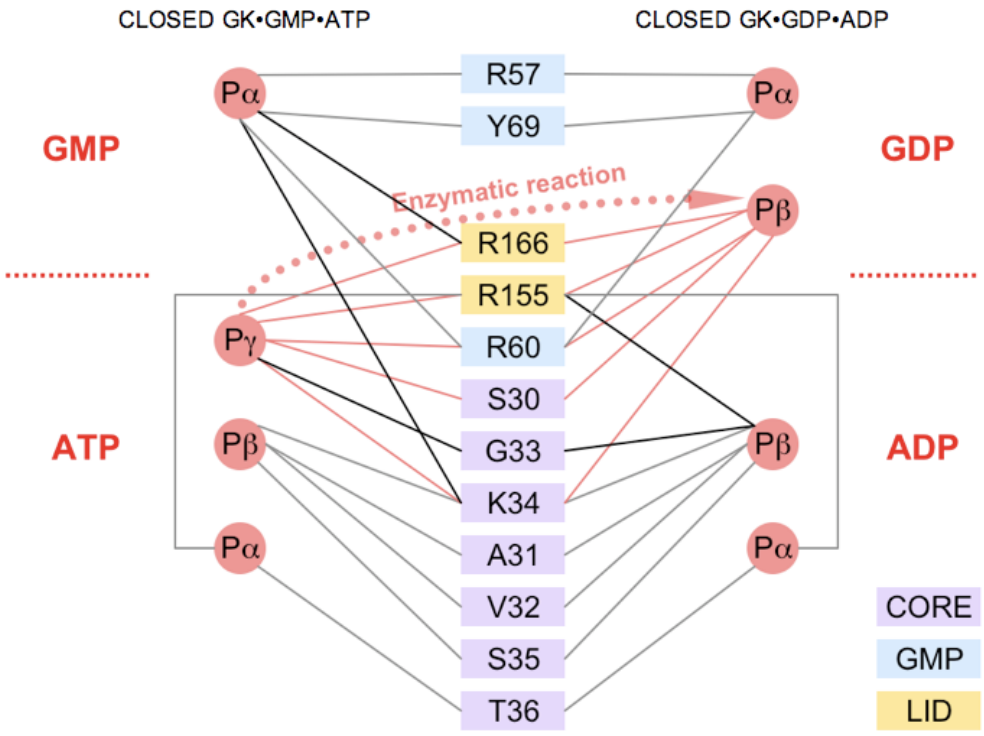


Figure S10: GK_{MT} residues interacting with phosphate groups in the $C_{GMP,\ ATP}$ sub-state model (left) and in the $C_{GDP,\ ADP}$ sub-state model (right). Red bonds concern the phosphate group transferred from ATP to GDP during the enzymatic reaction. Grey bonds do not change between the two states; bonding variations are plotted as black lines.