Phosphorylation Mechanism of N-acetyl-L-Glutamate Kinase, a QM/MM Study

*James McClory†, Gui-Xiang Hu,§ Jian-Wei Zou§, David J. Timson ‡, Meilan Huang†**

† School of Chemistry and Chemical Engineering, Queen's University Belfast, David Keir Building, Stranmillis Road, Belfast, BT9 5AG, Northern Ireland, United Kingdom

§ School of Biotechnology and Chemical Engineering, Ningbo Institute of Technology, Zhejiang University, 1 Xuefu Rd, Yinzhou Dist. Ningbo 315100, China

‡ School of Pharmacy and Biomolecular Sciences, The University of Brighton, Huxley Building, Lewes Road, Brighton, BN2 4GJ, United Kingdom

*Corresponding Author:

m.huang@qub.ac.uk

ABSTRACT

In microorganisms and plants, NAGK N-acetyl-L-glutamate kinase (NAGK) catalyses the second step in L-arginine synthesis, the phosphorylation of *N*-Acetyl-L-glutamate (NAG) to give *N*-acetyl-L-glutamate-5-phosphate (NAGP). NAGK is only present in microorganisms and plants but absent from mammals, which makes it an attractive target for antimicrobial or biocidal development. Understanding the substrate binding mode and reaction mechanism of NAGK is crucial for targeting the kinase to develop potential therapies. Here the substrate binding mode was studied by comparing the conformational change of NAGK in the presence and in the absence of the NAG substrate based on molecular dynamic simulations. We revealed that with substrate binding the catalytic site of the kinase involving three loops in NAGK exhibits a closed conformation, which is predominantly controlled by an interaction between Arg98 and the α-COO^- of NAG. Lys41 is found to guide phosphate transfer through the interactions with the β-,γ-, and γ- phosphate oxygen atoms of ATP surrounded by two highly conserved glycine residues (Gly44 and Gly76), while Arg98 helps to position the NAG substrate in the catalytic site, which facilitate the phosphate transfer. Furthermore, we elucidated phosphate transfer reaction mechanism using hybrid density functional theory-based QM/MM calculations (B97D/AMBER99) and found that the catalysis follows a dissociative mechanism.

1. INTRODUCTION

Arginine is one of the most metabolically versatile amino acid for the biosynthesis of proteins and serves as the precursor for the synthesis of a range of products such as urea, nitric oxide (NO), creatine, polyamine and agmatine.[1] L-Arginine/NO pathway mediates a large number of physiological and pathological processes such as neuronal signalling, immune response, inflammatory response and cardiovascular homeostasis.[2] Because of its NO-stimulating effects, L-Arginine has a variety of therapeutic applications [3] such as treatment of angina,[4] hypertension,[5,6] coronary heart disease,[7] preeclampsia,[8] and diabetes,[9] as well as improving cognitive function in senile patients with dementia.[10]

Arginine biosynthesis and metabolism have gained increasing attention over the past sixty years.[11-13] Understanding of arginine biosynthesis and metabolism would provide a novel approach for development of potential therapies. Plants and microorganisms produce arginine from glutamate via n-acetylated intermediates.[11,14,15] Among the enzymes that are involved in arginine biosynthesis, *N*-acetyl-L-glutamate kinase (NAGK), which catalyses the second step and usually rate-determining step in arginine synthesis, is of particular interest, because it is only present in bacteria and plants but absent in mammals and controls the nitrogen storage as arginine by signalling protein PII.[16,17]

NAGK is a representative kinase in amino acid kinase (AAK) family that can be further classified as phosphate sub-family of AAK and carboxylate sub-family of AAK. The "carboxylate" subdivision consisting of NAGK, carbamate kinase (CK), glutamate-5-kinase (G5K) and aspartokinase I (AK) catalyze the phosphate transfer from ATP to a carboxylate or carbamate group, while the "phosphate" subdivision comprising fosfomycin resistance kinase (FomA), isopentenyl phosphate kinase (IPK), UMP kinase (UMPK) catalyze phosphate transfer from ATP to a phosphate or phosphonate group. Two classes of

NAGK have been identified during the evolution of NAGKs, namely arginine-sensitive and arginine-insensitive/inhibitable enzymes. The only arginine-insensitive NAGKs with available crystal structure is *E. coli* NAGK, while a number of arginine-sensitive NAGKs have been structurally well characterised such as *Arabidopsis thaliana* (At-NAGK).

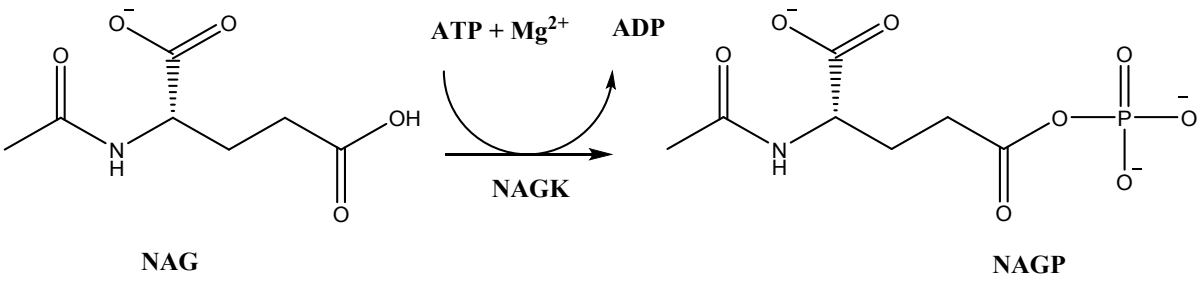


Figure 1. Phosphorylation of *N*-acetyl-glutamate to *N*-acetyl-glutamate-5-phosphate catalysed by *N*-acetyl-glutamate kinase in the presence of ATP and Mg^{2+}.

NAGK catalyzes the phosphorylation of N-Acetyl-L-glutamate (NAG) to generate the phosphorylated product *N*-acetyl-L-glutamate-5-phosphate (NAGP) (Figure 1). Intensive experimental and theoretical studies have been reported on the structures and catalytic mechanism for arginine-insensitive *Escherichia coli* NAGK, [18, 19] however, so far, no theoretical studies have been reported on the substrate binding and phosphoryl transfer mechanism of the arginine-sensitive plant NAGK. *Arabidopsis thaliana* has been a powerful tool for the study of the subdiscipline of plant pathology. Here we present the first theoretical study on At-NAGK in complex with ATP, NAG and Mg^{2+}. Using molecular dynamic (MD) simulations, we find that an interesting water network mediated by Asp216 helps to position the ATP molecule. Further, we identified the opening and close conformations of the protein structure during substrate binding, which is attributed to significant conformational changes around a couple of loops. Furthermore, a QM/MM study showed that the phosphate transfer catalyzed by At-NAGK proceeds through a dissociative mechanism, unlike *E.Coli* NAGK, which was proposed from experimental and theoretical studies to catalyze phosphoryl

transfer via the associative mechanism. [19,20] These findings would lay the basis for the rational metabolic engineering NAGK, the key enzyme in the arginine biosynthesis pathway and expanding its industrial applications for the synthesis of L-arginine.

2. MATERIALS AND METHODS

2.1 System Set-up

The crystal structure of *Arabidopsis thaliana* NAGK [21] (Chain A) in complex with ADP, Mg^{2+} and native substrate NAG (PDB Code: 4USJ) was used as the starting structure. The location of ATP in NAGK was obtained by superimposing the starting NAGK structure with the ATP-Mg^{2+}-*Arabidopsis thaliana* NAGK (PDB: 2RD5) (using Accelry's Discovery Studio. [22] The protonation states of titratable residues were determined using the ProPka along with visual inspection. [23] All mutants were generated in Accelrys Discovery Studio.

2.2 Molecular Dynamic Simulations

The parameters for ATP were attained from AMBER parameter database at the Bryce's group [24] and RESP charges for the substrate were obtained using single point energy calculation with HF/6-31 G(d) method implemented in Gaussian09 package (Table S1). [25]

All MD simulations were conducted using the GPU version of PMEMD engine in AMBER 14. [26] The FF14SB force field was utilized in the simulations and the Leap module implemented in AmberTools was used to add the missing hydrogen atoms and counter ions to neutralize the charge of the protein system. The system was soaked in an octahedral TIP3P water box and the periodic boundary conditions (PBC) were utilized. The particle mesh Ewald (PME) method [27] with a cut off distance of 10Å was used to calculate the long-range electrostatic interactions.

The system was first subjected to 1,250 steps of steepest descent and 1,250 steps of the conjugate gradient minimization. In the initial minimization all solute molecules were restrained using a force constant (50 kcal $mol^{-1}Å^{-2}$). The solvent water molecules and counter ions were allowed to move freely during the minimization. The system was then subjected to 1,250 steps of steepest descent and 1,250 steps of the conjugate gradient minimization.

The system was subsequently heated slowly from 0K to 300K in a duration of 50ps at a constant pressure using Langevin dynamics with a collision frequency of $1ps^{-1}$ utilizing the NVT canonical ensemble. The solute molecules were restrained using a force constant of 50 kcal $mol^{-1}A^{-2}$. The system was then equilibrated at 300K with the NPT ensemble and was allowed to relax freely at a constant pressure of 1atm. MD production simulations were run at 300K in a NPT ensemble for 100ns with a time step of 2fs. For the complex in absence of the NAG substrate, four replicas of 200-ns MD simulation were run. SHAKE constraints [28] were applied to all bonds involving hydrogen atoms. For each system, a total of four replica MD simulations were conducted with randomly assigned initial velocities.

The MD trajectories were analysed using CPPTRAJ and VMD. Root mean square deviation (RMSD) of the Cα atoms relative to the initially minimized crystal structure was calculated. Cluster analysis was performed to obtain the representative structures of the MD simulation trajectory.

2.3 QM/MM optimizations

The representative structures obtained from the MD simulations were visually inspected and selected for QM/MM calculations. QM/MM calculations were conducted using the ONIOM method implemented in the Gaussian09 package. [25] The QM region was calculated using the B97D functional [29] and the 6-31+G(d) basis set, while the MM region was calculated using the Amber Parm99 force field. The protocol proposed by Tao et al was adopted where the

residues within 6 Å of the active site of the ATP-Mg^{2+}-NAGK-NAG complex were allowed to move freely while the rest of the system was kept fixed. [30] Hydrogen link atoms were used to cap the bonds crossing the QM/MM boundary. [31,32]

The QM region consisted of 41 heavy atoms including the triphosphoryl side chain and truncated methyl group of ATP, the substrate *N*-acetyl-L-glutamate, Mg^{2+} and three coordinating water and two structural water molecules that hold Asp216 and ATP phosphate together. Lys41, Asp196 and Lys255 were truncated with the side chains kept, in order to comprise the demanding computation required by the QM calculations. The total charge of the QM region is -2.

The RESP charges of ATP and the substrate that were used in the molecular dynamic simulations were carried on in the QM/MM calculations. The selected representative cluster structures were subjected to an initial geometry optimization, serving as the starting point for a potential energy scan (PES). The reaction coordinate was defined as the distance between the γ-phosphate phosphorous atom of ATP and the carboxylate oxygen atom of NAG, which was scanned with a decrement size of 0.1 Å. The transition state structure obtained from the PES was optimized and validated by frequency calculations. The energies of the stationary points were corrected using the B97D/6-311++G (d,p) level of theory.

3. RESULTS AND DISCUSSION

3.1 Conformational change during NAG binding

Conformational change of proteins is often observed during substrate binding. [33] Kinases usually exhibit large plasticity that has emerged as the target for development of selective drug therapies. [34,35] Conformational dynamics is essential for the function of NAGKs. [36] For *E Coli* NAGK, the open conformation was suggested to be associated with the substrate entry and product release while the closed conformation is related to catalysis. [18] Here the

conformational change of At-NAGK induced by substrate binding was studied by comparing the MD simulated structures of NAGK bound with the native substrate NAG and in absence of the substrate. RMSD analysis of the Cα atoms was performed using the minimized crystal structure as the reference. The RMSD showed the ATP-Mg^{2+}-NAGK-NAG complex displays little deviation from the crystal structure, whereas that the ATP-Mg^{2+}-NAGK complex exhibited a larger deviation (Figure S1), indicating the conformational change induced by substrate binding.

The flexibility of the residues of the NAGK complexes was examined using RMSF analysis for the equilibrated trajectories. With time evolution, the residues 44-51 demonstrated rather large flexibility in the ATP-Mg^{2+}-NAGK complex compared to the ATP-Mg^{2+}-NAGK-NAG complex (Figure S3). Cluster analysis was performed based on the last 10ns of the equilibrated trajectories in order to obtain the representative structures of the MD simulations. In both the ATP-Mg^{2+}-NAGK-NAG and ATP-Mg^{2+}-NAGK complexes, Lys41 on the glycine-rich loop forms ionic interactions with the β- and γ-phosphates of ATP, maintaining the nucleotide tail of the nucleotide in a suitable pose of for phosphoryl transfer to occur.

Visual inspection of the representative structure of the substrate bound complex revealed notable changes around the loop 44-51 (Hereby denoted as Loop 1) compared to the NAGK in absence of the substrate (Figure 2c & 2d). In the ATP-Mg^{2+}-NAGK complex, the Loop 1 undergoes significant conformational change, moving far away from the nucleotide such that the hydrogen bond between the phosphate of ATP and Gly44 on the glycine-rich sequence observed in the substrate-bound ATP-Mg^{2+}-NAGK-NAG complex is lost (Figure 2d). Notably, Gly44 is located on the glycine-rich sequence IIKxGG, the signature of NAGK. This highly conserved glycine-rich sequence was also identified to play a crucial role in

catalysis of IPK, a homologous AAK kinase whereby highly conserved glycine residues (Gly7 and Gly8) form hydrogen bonds with ATP. [37]

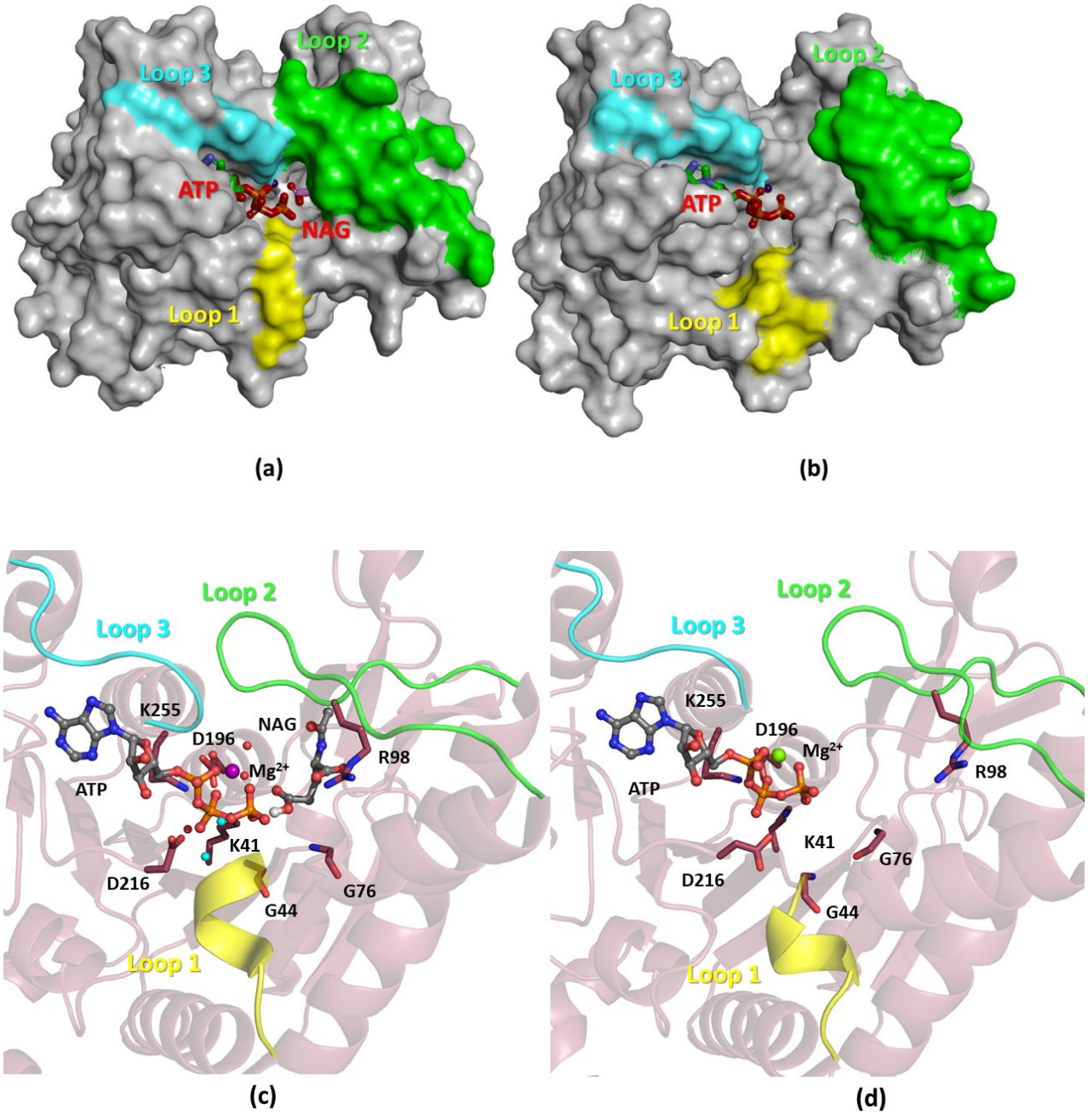


Figure 2. MD simulations of ATP-NAG-Mg^{2+}-NAGK and ATP-Mg^{2+}-NAGK complexes (a) Surface representation of closed conformation of NAGK in presence of NAG (b) Surface representation of open conformation of NAGK in absence of NAG (c) Secondary structure of

closed conformation of NAGK in presence of NAG (d) Secondary Structure of open conformation of NAGK in absence of NAG.

In addition, the RMSF plot showed that the region between Leu87 to Ala102 in the NAG bound complex displays a substantial increase in flexibility in comparison with the substrate bound complex (Figure S3). Comparison of the ATP-Mg^{2+}-NAGK-NAG complex and ATP-Mg^{2+}-NAGK complex reveals a significant conformational change around a loop on the surface of the protein between residues Leu87 to Ala102 (denoted as Loop 2 here, corresponding to the β3-β4-hairpin in *E Coli* NAGK) which adjoins two alpha helices αB and αC (Figure 2c & 2d). A third loop between residues Lys247 and Gly251 (denoted as Loop3) connecting αF and αG was also found to undergo a notable conformational change based on RMSF analysis and from visual inspection of the two superimposed complexes (Figure S3, Figure 2c & 2d). In the substrate bound complex, Loop2 sits directly above the both ATP and NAG while Loop 3 is above ATP. Such a conformation between the two loops locks the substrate into the active site in a 'closed' conformation (Figure 2a & 2c). In contrast, in the ATP-Mg^{2+}-NAGK complex in absence of the NAG substrate, the Loop2 and Loop3 fully retract apart creating a large opening to the active site (Figure 2b & 2d).

Arg98 on Loop2 that undergoes the most significant change acts as the anchoring point, forming an ionic interaction with the NAG α-COO^- and pulling the loop inward toward the active site. In contrast, in the ATP-Mg^{2+}-NAGK complex Loop2 falls away from the binding pocket due to the absence of the interaction between Arg98 and the substrate (Figure 2b & 2d). Thus in the absence of a bound substrate the enzyme is in an 'open' conformation ready to accommodate the entry of substrate and also in a conformation to allow for the product to be released. This is in accordance with *E Coli* NAGK in absence of substrate and one subunit of NAGK (the subunit A) in complex with the product, both of which exhibit a very open conformation, favourable for the substrate to enter and for the product to release. [18] Notably

MD simulations of AAK enzymes IPK [37] and FomA [38] have previously been used to elucidate their opening and closing mechanisms. In FomA two loops mediate the substrate induced conformational changes, one of the loops in FomA (Leu201-Gly211) [38] is analogous with Loop3 in NAGK, underlining the importance of this loop region for substrate binding in AAKs.

In the crystal structure of NAGK, Asp216 forms a hydrogen bond with the C3-OH of the nucleotide ribose. Notably, with time evolution of the MD simulations, a water network developed around Asp216, in addition to the H-bond interaction with the ribose hydroxyl (Figure 2c). Two structural water molecules are H-hydrogen-bonded to the α-phosphate and β,γ-bridging oxygen of ATP. Asp216 is highly conserved among the AAKs (Figure 3), indicating it may play a pivotal role in positioning the highly mobile phosphate tail of the nucleotide via the water-mediated network among the AAK kinase family. Interestingly, in the ATP-Mg^{2+}-NAGK complex the water network remains intact (Figure 2d), indicating that the presence of this network is independent of a bound substrate.

It is well known that water networks in kinases may play a prominent feature in catalysis. [39] For example, Aurora kinase A [40] and Bosutinib [41] have been reported to incorporate water networks in their active sites. In the phosphate sub-division of the AAK enzymes represented by FomA the active site is characterised by a catalytic lysine triad and a conserved histidine residue which interact with ATP molecule. [42] Unlike phosphate sub-division of AAKs that are characterized with a lysine triangle and histidine residue, in the carboxylate sub-division of the AAKs, the active site contains only two lysine residues (Lys41 and Lys255 in At-NAGK) and no histidine residue. [43] His58, which was suggested to position the fosfomycin substrate in phosphate sub-division of AAKs such as FomA, [42] was replaced by Asn81 in At-NAGK. Asn81 forms a H-bond with Arg98 on Loop 2, which in turn stabilizes the carboxylate group of the NAG substrate.

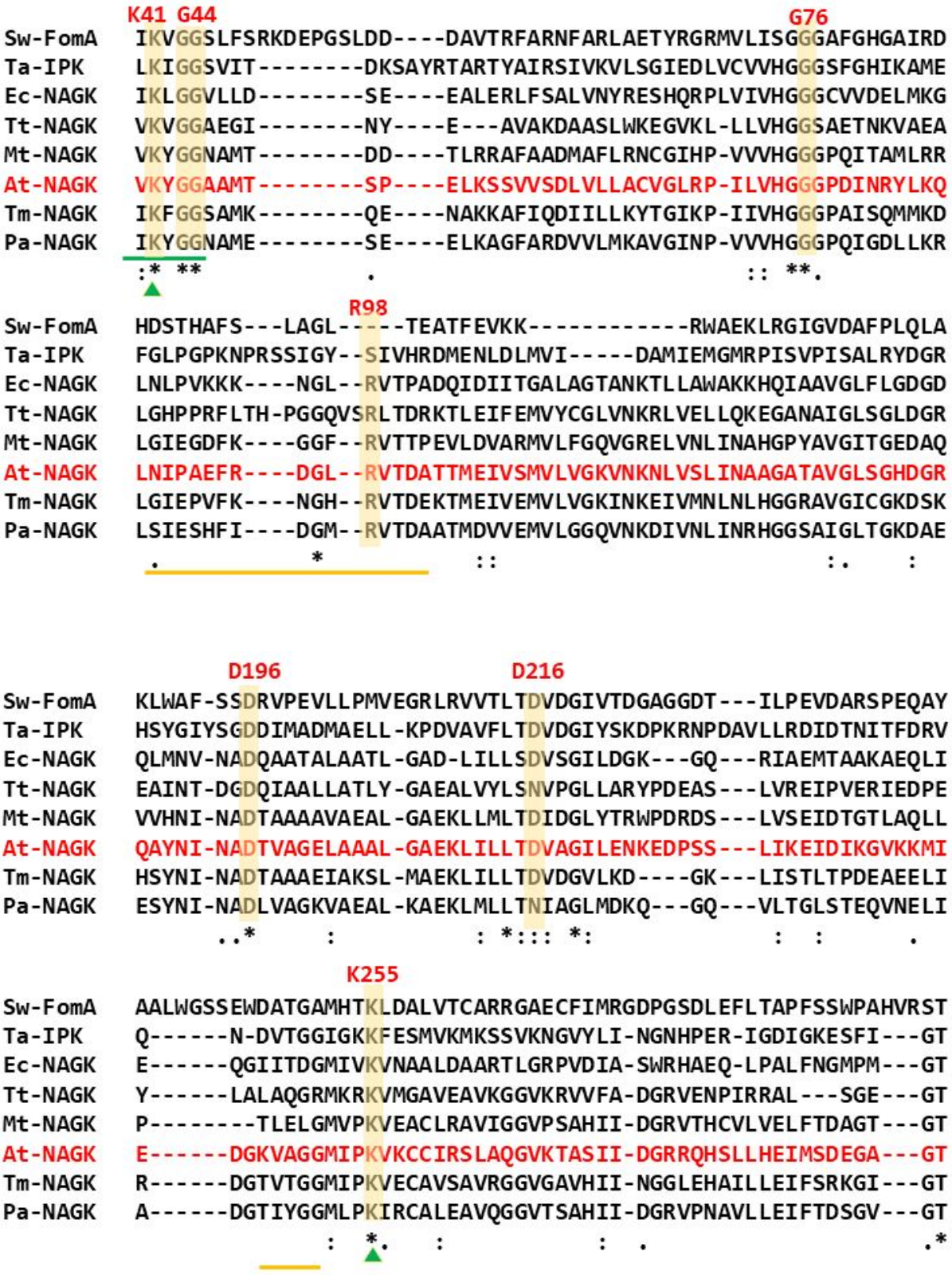


Figure 3. Sequence alignment of AAK enzymes NAGK, FomA and IPK using the CLUSTAL multiple sequence alignment by MUSCLE web server. Key residues among the kinases are

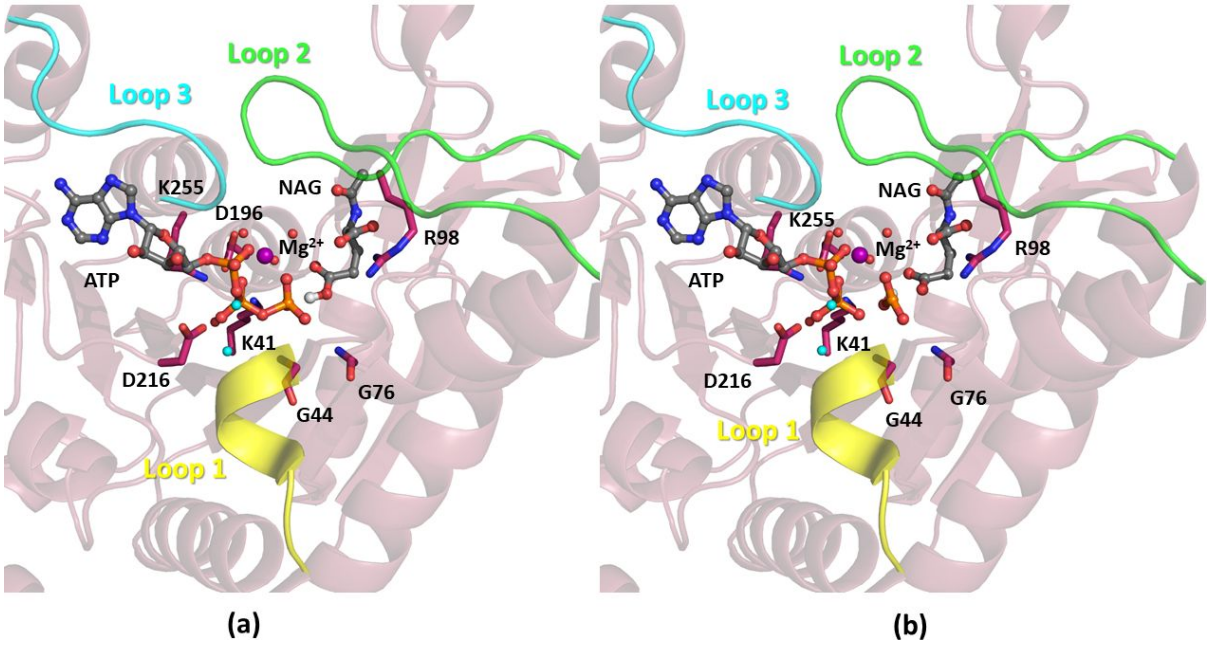
highlighted in yellow and labeled in red in reference to *Arabidopsis thaliana* NAGK. Species names are abbreviated as follows: Sw: *Streptomyces wedmorensis* (Sw-FomA, BAA32493.1); Ta: *Thermoplasma acidophilum* (Ta-IPK, CAC11251.1); Ec: Escherichia coli (Ec-NAGK, AAA23478.1); Tt, *Thermus thermophilus* (Tt-NAGK, Q5SH27) Tm: *Thermotoga maritima* (Tm-NAGK, Q9X2A4); Pa: *Pseudomonas aeruginosa* (Pa-NAGK, Q9HTN2); At: *Arabidopsis thaliana* (At-NAGK, OAP01684.1), Mt: *Mycobacterium tuberculosis* (Mt-NAGK, P9WQ01). The signature glycine-rich loop IIKxGG is underlined by green line, where residue isoleucine (I) can be replaced by Valine (V). The Loop 2 and Loop 3 are underlined by orange lines. The two highly conserved lysine residues Lys41 and Lys255 are highlighted by green arrows.

3.2 Mechanism of phosphate transfer in NAGK by QM/MM

In order to reveal how phosphoryl transfer catalyzed by NAGK, QM/MM calculations were performed. MD simulation studies and kinetic studies provide guidance on the important residues to be included in the QM region. The highly conserved residues Lys41 and Lys255 along with Asp196 that forms ionic bonds with the two positively charged residues were included in the QM layer. In addition, ATP, NAG, Mg^{2+} and three coordinating waters, two structural waters that form Hbond with the α-, β-phosphate oxygen atoms of ATP were contained in the QM region.

Based on the QM/MM study, it was found that an in-line phosphate transfer from ATP to the carboxylate group of NAG occurs via a S_N2-like concerted mechanism. The water mediated H-bond network identified from molecular dynamics remains in the QM/MM optimized stationary points of the reactant, TS and product (Figure 4a-c), where Asp216 forms an H-bond with the two structural water molecules which subsequently form hydrogen bonds with the α- and β- phosphate oxygen atoms of ATP, helping to position the nucleotide throughout

the reaction process. The highly conserved Lys41 (Figure 5) plays a vital role in the reaction mechanism by forming ionic interactions with β-, and γ- phosphate oxygen atoms of ATP in the reactants and TS. Evidently, Lys41 helps to align the nucleotide to facilitate in-line phosphoryl transfer. Furthermore the negative charge which accumulates at the β,γ-bridging oxygen of ATP in the phosphorylation process is alleviated through the interaction with Lys41. Interestingly, in the product, the ionic interaction with the transferred phosphate group Lys41 was replaced by Arg98 on Loop 2. As observed in the MD simulations, the conserved Gly44 on the glycine-rich loop helps substrate binding by forming an H-bond with both the β,γ-bridging, and γ- phosphate oxygen atoms of ATP. Notably a recent QM/MM study on the reaction mechanism of IPK also revealed two conserved glycine residues on a glycine-rich loop played a similar role in catalysis. [37] In addition, another conserved glycine residue among AAKs Gly76 (Figure 3) forms an H-bond with the γ- phosphate of ATP throughout the catalytic process. Thus these glycine residues play a critical role in directing the terminal phosphate of ATP throughout the reaction process.



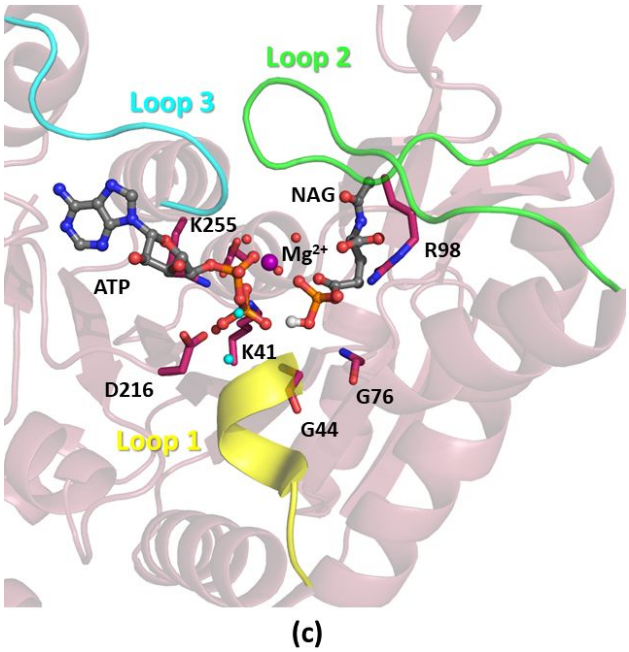
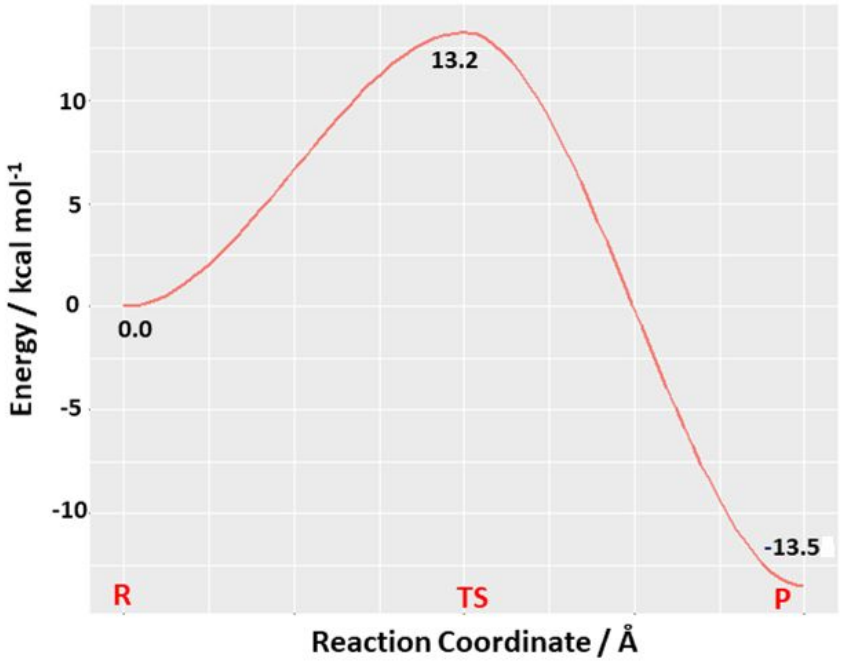


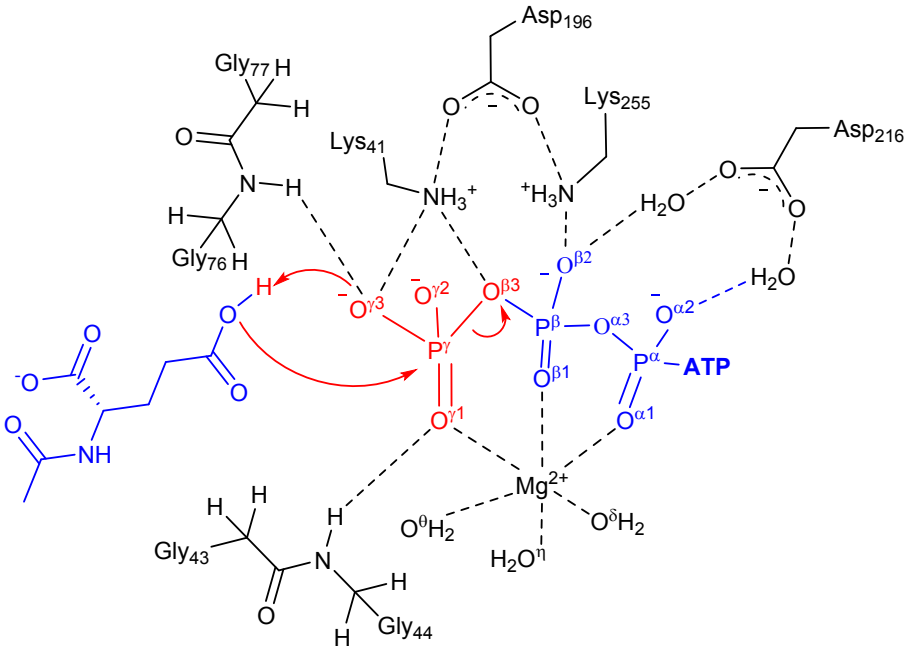
Figure 4. ATP-NAG-Mg^{2+}-NAGK complex (a) QM/MM optimised reactant structure (b) QM/MM optimised transition state structure (c) QM/MM optimised product structure. Loop 1, 2 and 3 are highlighted in yellow, green and cyan respectively

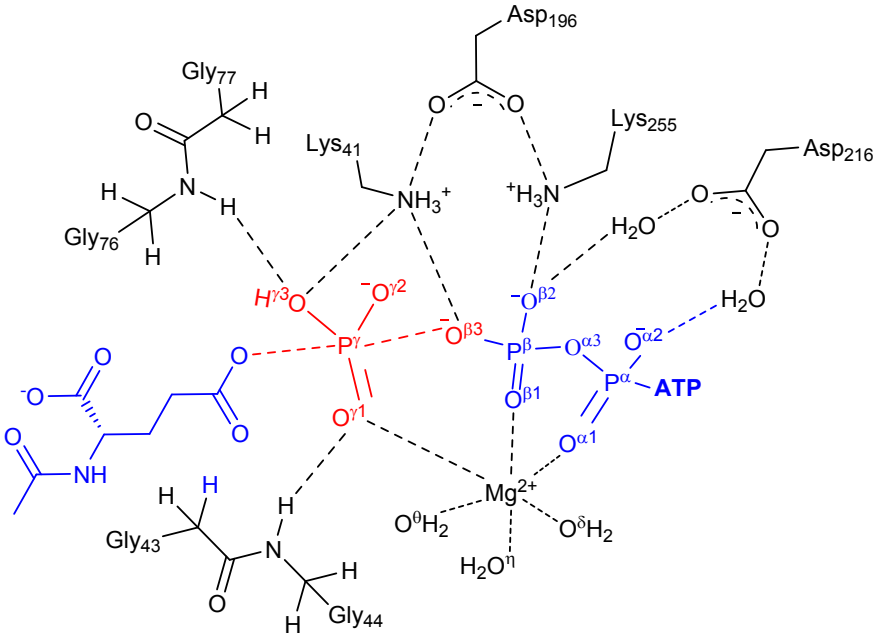
The associative or dissociative character of TS was measured based on Pauling's formula [44] $D(n) = D(1) - 0.6 \log n$, where $D(1)$ represents the average distance for a phosphate-oxygen bond, $D(n)$ is the average distance of the two phosphate-oxygen bonds in the TS and n represents the fractional bond number. Based on the optimized TS structure of At-NAGK from the QM/MM calculations (Figure 4b), the n is estimated as 0.17, indicating the phosphorylation has a dissociative character of 83%. This argues the previously proposed associative mechanism for *Escherichia coli* NAGK based on a structural study. [20] The dissociative mechanism disclosed for NAGK is in accordance with IPK and FomA in the other AAK sub-family, which was also found to catalyse the reaction via the transition state with a dissociative character. [37, 42]

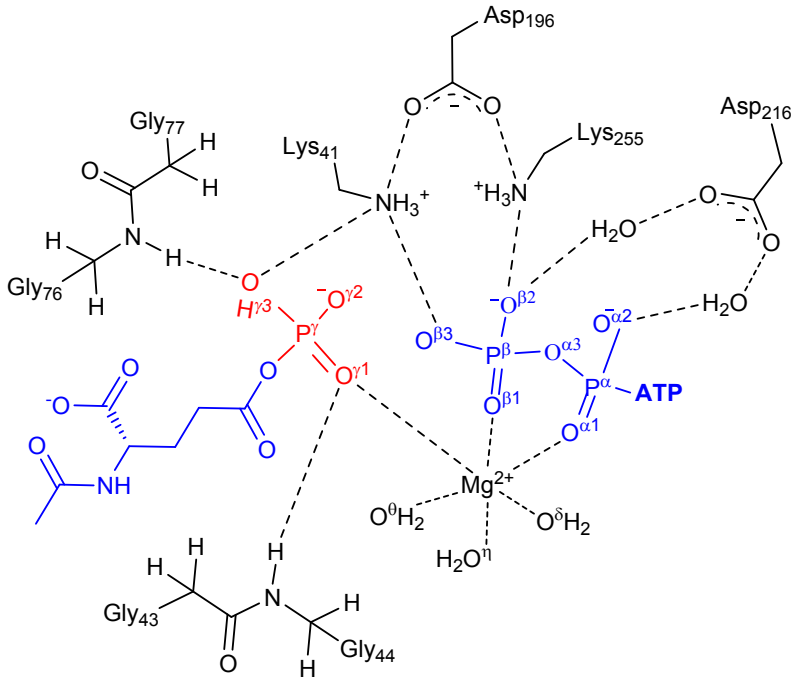
Based on the QM/MM calculations, the estimated barrier for phosphoryl transfer is 13.2 kcal mol^{-1}, in close agreement to the experimental barrier based on the previous kinetic studies on *Arabidopsis thaliana* NAGK, which revealed a catalytic turnover number of 115 s^{-1} corresponding to a reaction barrier of 14.7 kcal mol^{-1}. [45] The product *N*-acetyl-L-glutamate-5-phosphate had an energy of -13.5 kcal mol^{-1}, indicating an overall phosphorylation reaction catalyzed by At-NAGK is exothermic. Single point energy corrections were performed using B97D/ 6-311++G(d,p) giving a corrected reaction barrier of 14.0 kcal mol^{-1} and relative energy of the product being -14.1 kcal mol^{-1}. A unique imaginary frequency of -166.50 cm^{-1} was identified for the QM/MM optimized transition state corresponding to the phosphoryl transfer vibration.



(a)


Asp196
Gly77 H
Lys255
Asp216
Lys41
NH3+
+H3N
H2O
Gly76 H
Oβ2
Oγ3
Oγ2
Oβ3
Pβ
Oα3
Oα2
H2O
Pγ
Pα
ATP
Oβ1
Oγ1
Oα1
Mg2+
Oθ H2
Oδ H2
H2Oη
Gly43
Gly44
NH


Asp196
Gly77
Lys255
Asp216
Lys41
NH3+
+H3N
H2O
Gly76
Hγ3
Oγ2
Oβ2
Oβ3
Pβ
Oα3
Oα2
H2O
Pγ
Pα
ATP
Oβ1
Oγ1
Oα1
Mg2+
Oθ H2
Oδ H2
H2Oη
Gly43
Gly44
NH



(b)

Figure 5. (a) QM/MM energy profile for phosphate transfer in NAGK (b) 2-D sketch of phosphate transfer in NAGK. The reaction coordinates are defined as the distance between the carboxylate oxygen of NAG and the γ-phosphate phosphorous atom of ATP, and the distance between the P^{γ} atom of ATP and the β-,γ-bridging oxygen atom $O^{\beta 3}$. The red arrows denote the electron movement during the phosphorylation.

3.3 Validating the role of key residues in NAGK through MD simulations of the mutants

The key residues in NAGK identified by MD simulations were mutated. MD simulations were then performed to validate the roles of these residues in the binding and catalytic activity of the kinase.

K41A/R and K255A

MD simulations of the K41A mutation forfeited the strong ionic interaction with the phosphate tail of the nucleotide observed in the WT enzyme such that the γ-phosphate of

ATP moved away from the substrate ($d_{\gamma P\text{-}O}$ = 6.0 Å), resulting in a non-catalytic pose which would not facilitate phosphoryl transfer. (Figure 6a). In *Escherichia coli* NAGK, a K8R (equivalent to K41R) mutation resulted in diminished activity, [46] MD simulations of K41R showed that ATP and NAG remain in a suitable pose to facilitate phosphoryl transfer. In the WT the positively charged side chain of Lys41 forms an ionic interaction with ATP phosphate tail; in contrast, in the K41R mutant the arginine side chain is able to interact with both ATP and the carboxylate of NAG, in addition to the H-bond with Tyr42 on the glycine-rich Loop 1 (Figure 6b). Such strong interaction with NAG and conformational change of Loop 1 may be detrimental for the catalytic activity of NAGK.

MD simulations of K255A shows ATP still remains in a catalytically competent pose relative to NAG (Figure 6c). However, the mutation forfeited the ionic interaction between Lys255 and Asp196 in the wt enzyme, as a result, Asp196 pulled Asn194 toward it to form a H-bond, such that the stabilization of the substrate NAG by the H-bond between Asn194 and the NAG C5-carboxylate group observed in the WT enzyme is abolished in the mutant.

D216A

In the wt enzyme, Asp216 stabilizes the nucleotide phosphate tail via a water-mediated H-bond network. With the D216A mutation, the water network was disrupted. As a compensation, Gly76 moved in the proximity of the γ- phosphate of ATP to form a H-bond with it. In addition, Asn81 approached the catalytic site to stabilize the nucleotide phosphate tail via a two-water H-bond network, as a result, the displacement of the alpha helix where Asn81 is located (αB) was observed such that the adjacent Loop 2 also underwent significant conformational change (Figure S4a).

D196A/E

A D196A mutation caused the loss of the ionic interaction with Lys255, such that ATP moved away from NAG due to the unrestrained displacement of Lys255, Consequently, the interaction between the β-,γ-phosphate oxygen atoms of ATP and the conserved residue Lys41 is abolished. Loops 1, 2 and 3 all underwent substantial conformational changes. The conformational of Loop 2 caused a displacement of Arg98 and accordingly the displacement of NAG (Figure 6d). As a result, ATP phosphate is too far away from the substrate carboxylate ($d_{\gamma P\text{-}O}$ = 7.9 Å), making phosphoryl transfer impossible to occur.

Kinetic studies on *Escherichia coli* NAGK illustrated that a D162E mutation (equivalent to D196E in *Arabidopsis thaliana* NAGK) resulted in about 0.1% activity of the wild-type. [46] The MD simulations of D196E showed that ATP and NAG remain in a suitable orientation to facilitate phosphoryl transfer, however, the mutation caused the obvious displacement of Asn194, as a result, Asp194 cannot help to hold the substrate NAG via the H-bond interaction with the acetyl carbonyl group, which accounts for the reduced activity of the mutant (Figure 6e).

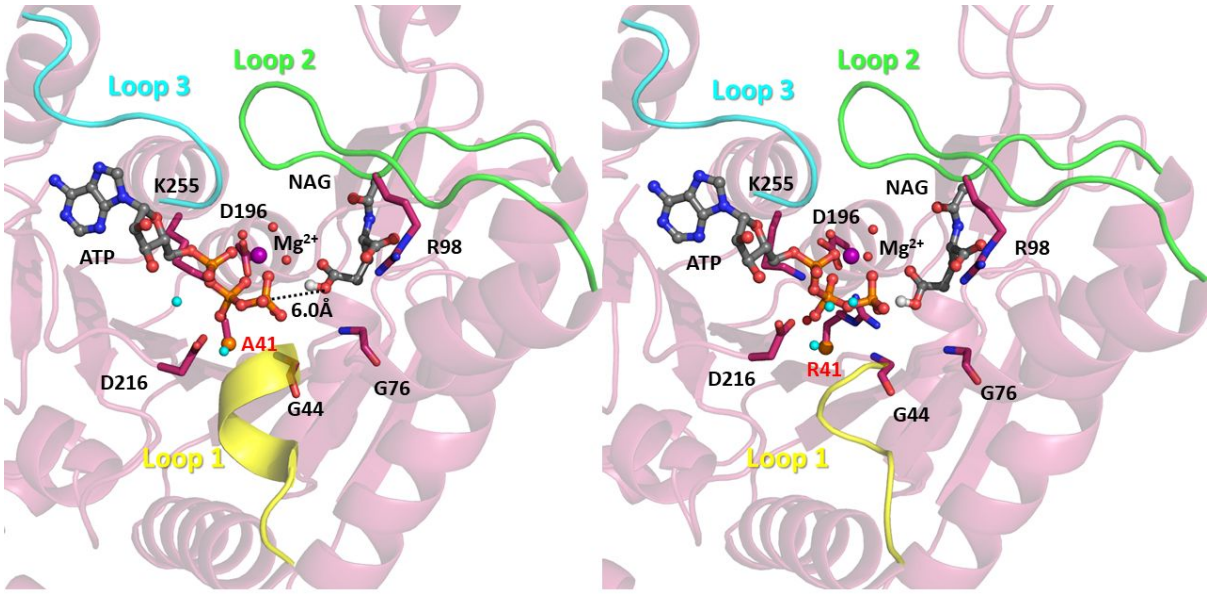
G44A and G76A

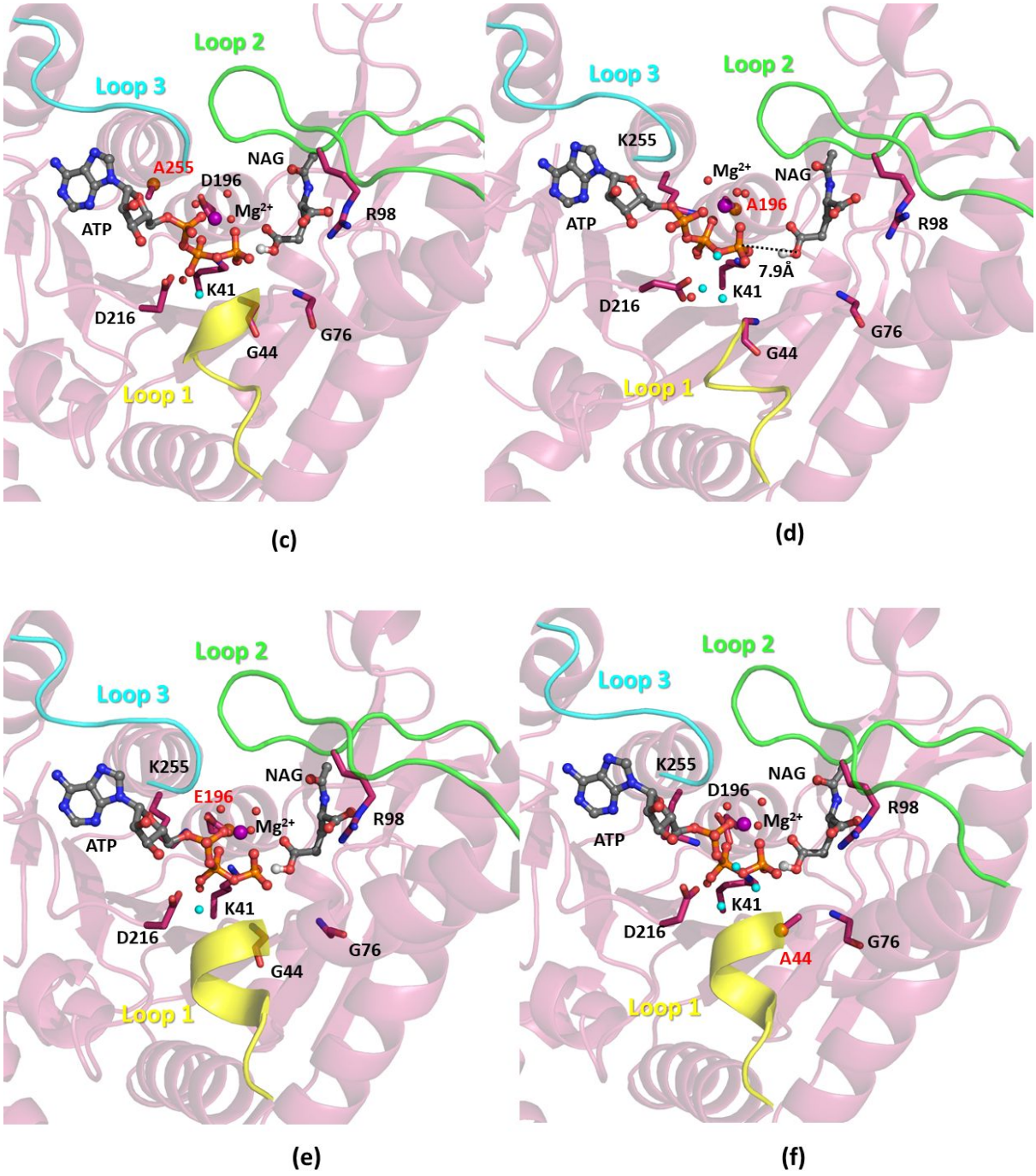
Kinetic studies on *Escherichia coli* NAGK revealed that a G11A mutation (Gly44 in *Arabidopsis thaliana* NAGK) causes 10-fold decrease in V_{max}, 8-fold increase in K_M for ATP and a subtle 3-fold increase in K_M for NAG. [46] From QM/MM studies on the reaction mechanism of NAGK we found that two glycine residues Gly44 and Glay76 play an important role in phosphoryl transfer by forming H-bonds with the β-,γ-, and γ- phosphate oxygens of ATP. MD simulations of the G44A mutant showed and the preceding it. G44A pushed Gly76 away from it due to the steric hindrance, which caused notable displacement of α-helix (αB) adjacent to Gly76 as well as notable conformational change around Loop 2. Similarly, G76A mutation pushed away Asn81 due to steric hindrance of the analine residue

such that the helical structure αB where Asn81 is located underwent significant conformational change (Figure 6f & 6g).

R98A/K

Kinetic characterization of *Escherichia coli* NAGK revealed a R66K mutant (Arg98 in *Arabidopsis thaliana* NAGK) caused diminished activity. [46] MD simulations of R98A/K revealed that the nucleotide and NAG remain in similar states to those found in the WT structure, however the interaction observed between Arg98 and Asn81 was lost due to the displacement of Asn81 caused by the mutation. In both mutants, substantial conformational changes were observed around Loop 1 and Loop 2 (Figure S4b & Figure 6h).




Loop 2
Loop 3
A255
D196
NAG
ATP
Mg2+
R98
K41
D216
G44
G76
Loop 1
(c)
Loop 3
Loop 2
K255
Mg2+
NAG
A196
ATP
R98
7.9Å
K41
D216
G76
G44
Loop 1
(d)
Loop 2
Loop 3
K255
NAG
E196
Mg2+
R98
ATP
K41
D216
G76
G44
Loop 1
(e)
Loop 2
Loop 3
K255
NAG
D196
Mg2+
R98
ATP
K41
D216
G76
A44
Loop 1
(f)

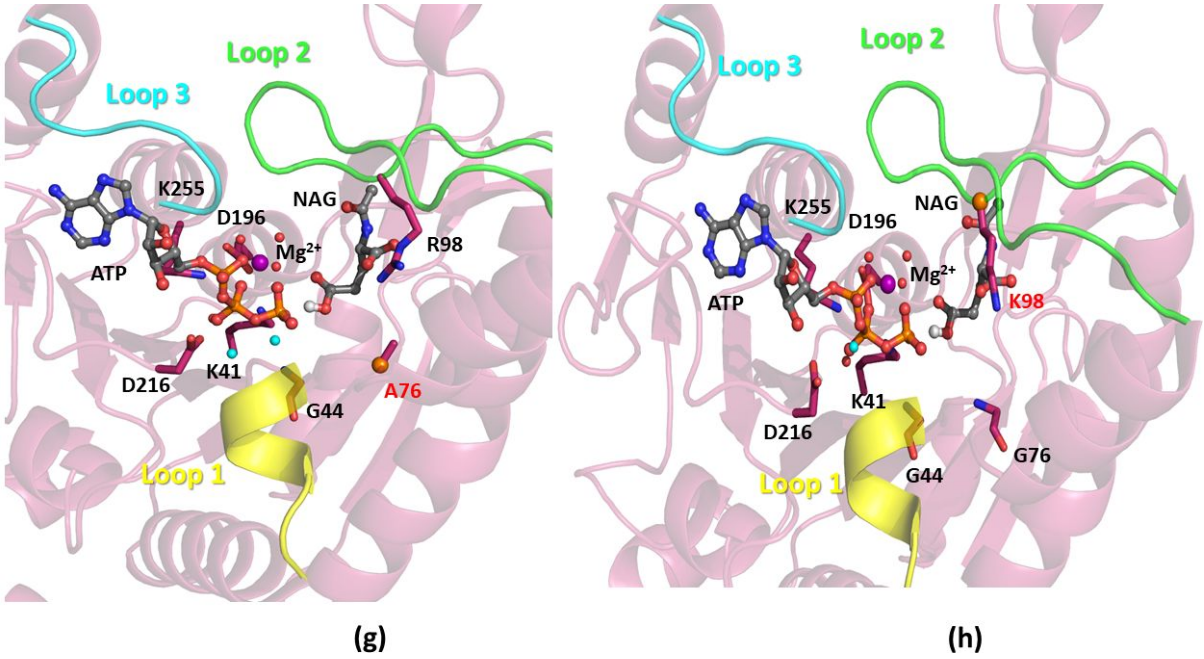


Figure 6. Representative structures obtained from cluster analysis of MD simulations of At-NAGK mutants (a) K41A (b) K41R (c) D255A (d) D196A e) D196E (f) G44A (g) G76E (h) R98K. Loop 1, 2 and 3 are highlighted in yellow, green and cyan respectively.

4. Conclusions

NAGK is an important enzyme in the acetylated route of ornithine biosynthesis for the production of arginine which due to its nitric oxide stimulating effects has a large number of pharmaceutical and medicinal applications. Although extensive studies have been conducted on *E Coli* NAGK, the substrate binding and catalytic mechanism on plant NAGK is scarce. Recently, QM/MM studies were reported on the reaction mechanism of the phosphate sub-family including IPK and FomA, however, no catalytic mechanism has been reported for the carboxylate AAK subfamily kinases.

In this study we studied the catalytic mechanism of At-NAGK, a kinase belonging to the carboxylate subfamily of AAK, using combined MD simulations and QM/MM calculations. From MD simulations, an important water network around ATP mediated by Asp216 was identified which orientates the nucleotide in the absence of a third active site lysine and

histidine residue found in the phosphate sub-division of the AAKs. Comparison of the MD simulated structures of ATP-NAG-Mg^{2+}-NAGK and the ATP-Mg^{2+}-NAGK complexes disclosed three important loops involved in the opening and closing of NAGK. Arg98 plays the pivotal role in substrate binding by interacting with the NAG α-COO^- group, pulling a loop inward to the active site to form the closed state of the complex, facilitating the catalysis to occur. In the absence of this interaction, the loop (Loop2) containing Arg98 moves away from the active site resulting in an open conformation of the kinase, which allows for the substrate entry and product release.

Similar to the phosphate AAK sub-family kinases such as FomA and IPK, which are characterised by a histidine residue and catalyse phosphorylation via a dissociative mechanism, QM/MM calculations revealed that At-NAGK also catalyses phosphoryl transfer via a dissociative mechanism, despite the absence of such histidine in the active site, arguing the previously opposed associative mechanism for Ec-NAGK. Our findings showed that a conserved catalytic lysine in AAK family, Lys41, plays a critical role in catalysis by stabilizing the accumulated negative charge at the β-,γ-bridging oxygen of the nucleotide during the phosphorylation process. Further, we revealed that two highly conserved glycine residues in AAK family Gly44 and Gly76 play an important role in catalysis through positioning the nucleotide via the H-bond interactions with the β,γ- and γ- phosphate oxygens of ATP. The result presents here would lay the basis for the rational engineering the biosynthesis pathway of arginine in the production of L-arginine.

SUOPPORTING INFORMATION

The RESP charges of the atoms in NAGK, RMSD plots of Cα carbon of the backbone of the NAGK complexes, comparison of the crystal structures of the At-NAGK and *E.Coli* NAGK,

RMSF analysis of NAGK and the representative structure of At-NAGK mutants from MD simulations.

ACKNOWLEDGEMENTS

The financial support from the Department of Education and Learning (DEL), Northern Ireland to J.M. is acknowledged, as well as the computing resources from QUB high performance computing centre.

REFERENCES

(1) Morris, S. M. Arginine: Beyond Protein. *Am. J. Clin. Nutr.* **2006**, *83* (2), 508S-512S.

(2) Tuteja, N.; Chandra, M.; Tuteja, R.; Misra, M. K. Nitric Oxide as a Unique Bioactive Signaling Messenger in Physiology and Pathophysiology. *J Biomed Biotechnol* **2004**, *2004* (4), 227–237.

(3) Appleton, J. Arginine: Clinical Potential of a Semi-Essential Amino Acid. *Altern Med Rev* **2002**, *7* (6), 512–522.

(4) Tripathi, P.; Chandra, M.; Misra, M. Oral Administration of L-Arginine in Patients with Angina or Following Myocardial Infarction May Be Protective by Increasing Plasma Superoxide Dismutase and Total Thiols with Reduction in Serum Cholesterol and Xanthine Oxidase. *Oxid. Med. Cell. Longev.* **2009**, *2* (4), 231–237.

(5) Gokce, N. L-Arginine and Hypertension. *J. Nutr.* **2004**, *134* (10 Suppl), 2807S–2811S; discussion. 2818S–2819S.

(6) Vasdev, S.; Gill, V. The Antihypertensive Effect of Arginine. *Int. J. Angiol. Off. Publ. Int. Coll. Angiol. Inc* **2008**, *17* (1), 7–22.

(7) Oomen, C. M.; Erk, M. J. van; Feskens, E. J. M.; Kok, F. J.; Kromhout, D. Arginine Intake and Risk of Coronary Heart Disease Mortality in Elderly Men. *Arterioscler. Thromb. Vasc. Biol.* **2000**, *20* (9), 2134–2139.

(8) Facchinetti, F.; Longo, M.; Piccinini, F.; Neri, I.; Volpe, A. L-Arginine Infusion Reduces Blood Pressure in Preeclamptic Women Through Nitric Oxide Release. *J. Soc. Gynecol. Investig.* **1999**, *6* (4), 202–207.

(9) Huynh, N. T.; Tayek, J. A. Oral Arginine Reduces Systemic Blood Pressure in Type 2 Diabetes: Its Potential Role in Nitric Oxide Generation. *J. Am. Coll. Nutr.* **2002**, *21* (5), 422–427.

(10) Ohtsuka, Y.; Nakaya, J. Effect of Oral Administration of L-Arginine on Senile Dementia. *Am. J. Med.* **2000**, *108* (5), 439.

(11) Cunin, R.; Glansdorff, N.; Piérard, A.; Stalon, V. Biosynthesis and Metabolism of Arginine in Bacteria. *Microbiol Rev* **1986**, *50* (3), 314–352.

(12) Xu, Y.; Labedan, B.; Glansdorff, N. Surprising Arginine Biosynthesis: A Reappraisal of the Enzymology and Evolution of the Pathway in Microorganisms. *Microbiol Mol Biol Rev* **2007**, *71* (1), 36–47.

(13) Winter, G.; Todd, C. D.; Trovato, M.; Forlani, G.; Funck, D. Physiological Implications of Arginine Metabolism in Plants. *Front Plant Sci* **2015**, *6*:534.

(14) Shargool, D.; Jain, J. C.; McKay, G. Ornithine Biosynthesis, and Arginine Biosynthesis and Degradation in Plant Cells. *Phytochemistry* **1988**, *27* (6), 1571–1574.

(15) Visek, W. J. Arginine Needs, Physiological State and Usual Diets. A Reevaluation. *J. Nutr.* **1986**, *116* (1), 36–46.

(16) Llácer, J. L.; Fita, I.; Rubio, V. Arginine and Nitrogen Storage. *Curr. Opin. Struct. Biol.* **2008**, *18* (6), 673–681.

(17) Burillo, S.; Luque, I.; Fuentes, I.; Contreras, A. Interactions between the Nitrogen Signal Transduction Protein PII and N-Acetyl Glutamate Kinase in Organisms That Perform Oxygenic Photosynthesis. *J. Bacteriol.* **2004**, *186* (11), 3346–3354.

(18) Gil-Ortiz, F.; Ramón-Maiques, S.; Fernández-Murga, M. L.; Fita, I.; Rubio, V. Two Crystal Structures of Escherichia Coli N-Acetyl-L-Glutamate Kinase Demonstrate the Cycling between Open and Closed Conformations. *J. Mol. Biol.* **2010**, *399* (3), 476–490.

(19) Sanchez-Martinez, M.; Marcos, E.; Tauler, R.; Field, M.; Crehuet, R. Conformational Compression and Barrier Height Heterogeneity in the N-Acetylglutamate Kinase. *J Phys. Chem. B* **2013**, *117* (46), 14261–14272.

(20) Ramón-Maiques, S.; Marina, A.; Gil-Ortiz, F.; Fita, I.; Rubio, V. Structure of Acetylglutamate Kinase, a Key Enzyme for Arginine Biosynthesis and a Prototype for the Amino Acid Kinase Enzyme Family, during Catalysis. *Structure* **2002**, *10* (3), 329–342.

(21) Chellamuthu, V.-R.; Ermilova, E.; Lapina, T.; Lüddecke, J.; Minaeva, E.; Herrmann, C.; Hartmann, M. D.; Forchhammer, K. A Widespread Glutamine-Sensing Mechanism in the Plant Kingdom. *Cell* **2014**, *159* (5), 1188–1199.

(22) Pakhomova, S.; Bartlett, S. G.; Doerner, P. A.; Newcomer, M. E. Structural and Biochemical Insights into the Mechanism of Fosfomycin Phosphorylation by Fosfomycin Resistance Kinase FomA. *Biochemistry* **2011**, *50* (32), 6909–6919.

(23) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

(24) Meagher, K. L.; Redman, L. T.; Carlson, H. A. Development of Polyphosphate Parameters for Use with the AMBER Force Field. *J. Comput. Chem.* **2003**, *24* (9), 1016–1025.

(25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, revision D. 01; Gaussian, Inc.: Wallingford, CT, **2009**.

(26) Case, D. A.; Babin, V.; Berryman, J.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T.E.; Darden, T. A.; Duke, R. E.; et al. *AMBER 14*; University of California, San Francisco. **2014**.

(27) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.

(28) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(29) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(30) Tao, P.; Schlegel, H. B. A Toolkit to Assist ONIOM Calculations. *J. Comput. Chem.* **2010**, *31*, 2363–2369.

(31) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103* (2), 227–249.

(32) Singh, U. C.; Kollman, P. A. A Combined Ab Initio Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the CH3Cl + Cl− Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.* **1986**, *7* (6), 718–730.

(33) Gutteridge, A.; Thornton, J. Conformational Change in Substrate Binding, Catalysis and Product Release: An Open and Shut Case? *FEBS Lett.* **2004**, *567* (1), 67–73.

(34) Huse, M.; Kuriyan, J. The Conformational Plasticity of Protein Kinases. *Cell* **2002**, *109* (3), 275–282.

(35) Tong, M.; Seeliger, M. A. Targeting Conformational Plasticity of Protein Kinases. *ACS Chem. Biol.* **2015**, *10* (1), 190–200.

(36) Yang, X. Conformational Dynamics Play Important Roles upon the Function of N-Acetylglutamate Kinase. *Appl. Microbiol. Biotechnol.* **2017**, *101* (9), 3485–3492.

(37) McClory, J.; Timson, D. J.; Singh, W.; Zhang, J.; Huang, M. Reaction Mechanism of Isopentenyl Phosphate Kinase: A QM/MM Study. *J Phys Chem B* **2017**, *121* (49), 11062–11071.

(38) Wu, Y.-J.; Zheng, Q.-C.; Zhang, J.-L.; Chu, W.-T.; Cui, Y.-L.; Wang, Y.; Zhang, H.-X. Fosfomycin Induced Structural Change in Fosfomycin Resistance Kinases FomA: Molecular Dynamics and Molecular Docking Studies. *J Mol Model* **2014**, *20* (5), 2236.

(39) Ball, P. Water Is an Active Matrix of Life for Cell and Molecular Biology. *Proc. Natl. Acad. Sci. U. S. A* **2017**, 114 (51), 13327-13335.

(40) Cyphers, S.; Ruff, E. F.; Behr, J. M.; Chodera, J. D.; Levinson, N. M. A Water-Mediated Allosteric Network Governs Activation of Aurora Kinase A. *Nat. Chem. Biol.* **2017**, *13* (4), 402–408.

(41) Levinson, N. M.; Boxer, S. G. A Conserved Water-Mediated Hydrogen Bond Network Defines Bosutinib's Kinase Selectivity. *Nat. Chem. Biol.* **2014**, *10* (2), 127–132.

(42) McClory, J.; Lin, J.-T.; Timson, D. J.; Zhang, J.; Huang, M. Water-Mediated Network in the Resistance Mechanism of Fosfomycin. *Phys. Chem. Chem. Phys.* **2018**, *20* (33), 21660–21667.

(43) Dellas, N.; Noel, J. P. Mutation of Archaeal Isopentenyl Phosphate Kinase Highlights Mechanism and Guides Phosphorylation of Additional Isoprenoid Monophosphates. *ACS Chem. Biol.* **2010**, *5* (6), 589–601.

(44) Mildvan, A. S. Mechanisms of Signaling and Related Enzymes. *Proteins: Struct., Funct., Genet*. **1997**, 29, 401−416.

(45) Beez, S.; Fokina, O.; Herrmann, C.; Forchhammer, K. N-Acetyl-l-Glutamate Kinase (NAGK) from Oxygenic Phototrophs: PII Signal Transduction across Domains of Life Reveals Novel Insights in NAGK Control. *J Mol. Biol.* **2009**, *389* (4), 748–758.

(46) Marco-Marín, C.; Ramón-Maiques, S.; Tavárez, S.; Rubio, V. 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.* **2003**, *334* (3), 459–476.

TOC Graphic