A Network of Conformational Transitions Revealed by Molecular Dynamics Simulations of the Binary Complex of *Escherichia coli* 6-Hydroxymethyl-7,8-dihydropterin Pyrophosphokinase with MgATP

Kaifu Gao, Ya Jia, and Minghui Yang

1Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan 430079, P. R. China
2Key Laboratory of Magnetic Resonance in Biological Systems, State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Wuhan Centre for Magnetic Resonance, Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences, Wuhan 430071, P. R. China

Supporting Information

ABSTRACT: 6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) catalyzes the first reaction in the folate biosynthetic pathway. Comparison of its X-ray and nuclear magnetic resonance structures suggests that the enzyme undergoes significant conformational change upon binding to its substrates, especially in three catalytic loops. Experimental research has shown that, in its binary form, even bound by analogues of MgATP, loops 2 and 3 remain rather flexible; this raises questions about the putative large-scale induced-fit conformational change of the HPPK–MgATP binary complex. In this work, long-time all-atomic molecular dynamics simulations were conducted to investigate the loop dynamics in this complex. Our simulations show that, with loop 3 closed, multiple conformations of loop 2, including the open, semiopen, and closed forms, are all accessible to the binary complex. These results provide valuable structural insights into the details of conformational changes upon 6-hydroxymethyl-7,8-dihydropterin (HP) binding and biological activities of HPPK. Conformational network analysis and principal component analysis related to the loops are also discussed.

Enzymes, or proteins generally, are not always static: X-ray crystallography and nuclear magnetic resonance (NMR) have revealed that multiple conformations can be adopted by different liganded states of the same enzyme or even by the multiple conformational equilibria of an enzyme in the same liganded state. Although the role of conformational dynamics in an enzymatic reaction is still under intense debate, it is clearly known that transitions between multiple conformations are critical for enzymatic catalysis because of the conflicting structural requirements for an enzyme through its catalytic cycle. An unbound enzyme must adopt an open conformation so that its substrate can enter its active site to form a Michaelis complex. Subsequently, to stabilize the transition state and prevent side reactions, the enzyme typically assumes a closed conformation with the substrate buried in its active site. After the chemical reaction, the enzyme must open its active site again to allow the product to exit. For a bisubstrate or multisubstrate enzyme, additional liganded forms, such as binary and ternary substrate complexes and product complexes, must be assumed, so as to undergo some additional conformational transitions between them. While ample evidence has indicated that, in a two-state equilibrium, an unbound enzyme can assume a bound as well as an unbound conformation, which is the current focus of experimental and computational studies; however, in a more complex situation for a bisubstrate or multisubstrate enzyme, it remains largely unknown whether one liganded form, such as binary or ternary substrate complexes, can assume multiple conformations that an enzyme must adopt throughout its catalytic cycle.

6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) catalyzes the transfer of Mg²⁺-dependent pyrophosphoryl from ATP to 6-hydroxymethyl-7,8-dihydropterin (HP), leading to the biosynthesis of folate, which is essential for life and must be synthesized de novo for most microorganisms. Because of the importance of this reaction in microbial growth, HPPK can be an ideal target for the development of antibiotics. Moreover, as a small (158 residues, 18 kDa), stable, monomeric protein that undergoes dramatic conformational changes during its catalytic cycle and is amenable to a variety of biochemical and biophysical analyses, *Escherichia coli* HPPK has emerged as an excellent model system for studying the role of protein conformational dynamics in enzymatic catalysis. The HPPK-catalyzed reaction follows an ordered bi-bi mechanism in which the nucleotide substrate MgATP binds to the enzyme first. During its catalytic cycle, at least five liganded forms must be assumed: the apo-HPPK form (without either of the substrates), the binary substrate complex with

Received: July 15, 2016
Revised: November 7, 2016
Published: November 16, 2016
MgATP, the ternary substrate complex (Michaelis complex) with MgATP and HP, the ternary product complex with AMP and 6-hydroxymethyl-7,8-dihydropterin pyrophosphate (HPPPP), and the binary product complex with HP.20,21 Atomic structures of these five forms have been resolved by X-ray crystallography and NMR using substrates (Protein Data Bank (PDB) structure 1RAO with AMP and HP [PDB entry 1HKA]; the binding of MgAMPCPP, an analogue of MgATP, results in an open conformation, with both loops extended from the active site [PDB entry 1EQ0 (Figure 1b)]. The binding of MgAMPCPP, another analogue of MgATP, and HP leads to a closed conformation: both loops move in, and a fully assembled active site is formed with the two substrates buried in it [PDB entry 1Q0N (Figure 1c)]; after the chemical reaction, the loops re-open so that the products can exit from the active site [PDB entries 1RAO and 1RB0 (Figure 1d,e)]. More details of these five experimental structures are listed in Table 1.

Our previous molecular dynamics (MD) simulations of apo-HPPK revealed that the conformations of loops 2 and 3 that are detected in different phases of the catalytic cycle (see Figure 1) as well as other multiple conformations exist prior to MgATP binding. Of five reported experimental structures, MgATP can enter the active site of structures 1EQ0 and 1RAO but cannot gain entry into the other three structures, 1HKA, 1Q0N, and 1RB0 (see Table 1). Therefore, the conformational changes of apo-HPPK that occur upon MgATP binding can be better described by the conformational-selection model.

Once MgATP binds to HPPK, the movement of loops 2 and 3 and the effect of MgATP in the active site on the motion of these loops remain unclear. The flexibility of the two catalytic loops in the HPPK–MgATP complex has been evidenced by the low intensities or disappearance of many loop residues in the 1H–15N heteronuclear single-quantum coherence spectrum27 and confirmed by the MD simulations with the locally enhanced sampling (LES) technique.24 However, even in the LES MD simulations, the motions of the loops with higher B factor values were not obviously observed. In this work, conventional MD simulations totaling 1 μs at room temperature (300 K), which represents a duration much longer than those of previous studies,24,25 were applied to investigate the conformational transitions of the HPPK–MgATP binary complex. Our simulations revealed that loop 3 of the HPPK–MgATP binary complex was stabilized in closed conformations; nevertheless, multiple conformations, including the open, semiopen, and closed forms, can be adopted by loop 2, providing additional insight into our understanding of the functional mechanism of the conformational change upon HP binding.23

### COMPUTATIONAL METHOD

**Initial Structure.** The initial coordinates of the binary complex of HPPK and MgATP were taken from the X-ray crystallographic structure of a ternary complex of HPPK with MgAMPCPP and HP [PDB entry 1Q0N (Figure 1c)];20 AMPCPP is an ATP analogue in which the O atom between β- and γ-phosphates is replaced by a -CH2- group to prevent hydrolysis. We remove HP from the structure and mutate AMPCPP to ATP by replacing the C atom between β- and γ-phosphates of AMPCPP with an O atom, resulting in the final

Figure 1. Snapshots of HPPK throughout the catalytic cycle. The steps of the catalytic cycle are indicated by the black arrows; the cycle begins with apo-HPPK, represented by structure 1HKA [here and below, structure designations are provided as Protein Data Bank (PDB) entries], and proceeds through the binary substrate complex represented by structure 1EQ0, the ternary substrate complex represented by structure 1Q0N, the ternary product complex represented by structure 1RAO, and the binary product complex represented by structure 1RB0. The protein molecule is drawn as a ribbon diagram; the ligands are drawn as stick models, and the two Mg2+ ions are drawn as spheres. The PDB entries are shown in parentheses. The structure of 1EQ0 was determined via NMR and therefore does not include the ligand, whereas the other structures were all determined via X-ray crystallography. The Cα atoms of Pro47, Ala86, and Phe101, used to monitor the conformational changes of the large loops, are marked with black dots.
Biochemistry

Table 1. PDB Entries, Ligands, and Conformational States of Loops 2 and 3 of the Five Experimental Structures Shown in Figure 1

<table>
<thead>
<tr>
<th>PDB entry</th>
<th>1HKA</th>
<th>1EQ0</th>
<th>1QSN</th>
<th>1RAO</th>
<th>1RB0</th>
</tr>
</thead>
<tbody>
<tr>
<td>ligand(s)</td>
<td>none</td>
<td>MgAMPPCP</td>
<td>MgAMPPCP and HP</td>
<td>AMP and HPPP</td>
<td>HPPP</td>
</tr>
<tr>
<td>loop 2</td>
<td>semiopen</td>
<td>open</td>
<td>closed</td>
<td>closed</td>
<td>closed</td>
</tr>
<tr>
<td>loop 3</td>
<td>semiopen</td>
<td>open</td>
<td>closed</td>
<td>open</td>
<td>semiopen</td>
</tr>
</tbody>
</table>

The DRMS model of the HPPK–MgATP complex. The partial charge of ATP reported by Duan et al.24 is adopted. The binary complex was placed in an octahedral TIP3P water box,29 with the edges of the water box located at least 10 Å from the solute atoms. Upon solvation, the system consisted of 2585 solute atoms and 14277 solvent atoms, neutralized by five Na\(^+\) ions, under periodic boundary conditions.

**MD Simulations.** The MD simulations were performed using the PMEMD simulation module for graphics processing units (pmemd.cuda.MPI) in AMBER1130 and the AMBER ff03 force field.33 To remove potential bad contacts, the initial model was subjected to two rounds of energy minimization: (1) 1000 steps of steepest descent followed by 1000 steps of a conjugate gradient with the protein atoms constrained by a harmonic potential with a force constant of 100 kcal/mol and (2) 2000 steps of steepest descent and 3000 steps of a conjugate gradient without any constraints. Five 200 ns MD trajectories at 300 K with different initial random velocities were obtained, resulting in a total simulation of 1 µs. The system was heated at a constant volume from 0 to 300 K over 60 ps. The MD simulations were then run at a constant temperature and a constant pressure with the temperature maintained at the temperatures set by Langevin’s thermostat,34 a collision frequency of 1.0 ps\(^{-1}\), and the pressure maintained at 1.0 atm using isotropic positional scaling with a pressure relaxation time of 2.0 ps. The particle mesh Ewald method35 was used to estimate the long-range electrostatic energies. The nonbonded cutoff was set to 10 Å. All bonds involving hydrogen atoms were fixed to their equilibrium values using the SHAKE algorithm to allow a 2.0 fs time step. The MD trajectories were recorded at 4 ps intervals.

**Clustering and Conformational Network Analysis.** The MD snapshots were clustered on the basis of the distance root mean square (DRMS) using the leader algorithm implemented in WORDOM.37,38 The DRMS between two structures or snapshots (a and b) is de\(\text{fi}c\)ned as follows:

\[
\text{DRMS} = \sqrt{\frac{1}{N} \sum_{i,j} (d_{ij} - \bar{d}_{ij})^2} \tag{1}
\]

where \(d_{ij}\) is the distance between atoms \(i\) and \(j\) and \(N\) is the total number of pairs of atoms.

Thus, the DRMS for \(n\) atoms takes into account the fluctuations of \(N = n(n - 1)/2\) distance pairs not only in one loop but also between two loops,19,46 which means the DRMS is sensitive to not only the conformations of individual loops but also the conformational coupling of the loops, whereas other methods such as root-mean-square deviation (RMSD) are simply based on coordinates of individual atoms and lose the conformational coupling information. Therefore, the DRMS was chosen for the conformational clustering analysis. Backbone atoms \(C_\alpha\), \(C\), and \(N\) of the three flexible loops (residues 8–15 in loop 1, residues 43–53 in loop 2, and residues 82–92 in loop 3) were used for the calculations. The DRMS cutoff for the conformational clustering was 3.5 Å. For the MD trajectories, each cluster was treated as a node of a conformational network, with the node size being proportional to the number of snapshots in the cluster and the edge thickness being proportional to the number of direct transitions between two connected clusters. The visual representations of the conformational networks were created with Pajek software.41

**Principal Component Analysis (PCA).** Using the ptraj module of AmberTools,30 PCA was performed on the \(C_\alpha\) atoms. The coordinates of each trajectory were superposed by means of a least-squares fit using the first snapshot as a reference. Then, a covariance matrix of atomic fluctuations, \(C\), was calculated, with each element \(c_{ij}\) for a given pair of atoms defined as follows:

\[
c_{ij} = \langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle = \langle r_{ij} \rangle - \langle r_i \rangle \langle r_j \rangle \tag{2}
\]

where \(r_i, r_j, \ldots, r_{IN}\) are the Cartesian coordinates of the \(C_\alpha\) atoms and the angular brackets denote an ensemble average calculated from all snapshots of the MD trajectory.

Eigenvalues \(\lambda_i\) (the diagonal elements of the diagonalized matrix) and eigenvectors \(L\) (the columns of the orthonormal transformation) were yielded by diagonalization of the covariance matrix. These eigenvectors represent the principal modes of motion, and the eigenvalues represent the mean-square fluctuations along the eigenvector coordinates. PCA is not restricted to harmonic motions; collective transitions between structures that differ greatly can also be described by it.

The percentage \(v_i\) of the total variance contained in a given eigenvector \(L_i\) is given by

\[
v_i = \frac{\lambda_i}{\sum_{k=1}^{IN} \lambda_k} \tag{3}
\]

where \(v_i\) measures the contribution of a given PC to the overall protein fluctuations. The cumulative contribution from a set of PCs is

\[
V_n = \sum_{i=1}^{n} v_i, \quad V_n = 1 \tag{4}
\]

Typically, the first few PCs capture most of the intramolecular fluctuation of a protein.34–51

**RESULTS**

**Stability of Trajectories.** Displayed in Figure 2a, the average RMSD value from the initial structure calculated using all \(C_\alpha\) atoms is \(\sim 1.5\) Å for all the trajectories, which means the protein was stable during each of the 200 ns MD simulations. As shown by the \(B\) factor values calculated for the MD trajectories (Figure 2b), the conformational changes during the simulations were mostly localized in loops 2 and 3, consistent with NMR order parameters derived from NMR structures (note there are no crystal structures for the binary complexes). Similar to those found for apo-HPPK,9 Pro47 and Ala86 have the highest \(B\) factor values in loops 2 and 3, respectively, and, Phe101 is located at the bottom of the protein with little
The reason for such choice is that the Cα Pro86 and Phe101 and the RMSD between loop 3 and 1Q0N discriminate di-Pro47 and Phe101 (Pro86 and Phe101) can reasonably minimizing and heating were performed before the production on the landscape are also shown in the graphs. Because the experimental structures 1Q0N, 1HKA, 1EQ0, 1RAO, and 1RB0 in the catalytic cycle (see Figure 1).

In our simulations, a significant difference is found between loops 2 and 3. Loop 2 has a wide conformational space, which includes the closed (1RAO and 1RB0), semiopen (1HKA), and open (1EQ0) experimental structures shown in Figure 1 and Table 1 as well as some “superopen” conformations. However, the conformational space of loop 3 is very limited: it consists of only closed conformations with the distance between Ala86 and Phe101 being <33 Å, so 1RB0 is inside it, but 1RAO and 1EQ0 are far beyond it [for loop 3, the representative points of 1RAO and 1EQ0 are (38.62, 13.27) and (37.61, 13.65), respectively, which are outside of range of Figure 3b].

Clustering and Conformational Network Analysis. To map out the conformational states as well as transitions between them, clustering and conformational network analysis was performed on the conformations of the three flexible loops based on the simulations, which is depicted in Figure 4. There are five clusters as shown in Figure 4a. Multiple conformations [closed (cluster 1), semiopen (clusters 2 and 3), and open (clusters 4 and 5); the centers of these clusters are marked in Figure 3] are observed for loop 2; loop 3 remains in the closed (clusters 1–3 and 5) or superclosed (cluster 4) conformations, and loop 1 changes only slightly.

Figure 4b shows the time series of the cluster assignments for our five trajectories. The clusters obtained from the five trajectories are different, which means different trajectories go through different conformational spaces: the transformations between open and closed conformations of loop 2 can be observed vividly. Although all trajectories start with the initial structure that belongs to cluster 1, cluster 2 with the semiopen conformations of loop 2 and the closed conformation of loop 3 is always the predominant one (>50%) in each trajectory.

The occurrence percentages of these clusters and conformational space network among them are shown in Figure 4c. Consistent with Figure 4b, cluster 2 has the largest percentage at 71.1%, implying it is the most stable. Cluster 1 with the closed conformation of loop 2 is metastable with the percentage being 24.1%. Moreover, the conformational space network indicates that cluster 2 is a key transition state as cluster 1 can connect to all the other clusters only via cluster 2. As loop 2 is semiopen in cluster 2, the closed to open transformation must pass through semiopen conformations. This point is more intuitively illustrated by Figure 3a: cluster 2 is “midway” between the closed and open conformations.

Principal Component Analysis. To intuitively characterize the dynamics of the protein, PCA is performed. All Cα atoms were used to define the backbone conformation of HPPK for PCA, resulting in 474 principal components (PCs) from the 158 Cα atoms in the form of 474 eigenvectors and their associated eigenvalues. Figure S2 shows that, consistent with earlier research,25 20 of the 474 PCs capture approximately 80% of the protein’s motion in all our trajectories. In particular, the first three PCs have eigenvalues much larger than the rest, and in every trajectory, they capture

![Figure 2](image-url)  
**Figure 2.** (a) Time evolution of Cα RMSDs of the MD simulations. (b) B factor distributions of the MD simulations.
no less than 50% of the motion. Hence, the first three PCs are sufficient for the analysis.

Figure 5 shows the vector field representations of the PCs obtained from the combined data of all trajectories. It clearly indicates that, quite consistent with the B factor graphs, loops 2 and 3 have the largest motions. In PCs 1 and 2 with the largest eigenvalues, the motion of loop 2 is moving up and down; however, for loop 3, there is no such motion. The motion of loop 3 is shearing (PC 1) and twisting (PC 2). Therefore, it is unsurprising that, even with high flexibility, loop 3 always remains in closed conformations in the HPPK–MgATP binary complex.

Hydrogen Bonds Related to HP Binding. The crystal structure of the ternary complex (HPPK–AMPCPP–HP, PDB entry 1Q0N) reveals that HP is bound to HPPK through six hydrogen bonds with residues Thr42, Pro43, Leu45, and Asn55, which altogether form a strong hydrogen bond network and stabilize the closed conformations of loop 2. However, because HP is removed in these simulations, the hydrogen bond network is lost and the closed conformations cannot be stabilized by the hydrogen bonds solely among these four residues, leading to a flexible loop 2 throughout our simulations. Therefore, only after HP binding can the hydrogen bond network stabilizing loop 2 be well formed; this is why loop 2 has large conformational changes in our simulations when HP is absent.

DISCUSSION

Dynamics of Loops 2 and 3 of the Binary Complex. The free energy and clustering analysis indicate that, even beginning from the initial crystal structure of the binary complex (PDB entry 1Q0N) with loop 2 closed, all the conformations of loop 2 in the five experimental structures
(1Q0N, 1HKA, 1EQ0, 1RAO, and 1RB0) depicted in Figure 1 and Table 1 as well as some superopen conformations can be observed in our simulations, and loop 3 is stabilized in closed conformations by MgATP, consistent with the PCA. Moreover, the clustering analysis shows that one particular semiopen conformation of loop 2 (cluster 2) is a key transition state as every transformation between open and closed conformations of loop 2 must pass through it (Figures 3a and 4c).

**Induced Fit** versus **Conformational Selection**. Until now, two different models, which are generally termed “induced fit” and “conformational selection”, have been proposed to describe the conformational changes that occur upon ligand binding.52−54 The induced-fit model presumes that

![Figure 4. DRMS-based clustering and network analysis of the MD simulations: (a) the center structures of each cluster, (b) the time evolution of the clusters, and (c) the distribution and interconversion network of the clusters. In panel b, the different clusters are represented by different colors. In panel c, the relative populations of the clusters are indicated by percentage values and the thickness of each connecting line is proportional to the corresponding frequency of interconversion.](image-url)
the binding of the ligand may induce a conformational change in the enzyme, resulting in an optimized geometry that exists only in the complex state.\textsuperscript{55} In this scenario, the protein energy landscape changes upon the binding of the ligand. By contrast, the conformational-selection model presumes that more than one structural conformation preexists in the conformational equilibrium before ligand binding.\textsuperscript{56–58} This model involves the stabilization of an accessible conformation, and the ligand binding causes a shift in the preexisting conformational equilibrium. It is important to note that these two models are not mutually exclusive: a survey of the recent literature indicates that many processes can simultaneously include certain elements of both the induced-fit and conformational-selection models.\textsuperscript{53,54,59–61}

Our MD simulations show that, even after the binding of MgATP, loop 2 still undergoes dramatic conformational changes: the five experimental structures in the catalytic cycle (see Figure 1 and Table 1) as well as multiple other conformations of loop 2 coexist. Of these five structures, structures 1HKA and 1EQ0 have large entrance channels for HP binding\textsuperscript{25} (the channels are at least 18 Å long, 8 Å wide, and 10 Å deep, while HP is only \( \sim 8 \times 2 \times 5 \) Å). However, the situation is quite different for 1Q0N, 1RAO, and 1RBO: the channels are blocked by the side chains of loop 2,\textsuperscript{20,21} so that HP can enter the binding pocket of only structures 1HKA and 1EQ0 but cannot gain entry into the other three structures (1Q0N, 1RAO, and 1RBO). Our simulations also include multiple other conformations of loop 2, such as “superopen”, etc., indicating that such transformations between active and inactive structures of HP binding are evident. In other words, the conformational changes of the HPPK–MgATP complex upon HP binding observed from our simulations favor the conformational-selection model as the conformation of loop 2 is variable, and its open conformations can exist prior to HP binding. When loop 2 is open, HP enters and binds to the HPPK–MgATP binary complex.

With a measure of the association rate constant of 11 ± 0.03 \( \mu M^{-1} s^{-1} \),\textsuperscript{62} kinetic studies have demonstrated that the addition of the second substrate (HP) to bind is much faster (40 times) than the binding of MgATP, but still much slower than nanosecond scale, the time frame in which the HP-binding conformations are formed in our MD simulations. As the HP-binding conformations are only transition state-like from the MD trajectories (the open conformations of loop 2 last \( \sim 100 \) ns in our 1000 ns simulations), they present only the possibilities that the HP can enter the binding pocket at these conformations. Previous research has indicated that gating by the flexible loops has little effect on the rate of substrate binding: opening and closing of the gate is always much faster than the characteristic diffusion time required for the substrate to explore the region around the enzyme.\textsuperscript{63,64} Moreover, the further process of HP binding at the pocket is expected to be even slower as specific hydrogen bonds between the HPPK–MgATP complex and HP\textsuperscript{20} have to form until the equilibrium is reached.

**Figure 5.** Vector field representations and corresponding eigenvalues of the first three PCs obtained from the combined data of the MD trajectories.
Conformational selection versus induced fit in kinases: The trajectory of pyrophosphoryl transfer catalyzed by 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as revealed by NMR spectroscopy. Biochemistry 45, 12573–12581.


Biochemistry


