Loop Conformation and Dynamics of the *Escherichia coli* HPPK Apo-Enzyme and Its Binary Complex with MgATP

Rong Yang,* Matthew C. Lee,* Honggao Yan,† and Yong Duan‡
*Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware; †Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan; and ‡Genome Center and Department of Applied Science, University of California, Davis, California

ABSTRACT Comparison of the crystallographic and NMR structures of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) suggests that the enzyme may undergo significant conformational change upon binding to its first substrate, ATP. Two of the three surface loops (loop 2 and loop 3) accounting for most of the conformational differences appear to be confined by crystal contacts, raising questions about the putative large-scale induced-fit conformational change of HPPK and the functional roles of the conserved side-chain residues on the loops. To investigate the loop dynamics in crystal-free environment, we carried out molecular dynamics and locally enhanced sampling simulations of the apo-enzyme and the HPPK·MgATP complex. Our simulations showed that the crystallographic B-factors underestimated the loop dynamics considerably. We found that the open-conformation of loop 3 in the binary complex is accessible to the apo-enzyme and is the favored conformation in solution phase. These results revise our previous view of HPPK-substrate interactions and the associated functional mechanism of conformational change. The lessons learned here offer valuable structural insights into the workings of HPPK and should be useful for structure-based drug design.

INTRODUCTION

The enzyme 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) is a 158-residue long, monomeric protein that catalyzes the reaction of pyrophosphate transfer from ATP to 6-hydroxymethyl-7,8-dihydropterin (HP) during the first step of the folate biosynthesis pathway (1). The folate pathway is an essential metabolic pathway for microorganisms because, unlike human and higher animals that can acquire folate from dietary sources, microbes must make their own folate through this pathway (2). Hence, understanding the structure-function relationships of HPPK has important implications for antibacterial drug design.

Currently, there are 16 experimentally determined structures of HPPK in various conformational states available in the Protein Data Bank (PDB) (3–12). The 3-D fold of HPPK is ferredoxinlike, which resembles a right-handed baseball glove. It consists of a six-stranded β-sheet sandwiched by two α-helices on either side of the β-core and three flexible surface loops connecting β_1-α_1 (loop 1, residues 8–15), β_2-β_3 (loop 2, residues 44–53), and β_4-α_2 (loop 3, residues 80–93). Structural and biochemical studies have shown that these surface loops are directly involved in the substrate binding (7,9) and the catalytic mechanisms of HPPK (3,4).

Based on the currently available thermodynamic and kinetic data, Blaszczyk et al. have proposed a reaction pathway consisting of the following seven steps: 1), apoenzyme → 2), binding of MgATP → 3), binding of HP → 4), formation of the reaction intermediate HPPK·MgATP·HP → 5), transfer of pyrophosphate to form HPPK·MgAMP·HPPP → 6), release of MgAMP → 7), release of HPPP. At each of these steps, a conformation was proposed from either an experimentally determined or a modeled conformation (13).

Although these experimental and modeled structures provide snapshots of the enzyme throughout its catalytic cycle, to understand how the enzyme works, one must also consider the dynamics involved in the conformational changes when the enzyme undergoes transitions from one state to the next. In general, two types of ligand-promoted conformational changes can be envisioned. The first type of conformational change is described by the classic ''induced-fit'' model in which the conformation of receptor enzyme changes upon binding of the ligand, resulting in an optimized geometry that exists only in the complex state. In this scenario, the protein energy landscape is changed upon binding of the ligand. The other type of conformational change is called ''selected-fit'' mechanism, which occurs when more than one structural conformation preexists in a conformational equilibrium before ligand binding. The selected-fit mechanism involves stabilization of an accessible conformation and ligand-binding shifts a preexisting conformational equilibrium. A practical difference between these two scenarios is that one may be able to observe the bound structure in apo state in the case of selected fit, whereas the bound structure exists only in the complex state in the case of induced fit.

Xiao et al. have carefully compared the crystal structures of the apo-enzyme and the binary complex. In their analysis, they found major conformational differences in all three loops between the apo and the binary states. In particular,

Submitted February 17, 2005, and accepted for publication April 5, 2005.

Address reprint requests to Yong Duan, Genome Center and Dept. of Applied Science, University of California, Davis, CA 95616. Tel.: 530-754-7632; Fax: 530-754-9658; E-mail: duan@ucdavis.edu.

0006-3495/05/07/95/12 $2.00 doi: 10.1529/biophysj.105.061556

loop 3, in an open conformation, deviates from its apo position by as much as 18.4 Å at the farthest point. Loop 2 also opens up in the binary complex. Together, the opening of loop 2 and loop 3 exposes the binding pocket of HP. Interpolation between apo and binary conformations suggested that the apo-to-binary transition might be through an induced-fit mechanism (7) that triggers the reorganization of the binding pocket and advances the enzyme to the next stage of the reaction pathway.

In the crystal structure of apo-enzyme, loop 3 forms crystal contacts with two adjacent symmetry copies from neighboring asymmetric units; a total of six residues from loop 3 (L^{78}, N^{79}, R^{82}, V^{83}, R^{88}, and R^{92}) are in direct contact with the crystallographic neighbors (Fig. 1). Within this crystalline environment, the range of loop 3′s flexibility is clearly hindered. Thus, loop 3 may exhibit a higher degree of conformational fluctuation in solution than is found in the crystal structure of the apo-enzyme. Loop 2 is also similarly restricted by the crystal. These were consistent with the observation by Shi et al. who reported, based on their NMR study of apo-HPPK, that there were seven unassigned ^{1}H-^{15}N crosspeaks involving main-chain amides and 12 missing main-chain amide crosspeaks, and all except one of these 19 unassigned or missing amides were located on loop 3 around residues 77–96 (14). Thus, loop 3 may exhibit a higher degree of conformational fluctuation than is found in the crystal structure of the apo-enzyme. Given its functional significance, an understanding of the dynamics of loop 3 in solution is needed.

In this article, we critically examine the conformational dynamics of the surface loops of HPPK in the apo and binary complex state via molecular dynamics (MD) simulations to resolve the above issues. More specifically, we ask the question whether the open conformation found in the binary complex is also accessible to the apo-enzyme. To address this question, we carried out 5 ns of molecular dynamics (MD) simulations of the apo-enzyme and the HPPK·MgATP binary complex. In addition to the conventional MD, we also applied the locally enhanced sampling (LES) technique. LES is a mean field method that has been shown to be effective in predicting and refining loop structures (15–19). In this case, it is particularly useful in sampling the available conformational states of the loops. Our results showed that loop 3 in the apo structure, free from its crystal contacts, exhibited a greater range of conformational dynamics and favored a conformation similar to that of the binary complex. In the following sections, we present the details of our simulations and propose a revised scenario of HPPK-substrate interaction and conformational change.

COMPUTATIONAL METHOD

Parameterization of ATP

The ATP molecule was divided into two parts: the triphosphate and the ribose with adenine. The charges of the ribose and adenine were directly taken from the AMBER force field (20). A methyl group was added to the end of the triphosphate group for charge calculation. Following the approach of Duan et al. (21), the charge set of the triphosphate was obtained by fitting to the electrostatic potentials calculated using the B3LYP/cc-pVTZ quantum mechanical methods in organic solvent ($\varepsilon = 4.0$) mimicked by the polarized continuum model (22,23) after geometry optimization using the HF/6-31G** level of theory. The charges of the phosphoryl group are listed in Table 1. The structure and the corresponding atom-naming scheme are illustrated in Fig. 2.

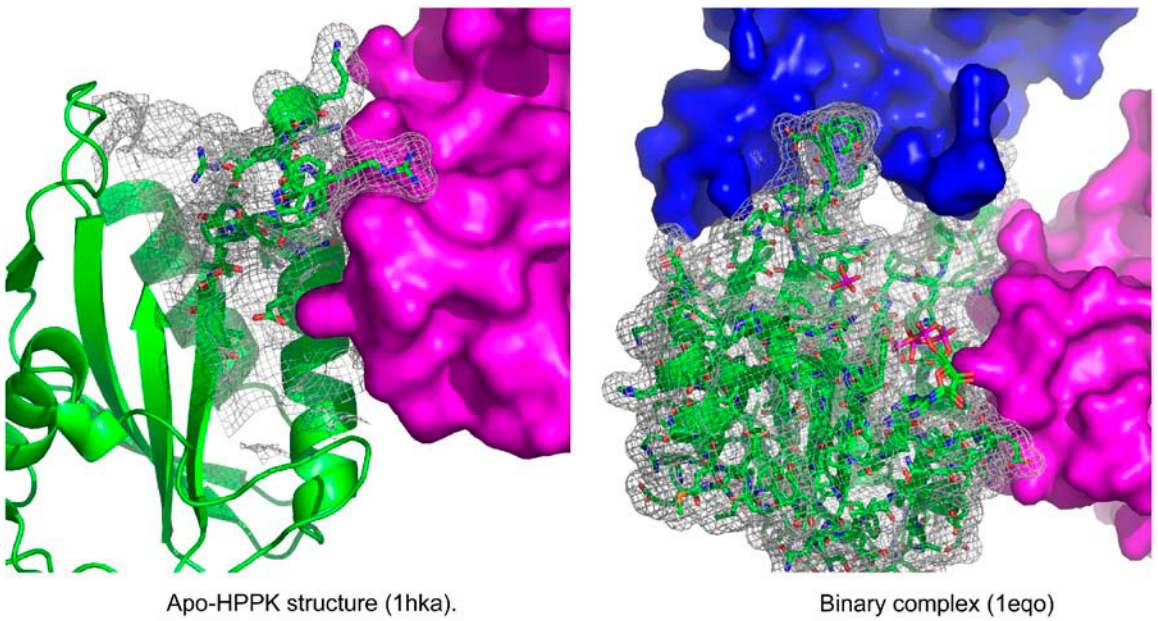

FIGURE 1 Crystal contacts between HPPK loop 3 and the neighboring asymmetric units. (*Left*) apo-HPPK structure (1hka). (*Right*) Binary complex (1eqo).

Initial structure coordinates

The initial coordinates of apo-HPPK were taken from the x-ray crystallographic structure (PDB id 1hka). The NMR structure (PDB id 1eq0) of HPPK·MgAMPCPP binary complex was used as a starting template for HPPK·MgATP. AMPCPP is an ATP analog in which the oxygen atom between β- and γ-phosphates is replaced by a $-CH_2-$ group to prevent hydrolysis. In the NMR structure, the two Mg^{2+} cofactors and the ATP substrate were not present. We superimposed the x-ray crystal structure of HPPK·MgAMPCPP·HP (PDB id 1eqo) to the binary NMR structure and extracted the coordinates of Mg^{2+} ions and AMPCPP to generate the structure of HPPK·MgAMPCPP complex. The AMPCPP in the complex model was then mutated to ATP by replacing the C atom between β- and γ-phosphates of AMPCPP with an O atom, resulting in the final model of HPPK·MgATP. Upon solvation, the apo system consisted of 2578 solute atoms (protein, 19 Cl^- ions and 24 Na^+ ions) and 27,483 solvent atoms for a rectangular box of 77.7 × 65.1 × 74.7 $Å^3$; the binary complex system contained a total of 2623 solute atoms (protein, ATP, two Mg^{2+} ions, 20 Cl^- ions, and 21 Na^+ ions) and 29,149 solvent atoms (86.3 × 62.9 × 73.7 $Å^3$). In both cases, the distances between the solute atoms and the box edges were no less than 10 Å.

Molecular dynamics

Two independent 5.0-ns MD simulations with 500-ps equilibration phase were carried out using the AMBER 7.0 modeling software package (19) and the Duan et al ff03 force-field parameter set (21) with identical simulation protocols. Water solvent was represented by TIP3P model (24) and in periodic boxes. The smooth particle-mesh Ewald method (25) was applied to achieve an accurate and efficient treatment of the long-range electrostatic energies and forces. The short-range van der Waals energy was truncated at 10 Å. The pressure was kept at 1.0 atm using isotropic positional scaling and the temperature was controlled by Berendsen's method (26). All bonds involving hydrogen atoms were constrained to their equilibrium values by means of the SHAKE algorithm (27) to allow a 2.0-fs integration time step. During the equilibration phase, the velocities were reassigned every 0.5 ps, conforming to the Maxwell-Boltzmann distribution.

TABLE 1 Partial charges of ATP molecule

Atom name	Atom type	Charge	Atom name	Atom type	Charge
PG	P	1.103139	C2′	CT	0.067000
O1G	O2	−0.901962	H2′1	H1	0.097200
O2G	O2	−0.901962	C1′	CT	0.039400
O3G	O2	−0.901962	O4′	OS	−0.354800
O3B	OS	−0.482101	H1′	H2	0.200700
PB	P	1.191788	O2′	OH	−0.613900
O1B	O2	−0.832034	HO′2	HO	0.418600
O2B	O2	−0.832034	N9	N*	−0.025100
O3A	OS	−0.506578	C4	CB	0.305300
PA	P	1.030667	C5	CB	0.051500
O1A	O2	−0.811423	C6	CA	0.700900
O2A	O2	−0.811423	N6	N2	−0.901900
O5′	OS	−0.536915	H62	H	0.411500
C5′	CT	0.055800	H61	H	0.411500
H5′1	H1	0.067900	C8	CK	0.200600
H5′2	H1	0.067900	H8	H5	0.155300
C4′	CT	0.106500	N7	NB	−0.607300
H4′	H1	0.117400	N3	NC	−0.699700
C3′	CT	0.202200	C2	CQ	0.587500
H3′	H1	0.061500	H2	H5	0.047300
O3′	OH	−0.654100	N1	NC	−0.761500
H3T	HO	0.437600	—	—	—

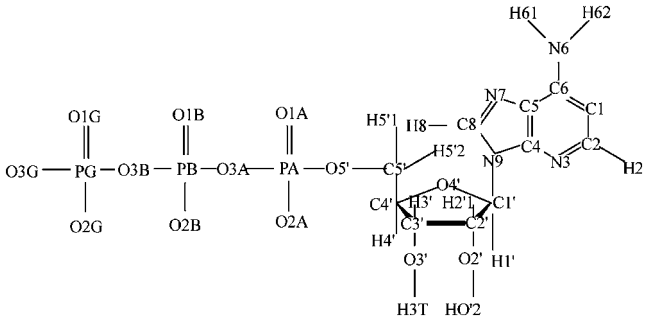

FIGURE 2 2-D structure and atom-naming scheme of ATP.

The simulations started with a 200-step energy minimization, followed by a 100-ps relaxation phase in which the system was heated from 0 to 300 K while the atomic positions of protein and ligands were subject to harmonic restraints with a force constant of 0.6 kcal/mol. This step ensured that the initial experimental structures were maintained while the solvent was allowed to relax. The system was kept at 300 K for another 400 ps for further equilibration without any constraints. A 5-ns production phase was done for each system. Snapshots of the production trajectories were saved at 2-ps intervals.

Locally enhanced sampling

After 5 ns normal MD run, the apo and binary complex simulations were switched to LES method (15–19). Loop 2 and loop 3 were divided into two (44–47 and 48–50) and four (80–83, 84–87, 88–91, and 92–93) LES segments, respectively. Each residue in the LES regions was assigned five copies. The topology file and restart file were generated by the ADDLES program. The LES simulations were started after 5 ns MD simulation and were performed for 4 ns.

Essential dynamics analysis

The essential dynamics (ED) analysis method (28) was applied to the trajectories to identify the concerted modes of motion. We limited our ED analysis to include only the 158 C_α atoms of the protein. In this analysis, a variance/covariance matrix was first constructed from the simulations by $v_{ij} = \langle \Delta x_i \Delta x_j \rangle$, where Δx_i and Δx_j are, respectively, the instantaneous deviation of the ith and jth dimensions of the $3N$-dimension coordinates from their respective averages ($N = 158$ particles involved in this calculation); brackets denote an average over the simulated trajectory. The eigenvectors (or principal components) were identified by diagonalizing the matrix. A total of $3N - 6$ (or 468) eigenvectors can be obtained, excluding the global translation and rotation. Protein movements in the essential space were identified by projecting the Cartesian trajectory coordinates along the most important eigenvectors resulting from the ED analysis. The analysis module implemented in the software package NWChem 4.1 (29) was used for the ED calculations.

RESULTS

Comparison with experimental data

The experimentally measurable macroscopic parameters, such as B-factors, nuclear Overhauser effect (NOE) violations, and NMR order parameters, were calculated from the simulation trajectories and compared to the experimentally measured values to provide an assessment of the

quality of the simulation. We discuss each of these comparisons.

We monitored the total energy and the root mean-square deviation (RMSD) from initial structure throughout the simulations to assess trajectory stability. In both conventional MD simulations, the total energies reached a steady state within 500 ps of the equilibration phase when the average energies remained constant (data not shown). The RMSD of all C_α atoms from the x-ray structure for each system is displayed in Fig. 3. The C_α RMSD of the protein core (excluding the three flexible loops and three residues from the N- and C-termini) are ~1.0 Å and ~1.5 Å for the apo protein and the binary complex, respectively. Clearly, the trajectories were stable. When the loops were also taken into consideration, the RMSD was higher than the core RMSD, which was expected. The RMSD in the apo complex rose from 1.5 Å to 2.0 Å at ~3 ns, whereas the core region maintained its C_α RMSD at 1.0–1.5 Å. Thus, the experimental structures were well maintained and the loop regions exhibited a higher degree of fluctuation.

B-factors, also commonly known as crystal temperature factors, are a composite measure reflecting the atomic thermal fluctuation, configurational disorder, and other experimental errors. They have often been used as a measure of structural mobility and flexibility. In simulations, this value can be calculated from the MD trajectories using the formula $B = 8\pi^2/3\langle\Delta r^2\rangle$, where $\langle\Delta r^2\rangle$ is the mean-square atomic position fluctuation. B-factors were calculated from both the last 1-ns and the entire 5-ns trajectories. The B-factors calculated from the 5-ns trajectories were generally higher than those calculated from the last 1 ns. This was due to the conformational changes.

It should be noted that although most x-ray data is now collected at low temperature (e.g., 100 K), the crystals are typically grown at 4°C or room temperature. They are brought rapidly to the lower temperature by flash-freezing technique, by which the liquid water around and inside the crystals reached a glassy state to prevent the water from becoming ice and destroying the crystals. At this low temperature, because of the frozen water in the glass state, proteins hardly move. The protein structures reflect an ensemble average structure of the ''frozen'' protein molecules. Yet, the thermal motion at the higher temperature (at which the crystal was grown) is captured by the slight variation between different structures of protein molecules. Therefore, the experimental B-factors measure the thermal fluctuation at the temperature at which the crystals are grown (which is close to room temperature).

Overall, the calculated and the experimentally measured B-factors are in good agreement with correlation coefficients of ~0.88 for both the apo-HPPK model and binary HPPK model. Most of the differences were found in the loop regions where discrepancies were expected largely due to crystal packing (we will discuss the dynamics of these surface loops in the absence of crystal packing later). Nevertheless, the calculation indicates that the simulations captured the main features of the system dynamics. It is interesting to note that the calculated B-factors from MD simulation are usually lower than the experimental data because of the exclusion of global translation and rotation in the B-factor calculations from the simulations. However, if the resolution of the crystal structure is very high (i.e., in our case, 1HKA is 1.50 Å), the calculated B-factor might be higher than the experimental data, even in the rigid core part. This is the case in HPPK, as shown in Fig. 4, an attribution to the extremely high-quality x-ray data.

The two major peaks in the apo structure B-factors (Fig. 4 *A*) correspond to loops 2 and 3. The higher B-factors in the simulation reflect the greater mobility of these two loops in the solution phase. Residues 102–110 form the turn connecting β_5 and β_6 strands and are at the bottom part of the ATP binding pocket making direct contacts to ATP through Arg^{110} and Tyr^{112}. In the crystal structure, residue Glu^{109} forms a salt-bridge with Arg^{41} of the neighboring symmetry copy, which accounts for the lower B-factors observed in this region when compared to simulation. In the

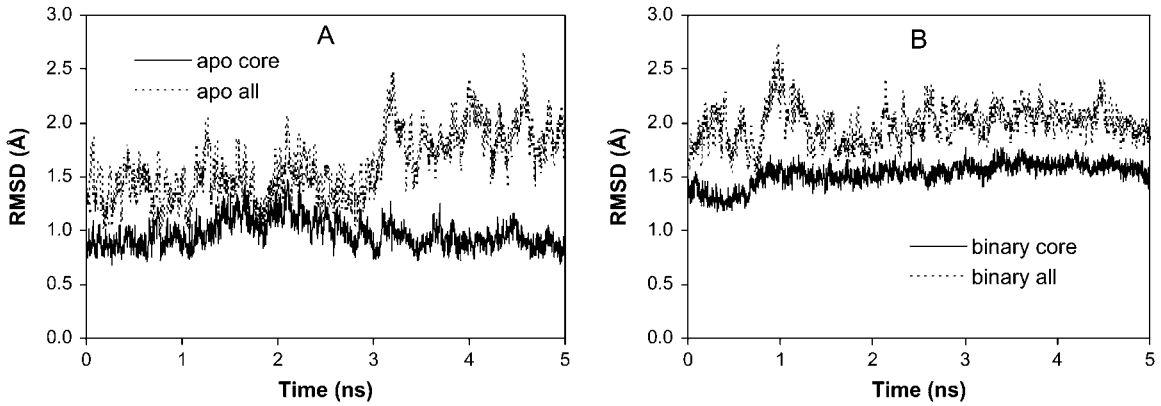

FIGURE 3 C_α RMSD versus time. (*Top*) The time evolution of C_α RMSD of the apo-enzyme. (*Bottom*) The time evolution of C_α RMSD for the HPPK·MgATP binary complex. ''Core'' means excluding the three surface loops, and ''all'' means all C_α atoms including the surface loops.

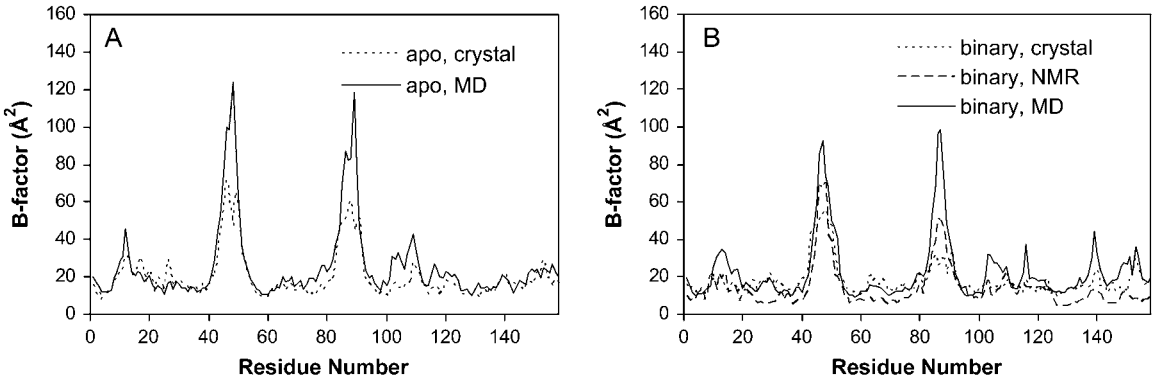

FIGURE 4 Comparisons of C_α-atom B-factors obtained from crystal structures and calculated from MD trajectories. In the case of the Binary_NMR, B-factors were estimated by using the 20 NMR structures reported in the PDB.

binary HPPK·MgADP complex crystal structure, this salt-bridge is not observed, but crystal packing still presents steric hindrance for this part of the structure.

Because the starting structure of the binary complex was modeled from the NMR structure template, we also compared our binary complex simulation to the experimental data by calculating the NOE distance restraint violations. In the PDB file 1eq0, there are twenty HPPK structures generated from a total of 3,523 NOE distance restraints. In analyzing our MD simulation of the binary complex, we divided the trajectory into five 1-ns time periods. It should be noted that on the short timescale of MD the influence of angular fluctuations can be neglected; thus, using $\langle 1/r^3 \rangle$ distance averaging in our case is more appropriate than the conventional $\langle 1/r^6 \rangle$ averaging (30).

Since we replaced AMPCPP with ATP in our simulation, we expected a higher level of NOE violations than what was observed experimentally due to the perturbation, particularly around the perturbed area. During the 5-ns production run, the violation increased from an initial 7.6% to 12.6%. When we ranked the NOE violations by residue (i.e., number of hydrogen atom distances that exceeded the limits obtained from the NMR NOE experiments), we found that a large portion of the violations were concentrated in the regions around loop 2, the V-shaped turn region (100 ~ 112) and the C-terminus (see Fig. 5). When these regions were excluded, the NOE violation was 5.6% initially and 7.8% at the end of the simulation. These values are well within range of previously reported NMR parameter correlations between MD and experiment (30,31). We take the higher level of NOE violation in the above-noted regions as an indication of perturbations due to substrate substitution.

The NMR order parameters (S^2) measure protein backbone flexibility and can be calculated from simulation by the formula

$$S^2 = \frac{1}{2}\left[3\sum_{\alpha=1}^{3}\sum_{\beta=1}^{3}\langle \mu_\alpha \mu_\beta \rangle - 1\right], \quad (1)$$

where μ_α or μ_β denotes the x, y, or z component of the NH unit vector and the angular brackets indicate an average over the trajectory. Here S^2 can take on values between 0 and 1, with 0 corresponding to completely disordered and 1 corresponding to completely frozen residues.

As shown in Fig. 6, NMR order parameters in loop 2, loop 3, and the turn region (104−110) in both the apo and the binary complex simulations are low. This reflects the high degree of internal flexibility in these regions. In contrast, the rigid part of the protein showed high values of order parameters. Overall, the patterns of order parameters are very similar in both the apo and the binary HPPK simulations. The major differences are found in the regions of loop 1 (residues 5−8), and the C-terminal: loop 1 was more disordered in the apo structure than in the binary complex, whereas the C-terminal was more stable in the apo system. The levels of rigidity in loop 2 and loop 3 are also different.

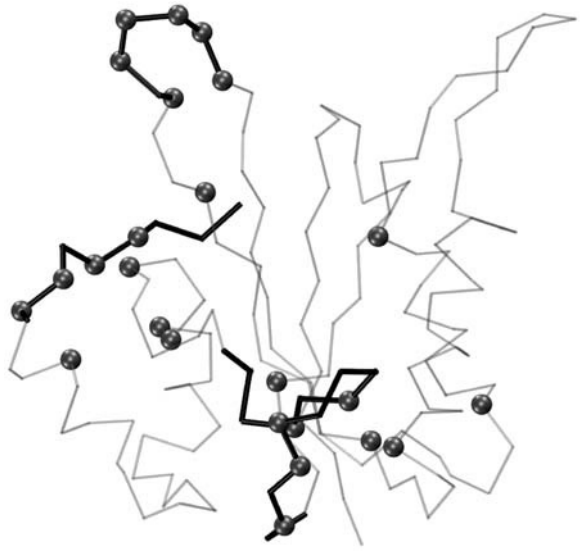

FIGURE 5 Wire-frame representation of the HPPK backbone. The highlighted areas indicate regions of high NOE violation during simulation.

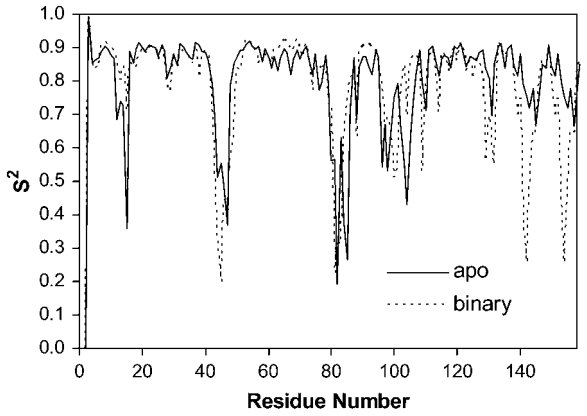

FIGURE 6 Comparison of NMR order parameters calculated from MD trajectories.

Compared to the apo simulation, the NMR order parameters in the binary complex simulation were ~0.3 lower in loop 2 and 0.5 higher in loop 3. We also analyzed the relative motions between the residues, which confirmed the observations by both B-factor and order parameter analyses (data not shown).

The NMR order parameters calculated from the MD trajectories and the experimental B-factors were also compared. They showed the expected anticorrelation with R values equal to -0.71 for the apo-HPPK system and -0.69 for the binary-HPPK system. The inherent correlation between these two measures is expected since both contain the dynamic information of the system. However, S^2 measures directly the dynamics in solution state, whereas B-factors measure disorder/fluctuation in crystalline environment, which includes crystal packing. Thus, discrepancies between the B-factors and NMR order parameters are also expected. Moreover, direct comparisons can be made between simulation and experiments on the S^2 that is more difficult for the B-factors.

Loop dynamics

In this section, we examine the dynamics of apo-HPPK and its binary complex, paying special attention to the three surface loops. To facilitate our analysis of the dynamics, we applied a method known as essential dynamics analysis (28). By comparing the major modes of motion observed in HPPK apo and binary complex states, we ascertained the general characteristics of HPPK's conformational dynamics. In the conventional MD trajectory of both apo-HPPK and its binary complex, loop 3, through its twisting and turning motions, exhibited a relatively large-scale movement away from its crystallographic positions. At the end of 5 ns, loop 3 in the apo simulation began to show signs of opening up, while loop 3 in the binary complex simulation opened up even further compared to the crystal structures. This general trend of loop opening was also observed in loop 2.

At room temperature, the barrier-crossing ability of MD is limited by the simulation time. A typical nanosecond simulation may only sample the conformations in the general ensemble around the initial structure. Although loop dynamics can, in principle, be explored by MD, in reality, reversible conformational changes of protein loops can take a much longer time than is presently practical. In the limited sampling time accessible to MD (5 ns), one cannot expect the loop to cross the energy barrier between the closed and open conformations. To fully explore the conformations accessible to the loops, we applied the LES method to circumvent the sampling limitation. Note that we only applied LES to loop 2 and loop 3; loop 1 was excluded because it is less flexible than the other two loops and its dynamics was adequately sampled by the conventional MD.

Essential dynamics analysis

This method has been applied in several previously published works by other authors (32–35) and has been shown to be an effective method for analyzing concerted motions in proteins. The central idea of this method is to calculate the correlated motions and recast the MD trajectory from the normal Cartesian space into the correlated essential space, unique to macromolecules and presumably containing motions that are essential for biological function (28). In effect, one can think of ED analysis as a motion filter that filters out the high-frequency, uncorrelated motions to expose the underlying modes of low-frequency, concerted motions. Since one of our primary objectives for this study is to understand the large-scale concerted motions in HPPK, this method of analysis is ideally suited to our purpose.

We used all C_α atoms to define the backbone conformation of HPPK for the ED analysis. This resulted in 468 essential modes from the 158 C_α atoms in the form of 468 eigenvectors and their associated eigenvalues. The eigenvectors are the directional vectors that point in the direction of each essential mode of motion and the associated eigenvalues are the amplitudes of motion. Previous studies have shown that $<10\%$ eigenvectors typically capture $>90\%$ of the protein's motion (28,36,37), hence, it is sufficient to analyze only the first few modes of essential motion.

In Fig. 7, we visualize the results of the ED analysis by projecting the MD trajectory onto the first three eigenvectors and superimposing the minimum and maximum structures delimiting the extent of the motion. From this figure, one can clearly see that most of the motions are concentrated in the loop regions in both the apo-HPPK and the binary complex.

In the first mode, loop 2 and loop 3 in the apo structure has a correlated motion moving from right (*cyan*) to left (*red*) in a loop-bending motion (Fig. 7 *A*). The first essential mode of the binary complex is also loop motion, although in this case the motion of loop 2 is a novel ''twisting'' motion, whereas loop 3 still moves in a typical loop-bending motion (Fig. 7

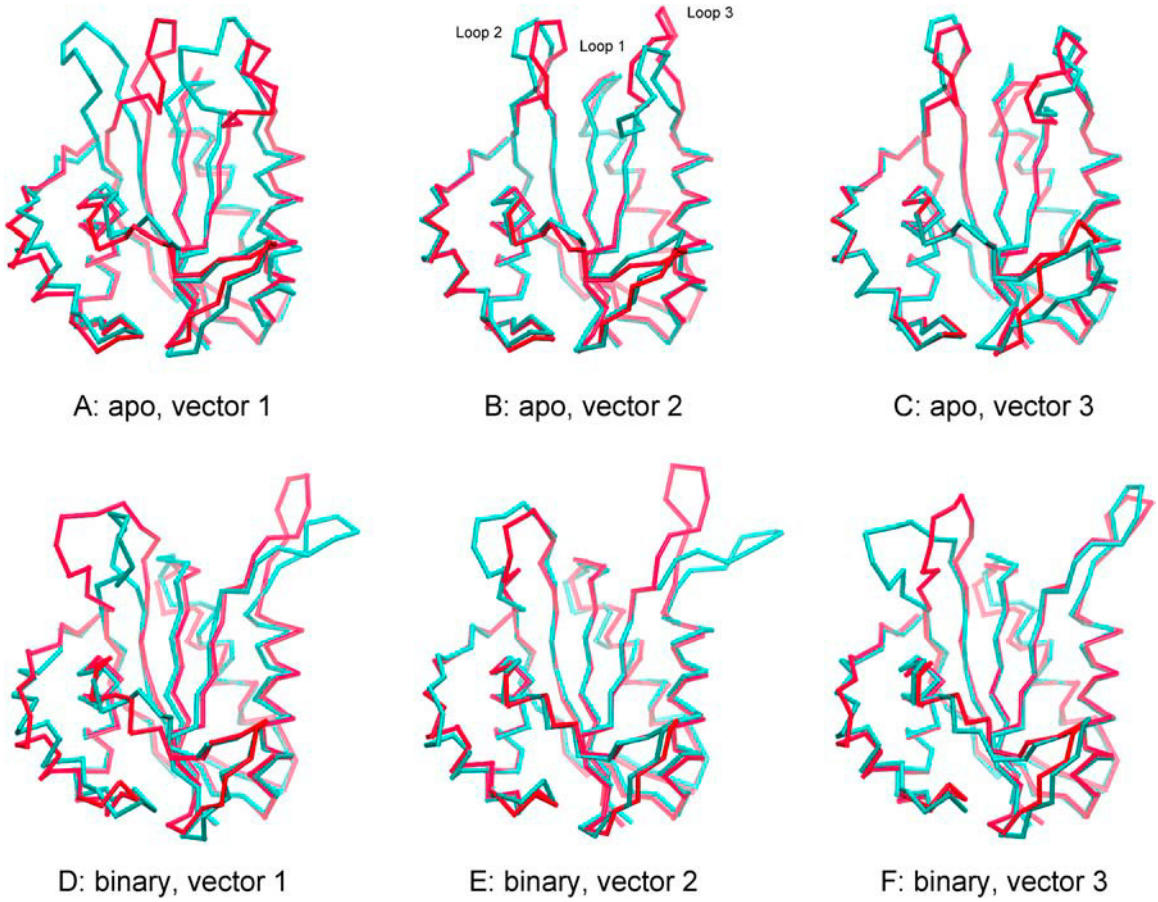

FIGURE 7 First three eigen modes from the simulations are illustrated as minimum (*cyan*) and maximum (*red*) structures. The first three eigen modes of the apo simulation are shown in *A*, *B*, and *C*, and those of the binary are in *E*, *F*, and *G*.

D). In the second essential mode, the motion of loop 2 and loop 3 in the apo structure is that of shear motion (Fig. 7 *B*). In the binary complex, loop 2 and loop 3 show an anti-correlated motion similar to a double-gated door swinging in-and-out (Fig. 7 *E*). In the third essential mode, all three loops in the apo structure move in a correlated fashion from right to left, whereas the binary complex shows a dramatic bending motion in loop 2 (Fig. 7, *C* and *D*).

Thus, at its resting state, the major internal motion of HPPK is basically loop 2 and loop 3 swinging in concert from left to right and right to left, with loop 1 sandwiched in between. Upon ATP binding, loop 1 moves toward the backside of the protein from in between loop 2 and loop 3, loop 3 extends away from the active site to assume an ''open'' conformation, and loop 2 and loop 3 swing back and forth like two panels of a gate. These motions are in nanosecond timescales, as seen from the simulation.

LES simulations

In our 5-ns conventional MD trajectory, we observed that loop 3 in the apo state showed signs of moving toward the open conformation seen in the binary complex even in the absence of substrate. To examine this further, we augmented our MD simulations with LES, a sampling technique that has been shown to be effective in facilitating conformational changes through a mean field approach (15,18). We conducted 4-ns LES simulations for each of the apo and binary complex models using the final structure obtained from MD simulations as the starting coordinates.

In the apo LES simulation, a dramatic conformational transition of loop 3 occurred at 200 ps of simulation time. The starting loop conformation at the beginning of LES simulation was ~14.8 Å away from the crystal structure (Fig. 8) and was in a closed form. The loop underwent substantial conformational change during the LES simulation. At the end of LES simulation, loop 3 reached an open conformation that was even more open than the x-ray structure of the binary complex. The C_α RMSD of the LES copies from the average structure of each snapshot converged to ~0.3 Å. The convergence of loop 2 also reached the same level at around ~0.3 Å, but no major conformation transition was observed.

It is interesting to note the major events of conformational transitions found in the LES simulations. In loop 3, the transition began with loop 3 bending over to the open

conformation, coupled by movement of the conserved side chain of Arg^{92}. As the loop opened up, the enlarged gap enabled the side chain of Arg^{92} to move closer to the active site without clashing into the Glu^{77} side chain. These residues that exhibited large-scale movement are in the region of Glu^{77}–Leu^{96} where HSQC spectrum showed no or missing crosspeaks in the apo-HPPK NMR study, indicating that the large-scale motion observed in the simulation is consistent with NMR data.

In the binary LES simulation, loop 2 and loop 3 both underwent conformational changes. The new conformations adopted in the final structures of the LES simulation are more open than those of the crystal structure. Furthermore, we found in this trajectory that several side chains showed considerable divergence among the copies. This divergence was a reflection of the enhanced speed of conformation sampling and an indication of elevated flexibility. As shown in the NMR ensemble, the Arg^{92} side chain is flexible and has several different conformations in the PDB. In the conventional MD trajectory, limited by the simulation length, we only observed a single conformational transition of the Arg^{92} side chain. In the binary LES simulation, the full ensemble of Arg^{92} conformations was sampled and better agreement with NMR data was observed.

In the crystal structure, residues 12–15 of loop 1 have two distinct conformations with an occupancy ratio of 54:46. In the conventional MD simulation, we started from the more populated conformation and observed transitions between these two conformations. In the LES simulation, more frequent transitions were observed and the ratio was 0.6:0.4, close to that observed experimentally.

Loop conformation change measured by interresidual distance

One experimental approach that may be utilized to quantify the magnitude of loop 3 movement is the fluorescence resonance energy transfer. By attaching fluorescent labels to Ala^{86}, Tyr^{107}, and Ala^{151}, one may treat the two pairs of distances among these three landmark residues as a measure of the horizontal and vertical movement of loop 3 relative to the rest of the molecule. These experiments are currently underway (preliminary data not shown). To parallel the experimental measurements, we directly calculated these two pairs of distances from our MD trajectories. The values obtained from MD trajectories may be regarded as predictions of the MD simulations until the experimental measures are unequivocally established to provide a direct comparison.

The changes in these two distances over time in both the apo and binary systems are shown in Fig. 9. The initial 5 ns represent data from the normal MD simulation. The data from the last 4 ns were calculated from the LES simulation. Since Ala^{86} was in the LES region, the relative distance was calculated for each LES copy and the average distance was used in the plot. It is clearly shown in Fig. 9 *A* that within the first 5-ns MD run, the C_α distances between Ala^{86} and Tyr^{107} in the apo and binary systems were distinct,l whereas in the 4-ns LES simulation, the plots of apo and binary systems, when overlain, converged at ~39 Å. A similar trend of convergence was also observed for the C_α distance between Ala^{86} and Ala^{151} (Fig. 9 *B*). This observation further indicates that in the apo system loop 3 shared conformational similarity with loop 3 in the binary system.

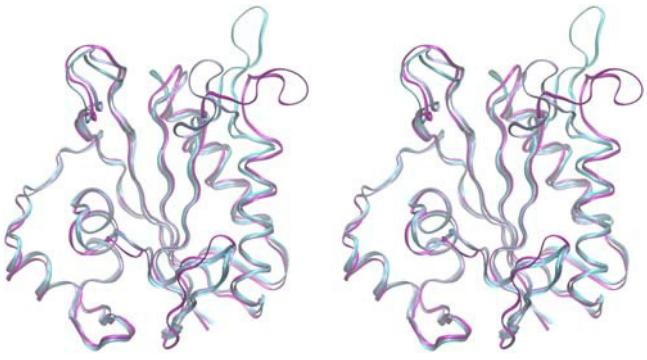

FIGURE 8 Ribbon stereo view of simulated and experimental structures. The structure at the end of LES simulation (*magenta*) is compared to the apo (*silver*) and the binary complex (*cyan*) experimental structures.

HPPK-substrate interactions

Finally, we turn our attention to the interactions between ATP and HPPK. In our simulation of the binary complex, we used the NMR-determined coordinates as our template. The substrate used in the NMR is the ATP analog AMPCPP, but its position is not revealed in the NMR structure. As we described in the Computational Method section, we placed ATP in the binding pocket by superimposing ATP with AMPCPP, found in the ternary-complex structure (PDB 1eqo now superseded by 1q0n). The binding-site side-chain arrangement in the ternary complex is very similar to the NMR structure, with a few notable exceptions: the orientation of the Arg^{92} side chain in the 20 NMR binary structures is not very well defined but Arg^{92} is found to coordinate with the oxygen atoms of the AMPCPP α-phosphate in the ternary complex. After 5 ns of MD, Arg^{92} reoriented to point in the direction of ATP and formed coordinations with the oxygen atoms of the ATP α-phosphate (Fig. 10). Thus, we believe that the ATP binding-site conformation generated from our MD and LES simulations should closely resemble the true binding-site conformation.

To get a better sense of how the binding site reorganized itself upon binding of ATP, we compared the side-chain orientations of the apo-enzyme and the binary complex. Among the residues that interacted with ATP, Arg^{92} showed the most dramatic conformational change. In the apo crystal structure, Arg^{92} points away from the active site, but in our MD and LES simulations of the apo-enzyme, Arg^{92} showed

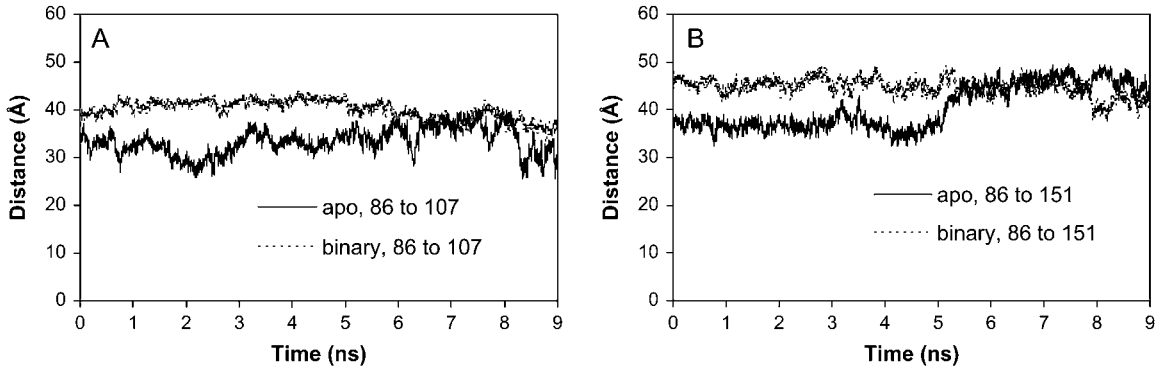

FIGURE 9 Interresidual C_α distances over time. (*A*) C_α distance between residues Ala^{86} and Tyr^{107}. (*B*) C_α distance between residues Ala^{86} and Ala^{151}. The first5 ns represent the initial normal MD productive run and the last 4 ns is the LES run.

considerable flexibility and changed orientation often. In the binary complex simulation, the coordination between Arg^{92} and ATP formed within 1 ns of MD simulation and was very stable throughout the trajectory.

Two other conserved residues that underwent major conformational changes upon ATP binding were Arg^{82} and Arg^{84}. Arg^{82} is a strictly conserved residue that has been shown to be critical for catalysis (4). In the apo structure, the side chain of this residue actually points to the backside of the enzyme, which makes it physically difficult for it to interact with the substrates. In both the MD and LES simulations of the apo-enzyme, this residue never changed orientation to point to the active site. But in the binary complex, Arg^{82} was found to point to the active site before simulation started and stayed in this orientation throughout the simulations.

Residue Arg^{84} is less strictly conserved and does not appear to be involved in catalysis or binding (6). The change of its side-chain position relative to the rest of the protein is a direct consequence of loop 3′s movement and, therefore, we would not comment any further.

DISCUSSION

The global molecular motion of HPPK was previously investigated by Keskin et al. using a coarse-grain method (38). In this method, the protein was represented by a harmonic network of residues restrained to the experimental structures. The results of that study predicted that the palmlike structure of HPPK was nearly rigid, and that loop 2 and loop 3 exhibited the most concerted motions for ligand recognition and, presumably, catalysis. It was observed that in the apo form, loop 2, loop 3, and the loop-β_5-loop regions (residues 102–111) showed significant flexibility, with high B-factors. These observations are in excellent agreement with what we observed in our MD simulation of the apo-HPPK. Interestingly, two of the dominant modes found in the coarse-grain model (see Fig. 5 of Keskin et al., motion *b* and *c*) correspond to the first and second eigen modes revealed in our essential dynamics analysis. This is not surprising, since the coarse-grain model was based on harmonic restraints and the essential dynamics also extracts harmonic motion from the trajectory. Thus, when the primary motion is (near) harmonic, these two analyses should agree.

However, there are several major differences between the MD results and the coarse-grain model. One such notable difference is that the two distinct conformations of loop 1 in the crystal structures were observed in our simulation to undergo a conformational switching dynamics, whereas this conformational switching was not observed in the coarse-grain model. Thus, when anharmonic motion plays an important role, such as in the functionally significant conformational changes, harmonic analysis can potentially fail and detailed modeling becomes necessary.

In this article, we studied the dynamics of the three flexible loops in HPPK. Our goal was to understand the ''unusual conformation change'' revealed by x-ray crystallography and NMR study of HPPK upon ATP binding (7). Although the conformational differences seen in the experimental structures suggested an induced fit, the crystal-packing interactions around the loops might cause them to behave differently when compared to the free enzyme in solution. By simulating the structural dynamics of the apo-enzyme and the binary complex in solution, we showed that the crystal-packing interactions did substantially influence the loop dynamics. Furthermore, we found that the range of loop motions accessible to the apo-enzyme substantially overlapped with that of the binary complex and that the favored loop conformations in both cases are also very similar.

Since we have shown that loop 3 and loop 2 in the apo structure can both open up in the absence of ATP, the mechanism of binding falls into the category of selected fit in which the conformation is readily accessible and binding serves to stabilize the flexible (or disordered) regions. In our current understanding, the effect of ATP binding is mostly to add stability as opposed to trigger major conformational changes. This appears also true for the crystal structures in

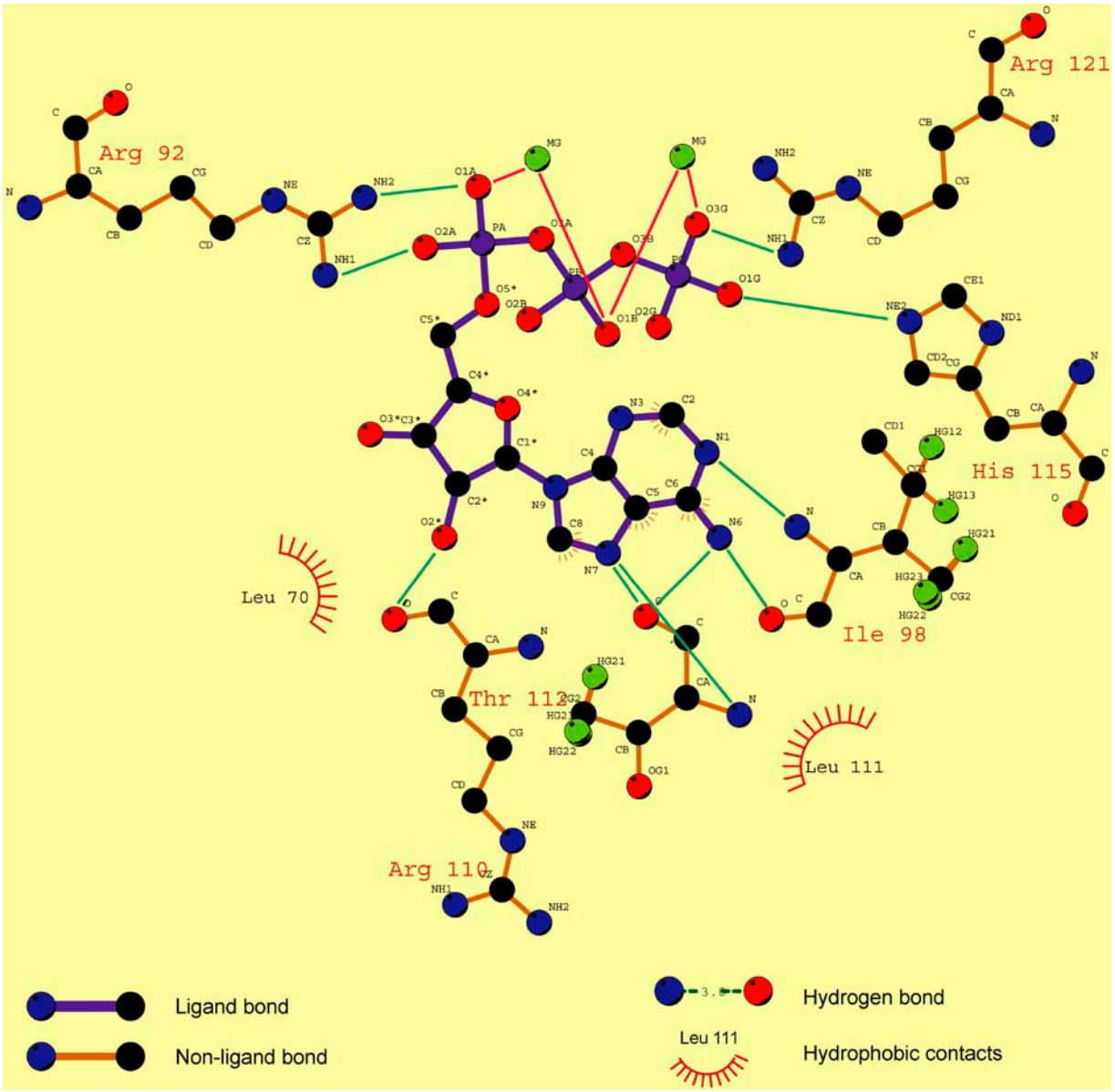

FIGURE 10 Schematic drawing of the interactions between ATP and side chains in the binding pocket. The figure was generated using the software LIGPLOT (39).

which the crystal contacts stabilize the loops but to a conformation that is different from that in the complex. These two conformations appear to fall into the same general free-energy basin, allowing the loops to demonstrate a bimodular behavior manifested by the two structures. We hypothesize that the transition of HPPK from apo-enzyme to the binary and ternary complex might unfold in the following sequence of events:

1. In the apo state, the loops of HPPK undergo diffusive thermal motion and readily interconvert among a wide range of conformations.
2. Upon binding of ATP and the associated 2 Mg^{2+} ions, residue Arg^{92} on loop 3 forms a stable coordination interaction with ATP's phosphate oxygen atoms. This interaction stabilizes loop 3 in the open conformation. The presence of ATP also brings a second catalytically important residue on this loop, Arg^{82}, into the binding site, where it can interact with the substrates in subsequent steps (before this, the side chain of Arg^{82} is situated outside the binding pocket). These two residues are essential for catalytic transfer of pyrophosphate. The reorganization of the binding pocket primes HPPK for binding of HP.
3. HP binds to the HPPK·MgATP complex and triggers a conformational change in loop 2 and loop 3 to envelop the reaction center.

The above scenario can be best summarized as a selected-fit mechanism. One implication for this revised model is that when pursuing structure-based drug design, the open conformations of HPPK apo-enzyme could also be a sensible target.

Insightful comments of the reviewers and use of Visual Molecular Dynamics, Pymol, and SwissPDB Viewer are gratefully acknowledged.

This work has been supported by research grants from the National Instistutes of Health (GM64458 and GM67168 to Y.D. and GM51901 to H.Y.).

REFERENCES

1. Talarico, T. L., I. K. Dev, W. S. Dallas, R. Ferone, and P. H. Ray. 1991. Purification and partial characterization of 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase and 7,8-dihydropteroate synthase from Escherichia coli MC4100. *J. Bacteriol.* 173:7029–7032.
2. Brown, G. M. 1971. The biosynthesis of pteridines. *Adv. Enzymol. Relat. Areas Mol. Biol.* 35:35–77.
3. Blaszczyk, J., Y. Li, Y. Wu, G. Shi, X. Ji, and H. Yan. 2004. Essential roles of a dynamic loop in the catalysis of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Biochemistry.* 43:1469–1477.
4. Li, Y., Y. Wu, J. Blaszczyk, X. Ji, and H. Yan. 2003. Catalytic roles of arginine residues 82 and 92 of Escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: site-directed mutagenesis and biochemical studies. *Biochemistry* 42:1581–1588.
5. Blaszczyk, J., Y. Li, G. Shi, H. Yan, and X. Ji. 2003. Dynamic roles of arginine residues 82 and 92 of Escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: crystallographic studies. *Biochemistry* 42:1573–1580.
6. Yan, H., J. Blaszczyk, B. Xiao, G. Shi, and X. Ji. 2001. Structure and dynamics of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *J. Mol. Graph. Model.* 19:70–77.
7. Xiao, B., G. Shi, J. Gao, J. Blaszczyk, Q. Liu, X. Ji, and H. Yan. 2001. Unusual conformational changes in 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as revealed by X-ray crystallography and NMR. *J. Biol. Chem.* 276:40274–40281.
8. Shi, G., J. Blaszczyk, X. Ji, and H. Yan. 2001. Bisubstrate analogue inhibitors of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: synthesis and biochemical and crystallographic studies. *J. Med. Chem.* 44:1364–1371.
9. Blaszczyk, J., G. Shi, H. Yan, and X. Ji. 2000. Catalytic center assembly of HPPK as revealed by the crystal structure of a ternary complex at 1.25 Å resolution. *Struct. Fold. Des.* 8:1049–1058.
10. Xiao, B., G. Shi, X. Chen, H. Yan, and X. Ji. 1999. Crystal structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, a potential target for the development of novel antimicrobial agents. *Struct. Fold. Des.* 7:489–496.
11. Stammers, D. K., A. Achari, D. O. Somers, P. K. Bryant, J. Rosemond, D. L. Scott, and J. N. Champness. 1999. 2.0 A X-ray structure of the ternary complex of 7,8-dihydro-6-hydroxymethylpterinpyrophosphokinase from Escherichia coli with ATP and a substrate analogue. *FEBS Lett.* 456:49–53.
12. Hennig, M., G. E. Dale, A. D'Arcy, F. Danel, S. Fischer, C. P. Gray, S. Jolidon, F. Muller, M. G. Page, P. Pattison, and C. Oefner. 1999. The structure and function of the 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from Haemophilus influenzae. *J. Mol. Biol.* 287: 211–219.
13. Blaszczyk, J., G. Shi, Y. Li, H. Yan, and X. Ji. 2004. Reaction trajectory of pyrophosphoryl transfer catalyzed by 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Structure* (*Camb.*). 12:467–475.
14. Shi, G., J. Gao, and H. Yan. 1999. 1H, 13C and 15N resonance assignments of Escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase and its complex with MgAMPPCP. *J. Biomol. NMR.* 14:189–190.
15. Elber, R., and M. Karplus. 1990. Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* 112:9161–9175.
16. Roitberg, A., and R. Elber. 1991. Modeling side-chains in peptides and proteins: application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *J. Chem. Phys.* 95:9277–9287.
17. Simmerling, C., and R. Elber. 1994. Hydrophobic ''collapse'' in a cyclic hexapeptide: computer simulations of CHDLFC and CAAAAC in water. *J. Am. Chem. Soc.* 116:2534–2547.
18. Hornak, V., and C. Simmerling. 2003. Generation of accurate protein loop conformations through low-barrier molecular dynamics. *Proteins.* 51:577–590.
19. Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, J. Wang, W. S. Ross, C. Simmerling, T. Darden, K. M. Merz, R. V. Stanton, A. Cheng, J. J. Vincent, M. Crowley, V. Tsui, H. Golhke, R. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. Weiner, and P. A. Kollman. 2002. AMBER 7. University of San Francisco, San Francisco.
20. Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Goulg, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.
21. Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24:1999–2012.
22. Tomasi, J., B. Mennucci, and E. Cances. 1999. The IEF version of the PCM solvation method: an overview of a new method addressed to study molecular solutes at the QM ab initio level. *J. Mol. Struct. Theochem.* 464:211–226.
23. Pomelli, C. S., J. Tomasi, and V. Barone. 2001. An improved iterative solution to solve the electrostatic problem in the polarizable continuum model. *Theor. Chem. Acc.* 105:446–451.
24. Jorgensen, W. L., J. Chandrasekhar, J. Madura, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
25. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
26. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
27. Ryckaert, J. P., G. Ciccotti, and H. J. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23: 327–341.
28. Amadei, A., A. B. Linssen, and H. J. Berendsen. 1993. Essential dynamics of proteins. *Proteins.* 17:412–425.
29. Harrison, R. J., J. A. Nichols, T. P. Straatsma, M. Dupuis, E. J. Bylaska, G. I. Fann, T. L. Windus, E. Apra, J. Anchell, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, B. de Jong, M. Deegan, K. Dyall, D. Elwood, H. Fruchtl, E. Glendenning, M. Gutowski, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kendall, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, J. Nieplocha, S. Niu, M. Rosing, G. Sandrone, M. Stave, H. Taylor, G. Thomas, J. van Lenthe, K. Wolinski, A. Wong, and Z. Zhang. 2000. NWChem, A Computational Chemistry Package for Parallel Computers. Pacific Northwest National Laboratory, Richland, WA.

30. Smith, L. J., A. E. Mark, C. M. Dobson, and W. F. van Gunsteren. 1995. Comparison of MD simulations and NMR experiments for hen lysozyme. Analysis of local fluctuations, cooperative motions, and global changes. *Biochemistry* 34:10918–10931.

31. Stocker, U., and W. F. van Gunsteren. 2000. Molecular dynamics simulation of hen egg white lysozyme: a test of the GROMOS96 force field against nuclear magnetic resonance data. *Proteins*. 40:145–153.

32. Daidone, I., A. Amadei, D. Roccatano, and A. D. Nola. 2003. Molecular dynamics simulation of protein folding by essential dynamics sampling: folding landscape of horse heart cytochrome *c*. *Biophys. J*. 85:2865–2871.

33. Crabbe, M. J., L. R. Cooper, and D. W. Corne. 2003. Use of essential and molecular dynamics to study gammaB-crystallin unfolding after non-enzymic post-translational modifications. *Comput. Biol. Chem.* 27:507–510.

34. de Groot, B. L., X. Daura, A. E. Mark, and H. Grubmuller. 2001. Essential dynamics of reversible peptide folding: memory-free conformational dynamics governed by internal hydrogen bonds. *J. Mol. Biol.* 309:299–313.

35. Arcangeli, C., A. R. Bizzarri, and S. Cannistraro. 2001. Concerted motions in copper plastocyanin and azurin: an essential dynamics study. *Biophys. Chem.* 90:45–56.

36. Duan, Y., P. Wilkosz, and J. M. Rosenberg. 1996. Dynamic contributions to the DNA binding entropy of the *Eco* RI and *Eco* RV restriction endonucleases. *J. Mol. Biol.* 264:546–555.

37. Kazmierkiewicz, R., C. Czaplewski, B. Lammek, and J. Ciarkowski. 1999. Essential dynamics/factor analysis for the interpretation of molecular dynamics trajectories. *J. Comput. Aided Mol. Des.* 13:21–33.

38. Keskin, O., X. Ji, J. Blaszcyk, and D. G. Covell. 2002. Molecular motions and conformational changes of HPPK. *Proteins*. 49:191–205.

39. Wallace, A. C., R. A. Laskowski, and J. M. Thornton. 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*. 8:127–134.