

ROLE OF PROTEIN DYNAMICS IN THE CATALYSIS BY 6-
HYDROXYLMETHYL-7,8-DIHYDROPTERIN
PYROPHOSPHOKINASE: A COMBINED NMR AND MOLECULAR
DYNAMICS SIMULATION STUDY

By

Zhenwei Lu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Biochemistry and Molecular Biology

2011

Abstract

ROLE OF PROTEIN DYNAMICS IN THE CATALYSIS BY 6-HYDROXYLMETHYL-7,8-DIHYDROPTERIN PYROPHOSPHOKINASE: A COMBINED NMR AND MOLECULAR DYNAMICS SIMULATION STUDY

By

Zhenwei Lu

Proteins are intrinsically dynamic and it is believed that the dynamics has functional roles. While it is still a debate whether the protein dynamics is important for the chemical step of enzymatic function, there is an increasing consensus that protein dynamics plays important roles in the substrate binding and product release.

6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) catalyzes the pyrophosphate transfer from ATP to 6-hydroxymethyl-7,8-dihydropterin (HP), a key step in folate biosynthetic pathway. Evidence has indicated that HPPK dynamics is important for its function. Atomic structures have been determined for nearly every stage of its catalytic cycle. Comparison of those structures clearly shows that HPPK goes through dramatic conformational changes during the catalytic cycle, especially in the three catalytic loop regions.

In this thesis, the HPPK dynamics along the first half of the catalytic cycle, substrate binding, was studied using Nuclear Magnetic Resonance (NMR) and molecular dynamics (MD) simulation. The importance of protein dynamics was addressed by studies on two mutant proteins: Q50A and N10A HPPK. Q50 and N10 are key residues in the hydrogen-bond network found in the x-ray crystal structure of HPPK ternary complex, which couple the three catalytic loops together.

My results show that ligand-free (apo) HPPK is highly dynamic on a timescales ranging from picosecond to second, and binding of the first substrate does not reduce the internal dynamics but rather enhances it moderately, especially in the catalytic loop region. HPPK dynamics is largely quenched upon the binding of the second substrate, however, some mobility remains. The remaining dynamics may help the optimization of the active site interaction. Both N10 and Q50 are important for connecting the three catalytic loops and the loop coupling is important for the binding of the second substrate HP and the full assembling and stabilization of the active center and catalysis. MD simulation studies show that the active site residues pre-sample the side-chain conformation for substrate binding even without substrates, indicating that HPPK follows a selected-fit mechanism. One possible HPPK conformational transition pathway during the substrate binding is identified through targeted MD simulation.

ACKNOWLEDGEMENT

First and foremost, I would like to thank my advisor Dr. Honggao Yan for providing me this great project and always being available whenever needed: both in my thesis research and in my personal life.

Secondly, I would like to thank all members of my guidance committee, Dr. Aizhuo Liu, Dr. Michael Feig, Dr. Robert Cukier, Dr. Robert P. Hausinger, and Dr. Zachary F. Burton for their kind guidance and critical review of my thesis.

Thirdly, I would like to thank Dr. Yue Li and Yan Wu for teaching me molecular biology and protein purification techniques. Thanks also go to Dr. Daniel Holmes and Dr. Kermit Johnson for training and help on the operation of NMR, Dr. Kaillathe Padmanabhan for help on computation, and Mr. Joseph Leykam for training on the analytical ultracentrifuge. I would also like to thank Dr. Lishan, Yao, Dr. Jaroslaw Blaszczyk, Dr. Yi Wang, Dr. Bing Chen, Dr. Hongfeng Lou, Dr. Li Su, Jifeng Wang, Rahul Banerjee and Yichen Tang. It is a great pleasure to work with them. And luckily enough to me, they are my friends now.

Finally, I would like to thank my wife Shujuan, my parents, and my friends for all their love and support. I could not get anywhere without them.

TABLE OF CONTENTS

List of Tables	vi
List of Figures	vii
List of Abbreviations.....	viii
CHAPTER 1: Introduction	1
Enzyme catalysis and dynamics.....	2
The timescales of protein dynamics.....	3
NMR method to study protein dynamics.....	3
NMR study of picosecond to nanosecond timescale dynamics.....	4
NMR study of microsecond to millisecond timescale dynamics.....	8
Molecular dynamics simulation.....	10
Literature review of HPPK.....	13
Figures.....	21
References.....	27
CHAPTER 2: Dynamics of the Conformational Transitions in the Assembling of the Michaelis Complex of a Bi-substrate Enzyme: A ¹⁵ N Relaxation Study of <i>Escherichia coli</i> 6-Hydroxymethyl-7,8-dihydropterin Pyrophosphokinase.....	34
Abstract.....	35
Introduction.....	36
Materials and methods.....	39
Results.....	44
Discussion.....	53
Figures.....	61
References.....	72
CHAPTER 3: Nuclear Magnetic Resonance and Molecular Dynamics studies of the N10A and Q50A HPPKs.....	78
Introduction.....	79
Materials and methods.....	81
Results.....	88
Discussion.....	96
Tables and figures.....	100
References.....	116
CHAPTER 4: Insight into <i>E. coli</i> HPPK Substrate Binding Mechanism and Conformational Transition Pathway: Molecular Dynamics Simulation Studies.....	120
Introduction.....	121
Methods.....	122
Results.....	125
Discussion.....	129

Figures.....	131
References.....	141
CHAPTER 5: Conclusion and Perspective	143

LIST OF TABLES

3.1 Thermodynamic and kinetic parameters of WT, Q50A and N10A HPPK.....	100
3.2 Hydrogen bond between the three catalytic loops of WT, Q50A and N10A HPPK during the MD simulations.....	101
3.3 $^3J_{\text{HNHA}}$ coupling constants (Hz) of loop residues of the WT, the Q50A, and the N10A HPPK ternary complex	102

LIST OF FIGURES

1.1. Functional roles of protein conformational changes in enzymatic catalysis.....	21
1.2 Kinetic mechanism of HPPK	22
1.3 Crystal structure of the apo HPPK	23
1.4 protein•HP (a) and protein•MgAMPCPP (b) interactions	24
1.5 HPPK reaction trajectory	26
2.1 The HPPK-catalyzed reaction (a) and a ribbon diagram of the crystal structures of HPPK showing the conformational changes of the three catalytic loops during several stages of its catalytic cycle (b)	61
2.2 The ^1H - ^{15}N HSQC spectra of HPPK in the apo state (black) and in complex with MgAMPCPP (red)	63
2.3 ^{15}N relaxation parameters obtained for the apo HPPK at 25 $^{\circ}\text{C}$	64
2.4 ^{15}N relaxation parameters obtained for the binary complex HPPK•MgAMPCPP at 25 $^{\circ}\text{C}$	65
2.5 ^{15}N relaxation parameters obtained for the ternary complex HPPK•MgAMPCPP•DMHP at 25 $^{\circ}\text{C}$	66
2.6 Comparison of the model-free parameters versus residue number between the apo HPPK (red) and the binary complex with MgAMPCPP (blue) at 0.5 mM protein concentration.....	67
2.7 Comparison of the model-free parameters versus residue number between the binary complex HPPK•MgAMPCPP (blue) and the ternary complex HPPK•MgAMPCPP•DMHP (black) at 0.5 mM protein concentration.....	69
2.8 Mapping of the dynamic parameters S^2 (left panel) and R_{ex} (right panel) to the structures of the apo HPPK (top), the binary complex (HPPK•MgAMPCPP, middle), and the ternary complex (HPPK•MgAMPCPP•DMHP, bottom).....	70
3.1 The crystal structure of the apo HPPK and representative structure of the HPPK substrate ternary complex.....	103

3.2 Schematic illustration of the coupling of three flexible loops through a hydrogen bond network in the HPPK ternary complex.....	104
3.3 Overlay of the ^1H - ^{15}N HSQC spectra of the wild-type and Q50A HPPK mutant enzymes.....	105
3.4 Overlay of the ^1H - ^{15}N HSQC spectra of the wild-type and N10A HPPK mutant enzymes.....	107
3.5 Chemical shift mapping for HPPK mutant enzymes.....	109
3.6 Correlation of $^3J_{\text{HNHA}}$ coupling constants between mutant and WT HPPK ternary complexes.....	111
3.7 Comparison of ^{15}N relaxation parameters versus residue number between WT, Q50A, and N10A HPPK ternary complexes.....	112
3.8 Comparison of the model-free parameters versus residue number between WT, Q50A and the N10A HPPK ternary complex.....	113
3.9 Correlation of the order parameter (S^2) from NMR spectroscopy and MD simulation.....	114
3.10 Comparison of RMSFs of WT (blue), N10A (red) and Q50A HPPK enzymes (green).....	115
4.1 The structures of the apo form, binary complex, and ternary complex of HPPK.....	131
4.2: C α RMSDs of the apo HPPK (black), the HPPK•MgATP complex (red), and the HPPK•MgATP•HP complex (blue).....	132
4.3 Side-chain conformations sampled during the MD simulations of the apo HPPK (black), the HPPK•MgATP complex (red), and the HPPK•MgATP•HP complex (green).....	133
4.4 The involvement coefficients (thin bar; left axis) and the cumulative involvement coefficients (thick line, right y-axis) of the first 50 PCs.....	134
4.5 The characteristic motions and reaction coordinates from the apo HPPK to the HPPK•MgATP complex.....	135

4.6 The characteristic motions and reaction coordinates from the HPPK•MgATP to the HPPK•MgATP•HP complex.....	136
4.7 The conformations of the active site residues in the crystal structures of the apo HPPK, the HPPK•MgATP, and the HPPK•MgATP•HP complex.....	137
4.8 The snapshot of the HPPK ^{apo} •Mg ATP complex (left) and the HPPK ^{bin} •Mg ATP•HP complex (right) after 2 ns MD simulations.....	138
4.9 The positioning of R92 and D95 in the HPPK•MgATP binary complex as shown in the targeted MD simulation.....	139
4.10 The positioning of L45, W89 and R92 in the HPPK•Mg ATP•Mg HP complex as shown in the targeted MD simulation	140

LIST OF ABBREVIATIONS

Apo:	ligand free
ADP:	adenosine diphosphate
AMP:	adenosine diphosphate
AMPCPP:	α,β -methyleneadenosine 5'-triphosphate
AMPPCP:	β,γ -methyleneadenosine 5'-triphosphate
ATP:	adenosine triphosphate
BMD	biased molecular dynamics
CPMG:	Carr-Purcell-Meiboom-Gill
DMHP:	7,7-dimethyl-6-hydroxymethylpterin
<i>E. coli</i> :	<i>Escherichia coli</i>
EcHPPK:	<i>Escherichia coli</i> 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase
fs:	femtosecond
HP:	6-hydroxymethyl-7,8-dihydropterin
HPPK:	6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase
HPPP:	6-hydroxymethyl-7,8-dihydropterin pyrophosphate
HSQC:	heteronuclear single-quantum coherence
IPTG:	isopropyl β -D-thiogalactoside
MD:	molecular dynamics
ms:	millisecond
ns:	nanosecond
NMR:	nuclear magnetic resonance
NOE:	nuclear Overhauser effect
ps:	picosecond
R_1 :	spin-lattice relaxation rate constant

R_2 :	spin-spin relaxation rate constant
SMD	steered molecular dynamics
T_1 :	spin-lattice relaxation time constant
T_2 :	spin-spin relaxation time constant
S^2 :	generalized order parameter
τ_c :	correlation time.
TMD:	targeted molecular dynamics
WT:	wild type
PCA:	principal component analys

CHAPTER 1

Introduction

Enzyme Catalysis and Dynamics

Life is impossible without enzymes. The power of enzymes makes it possible for the reactions needed for life to take place much faster to meet the life cycle. Rate enhancement by enzymes could be up to 10^{20} fold (1), which is amazing, considering that rate enhancement by non-enzyme catalysts is in the range of 10^2 - 10^4 fold (2). How enzymes achieve this enormous rate enhancement is an ultimate question attracting biochemists.

It is generally accepted that an enzyme achieves its catalytic power by lowering the activation energy of a reaction and there are conflicting structural requirements as the enzyme goes through its catalytic cycle (Figure 1.1). As a rule of thumb, the active center of the enzyme must be in an open conformation for substrate binding, in a closed conformation for maximal transition state stabilization in the chemical step, and in an open conformation again for product release. Consequently, enzyme dynamics is an important aspect of enzymatic catalysis. Although whether protein dynamics enhances the chemical step is still not fully understood and remains controversial (3-5), there is consensus that protein dynamics plays an important role in substrate binding and product release.

To understand the catalytic mechanism of an enzyme, one has to have both structural and dynamics information of the enzyme. However, the current focus on structural biology is to use static structures from either X-ray crystallography or solution nuclear

magnetic resonance (NMR) spectroscopy to understand the underlying mechanism of a biological function. Although the structures have greatly enhanced our understanding of structure-function relationships of enzymes, their usage in understanding the role of protein dynamics in enzymatic catalysis is limited.

The Timescales of Protein Dynamics

In general, protein motion can be divided into two categories: the transition from one conformation to a different conformation and the dynamic fluctuation within one conformation (6, 7). The transition between different conformations usually occurs on the μ s to second timescale, similar to a timescale for an enzyme to complete its catalytic cycle. Therefore, it is generally believed that the dynamics on this timescale is more relevant to enzyme function. The dynamic fluctuation within one conformation usually occurs on fs to ns timescale. This kind of motion may assist the alignment of the active site residues for maximal transition state stabilization. It is also proposed that the dynamic fluctuation within one conformation might facilitate the conformational transition (8).

NMR Method to Study Protein Dynamics

While many spectroscopic methods can be used to study protein dynamics, NMR, the method used in this thesis, is the most powerful method in term of both spatial and time resolution (9). The power of NMR lies in its atomic resolution, the wide range of

timescale it covers, from ps to days, and the large number of natural probes available in proteins, including hydrogen, carbon and nitrogen atoms. Specific NMR techniques have been developed for characterizing protein motions on different timescales. NMR relaxation measurements can be used to study motions on ps to ns timescales, while NMR relaxation dispersion measurements can be used to study motions on μ s to ms timescales. ZZ exchange measurements can be used to study motions on ms to second timescales. Hydrogen/deuterium exchange measurements can be used to study motions on the timescales of seconds to days. Combining these different NMR measurements, a near-complete characterization of protein motions at the atomic resolution could be obtained.

NMR Study of Picosecond to Nanosecond Timescale Dynamics

Ps to ns timescale protein motions can be studied through NMR relaxation measurements. How do NMR relaxation parameters relate to protein motion on a given timescale? Generally speaking, molecular motion causes time-dependent fluctuations of the magnetic field through two mechanisms: chemical shift anisotropy (CSA) and dipolar coupling interactions. These fluctuations cause spin transitions between different energy levels, which leads to NMR spin relaxation. Studying the relaxation processes can provide information of motions within a molecule at the atomic level.

The nuclear spin most frequently used for NMR relaxation study of protein dynamics is backbone ^{15}N , because its mechanism of relaxation is simpler than those of other

spins. First, backbone ^{15}N relaxation is dominated by CSA of ^{15}N and the dipolar coupling interactions with the attached proton, and the effects of other protons could be considered negligible. Second, the ^{15}N - ^1H distance is fixed by the covalent bond. Thus, fluctuation in the angular term is the main cause of relaxation.

Backbone ^{15}N is used as an example for the following discussion. The relaxation processes can be described through three parameters that can be measured by NMR: spin-lattice relaxation time (T_1), spin-spin relaxation time (T_2), and hetero-nuclear nuclear Overhauser effect (NOE). They are related to the spectral density function through the following equations (10):

$$\frac{1}{T_1} = \frac{d^2}{10} [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + \frac{2}{15} \omega_N^2 \Delta\sigma^2 J(\omega_N) \quad (1.1)$$

$$\begin{aligned} \frac{1}{T_2} = & \frac{d^2}{20} [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] \\ & + \frac{1}{45} \omega_N^2 \Delta\sigma^2 [4J(0) + 3J(\omega_N)] + R_{\text{ex}} \end{aligned} \quad (1.2)$$

$$\text{NOE} = 1 + \frac{\gamma_H}{\gamma_N} d^2 \frac{1}{10} [6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] T_1^N \quad (1.3)$$

Where $d^2 = \gamma_H^2 \gamma_N^2 \hbar^2 / r^6$, $\Delta\sigma$ is the CSA, and the R_{ex} term is a compensation for μs to ms timescale motions. The spectral density function $J(\omega)$ is the Fourier transform of the normalized autocorrelation function $g(\tau)$. It is defined as:

$$J(\omega) = \int_0^\infty g(\tau) e^{-i\omega\tau} d\tau \quad (1.4)$$

For random isotropic tumbling due to Brownian motions:

$$J(\omega) = \frac{\tau_c}{1 + \omega^2 \tau_c^2} \quad (1.5)$$

$$g(\tau) = e^{-\tau/\tau_c} \quad (1.6)$$

where τ_c is the rotational correlation time.

The spectral density function is determined by the nature of the molecular motion. As the complexity of motion increases, the number of parameters to define the spectral density function increases. To simplify this issue, the model-free approach was developed, initially by Lipari and Szabo (11, 12) and extended by Clore and coworkers (13). The details of motion are not considered in this approach. Instead, it is assumed that the overall motion can be separated from fast internal motion, and the internal motion can be described by a generalized order parameter (S^2) and the correlation time (τ_i).

In the extended model-free analysis, the motion is described as (13):

$$J(\omega) = \frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{(S_f^2 - S^2) \tau_m}{1 + \omega^2 \tau_m^2} \quad (1.7)$$

and the correlation function is:

$$g(\tau) = S^2 + (1 - S_f^2) e^{-\tau/\tau_f} + (S_f^2 - S^2) e^{-\tau/\tau_s} \quad (1.8)$$

where $\tau_m = \frac{\tau_c \tau_i}{\tau_c + \tau_i}$, S_f^2 is the amplitude for the faster motion. τ_f and τ_e are the correlation time for fast and slow internal motion, respectively.

In addition to these four parameters, τ_f , τ_e , S^2 and S_f^2 , the extended model-free

analysis also includes an R_{ex} term to account for the chemical exchange. Thus, there are in total five possible parameters to define the internal motion. A subset of these parameters might be sufficient for fitting, considering that only three experimental observables (T_1 , T_2 , and NOE) are measured in a single field NMR relaxation experiment. One widely used approach has five sets of parameters to define the spectral density function for the internal motion: S^2 , (S^2, τ_f) , (S^2, R_{ex}) , $(S^2, R_{\text{ex}}, \tau_f)$, and (S^2, S_f^2, τ_s) (14). The fitting starts from the simplest parameter set then moves onto the more complicated ones, with statistic tests, the χ^2 test and the F-test, to determine which model is appropriate. There are other methods available for model selection, most of which are based on either Akaike's information criteria or Bayesian information criteria (15, 16).

Prior to the model-free analysis, the global correlation time has to be determined from the T_1/T_2 ratio. Since the presence of significant internal motion and chemical exchange distorts the T_1/T_2 ratio, only residues without significant internal motion and chemical exchange are chosen for the calculation. For these residues, the T_1/T_2 ratio is determined by the rotational correlation time. The global correlation time can be obtained by minimizing the difference between the left and right side of the following equation (17):

$$\frac{T_1}{T_2} = \frac{\frac{d^2}{20}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{1}{45}\omega_N^2\Delta\sigma^2[4J(0) + 3J(\omega_N)]}{\frac{d^2}{10}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + \frac{2}{15}\omega_N^2\Delta\sigma^2J(\omega_N)}$$

(1.9)

NMR Study of Microsecond to Millisecond Timescale Dynamics

Motions on the μs to second timescales are related to chemical exchange, as it involves a change in the chemical environment of a spin and leads to a chemical shift change. Consider a simple model for an exchange process:



The rate constant for the conversion from A to B is k_1 and the rate constant for the reverse reaction is k_2 . The overall equilibrium constant $K_{\text{eq}}=k_1/k_2$. The population fraction for each rate is:

$$p_A = \frac{k_2}{k_1 + k_2} = \frac{1}{1 + K_{\text{eq}}} \quad (1.11)$$

$$p_B = \frac{k_1}{k_1 + k_2} = \frac{K_{\text{eq}}}{1 + K_{\text{eq}}} \quad (1.12)$$

The exchange rate between A and B is defined as:

$$k_{\text{ex}} = k_1 + k_2 = \frac{k_1}{p_B} = \frac{k_2}{p_A} \quad (1.13)$$

The frequency difference between A and B is:

$$\Delta\omega = \omega_A - \omega_B \quad (1.14)$$

The exchange process can be categorized based on the relative magnitudes of k_{ex} and $\Delta\omega$:

slow exchange: $k_{\text{ex}} < \Delta\omega$;

intermediate exchange: $k_{\text{ex}} \approx \Delta\omega$;

fast exchange: $k_{\text{ex}} > \Delta\omega$;

Our focus is on the fast exchange process, which for proteins is on the μs to ms timescale. Motions on this timescale are considered to be very important for enzymes, since most enzymatic reactions occur on this timescale. The fast exchange process can be studied using Carr-Purcell-Meiboom-Gill (CPMG) NMR relaxation dispersion method (18).

In CPMG NMR measurements, the observed relaxation rate for all exchange processes is given by the equation below (19):

$$R_2 = \frac{1}{2}[R_{2A} + R_{2B} + k_1 + k_2] - \frac{1}{\tau_{\text{cp}}} \ln \lambda^+ \quad (1.1)$$

$$\begin{aligned} \lambda^+ &= \ln[(D_+ \cosh^2 \xi - D_- \cos^2 \eta)^{1/2} + (D_+ \sinh^2 \xi - D_- \sin^2 \eta)^{1/2}] \\ &= (1/2) \cosh^{-1}[D_+ \cosh 2\xi - D_- \cosh 2\eta] \end{aligned} \quad (1.1)$$

$$\text{where: } D_{\pm} = \frac{1}{2}[\pm 1 \pm \frac{\psi + 2(\Delta\omega)^2}{\sqrt{\psi^2 + \varsigma^2}}]$$

$$\xi = \frac{\tau_{\text{cp}}}{\sqrt{8}}[+\psi + \sqrt{\psi^2 + \varsigma^2}]^{1/2}$$

$$\eta = \frac{\tau_{\text{cp}}}{\sqrt{8}}[-\psi + \sqrt{\psi^2 + \varsigma^2}]^{1/2}$$

$$\varsigma = 2\Delta\omega(R_{2A} - R_{2B} + k_1 - k_2)$$

$$\psi = (R_{2A} - R_{2B} + k_1 - k_2)^2 - (\Delta\omega)^2 + 4k_1 k_2$$

For the fast exchange process, this could be simplified to the following equation

(20):

$$R_2 = \frac{R_2^A + R_2^B}{2} + \frac{p_A p_B (\Delta\omega)^2}{k_{\text{ex}}} \left[1 - \frac{2}{k_{\text{ex}} \tau_{\text{cp}}} \tanh \frac{k_{\text{ex}} \tau_{\text{cp}}}{2} \right] \quad (1.17)$$

In general, the exchange rate could be obtained by fitting the measured R_2 to these equations. In favorable cases in which the population of each state can be obtained, it is possible to get all the information about the chemical exchange, including forward and reverse rate constants and the chemical shift difference between the two exchange species.

Molecular Dynamics Simulation

Molecular dynamics (MD) uses classical Newtonian mechanics to explain the behaviors of a molecular system. Given the potential V and the initial configuration r , MD simulation can provide the precise position of each atom i at any time by solving the Newtonian equation of motion.

$$m \frac{d^2 r_i}{dt^2} = F_i \quad (1.18)$$

$$F_i = -\nabla_i V \quad (1.19)$$

The general form of the total energy in molecular mechanics is:

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{non-bonded}} \quad (1.20)$$

in which the bonded and non-bonded potentials are given by the following equation:

$$E_{\text{bonded}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} \quad (1.21)$$

$$E_{\text{nonbonded}} = E_{\text{electrostatic}} + E_{\text{vanderWaals}} \quad (1.22)$$

The force field used in the Amber molecular dynamics package we used for the MD simulations is:

$$E_{\text{total}} = \sum_{\text{bonds}} k_r (l - l_{\text{eq}})^2 + \sum_{\text{angles}} k_{\theta} (l - l_{\text{eq}})^2 + \sum_{\text{torsions}} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \\ + \sum_{i < j}^N \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon r_{ij}} \right)$$

The first three terms refer to the potential energies of bonded atoms (atoms connected by three or less chemical bonds) and the fourth is for non-bonded interactions. The first term is the potential energy of bond stretching, the second the potential energy of bending of bond angles, the third the potential energy of torsion angles, and the fourth includes both van der Waals and electrostatic interactions.

Given the initial coordinates and velocities, the trajectories of the system can be calculated by integrating Newton's equations of motion. Verlet-type algorithms are the most widely used due to low computational cost and high accuracy. The velocity Verlet algorithm is given below for updating the coordinates and velocities:

$$r(t + \Delta t) = r(t) + v(t)\Delta t + (1/2)a(t)\Delta t^2 \quad (1.24)$$

$$v(t + \Delta t / 2) = v(t) + (1/2)a(t)\Delta t \quad (1.25)$$

$$a(t + \Delta t) = -(1/m)\nabla V(r(t) + \Delta t) \quad (1.26)$$

$$v(t + \Delta t) = v(t + \Delta t / 2) + (1/2)a(t + \Delta t)\Delta t \quad (1.27)$$

The application of all-atom molecular dynamics is limited to sub-microsecond timescale with current algorithms and computer resources. However, μ s to ms timescale dynamics is probably most relevant to protein function, since protein functions mostly on this timescale. Various methods to overcome this issue have been developed. Among these methods, molecular dynamics with external potential to guide the conformation search is particularly useful to identify the possible transition pathways when the conformations of both starting and ending states are known. Three most commonly used such methods are targeted MD (TMD), steered MD (SMD) and biased MD (BMD).

These methods differ in how the progress variable is controlled. TMD applies a time-dependent holonomic geometrical constraint to a target structure. Thus, the conformational transition is enforced independent of the energy barrier (21, 22). SMD applies a harmonic restraint based on a reference point to move the system toward the target in a way similar to atomic force microscopy (23-25). BMD applies a harmonic potential to guide the conformational transition by penalizing the move to the undesired direction (26, 27). Comparison of the paths generated from these three methods has shown that the paths are similar for a given progress variable in most cases but depend strongly on the choice of progress variable (28). In this thesis, we use TMD to study the conformation transitions in the formation of the substrate complexes of 6-hydroxymethyl-7, 8-dihydropterin pyrophosphokinase (HPPK).

Literature Review of HPPK

Tetrahydrofolate cofactors are required for the synthesis of certain amino acids, purine and thymidine and therefore essential for life. Bacteria must synthesize folates *de novo*, because they lack the transport system. In contrast, mammals have to obtain folate from their diet, as they lack the enzymes such HPPK in the mid folate pathway. This difference between mammals and bacteria makes enzymes in this pathway attractive targets for the development of antibacterial agents. For example, inhibitors of dihydropteroate synthase (DHPS) and dihydrofolate reductase (DHFR) have been used as antibiotics for the treatment of bacterial infections.

HPPK catalyzes the transfer of the pyrophosphate group from ATP to 6-hydroxymethyl-7,8-dihydropterin (HP) and is one of the key enzymes that catalyze folate biosynthesis. It belongs to a class of enzymes that catalyzes pyrophosphoryl transfer. Other members of this class of enzymes include ribose-phosphate pyrophosphokinase, nucleotide pyrophosphokinase, GTP pyrophosphokinase, and thiamine pyrophosphokinase. Although the mechanism of many kinases that catalyze phosphoryl transfer has been extensively studied, much less is known about the mechanism of pyrophosphokinase.

Escherichia coli HPPK (EcHPPK) has high sequence similarity to HPPKs from many pathogens. For example, it shares ~56% sequence identity with *Haemophilus influenza* HPPK, ~40% identity with *Pneumocystis carinii* HPPK, ~35% identity with

Staphylococcus aureus HPPK and ~33% identity with *Mycobacterium tuberculosis*

HPPK. As a small (158 residues, ~18 kDa), monomeric protein, EcHPPK is an excellent model system to study the mechanism of pyrophosphate transfer.

Most of the studies on HPPK have been focused on EcHPPK, although there are some studies on HPPKs from other species (30-32). The EcHPPK-catalyzed pyrophosphoryl transfer follows an ordered bi-bi mechanism with MgATP binding to the enzyme first (33, 34). Mg^{2+} is required for the reaction. The enzyme is very specific to ATP: the affinity of MgATP for HPPK is 260-fold higher than that of MgGTP, and more than 1000-fold higher than those of any other nucleoside triphosphates. The binding of MgATP is driven by favorable enthalpy and entropy (35). The individual rate constants for the HPPK-catalyzed reaction have been determined through combined stopped-flow and quench-flow analyses (34) (Figure 1.2). The results show that the chemical step is not the rate-limiting step for the EcHPPK-catalyzed reaction. This finding is a clear indication that protein dynamics play important roles during the catalytic cycle. The high affinity of HPPK for HPPP (6-hydroxymethyl-7,8-dihydropterin pyrophosphate) (36) and low affinity of AMP (35) suggest that the release of HPPP could be rate-limiting, and AMP is preferentially released first.

Along with the kinetic studies, extensive structural studies have been carried out on EcHPPK, using both X-ray crystallography and solution NMR spectroscopy (37-42). Both commercially available and homemade substrate analogues have been used for the

structural studies, including α,β -methyleneadenosine 5'-triphosphate (AMPCPP) and β,γ -methyleneadenosine 5'-triphosphate (AMPPCP) for ATP and DMHP (7,7-dimethyl-6-hydroxymethylpterin) for HP. Figure 1.3 shows the crystal structure of the ligand-free (apo) HPPK (39). It has a three layered α - β - α fold formed by six β strands and four α -helices. Three flexible loops are formed between β 1- α 1 (loop 1), β 2- β 3 (loop 2) and α 2- β 4 (loop 3). The substrate binding site has been identified by NMR (43) and revealed by the crystal structure of the ternary complex of HPPK with AMPCPP, HP and two Mg^{2+} ions at 1.25 Å resolution (38). As shown in Figure 1.4, HP is sandwiched between the two aromatic rings of F123 and Y53 and forms six hydrogen bonds with residues T42, P43, L45 and N55. Twelve residues are involved in the binding of AMPCPP, including Q74, E77, R84, R88, W89, R92, I98, R110, T112, H115, Y116, and R121, among which E77, R92, H115, and R121 are conserved. Two Mg^{2+} ions are found in the active site and both are six-coordinated. One is coordinated with oxygens from the α , β phosphate groups of ATP, two water molecules, and the carboxyl oxygens from D95 and D97. The other is coordinated with the remaining carboxyl oxygens from D95 and D97, the β , γ phosphate groups of ATP, the hydroxyl group from HP, and one water molecule. Thus, the two substrates are well aligned for the pyrophosphoryl transfer.

EcHPPK undergoes dramatic conformational changes during the catalytic cycle, particularly in loops 2 and 3, as shown by representative HPPK structures along the catalytic cycle that have been determined using either substrate or substrate analogues

(Figure 1.5) (37, 38, 40, 42). The large conformational changes are not due to crystal packing effects and are confirmed by several NMR solution structures, including the HPPK•MgAMPCPP binary complex and the HPPK•MgAMPCPP•DMHP ternary complex. In the apo form, loop 3 is open, but upon binding AMPPCP or ADP, loop 3 further opens up and moves away from the active center with the C α atoms moving by as far as ~17 Å. Loop 3 is closed upon the binding of the second substrate HP, as revealed by the crystal structure of the HPPK•AMPCPP•HP ternary complex (38, 42). It opens again after the pyrophosphoryl transfer reaction as revealed by the crystal structure of the ternary product complex HPPK•AMP•HPPP (37), with a conformation similar to that of the HPPK•MgAMPPCP HPPK•MgADP complex. After the release of AMP, loop 3 undergoes yet another conformational change as revealed by the crystal structure of the HPPK•HPPP complex, with a conformation similar to that of the HPPK•MgAMPCPP complex (37, 40). Loop 3 opens after the release of the second product HPPP and returns to the conformation of the apo form.

HPPK is a well-established system for NMR dynamics study. NMR structures of HPPK have been determined for the apo form (42), the HPPK•AMPCPP binary complex and the HPPK•AMPCPP•DMHP ternary complex (40). Unlike kinase which catalyzes the single phosphate transfer of γ -phosphate of ATP, pyrophosphate kinase catalyzes pyrophosphate transfer at β phosphorus. Thus, AMPCPP is an excellent mimic of ATP for pyrophosphate transfer reaction. DMHP is a stable mimic of the second substrate HP. So

these structures represent the snapshots of the first half trajectory of reaction. The NMR structural work discussed here has laid a solid foundation for the dynamics studies of the HPPK.

The structure-function relationship of HPPK has been investigated by the combination of site-directed mutagenesis, biochemical analysis, and X-ray crystallography. The functional and dynamic roles of R82 and R92, which interact with the nucleotides, have been revealed by such combined studies (44, 45). Substitution of R82 and R92 with alanine reduces the rate constant for the chemical step by a factor of ~ 380 and 3.5×10^4 , respectively. However, there is no significant change in either binding energy or binding kinetics of ATP/HP for either mutant, except that the R92A mutant reduces the association rate constant of HP by a factor of 4.5 and the dissociation rate constant by a factor of 10. The results suggest that both R82 and R92 play important roles in the chemical step, and R92 is particularly important for the transition state stabilization. However, neither residue is important for ATP and HP binding, except that R92 might facilitate the binding of HP. X-ray crystal structure studies on the wild type (WT) and mutant enzymes suggest that R92 first binds to the α -phosphate group of ATP and then shifts to interact with the β -phosphate, whereas R82, which initially does not interact with ATP, moves in and binds to the α -phosphate when the pyrophosphoryl transfer is about to occur. The conformational changes of loop 3 as described earlier are necessary to bring the catalytic residues R82 and R92 into the active center (44,45).

As described earlier, loop 3 undergoes dramatic conformational changes during the catalytic cycle of HPPK. The roles of loop 3 in catalysis have been investigated also through combined mutagenesis, biochemical, and X-ray crystallographic studies. A deletion mutant (V83G Δ 84-89) has been designed for this purpose (46). The kinetics results of the mutant enzyme show that the deletion mutation causes no significant effects on either the rate constants or the dissociation constant for ATP binding and ~ 100 times increase in the dissociation constant of HP due to the increase in the dissociation rate constant. It reduces the rate constant of the chemical step by a factor of $\sim 1.1 \times 10^5$ and shifts the rate limiting step to the chemical step. X-ray crystal structure studies show that the mutant enzyme in the apo form is well folded with a conformation similar to that of the WT. However, the mutant ternary complex does not have a fully assembled active center. Together, these results suggest that the long loop is not important for the folding of the enzyme but plays an important role in the active center assembling and is critical for catalysis.

The functional importance of loop 3 is not due to the interactions of the guanidinium group on R84 and the indole ring of W89 with the substrates (47). Substitution of R84 with alanine causes no significant change in either dissociation or kinetic constant of the reaction, resulting in only ~ 4 times decrease in the forward rate constant of the chemical step. Substitution of W89 with alanine causes no significant change in ATP binding, a six fold increase in the dissociation constant for HP ~ 6.5 times, and a ~ 20 -fold decrease in

the forward rate constant and a ~30-fold decrease in the reverse rate constant of the chemical step. These results indicate that R84 is not important for HPPK function while W89 plays a minor role in HP binding and a more significant role in catalysis.

Computational studies of HPPK conformational changes and dynamics have been carried out. Keskin et al. used a coarse-grained model to study the molecular motions and conformational changes of HPPK (48). Six domains of collective motions were identified and two of those domains are loop 2 and loop 3, which exhibit the most concerted motions upon ligand binding. Yang et al. used MD simulations to study the apo HPPK, and HPPK•ATP binary complex based on the NMR structure (49). The local enhanced sampling method was used to improve the loop sampling. The results show 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. Su & Cukier used MD-based methods to study the conformational exploration of the HPPK•ATP complex and the mechanism by which HPPK traps and binds HP (50). Their results indicate that the conformational exploration of HPPK in binding ATP spans from the apo to more ligand-bound-like conformations. Thus, the mechanism of substrate binding is better described as selected fit, in which a pre-existing substrate binding conformation is readily accessible, instead of induced fit. They also use HERM (Hamiltonian Replica Exchange Method) MD simulation to study both EcHPPK and *Y. pestis* HPPK (51). Their results show that the HP binding sites of HPPKs from these two species have different conformations. More importantly, this

method can enlarge the sample space substantially with a small number of systems, which makes it a good method for future computational studies on HPPK dynamics.

The focus of my thesis research is on the roles of protein dynamics in the catalysis by HPPK. While EcHPPK has been extensively studied with respect to reaction kinetics, structure, and structure-function relationships, the roles of protein dynamics in HPPK catalysis are only explored by MD simulations as described above. A combined experimental and computational approach is employed in my research. The dynamical properties of HPPK at the different stages of the catalytic cycle are investigated by both NMR relaxation and MD simulation studies. The dynamical properties are then correlated with catalysis by mutagenesis and combined analyses of the mutant enzymes by biochemical, NMR, and MD methods.

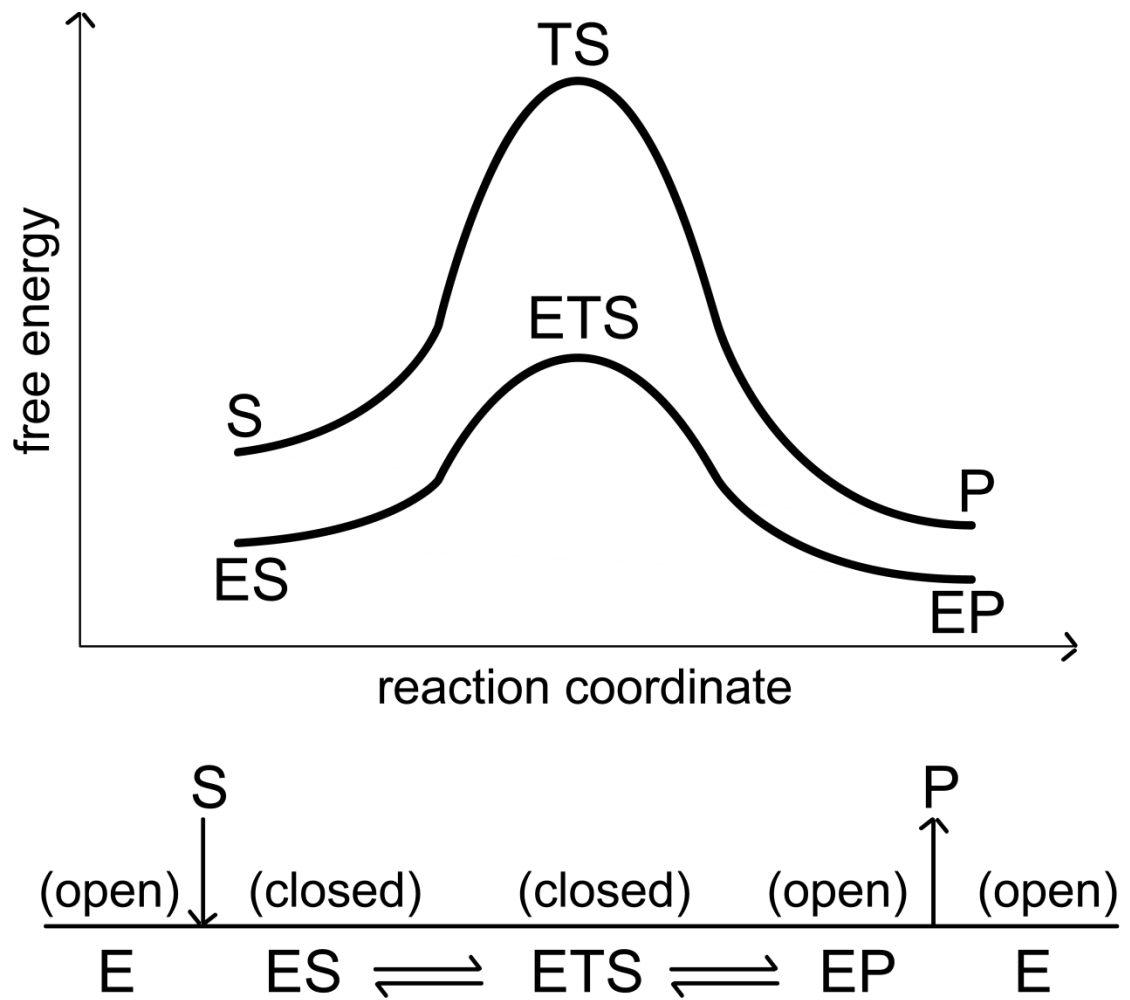


Figure 1.1. Functional roles of protein conformational changes in enzymatic catalysis.

Top, free energy profiles of non-catalytic reaction and the corresponding enzymatic reaction. Bottom, the conflicting structural requirements during enzymatic catalysis.

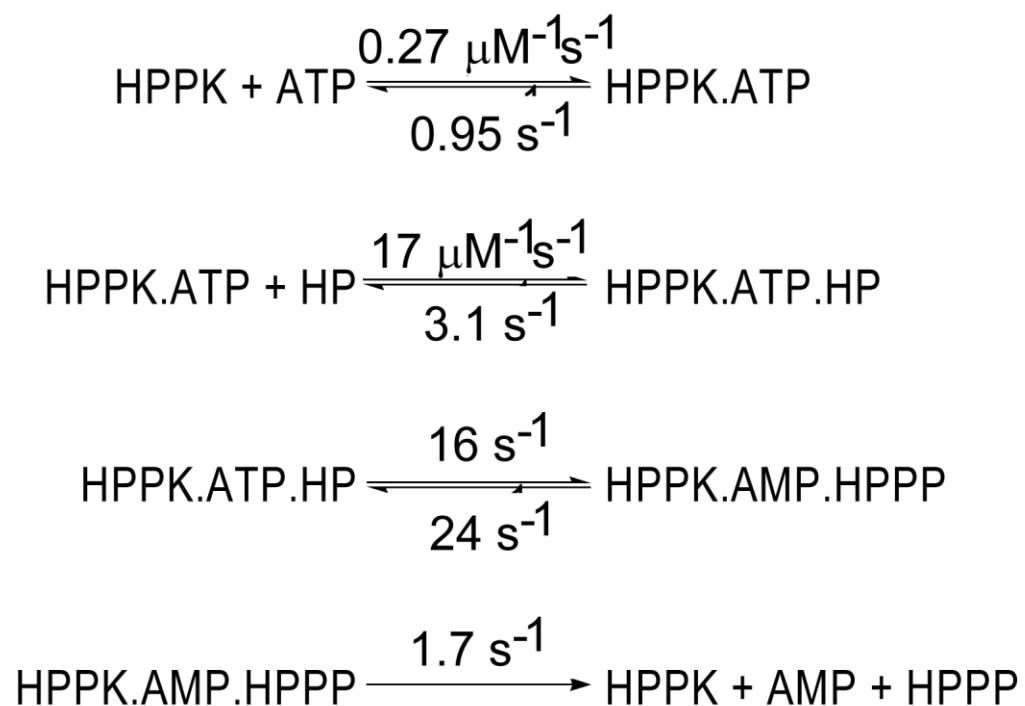


Figure 1.2: Kinetic mechanism of HPPK. Mg^{2+} is omitted for simplicity.

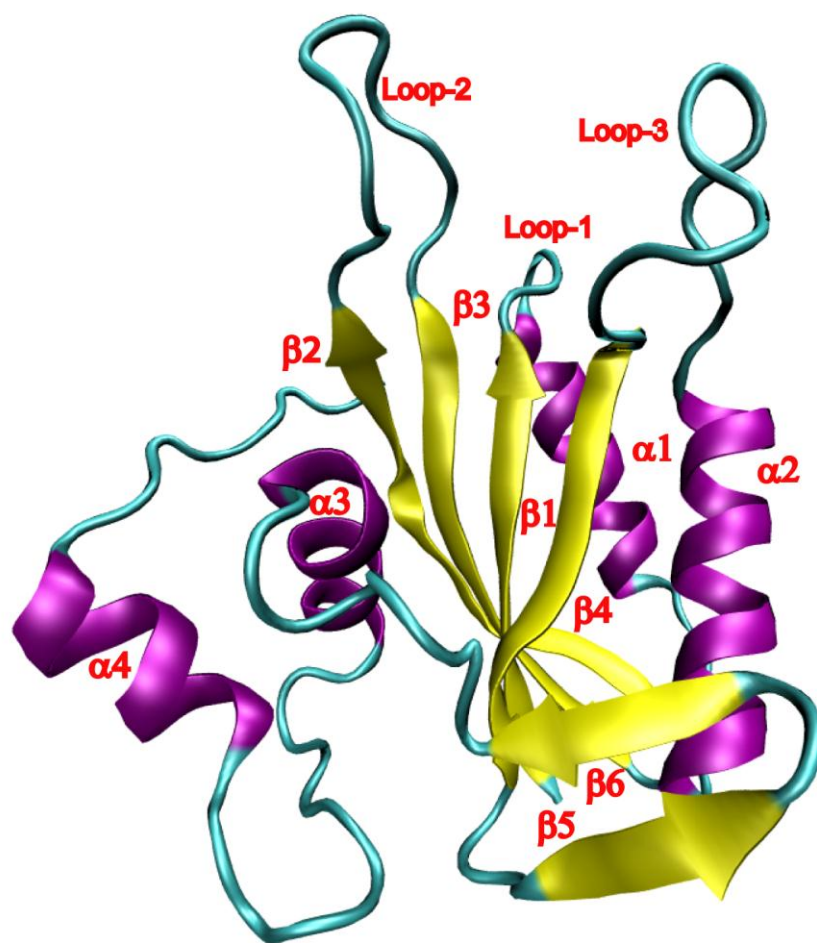
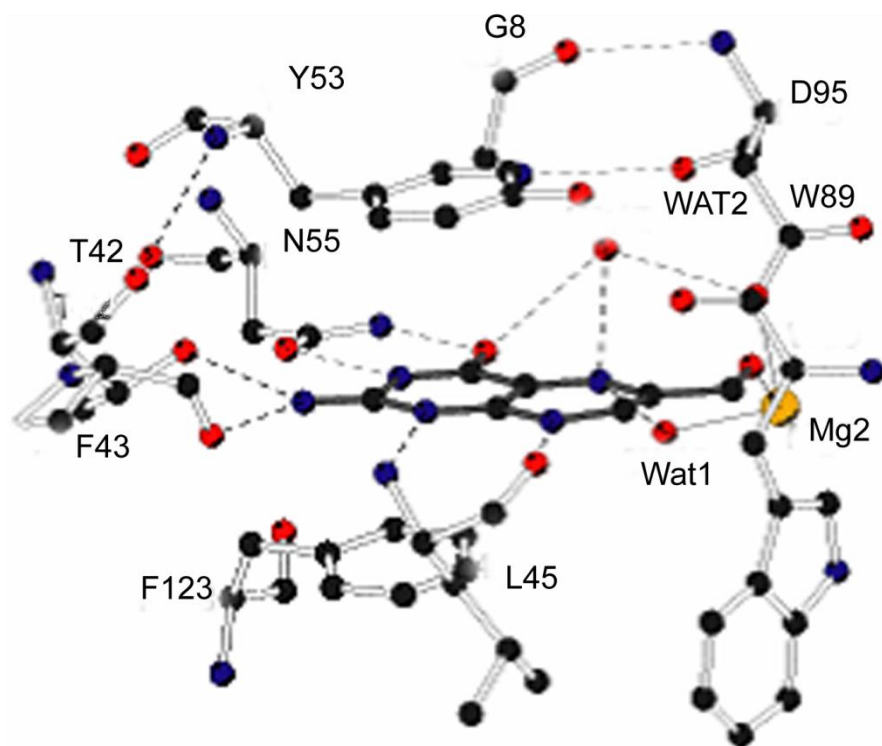


Figure 1.3. Crystal structure of the apo HPPK (pdb ID: 1HKA). The α -helices are in purple and the β -sheet is in yellow. This figure is prepared by the program VMD, version 1.8.6. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

Figure 1.4



A

Figure 1.4 (cont'd)

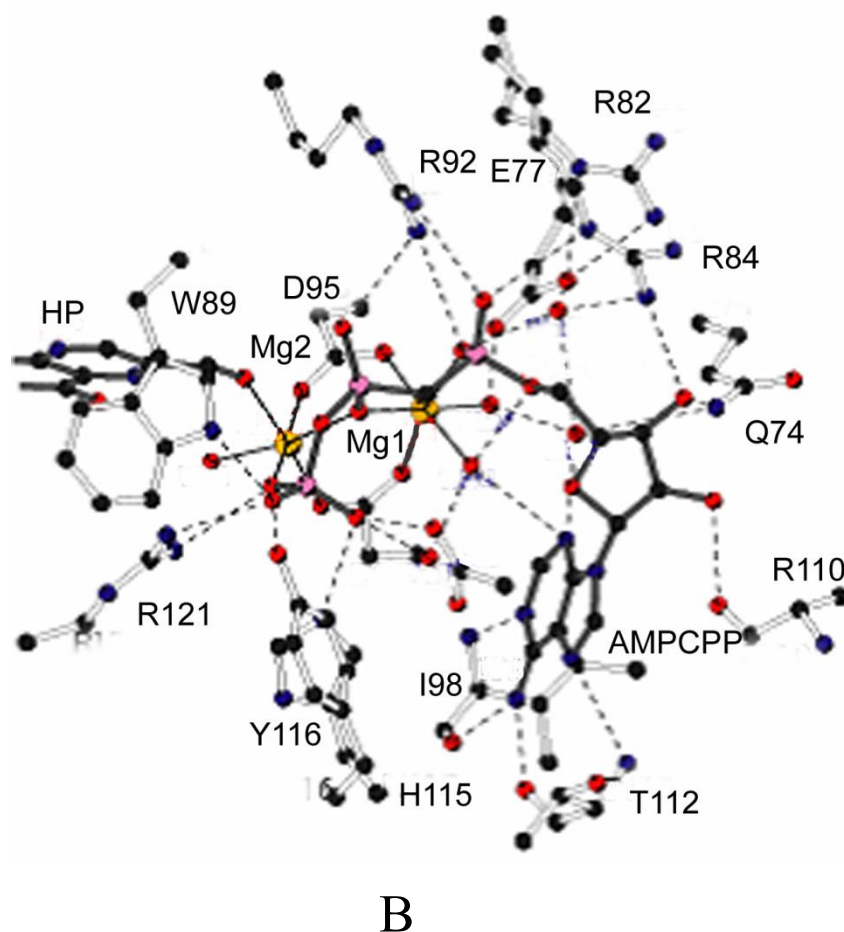


Figure. 1.4. protein•HP (a) and protein•MgAMPCPP (b) interactions. Ball-and-stick model with open bonds represent protein residues and those with filled bonds represent substrate molecules. Dotted lines represent electrostatic interactions. This figure is modified from figures 5 and 6 of ref. 38.

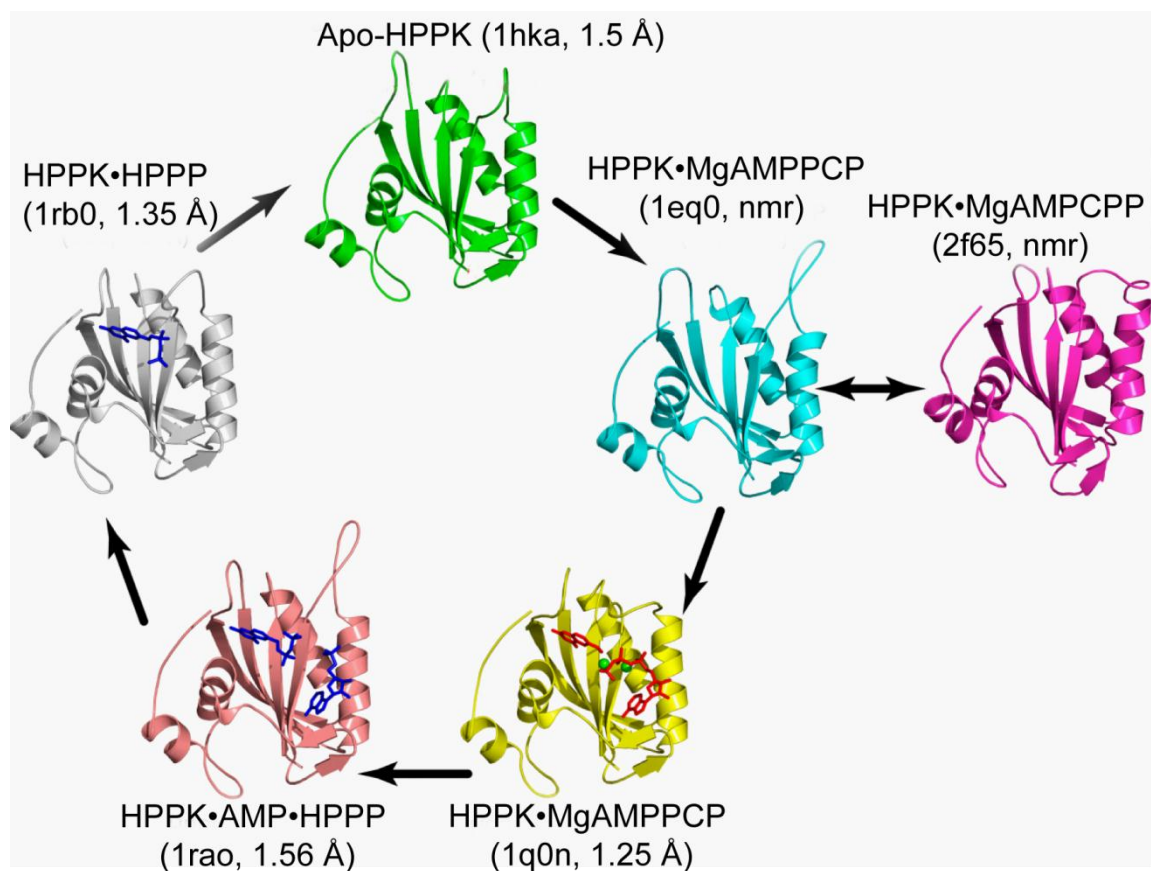


Figure 1.5. HPPK reaction trajectory. The HPPK molecule is illustrated as a ribbon diagram (arrows, β -strands; spirals, helices; tubes, loops), the ligands in the crystal structures are drawn as stick models (substrates in red, products in blue), and the two Mg^{2+} ions as spheres (in green). In parenthesis are the PDB accession codes and the resolution of the structures (when applicable). The black arrows indicate the conformational changes of HPPK during the catalytic cycle. This figure is from a review manuscript by Yan and Ji (personal communication).

References

REFERENCES

1. Lad, C., Williams, N.H. and Wolfenden, R. (2003). The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5607-10.
2. Campbell, M. K. and Farrell, S.O. (2007). *Biochemistry*. 6th ed., Brooks Cole Press
3. Pisliakov, A. V., Cao, J., Kamerlin, S. C. L. & Warshel, A. (2009). Enzyme millisecond conformational dynamics do not catalyze the chemical step. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17359-17364.
4. Karplus, M. (2010). Role of conformation transitions in adenylate kinase. *Proc. Natl. Acad. Sci. U. S. A.* **107**, E71-E71.
5. Kamerlin, S. C. L. & Warshel, A. (2010). Reply to Karplus: Conformational dynamics have no role in the chemical step. *Proc. Natl. Acad. Sci. U. S. A.* **107**, E72-E72.
6. Falke, J. J. (2002). A moving story. *Science* **295**, 1480-1481.
7. Henzler-Wildman, K. & Kern, D. (2007). Dynamic personalities of proteins. *Nature* **450**, 964-972.
8. Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M., and Kern, D. (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis, *Nature* **450**, 913-916.
9. Boehr, D. D., Dyson, H. J. & Wright, P. E. (2006). An NMR perspective on enzyme dynamics. *Chem. Rev.* **106**, 3055-3079
10. Abragham, A. (1961) *The Principles of Nuclear Magnetism*, Clarendon Press, Oxford, England
11. Lipari, G. and Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546-4559.
12. Lipari, G. and Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 2. Analysis of

- experimental results. *J. Am. Chem. Soc.* **104**, 4559-4570.
13. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C. & Gronenborn, A. M. (1990). Deviations from the simple 2-parameter model-free approach to the interpretation of n-15 nuclear magnetic-relaxation of proteins. *J. Am. Chem. Soc.* **112**, 4989-4991.
 14. Mandel, A. M., Akke, M., and Palmer, A. G. (1995) Backbone dynamics of *Escherichia-coli* ribonuclease Hi - correlations with structure and function in an active enzyme, *J. Mol. Biol.* **246**, 144-163.
 15. Chen, J. H., Brooks, C. L. & Wright, P. E. (2004). Model-free analysis of protein dynamics: Assessment of accuracy and model selection protocols based on molecular dynamics simulation. *J. Biomol. NMR* **29**, 243-257.
 16. d'Auvergne, E. J. & Gooley, P. R. (2003). The use of model selection in the model-free analysis of protein dynamics. *J. Biomol. NMR* **25**, 25-39.
 17. Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by N-15 Inverse detected heteronuclear nmr spectroscopy - Application to Staphylococcal Nuclease, *Biochemistry* **28**, 8972-8979.
 18. Mittermaier, A. & Kay, L. E. (2006). Review - new tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224-228.
 19. Luz, Z.; Meiboom, S. (1963) Nuclear magnetic resonance study of the protolysis of trimethylammonium ion in aqueous solution-order of the reaction with respect to solvent *J. Chem. Phys.* **39**, 366-370
 20. Carver, J. P., and Richards, R. E. (1972) General two-site solution for chemical exchange produced dependence of t₂ upon carr-purcell pulse separation, *J. Magn. Reson.* **6**, 89-105
 21. Schlitter, J., Engels, M. & Kruger, P. (1994). Targeted molecular-dynamics - a new approach for searching pathways of conformational transitions. *J. Mol. Graph.* **12**, 84-89.
 22. Schlitter, J., Engels, M., Kruger, P., Jacoby, E. & Wollmer, A. (1993). Targeted molecular-dynamics simulation of conformational change - application to the t $\leftrightarrow$ r transition in insulin. *Mol. Simul.* **10**, 291-308.

23. Park, S. & Schulten, K. (2004). Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **120**, 5946-5961.
24. Lu, H., Isralewitz, B., Krammer, A., Vogel, V. & Schulten, K. (1998). Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* **75**, 662-671.
25. Leech, J., Prins, J. F. & Hermans, J. (1996). Smd: Visual steering of molecular dynamics for protein design. *IEEE Comput. Sci. Eng.* **3**, 38-45.
26. Marchi, M. & Ballone, P. (1999). Adiabatic bias molecular dynamics: A method to navigate the conformational space of complex molecular systems. *J. Chem. Phys.* **110**, 3697-3702.
27. Paci, E. & Karplus, M. (1999). Forced unfolding of fibronectin type 3 modules: An analysis by biased molecular dynamics simulations. *J. Mol. Biol.* **288**, 441-459.
28. Huang, H., Ozkirimli, E. & Post, C. B. (2009). Comparison of three perturbation molecular dynamics methods for modeling conformational transitions. *J. Chem. Theory Comput.* **5**, 1304-1314.
29. Jeremy M. Berg, L. S., John L. Tymoczko (2002). *Biochemistry*.
30. Hennig, M., Dale, G. E., D'Arcy, A., Danel, F., Fischer, S., Gray, C. P., Jolidon, S., Muller, F., Page, M. G. P., Pattison, P. & Oefner, C. (1999). The structure and function of the 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from haemophilus influenzae. *J. Mol. Biol.* **287**, 211-219.
31. Lawrence, M. C., Iliades, P., Fernley, R. T., Berglez, J., Pilling, P. A. & Macreadie, I. G. (2005). The three-dimensional structure of the bifunctional 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase/dihydropteroate synthase of saccharomyces cerevisiae. *J. Mol. Biol.* **348**, 655-670.
32. Blaszczyk, J., Li, Y., Cherry, S., Alexandratos, J., Wu, Y., Shaw, G., Tropea, J. E., Waugh, D. S., Yan, H. G. & Ji, X. H. (2007). Structure and activity of yersinia pestis 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as a novel target for the development of antiplague therapeutics. *Acta. Crystallogr. D* **63**, 1169-1177.

33. Bermingham, A., Bottomley, J. R., Primrose, W. U. & Derrick, J. P. (2000). Equilibrium and kinetic studies of substrate binding to 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from escherichia coli. *J. Biol. Chem.* **275**, 17962-17967.
34. Li, Y., Gong, Y. C., Shi, G. B., Blaszczyk, J., Ji, X. H. & Yan, H. G. (2002). Chemical transformation is not rate-limiting in the reaction catalyzed by escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Biochemistry* **41**, 8777-8783.
35. Shi, G. B., Gong, Y. C., Savchenko, A., Zeikus, J. G., Xiao, B., Ji, X. H. & Yan, H. G. (2000). Dissecting the nucleotide binding properties of escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase with fluorescent 3'-(2')-o-anthraniloyladenine 5'-triphosphate. *Biochim. Biophys. Acta.* **1478**, 289-299.
36. Garcon, A., Bermingham, A., Lian, L. Y. & Derrick, J. P. (2004). Kinetic and structural characterization of a product complex of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from escherichia coli. *Biochem. J.* **380**, 867-873.
37. Blaszczyk, J., Shi, G. B., Li, Y., Yan, H. G. & Ji, X. H. (2004). Reaction trajectory of pyrophosphoryl transfer catalyzed by 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Structure* **12**, 467-475.
38. Blaszczyk, J., Shi, G., Yan, H. & Ji, X. (2000). Catalytic center assembly of hppk as revealed by the crystal structure of a ternary complex at 1.25 angstrom resolution. *Structure* **8**, 1049-1058.
39. Xiao, B., Shi, G. B., Chen, X., Yan, H. G. & Ji, X. H. (1999). Crystal structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, a potential target for the development of novel antimicrobial agents. *Structure* **7**, 489-496.
40. Li, G. Y., Felczak, K., Shi, G. B. & Yan, H. G. (2006). Mechanism of the conformational transitions in 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as revealed by NMR spectroscopy. *Biochemistry* **45**, 12573-12581.
41. Stammers, D. K., Achari, A., Somers, D. O., Bryant, P. K., Rosemond, J., Scott, D. L. & Champness, J. N. (1999). 2.0 angstrom 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.

42. Xiao, B., Shi, G. B., Gao, J. H., Blaszczyk, J., Liu, Q., Ji, X. H. & Yan, H. G. (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.
43. Shi, G. B., Gao, J. H. & Yan, H. G. (1999). H-1, C-13 and N-15 resonance assignments of escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase and its complex with mgamppecp. *J. Biomol. NMR* **14**, 189-190.
44. Blaszczyk, J., Li, Y., Shi, G. B., Yan, H. G. & Ji, X. H. (2003). Dynamic roles of arginine residues 82 and 92 of escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: Crystallographic studies. *Biochemistry* **42**, 1573-1580.
45. Li, Y., Wu, Y., Blaszczyk, J., Ji, X. H. & Yan, H. G. (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.
46. Blaszczyk, J., Li, Y., Shi, G. B., Yan, H. G. & Ji, X. H. (2004). Essential roles of a dynamic loop in the catalysis of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Biochemistry* **43**, 1469-1477.
47. Li, Y., Blaszczyk, J., Wu, Y., Shi, G. B., Ji, X. H. & Yan, H. G. (2005). Is the critical role of loop 3 of escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase in catalysis due to loop-3 residues arginine-84 and tryptophan-89? Site-directed mutagenesis, biochemical, and crystallographic studies. *Biochemistry* **44**, 8590-8599.
48. Keskin, O., Ji, X., Blaszczyk, J. & Covell, D. G. (2002). Molecular motions and conformational changes of hppk. *Proteins: Struct. Funct. Genet.* **49**, 191-205.
49. Yang, R., Lee, M. C., Yan, H. G. & Duan, Y. (2005). Loop conformation and dynamics of the escherichia coli hppk apo-enzyme and its binary complex with mgatp. *Biophys. J.* **89**, 95-106.
50. Su, L. & Cukier, R. I. (2009). An enhanced molecular dynamics study of hppk-atp

conformation space exploration and atp binding to hppk. *J. Phys. Chem. A* **113**, 2025-2035.

51. Su, L. & Cukier, R. I. (2009). Hamiltonian replica exchange method study of escherichia coli and yersinia pestis HPPK. *J. Phys. Chem. B* **113**, 16197-16208.

Chapter 2

Dynamics of the Conformational Transitions in the Assembling of the Michaelis Complex of a Bi-substrate Enzyme: A ^{15}N Relaxation of *Escherichia coli* 6-Hydroxymethyl-7,8-dihydropterin Pyrophosphokinase

The research presented in this chapter was published as the following manuscript:

Ewen Lescop, Zhenwei Lu, Qin Liu, Huimin Xu, Guangyu Li, Bin Xia, Honggao Yan and Changwen Jin. Dynamics of the Conformational Transitions in the Assembling of the Michaelis Complex of a Bi-substrate Enzyme: A ^{15}N Relaxation Study of *Escherichia coli* 6-Hydroxymethyl-7,8-dihydropterin Pyrophosphokinase, *Biochemistry*, 2009, 48 (2), pp 302–312. **DOI:** 10.1021/bi8016262

My contribution is the relaxation studies of HPPK substrate ternary complex.

Dynamics of the Conformational Transitions in the Assembling of the Michaelis Complex of a Bisubstrate Enzyme: A ^{15}N Relaxation Study of *Escherichia coli* 6-Hydroxymethyl-7,8-dihydropterin Pyrophosphokinase[†]

ABSTRACT

6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) catalyzes the transfer of pyrophosphate from ATP to 6-hydroxymethyl-7,8-dihydropterin (HP), which follows an ordered bi-bi kinetic mechanism with ATP binding to the enzyme first. HPPK undergoes dramatic conformational changes during its catalytic cycle as revealed by X-ray crystallography, and the conformational changes are essential for the enzymatic catalysis as shown by site-directed mutagenesis and biochemical and crystallographic analysis of the mutant enzymes. However, the dynamic properties of the enzyme have not been measured experimentally. Here, we report the ^{15}N NMR relaxation studies of the dynamic properties of *Escherichia coli* HPPK from the ligand-free (apo) form to the binary substrate complex with MgATP (represented by MgAMPCPP, an ATP analogue) to the Michaelis complex (ternary substrate complex) with MgATP (represented by MgAMPCPP) and HP (represented by 7,7-dimethyl-6-hydroxypterin, an HP analogue). The results show that the binding of the nucleotide to HPPK does not cause major changes in the dynamic properties of the enzyme. Whereas enzymes are often more rigid when bound to the ligand or the substrate, the internal mobility of HPPK is not reduced and is even moderately increased in the binary complex, particularly in the catalytic

loops. The internal mobility of the catalytic loops is significantly quenched upon the formation of the ternary complex, but some mobility remains. The enhanced motions in the catalytic loops of the binary substrate complex may be required for the assembling of the ternary complex. On the other hand, some degrees of mobility in the catalytic loops of the ternary complex may be required for the optimal stabilization of the transition state, which may need the instantaneous adjustment and alignment of the side-chain positions of catalytic residues. Such dynamic behaviors may be characteristic of bisubstrate enzymes.

Introduction

6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) catalyzes the transfer of pyrophosphate from ATP to 6-hydroxymethyl-7,8-dihydropterin (HP, Figure 2.1a), leading to the biosynthesis of folate cofactors (1). Folate cofactors are essential for life (2). Mammals have an active transport system for deriving folates from the diet. In contrast, most microorganisms must synthesize folates *de novo*, because they lack the active transport system. Therefore, like other enzymes in the folate biosynthetic pathway, HPPK is an attractive target for developing antimicrobial agents.

Protein dynamics is thought to play an important role in enzymatic catalysis (3-7). While the role of protein dynamics in the chemical step of bond cleavage and formation is still controversial (8, 9), it is no doubt that protein dynamics plays an important role in the physical steps of the assembling of the Michaelis complex and product release.

Although structural studies have provided ample evidence for protein conformational changes through a catalytic cycle, there are only a few systematic experimental studies of the dynamic properties of enzymes through their catalytic cycles and their roles in the conformational transitions that are required for enzymatic catalysis (see ref (4) for an excellent recent review). Our understanding of the role of protein dynamics in enzymatic catalysis lags far behind our knowledge of the structures and chemical mechanisms of enzymes. With great progress in structural genomics and structure-function relationship studies, elucidating the role of protein dynamics in enzymatic catalysis becomes one of the greatest remaining challenges for biochemists and biophysicists. *E. coli* HPPK has emerged as an excellent model for studying the role of protein dynamics in enzymatic catalysis, because the enzyme is small (~18 kDa), stable, and amenable to both X-ray crystallographic and NMR analysis. Atomic structures have been determined for nearly every stage of the catalytic cycle, and protein dynamics has been shown to be important for HPPK catalysis (10-24).

The HPPK-catalyzed reaction follows an ordered kinetic mechanism with ATP binding to the enzyme first (15, 18). Comparative analysis of the crystal structures of apo HPPK and its ternary complex has revealed that the complete active center of HPPK is assembled only after both substrates bind to the enzyme (14, 21). The assembling of the active center involves large conformational changes, particularly in the three catalytic loops (Figure 2.1b). Of the three catalytic loops, loop 3 undergoes the most dramatic

conformational changes. As expected, the loop moves in to close the active center upon the formation of the ternary complex (14). However, the loop moves away from the active center upon the binding of MgADP or an MgATP analogue (16). Recently, it has been shown that loop 3 also moves away from the active center upon completion of pyrophosphoryl transfer, to the same extent as upon the binding of MgADP or an MgATP analogue (21). By site-directed mutagenesis, it has been shown that loop 3 is required for the assembly of the full active center, plays an important role in the stabilization of the ternary complex and the transition state of the reaction, and is essential for catalysis (22, 23).

While the crystal and NMR structures have provided snapshots of the catalytic cycle, the dynamic properties of HPPK at various stages of the catalytic cycle have to be studied in order to fully understand the relationship between protein dynamics and catalysis in HPPK. In the present work, we investigate the backbone dynamic properties of *E. coli* HPPK by ^{15}N relaxation measurements focusing on the assembling of the Michaelis complex, i.e., from the apo enzyme to the binary substrate complex and then to the ternary Michaelis complex. Model-free analysis of the ^{15}N NMR relaxation data together with the observation of weak or missing signals from loop 3 revealed that the three catalytic loops are mobile on different timescales in the absence of ligand. Surprisingly, the binding of the nucleotide to HPPK does not cause major changes in the dynamic properties of the enzyme. Whereas enzymes are often more rigidified upon the

binding of a substrate or a ligand, the internal mobility of HPPK is not reduced and is even moderately increased in the binary complex, particularly in the catalytic loops. The enhanced motions in the three loops may be required for the assembling of the ternary substrate complex. The internal mobility of the catalytic loops is significantly reduced upon formation of the ternary complex, but some mobility remains. Some degrees of mobility in the catalytic loops of the ternary complex may be required for the optimal stabilization of the transition state, which may need the instantaneous adjustment and alignment of the side-chain positions of catalytic residues. Such dynamic behaviors may be characteristic of bisubstrate enzymes.

Materials and Methods

Sample Preparation. The overexpression and protein purification protocol was similar to that previously described (13). Briefly, BL21(DE3) *E. coli* cells containing pET17b-HPPK were grown in LB media. When OD₆₀₀ reached 1.0, the cells were collected and resuspended in M9 media containing ¹⁵NH₄Cl and [¹³C₆]-D-glucose as the sole nitrogen and carbon sources, respectively, when necessary (25). Induction by 100 mg/L IPTG was started 1 h later and lasted 4 h. The cells were harvested by centrifugation (6000 rpm, 20 min), washed with 20 ml buffer A (20 mM Tris-HCl, pH 8.0), and kept at -80 °C until use. The frozen bacterial paste was thawed at 42 °C and sonicated on ice. Following the initial purification by ion-exchange chromatography, fractions containing HPPK were concentrated to ~3 mL by an Amicon concentrator and

then applied to a gel filtration column (Sephadex G-75) equilibrated with buffer B (20 mM sodium phosphate, pH 7.4). The column was developed with the same buffer. The purified HPPK fractions were lyophilized. NMR samples were prepared by dissolving the required amount of protein powder in buffer B made with 5% D₂O/95% H₂O. Great care was taken to adjust and to keep the pH at 7.4 during NMR experiments. Five ¹⁵N-labeled apo HPPK samples were prepared with the protein concentrations at 0.1, 0.25, 0.5, 1.0, and 2.0 mM. One ¹⁵N/¹³C-labeled apo HPPK sample was prepared with the protein concentration at ~0.5 mM for sequential resonance assignment. Two samples of the binary complex of ¹⁵N-labeled HPPK and MgAMPCPP were prepared, one containing 0.5 mM HPPK, 5 mM AMPCPP, and 20 mM MgCl₂, and the other containing 1.0 mM HPPK, 10 mM AMPCPP, and 40 mM MgCl₂. The dissociation constant of the HPPK•MgAMPCPP complex (*K_d* of 77 nM) suggests that, under this condition, the protein is saturated with MgAMPCPP (18). The ternary complex of HPPK was prepared using AMPCPP, 7,7-dimethyl-6-hydroxypterin (DMHP, an HP analogue), ¹⁵N-labeled HPPK, and Mg²⁺ as previously described (24). The NMR sample contained 0.5 mM HPPK, 5 mM AMPCPP, ~2 mM DMHP, and 20 mM MgCl₂.

Sequential Resonance Assignments. The backbone NMR assignment of the HPPK protein was previously carried out in the apo state (12) and in complex with MgAMPCPP (24). The backbone resonance assignment of the apo HPPK was further extended at a lower concentration of protein (0.5 mM). HNCA and CBCA(CO)NH

experiments (26) were acquired for the $^{15}\text{N}/^{13}\text{C}$ -labeled apo HPPK sample on a Bruker Avance 800 MHz and a Bruker Avance 500 MHz spectrometer equipped with a cryoprobe. The NMR data were processed with NMRPipe (27) and analyzed with NMRView (28).

^{15}N Relaxation Measurements. ^{15}N relaxation experiments were carried out at 25 °C on a Bruker Avance 600 MHz spectrometer for the apo HPPK and the binary complex HPPK•MgAMPCPP and on a Varian INOVA 600 MHz spectrometer for the ternary complex HPPK•MgAMPCPP•DMHP. Standard pulse sequences were used to obtain the longitudinal and transverse relaxation rates R_1 , R_2 , as well as $\{^1\text{H}\}$ - ^{15}N heteronuclear NOEs (29). The decays of the ^{15}N longitudinal and transverse magnetizations for the apo HPPK and the binary complex were sampled by typically 9 (10, 100, 300, 500, 800, 1200, 1500, and 2000 ms in random order) and 10 points (6, 14, 42, 62, 70, 82, 102, 122, 142, and 162 ms in random order), respectively. The first points of both experiments were duplicated at the end of the experiments to verify the stability of the sample and to estimate the intensity error. The time points were 11, 44, 167, 389, 666, 944, 110, 1443, and 1998 ms for the R_1 measurement of the ternary complex and 17, 35, 52, 70, 87, 104, 122 and 139 ms for the R_2 measurement of the ternary complex. The third points of both experiments were repeated for error estimation. The delay τ_{cp} between ^{15}N inversion pulses in the CPMG module in R_2 experiments was set to 1 ms. A recycle delay of 2.5 s was used for both R_1 and R_2 experiments. The residue-specific $\{^1\text{H}\}$ - ^{15}N heteronuclear

NOE was measured by dividing the peak intensity in two 2D spectra recorded with and without a 3-s proton pre-saturation achieved by 120 ° ^1H pulses following a 5-s delay (30). The experiments were duplicated at each protein concentration to estimate the uncertainty. All the 2D spectra were recorded with 1024 complex data points and 128 complex increments for the apo HPPK and the binary complex and with 1994 complex data points and 100 complex increments for the ternary complex. The number of transients was 16 for all relaxation experiments except the $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE experiment of the ternary complex, which was 64. The spectral widths were set to 8389 and 1460 Hz for the apo HPPK, 8389 and 1581 Hz for the binary complex, and 9000 and 1720 Hz for the ternary complex in the direct ^1H and the indirect ^{15}N dimensions, respectively.

The relaxation data were processed with NMRPipe (27) and analyzed with NMRView (28). For the relaxation data of the apo HPPK, a Lorenz-to-Gauss window function was applied in both dimensions and for those of the binary complex, a cosine-bell window function was applied. A forward-and-backward linear prediction and zero filling were typically used to extend the indirect dimension by a factor of 2. The peak intensities of the relaxation rate measurements were extracted using the NvRatePrintHeight subroutine of NMRView (28). Intensities below the noise level were typically removed and the exponential decay curves were fitted to a two-parameter exponential equation using an in-house modified version of RELAXFIT (31). The

uncertainties on the relaxation rates were estimated by 500 Monte-Carlo simulations. The HetNOE subroutine of NMRView was used to calculate the heteronuclear NOEs. The final values were extracted from the duplicates.

Rotational diffusion and model-free analysis. Rotation diffusion was analyzed using the R_2/R_1 values of well-ordered regions of the structure deemed free of either fast internal motion or chemical exchange. Hydrogen atoms were added to the crystal structure (PDB code 1HKA) of the apo HPPK (10) by using MOLMOL (32). The mean structures of the NMR ensembles of the binary complex HPPK•MgAMPCPP (PDB code 2F65) (24) and the ternary complex HPPK•MgAMPCPP•DMHP (PDB code 2F63) (24) were taken as the representative structures of the complexes. The optimization of the rotational diffusion tensor against the experimental data were achieved by using ROTFIT (33) and TENSOR2 (34). HYDRONMR (35) was used to predict the rotational diffusion tensors of the three forms of HPPK based on their structures.

The microdynamic parameters under the Lipari-Szabo formalism (36-38) were extracted using the TENSOR2 program (34). Briefly, the ^{15}N relaxation data were interpreted in terms of motion of the N-H bond. Five models of increasing complexity were tested including model 1 (S^2), model 2 (S^2, τ_e), model 3 (S^2, R_{ex}), model 4 (S^2, τ_e, R_{ex}), and model 5 (S^2, τ_e, S_f^2), where S^2 is the squared order parameter, τ_e a correlation time describing the internal motion (ps to ns timescale), assumed to be independent of the overall tumbling, S_f^2 is a second squared order parameter describing the fast internal

motion on the ps to ns timescale, R_{ex} is the chemical exchange contribution to R_2 . The F -test statistics was then used to select the model that satisfies the data with the lowest number of parameters. Standard errors in the model-free parameters were evaluated by 100 Monte Carlo simulations.

Results

Extension of Backbone Resonance Assignments of the Apo HPPK. HPPK

consists of 158 amino acid residues, 12 of which are prolines. The sequential resonance assignment of the apo HPPK has been previously achieved at 1.5 mM protein concentration (12). Several residues could not be assigned due to missing $^1\text{H}/^{15}\text{N}$ cross-peaks in the $^1\text{H}-^{15}\text{N}$ HSQC spectra of apo HPPK under these conditions. Additional $^1\text{H}-^{15}\text{N}$ HSQC spectra were collected at lower concentrations (0.1, 0.25, 0.5, 1 mM). Overall, the $^1\text{H}-^{15}\text{N}$ HSQC cross-peaks were sharpened with decreasing the protein concentration. Several additional cross-peaks could be observed at 0.5 mM that were absent at 1 mM concentration, and concomitantly, 10 cross-peaks showed small chemical shift variations ($|\Delta ^1\text{H}|_{\text{max}} < 0.04$ ppm and $|\Delta ^{15}\text{N}|_{\text{max}} < 0.5$ ppm). Below 0.5 mM protein concentration, the spectra were exactly superimposable. Two triple resonance experiments (HNCA and CBCACONH) were acquired on a 0.5 mM $^{13}\text{C}/^{15}\text{N}$ -labeled apo HPPK sample to extend the backbone resonance assignment. Seven additional residues could be assigned (Figure 2.2): E77, L78, Q79, and Q80 in the C-terminal extremity of helix $\alpha 2$, D95 and L96 in strand $\beta 4$, and E109 at the turn of the β -hairpin

$\beta 5$ - $\beta 6$. About 10 weak ^1H - ^{15}N correlation peaks in the ^{15}N HSQC could not be unambiguously assigned. They likely correspond to residues Q48 in loop 2, R82, V83, A86, W89, G90, R92 and T93 in loop 3 and Y116 between strand $\beta 6$ and helix $\alpha 3$, which remain unassigned.

^{15}N Relaxation Data and Indications of Weak Self-Association of the Apo HPPK. Initial NMR studies indicated that the ^{15}N relaxation parameters of the apo HPPK were protein concentration dependent, which is usually indicative of self-association. To ascertain the effects of the self-association on the relaxation parameters, we collected the residue-specific ^{15}N R_1 , R_2 and $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE data of apo HPPK at four protein concentrations: 0.1, 0.5, 1, and 2 mM. Because of peak overlaps, unassigned residues, and weak signals leading to decay curves of poor quality, ^{15}N relaxation data could be obtained for 118, 118, 112, and 104 residues out of the 146 non-proline residues for 0.1, 0.5, 1, and 2 mM samples, respectively. Fewer data were obtained at high concentrations, because of line broadening and peak disappearance. Figure 2.3 shows the distribution of these data as well as the R_2/R_1 ratio versus the amino acid sequence at the four concentrations. The mean R_1 values were 1.12 ± 0.06 , 1.14 ± 0.06 , 1.06 ± 0.05 , and $0.76 \pm 0.06 \text{ s}^{-1}$ at 0.1, 0.5, 1, and 2 mM, respectively, and the corresponding mean R_2 values were 13.15 ± 1.82 , 14.32 ± 2.38 , 16.24 ± 2.90 , and $26.30 \pm 4.21 \text{ s}^{-1}$. The mean R_1 value decreased significantly whereas the mean R_2 value

increased significantly at 1 and 2 mM HPPK compared with the values at 0.1 and 0.5 mM, suggesting a significant self-association at higher concentrations.

On the individual residue basis, all the secondary structure elements had R_1 , R_2 , and NOE values near the mean values at each concentration, except the C-terminal extremity of the helix α_2 (residues R75 to Q80) and the N-terminal extremity of strand β_4 (residues L94 and D97), both of which are linked to loop 3 and had higher R_2 values. Loop 1 also contained higher than average R_2 values. In contrast, residues in loops-2 and -3 have a lower than average R_2 together with low NOE values. The NOE values at different concentrations were quite similar, except some variations in loops 1, 2 and 3, indicating that protein concentration does not have significant effects on the internal motions of the protein on the ps to ns timescale.

The average R_2/R_1 values were used to obtain the estimate of the isotropic correlation times (τ_c) of the protein. At 0.1 mM, the uncertainties in the relaxation data were relatively high so that the isotropic diffusion model ($\tau_c = 10.48 \pm 0.25$ ns) was statistically sufficient to describe the rotational diffusion properties of HPPK. At 0.5, 1, and 2 mM concentrations, the rotational diffusion of HPPK was estimated to be fully anisotropic. The ratios of the estimated principal components were quite similar at each concentration (1:0.83:0.76, 1:0.79:0.73, and 1:0.81:0.75, at 0.5, 1, and 2 mM, respectively) and the corresponding τ_c values were 10.87 ± 0.31 , 12.28 ± 0.42 , and 19.30 ± 0.71 ns. The rotational diffusion tensors at 0.1 and 0.5 mM concentrations were in

excellent agreement with that predicted by HYDRONMR (35) on the crystal structure of the monomeric form of the apo HPPK (PDB code 1HKA, ratios of the principal components of 1:0.80:0.70 and average correlation time of 10.22 ns). The significant increases in τ_c at 1 and 2 mM are likely due to the self-association of HPPK at high concentrations.

The changes of ^1H and ^{15}N chemical shifts and peak intensity in the ^1H - ^{15}N HSQC spectra of HPPK were also monitored at five concentrations: 0.1, 0.25, 0.5, 1, and 2 mM. The chemical shifts changes were not significant in the 0.1 - 0.5 mM concentration range whereas the chemical shifts of quite a few residues were found to vary at protein concentrations higher than 1 mM. The changes were relatively small for most of these residues, however. Attempts were made to determine the K_d value for the self-association from the cross-peaks with the largest chemical shift changes. However, as the ^1H and ^{15}N chemical shifts changed in a linear fashion with the protein concentration (not shown), it was not possible to estimate the K_d value for the self-association. These results nevertheless indicated that the self-association of HPPK is weak in the studied concentration range.

Taken together, the results of the relaxation, rotational diffusion tensors and chemical shift analysis over the four studied samples suggested that the apo HPPK is prone to self-associate in the millimolar range concentration. However, the self-association of the apo HPPK is insignificant at 0.1 and 0.5 mM protein concentrations.

¹⁵N Relaxation Data Analysis of the Binary Complex HPPK•MgAMPCPP.

MgAMPCPP was used as an analogue for the substrate MgATP, as HPPK had a low level of ATPase activity. MgAMPCPP is an excellent MgATP analogue for HPPK with respect to both structure (14) and binding affinity (18). Binding of MgAMPCPP caused significant changes in the ¹H-¹⁵N HSQC spectrum (Figure 2.2). All expected backbone resonances were previously assigned at 1.6 mM protein concentration, except those of residues T1, S13, E16, R82, W89, G90, and L94, which were missing (24). Several residues in loops-2 and -3 showed two or more sets of cross-peaks, including L45-Q50 in loop 2 and A86-E87 in loop 3 (24). Additional ¹H-¹⁵N HSQC spectra were acquired at both 0.5 mM and 1 mM concentrations of the HPPK•MgAMPCPP complex. No significant change in chemical shifts was observed. For the relaxation study, we extracted the relaxation data for the most populated isomer. ¹⁵N R_1 , R_2 and {¹H}-¹⁵N heteronuclear NOEs were measured for 130 and 129 residues at the 0.5 and 1 mM concentrations of the complex, respectively. The results are shown in Figure 2.4. At 0.5 mM, the mean R_1 and R_2 were 1.29 ± 0.20 and $12.64 \pm 3.50 \text{ s}^{-1}$, respectively, and the corresponding values at 1 mM were 1.16 ± 0.10 and $14.83 \pm 4.60 \text{ s}^{-1}$. Similar to the apo protein, a small increase in R_2 was accompanied by a decrease in R_1 , suggesting that the binding of MgAMPCPP does not completely prevent the weak self-association of HPPK, as the derived τ_c values slightly increased between 0.5 mM ($9.26 \pm 0.30 \text{ ns}$) and 1 mM ($10.25 \pm 0.20 \text{ ns}$) concentrations. However, these τ_c values were shorter than the τ_c value

predicted by HYDRONMR (10.42 ns) on the basis of the mean structure of the HPPK•MgAMPCPP complex, suggesting that the monomeric form of the complex is highly predominant at both concentrations. The R_2/R_1 -derived anisotropic rotational diffusion model at 0.5 mM complex concentration had ratios of the principal components of 1:0.95:0.78, similar to those predicted based on the mean NMR structure of the HPPK•MgAMPCPP complex (1:0.89:0.81). Taken together, both the chemical shift and the rotational diffusion tensor data indicated that the self-association of HPPK is weakened upon the binding of MgAMPCPP and that the self-association can be regarded as insignificant for the HPPK•MgAMPCPP complex at 0.5 mM concentration.

¹⁵N Relaxation Data Analysis of the Ternary Complex

HPPK•MgAMPCPP•DMHP. Except for the first residue, sequential resonance assignment has been made for all residues, including those from the three catalytic loop region (24). ¹⁵N relaxation data could be obtained for 135 residues (Figure 2.5), 5 more residues than the binary complex. The mean R_1 and R_2 were 1.32 ± 0.06 and 11.51 ± 0.94 s⁻¹, respectively. While the mean R_1 was similar to that of the binary complex, the mean R_2 of the ternary complex was significantly smaller, with a much smaller standard deviation as well. The uniformity of the R_2 values was consistent with the uniformity of the intensities of the ¹H-¹⁵N HSQC cross-peaks of the ternary complex, indicating that the protein is rigidified upon the formation of the ternary complex. The τ_c value of the ternary complex was estimated to be 8.88 ± 0.01 ns based on the average R_2/R_1 ratio,

shorter than the τ_c value predicted by HYDRONMR (10.42 ns) based on the mean structure of the ternary complex. The R_2 data and the correlation times indicated that the ternary complex has a more compact structure than the apo HPPK and the binary complex. The ratios of the estimated principle components were 1:0.87:0.89, similar to those calculated using the mean NMR structure of the ternary complex (1:0.86:0.92).

Model-Free Analysis of the Three Forms of HPPK. The microdynamic parameters of the apo HPPK were extracted in the model-free framework (36-38). HPPK is in rather low abundance *in vivo* and therefore most likely functions in a monomeric form (39-41). Therefore, our analysis was performed on the relaxation data obtained at 0.5 mM protein concentration, because the data acquired at this concentration had excellent quality and the self-association of the protein was insignificant. The model-free parameters S^2 , τ_e , and R_{ex} could be extracted for 118 residues and are presented in Figures 2.6 and 2.8. The mean S^2 was 0.88 ± 0.08 . All residues in the secondary structure, except T35 in the distorted strand β_2 , had S^2 values higher than 0.8, indicating that the secondary structure elements are rigid. The four helices were slightly more rigid than the central β -sheet with mean S^2 values of 0.92 and 0.86 respectively. In addition to T35, the following residues had significantly lower S^2 values than the mean: L45, G46, Q50, R85, N103, V105, R110, F139, F152, L155, and N156. Residues L45, G46, and Q50 are located in loop 2 (Figure 2.1); R85 is located in loop 3; N103, V105, R110, and D139 are located in short turns; and F152, L155, and N156 in the C-terminal region.

Twelve residues required the introduction of a significant R_{ex} term (higher than 2 s^{-1}) for the model-free analysis: S9, L11, A12, Q17, F39, T42, L78-Q80, N103, R121, and F123. Residues S9, L11, and A12 are located in loop 1; Q17 is located in the N-terminal region of helix $\alpha 2$ connected to loop 1; F39 is in the C-terminal region of the distorted strand $\beta 2$ connected to loop 2; T42 is in loop 2; L78-Q80 are in the C-terminal region of helix $\alpha 2$ connected to loop 3; R121 is in the loop connected to helix $\alpha 3$; and F123 is in the N-terminal region of helix $\alpha 3$.

The microdynamic parameters of the binary complex HPPK•MgAMPCPP are also summarized in Figures 2.6, 2.7, and 2.8 for comparison with those of both the apo HPPK and the ternary complex. The mean S^2 was 0.84 ± 0.07 , similar to that of the apo HPPK. Significant chemical exchange (higher than 1.5 s^{-1}) was introduced for 17 residues: G8, S9, L11, A12, L15, and Q17 around loop 1; T42, D49, and L54 around loop 2; E77, L78, Q79, E87, R92, T93, and L96 around loop 3; and N103 in loop $\beta 4$ - $\beta 5$. The dynamic properties of the HPPK•MgAMPCPP complex are similar to those of the apo enzyme with some regions becoming even more mobile upon the binding of the nucleotide.

The microdynamic parameters of the ternary complex HPPK•MgAMPCPP•DMHP are summarized in Figures 2.7 and 2.8. The mean value for S^2 was 0.88 ± 0.04 , similar to those of the apo HPPK and the binary complex. However, only five residues were below two standard deviations of the mean value: Q48, A86,

A151, F152, and D153. Q48 and A86 are located in loops-2 and 3, respectively, and residues 151-153 in the C-terminal region. Only four residues required a significant R_{ex} term (higher than 1.5 Hz) for the model-free analysis: D49, V83, K85, and R92; the R_{ex} values were significantly smaller than those of the binary complex. Residue 49 is located in loop 2 and residues 83, 85, and 92 are located in loop 3. The results indicated that the C-terminal region is mobile as in the apo HPPK and the binary complex; the catalytic loops are significantly rigidified upon the formation of the ternary complex. However, loops-2 and -3 retain some degrees of mobility.

In general, several physical phenomena are prone to give rise to chemical exchange. In the case of HPPK, self-association represents a potential source of chemical exchange. Although the results from model-free analysis are known to be affected when self-association is severe (42), we estimated the chemical exchange R_{ex} from the data obtained on HPPK in the apo form at all concentrations (data not shown). We did not observe any significant systematic trend in R_{ex} over the four concentrations that could be associated with self-association (as discussed in ref (43)). As a consequence, the contribution of self-association to the chemical exchange was judged to be insignificant. In addition, the fast exchange between the apo and the bound states of the protein can also induce chemical exchange, if the chemical shifts of a given spin differ significantly in the two states. In such a situation, the chemical exchange can be estimated to be proportional to the product of the two populations in exchange (44). Given the

dissociation constant of the binary complex (77 nM) (18) and the high concentration of AMPCPP, the contribution of the dissociation of the binary complex to the chemical exchange was judged to be insignificant. Taken together, we conclude that the chemical exchanges observed in both the apo and the nucleotide-bound states were mainly associated with the intrinsic internal mobility of the monomeric forms of the apo HPPK and of the binary complex.

Discussion

HPPK undergoes dramatic conformational changes during its catalytic cycle and the conformational changes play important roles in HPPK catalysis (Figure 2.1b) (10-12, 14, 16, 21-24). In order to fully understand the roles of conformational dynamics in HPPK catalysis, we carried out the first direct experimental measurement of backbone dynamics of HPPK by NMR, focusing on the formation of the Michaelis complex. Our analysis was, however, complicated by the weak self-association of HPPK at high concentrations. Therefore, we made ^{15}N relaxation measurements at four protein concentrations for the apo HPPK and at two concentrations for the binary complex HPPK•MgAMPCPP. Based on the chemical shifts, relaxation parameters, and correlation time analysis, we concluded that self-association is insignificant for the apo HPPK and for the binary complex HPPK•MgAMPCPP at 0.5 mM protein concentration. Therefore, we focused our analysis on the relaxation data acquired at this concentration on both states of HPPK using the model-free formalism (36-38).

In the model-free framework, the squared-order parameter S^2 reflects the amplitude of internal motions on the ps to ns timescale, the internal correlation time τ_c gives an estimation of the timescale for these motions, and R_{ex} reflects conformational exchange on the μ s to ms timescale. Overall, the core structure of the enzyme remains highly rigid on the ps to ns and μ s to ms timescales in all three forms. This region includes one half of the cleft forming the active site, which contains the residues in strands $\beta 4$, $\beta 5$, and $\beta 6$ that are in contact with the nucleotide and was previously referred as the rigid wall (10). This observation is consistent with the published structures, which indicate that the core structure of HPPK region is mostly structurally invariant (with very small backbone rms deviations) with respect to ligand binding or mutations (10-12, 14, 16, 21-24). We also notice that the C-terminal region of the protein is flexible on the ps to ns timescale in all three forms, in consistence with elevated rms deviations of this region when the structures of the various forms of the wild-type and mutant HPPKs are superimposed. Since the C-terminal extremity is in contact with strand $\beta 2$, near the HP substrate binding site, mobility in this region may be required for the fine adjustment of the local conformation of the HP-binding site during the catalytic reaction. Because the major conformational changes during the catalytic cycle are localized in the three catalytic loops, we will focus our discussion on the dynamic properties of these loops and their changes in the conformational transitions from the apo HPPK to binary substrate complex to the Michaelis complex.

Based on the S^2 values, loop 1 of the apo HPPK is rigid on the ps to ns timescale (Figures 2.6a and 2.8). Its mean S^2 value is 0.90 ± 0.06 , comparable to those of the four helices and the central β -sheet, which have mean S^2 values of 0.92 and 0.86, respectively. On the other hand, the model-free analysis of the relaxation data introduced a chemical exchange term for three residues of this short loop (S9, L11, and A12) (Figures 2.6c and 2.8), suggesting that loop 1 undergoes conformational exchange on the μ s to ms timescale. The crystal structure of the apo HPPK reveals two conformations for this loop with moderate B-factor values (Figure 2.1b) (10). The dynamic properties of loop 1 determined in this NMR work therefore suggest that the two conformations are likely in exchange on the μ s to ms timescale. Upon the binding of the nucleotide, loop 1 undergoes conformational exchange on the μ s to ms timescale to a similar extent to that of the apo HPPK. Nevertheless, loop 1 becomes significantly more mobile on the ps to ns timescale. Indeed, the mean S^2 value of the loop is 0.71 for the binary complex and 0.90 for the apo HPPK.

The model-free parameters indicate that loop 2 is significantly mobile on the ps to ns timescale in the apo HPPK, with a mean S^2 value of 0.62 ± 0.20 (Figure 2.6a and 2.8). Due to the presence of four proline residues (P43, P44, P47, and P51), missing or overlapping cross-peaks, the dynamic properties of only four residues of loop 2 (P43-D52) were accessible. The S^2 values of L45, G46, and Q50 are 0.72, 0.36, and 0.59, respectively, and that of D52 (0.82) is close to the average S^2 value of the enzyme (0.88).

The S^2 values of the loop residues vary greatly with the middle region (around G46) having the lowest S^2 values. It is also noticed that the two residues that delimit the loop are rigid on the ps to ns timescale with S^2 values of 0.93 and 0.98 for T42 and Q53, respectively. Surprisingly, although two residues (F39 and T42) in the twisted strand $\beta 2$ that precedes loop 2 required an R_{ex} term for the Model-free analysis, none of the loop 2 residues required an R_{ex} term (Figure 2.6c and 2.8), suggesting that loop 2 does not undergo significant conformational exchange on the μ s to ms timescale.

Upon the binding of MgAMPCPP, loop 2 remains highly flexible on the ps to ns timescale with an only minor increase in S^2 compared with the apo protein. In contrast to the apo HPPK, mobility on slow timescales (slower than the μ s timescale) could be observed in loop 2 in the complex. Indeed, significant chemical exchange was detected in loop 2 (for residue D49, R_{ex} of $6.1 \pm 1.5 \text{ s}^{-1}$), and the existence of (at least) two sets of cross-peaks for many consecutive residues in loop 2 (24) may suggest that this loop also undergoes conformational isomerization at an even slower exchange rate (at the NMR timescale). Furthermore, the intensity ratios of the major and the minor peaks are significantly different for some residues, suggesting that loop 2 may assume multiple conformations. It is possible that the slow conformational exchanges are due to a *cis-trans* isomerization of the four proline residues (P43, P44, P47, and P51), but no NMR indication for a *cis* configuration was observed for any of the four proline residues and all are in a *trans* configuration in the NMR and crystal structures. The mobility of loop 2 in

the HPPK•MgAMPCPP complex is consistent with the structures of HPPK in complex with nucleotides. Indeed, the backbone rms deviation of loop 2 in the NMR structure of the HPPK•MgAMPCPP complex is significantly higher than other regions of the protein (24) (Figure 2.6e). The increased mobility of loop 2 is further supported by the loss of electron density for a significant part of the loop (residues 44-48) upon binding of MgADP (16).

The dynamic properties could be directly obtained only for two residues in loop 3 for the apo HPPK. This is mainly due to the weakness or absence of NMR signals for the NH groups of residues R82, V83, A86, W89, G90, and R92, which clearly indicates that loop 3 is flexible and undergoes conformational exchange at an intermediate rate (at the chemical shift timescale). In addition, conformational exchange occurring on the μ s to ms timescale was detected for several residues at the extremity of helix α 2 linked to loop 3 (Figure 2.8). Similar to the apo HPPK, the weak intensity of most cross-peaks corresponding to residues in loop 3 precluded the accurate measurements of the relaxation parameters in loop 3 in the HPPK•MgAMPCPP complex. This indicates that loop 3 remains highly mobile on the intermediate timescale upon the binding of the nucleotide. The existence of internal mobility of loop 3 is also enforced by the presence of chemical exchange on the μ s to ms timescale for residues at the bottom of the loop and for E87 in the middle of the loop (Figures 2.6c and 2.8), and by the multiple chemical

shifts observed for A86 and E87, suggesting that like loop 2, loop 3 also undergoes conformational exchange on an even slower timescale.

In summary, the binding of the nucleotide does not quench the internal motions of HPPK on the ps to ns, μ s to ms, and ms to second timescales, as evidenced by the S^2 and R_{ex} values, weak and missing cross-peaks, and multiple cross-peaks with different intensity ratios. The mobilities of the three catalytic loops remain the same or are even enhanced upon the binding of the nucleotide. The result is consistent with the biochemical data indicating that the entropy of binding is in favor of the formation of the HPPK•MgATP complex (13). It is also consistent with the structural data revealing that the catalytic loops of the enzyme can assume different conformations upon the binding of a nucleotide (16, 24).

The motions of the three catalytic loops are significantly quenched upon the formation of the ternary complex (Figures 2.7 and 2.8). Loop 1 is represented by S9, L11, and A12. Although the S^2 values of S9 in the binary and the ternary complexes (0.83 and 0.88, respectively) are similar, S9 has a significant R_{ex} term (4.3) in the binary complex but not in the ternary complex. The S^2 values of L11 and A12 increase significantly from the binary complex (0.71 and 0.72, respectively) to the ternary complex (0.88 and 0.87, respectively), and meanwhile, both residues have a significant R_{ex} term (5.5 and 17.6, respectively) in the binary complex but not in the ternary complex. For loop 2, T42, L45, G46, and Q48 have significantly higher S^2 values in the ternary complex (0.86, 0.89,

0.85, and 0.76, respectively) than in the binary complex (0.66, 0.77, 0.52, and 0.64, respectively). Other residues have similar S^2 values in the two states. Two residues have a significant R_{ex} term (5.6 for T42 and 6.1 for D49) in the binary complex, whereas only one residue (D49) has a significant R_{ex} term (4.8 s^{-1}). For loop 3, relaxation data could be obtained for only three out of ten non-proline residues in the binary complex, because of weak or missing cross-peaks, but for all non-proline residues in the ternary complex. Of the three residues measured for both complexes, the S^2 value of R92 increases significantly from the binary complex (0.71) to the ternary complex (0.89), whereas those of the other two residues (R84 and E87) are similar between the two complexes. E87 and R92 in the binary complex have R_{ex} values of 14.7 and 3.3 s^{-1} , respectively, whereas only R92 has a significant R_{ex} term of 2.0 s^{-1} in the ternary complex.

While the motions of the three catalytic loops are significantly quenched upon the formation of the ternary complex, the two large loops that are directly involved in the binding of the substrates retain some degrees of mobility. The S^2 values of Q48 in loop 2 (0.76) and A86 in loop 3 (0.59) are lower than the mean value by more than two standard deviations. Furthermore, D49 in loop 2 and V83, K85, and R92 in loop 3 have a significant R_{ex} term, ranging from 2.0 to 4.8 s^{-1} .

The dynamic properties of HPPK may be characteristic of bisubstrate enzymes. The hallmarks of enzymatic catalysis are the formation of enzyme-substrate(s) complexes and the maximization of transition state stabilization, which have conflicting structural

requirements. The formation of an enzyme-substrate complex requires an open active center so that the substrate(s) can get in. On the other hand, the maximization of transition state stabilization requires a closed active center that maximizes the favorable interactions between the enzyme and the transition state. The conflicting structural requirements can be resolved by a flexible active center that can sample both open and closed conformational states. For a bisubstrate enzyme like HPPK, the Michaelis complex consists of two substrates in addition to the enzyme. The enzyme must remain flexible upon the binding of the first substrate so that the second substrate can get into the active center. The active center is fully assembled and stabilized only when both substrates bind to the enzyme. However, the side-chain positions of the catalytic residues in the Michaelis complex are still not optimally aligned for the stabilization of the transition state, which lasts only approximately 10^{-13} s. The instantaneous and optimal alignment of catalytic groups for the transition state stabilization requires a dynamic enzyme, not an enzyme which undergoes a large scale of movements but an enzyme which permits at least a small scale of adjustment of catalytic group positions. The dynamic properties of the three forms of HPPK as determined by the ^{15}N relaxation measurements are consistent with the conformational transitions for a bisubstrate enzyme from the apo form to the binary substrate complex and the ternary Michaelis complex.

Figure 2.1

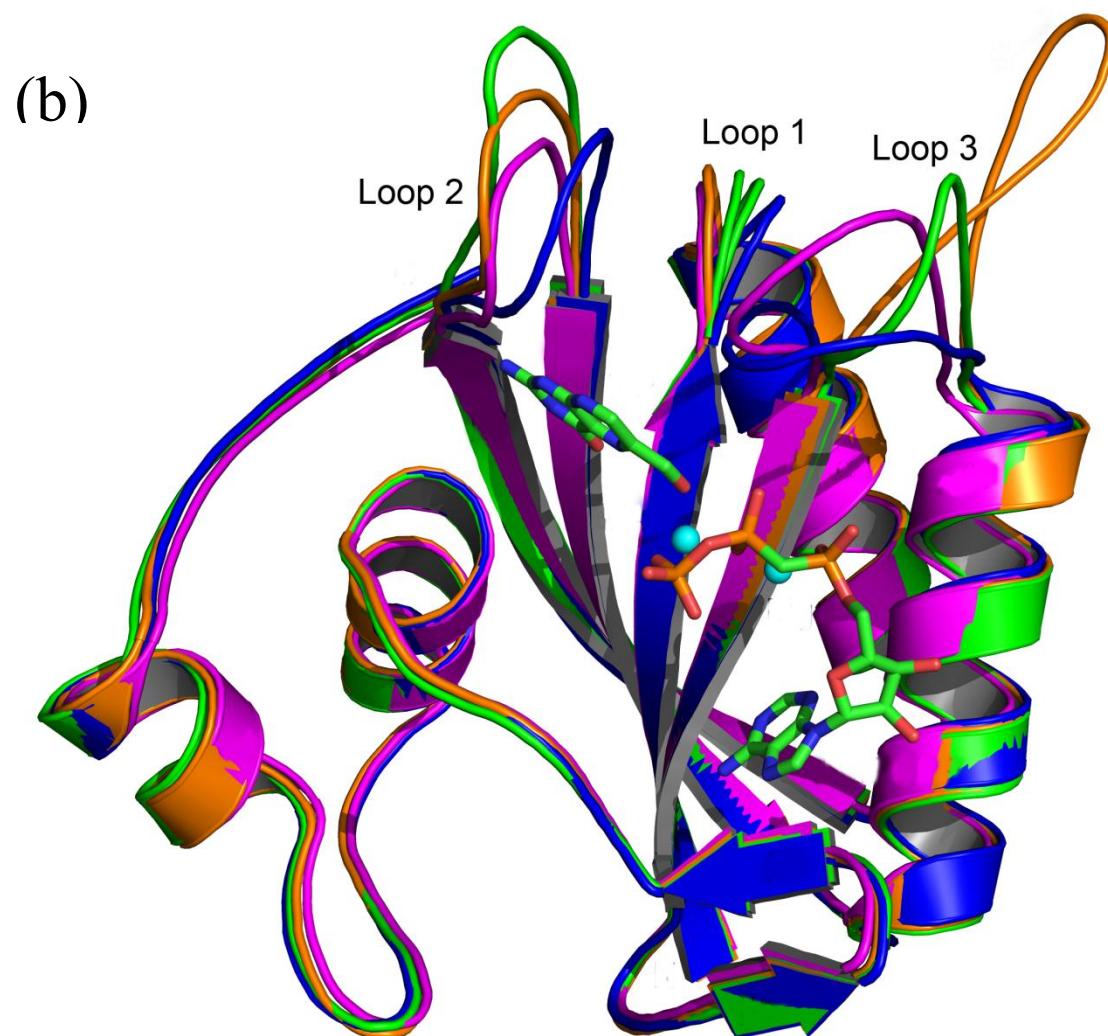
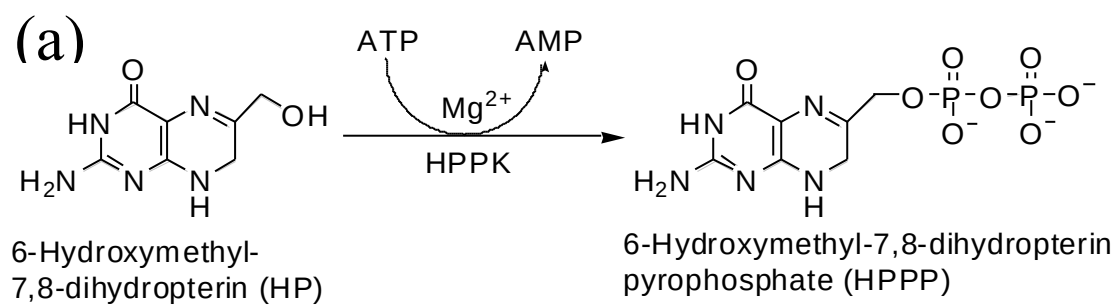


Figure 2.1 (cont'd)

Figure 2.1. The HPPK-catalyzed reaction (a) and a ribbon diagram of the crystal structures of HPPK showing the conformational changes of the three catalytic loops during several stages of its catalytic cycle (b). The apo enzyme (PDB code 1HKA) is in green, the ternary substrate complex (PDB code 1Q0N) with HP, AMPCPP (a substrate ATP analogue), and Mg^{2+} in blue, the ternary product complex (PDB code 1RAO) with HPPP and AMP in orange, and the binary product complex (PDB code 1RB0) with HPPP in magenta. Loop 1 of the apo enzyme has two conformations. AMPCPP, HP, and two Mg^{2+} ions are drawn to illustrate the location of the active center. Panel b was made with PyMOL (45).

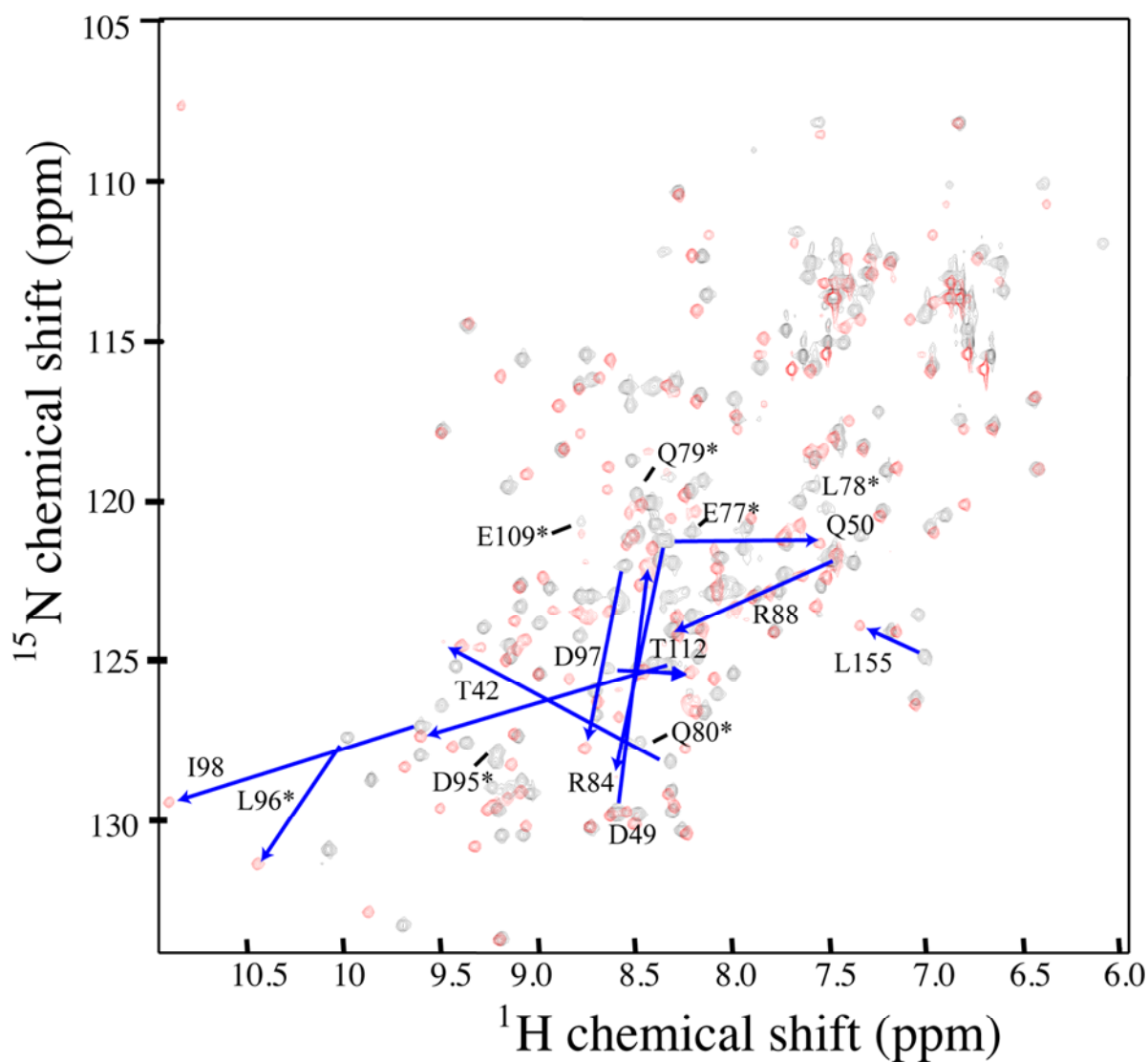


Figure 2.2. The ^1H - ^{15}N HSQC spectra of HPPK in apo-state (black) and in complex with MgAMPCPP (red). The spectra were acquired on a Bruker 600 MHz spectrometer at 25 $^{\circ}\text{C}$ and 0.5 mM protein concentration. The most shifted cross-peaks upon MgAMPCPP complexation are labeled according to their corresponding residue number in the sequence and the arrow joins the cross-peaks in apo- and bound states. In addition, newly assigned cross-peaks of the apo HPPK are labeled with a star.

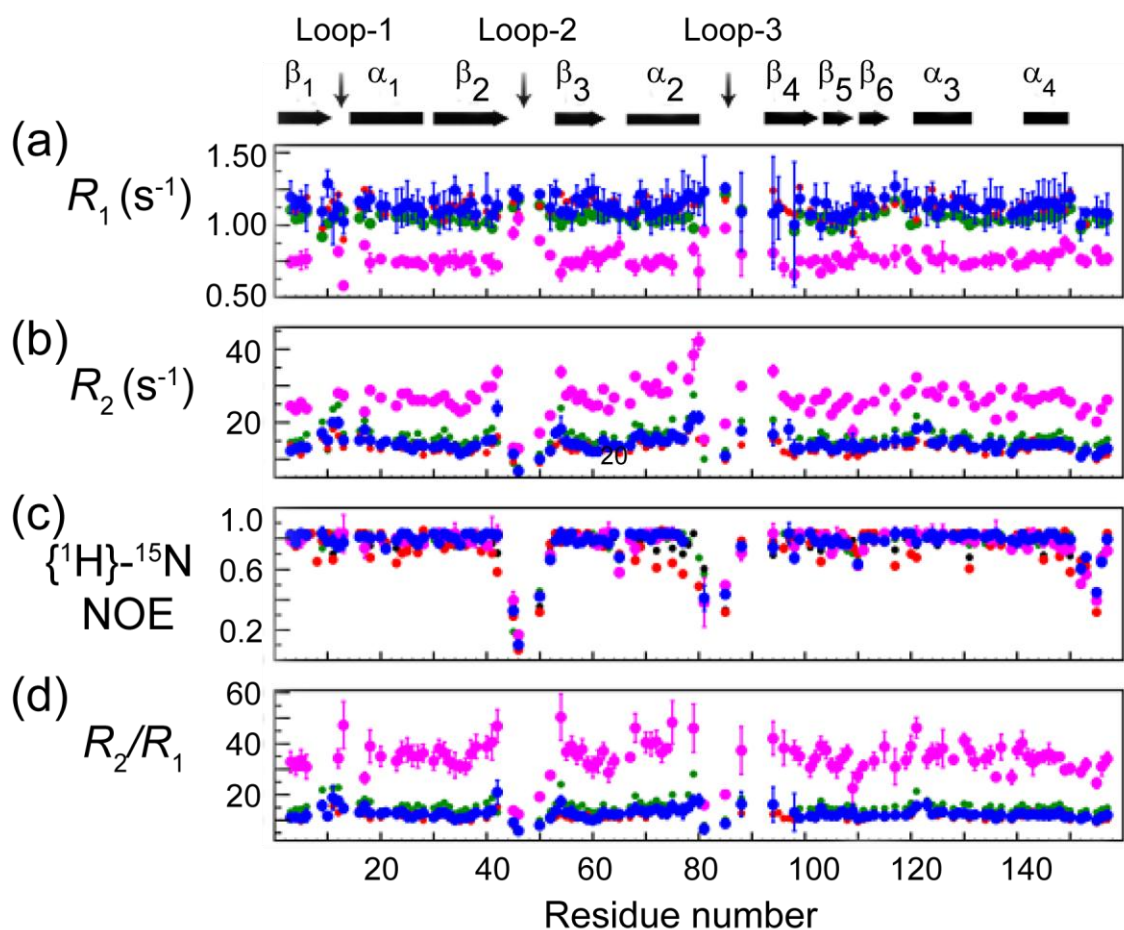


Figure 2.3. ^{15}N relaxation parameters obtained for the apo HPPK at 25 $^{\circ}\text{C}$ at four concentrations: 0.1 mM (red), 0.5 mM (blue), 1 mM (green) and 2 mM (magenta). For clarity, the estimated uncertainties are not represented at 0.1 and 1 mM concentrations. The secondary structure of the apo protein is represented at the top of the figure. (a) R_1 relaxation rates; (b) R_2 relaxation rates; (c) $\{^1\text{H}\}\text{-}^{15}\text{N}$ NOE values; and (d) R_2/R_1 ratio.

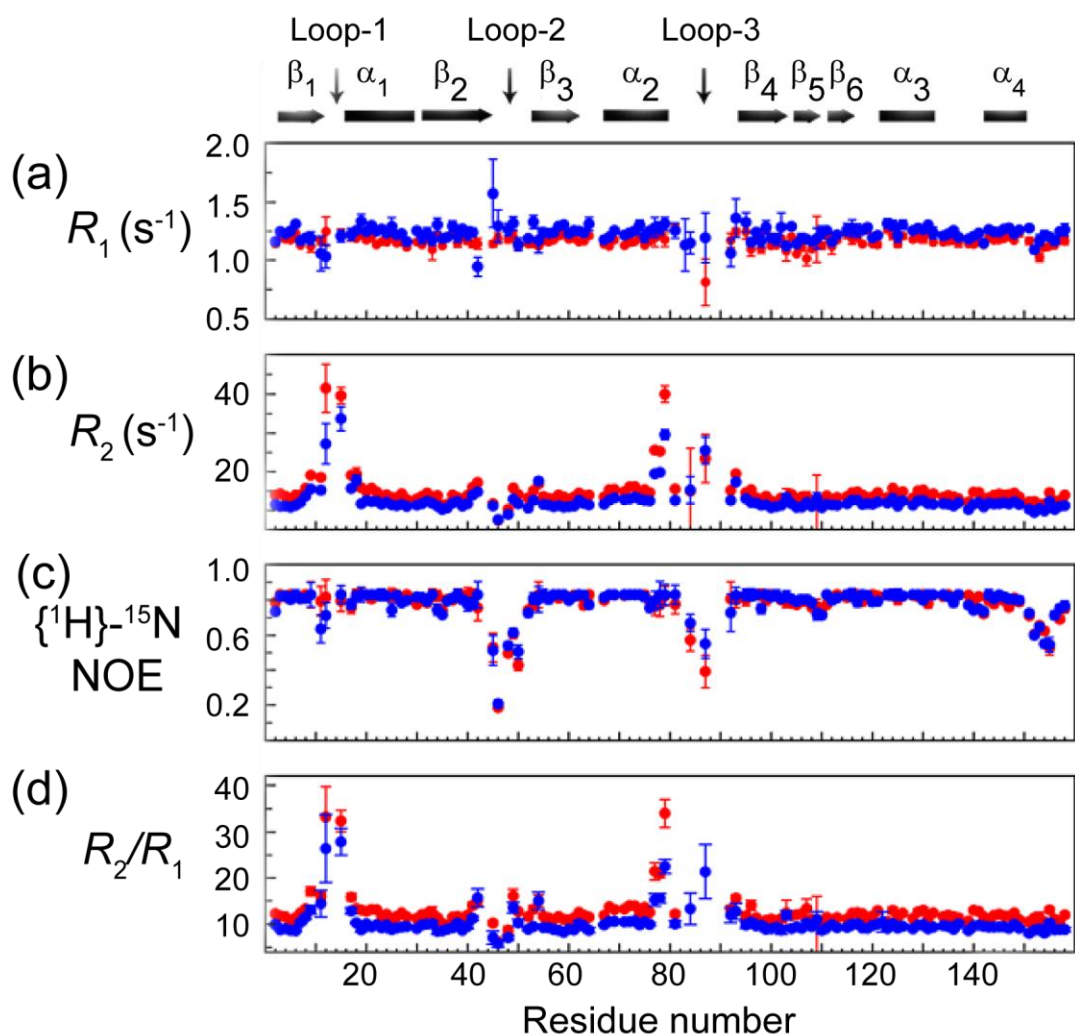


Figure 2.4. ^{15}N relaxation parameters obtained for the binary complex

HPPK•MgAMPCPP at 25 $^{\circ}\text{C}$ at two concentrations: 0.5 mM (blue) and 1 mM (red). The secondary structure is represented at the top of the figure. (a) R_1 relaxation rates; (b) R_2 relaxation rates; (c) $\{^1\text{H}\}\text{-}^{15}\text{N}$ NOE values; and (d) R_2/R_1 ratio.

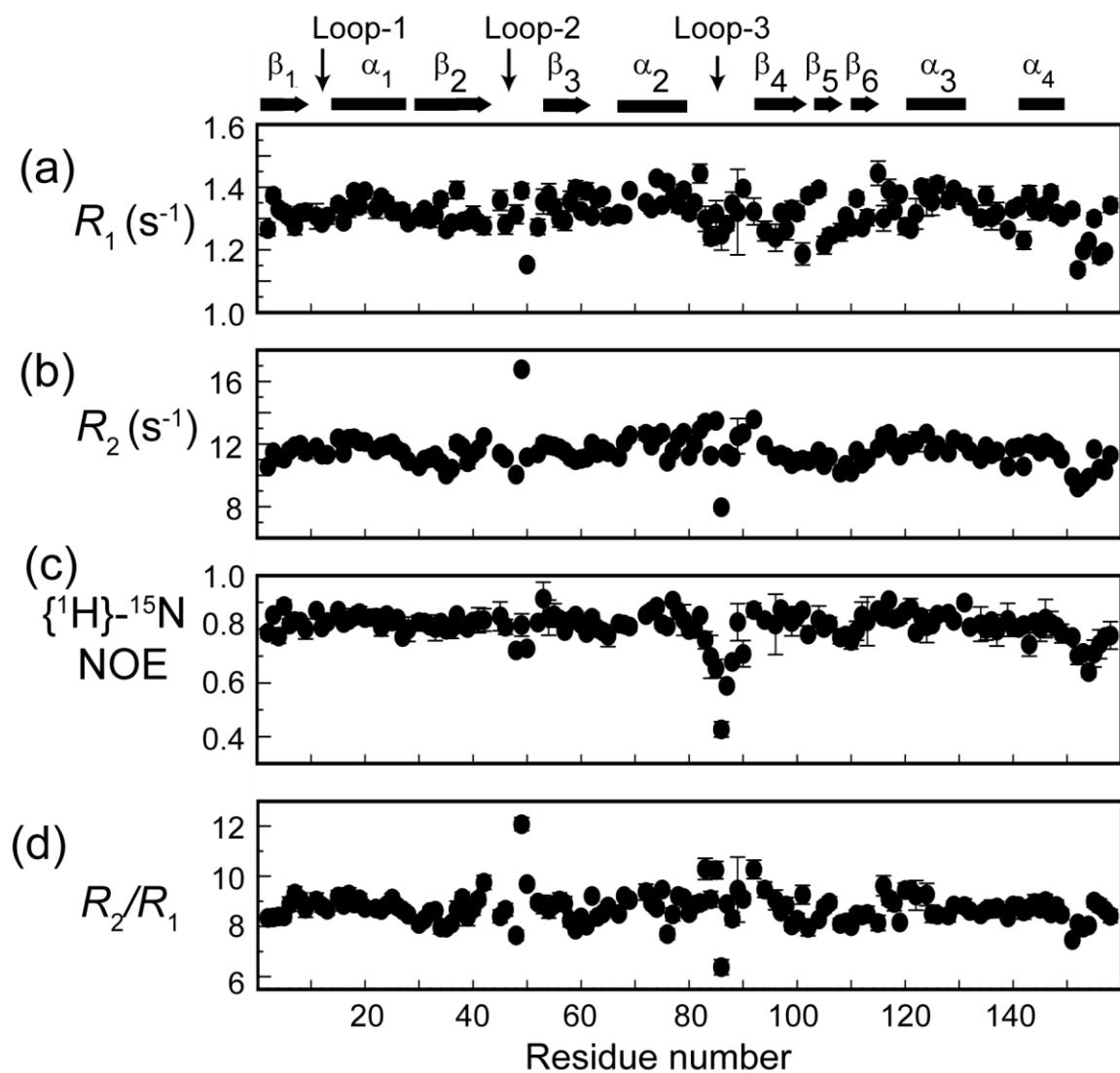


Figure 2.5. ^{15}N relaxation parameters obtained for the ternary complex

HPPK•MgAMPCPP•DMHP at 25 °C. The secondary structure is represented at the top of the figure. (a) R_1 relaxation rates; (b) R_2 relaxation rates; (c) $\{^1\text{H}\}\text{-}^{15}\text{N}$ NOE values; and (d) R_2/R_1 ratio.

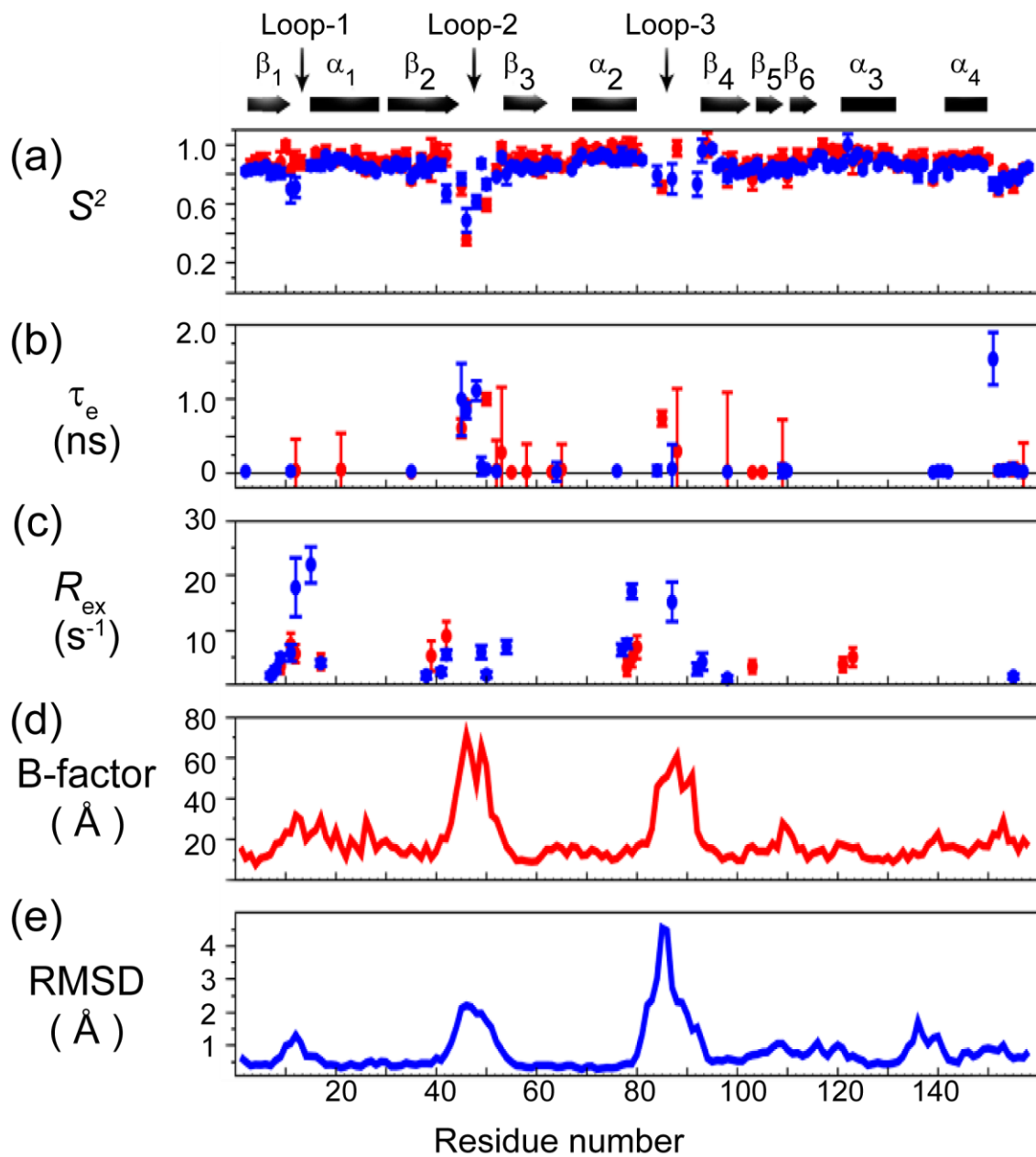


Figure 2.6. Comparison of the model-free parameters versus residue number between the apo HPPK (red) and the binary complex with MgAMPCPP (blue) at 0.5 mM protein concentration. The secondary structure is represented at the top of the figure. (a) Squared order parameter S^2 ; (b) correlation time (τ_e) of internal motion; (c) chemical exchange

Figure 2.6 (cont'd)

contribution R_{ex} ; (d) C α atom B-factors of the crystal structure of the apo HPPK (in red); and (e) backbone rms deviations of the NMR structure of the binary complex with MgAMPCPP (in blue).

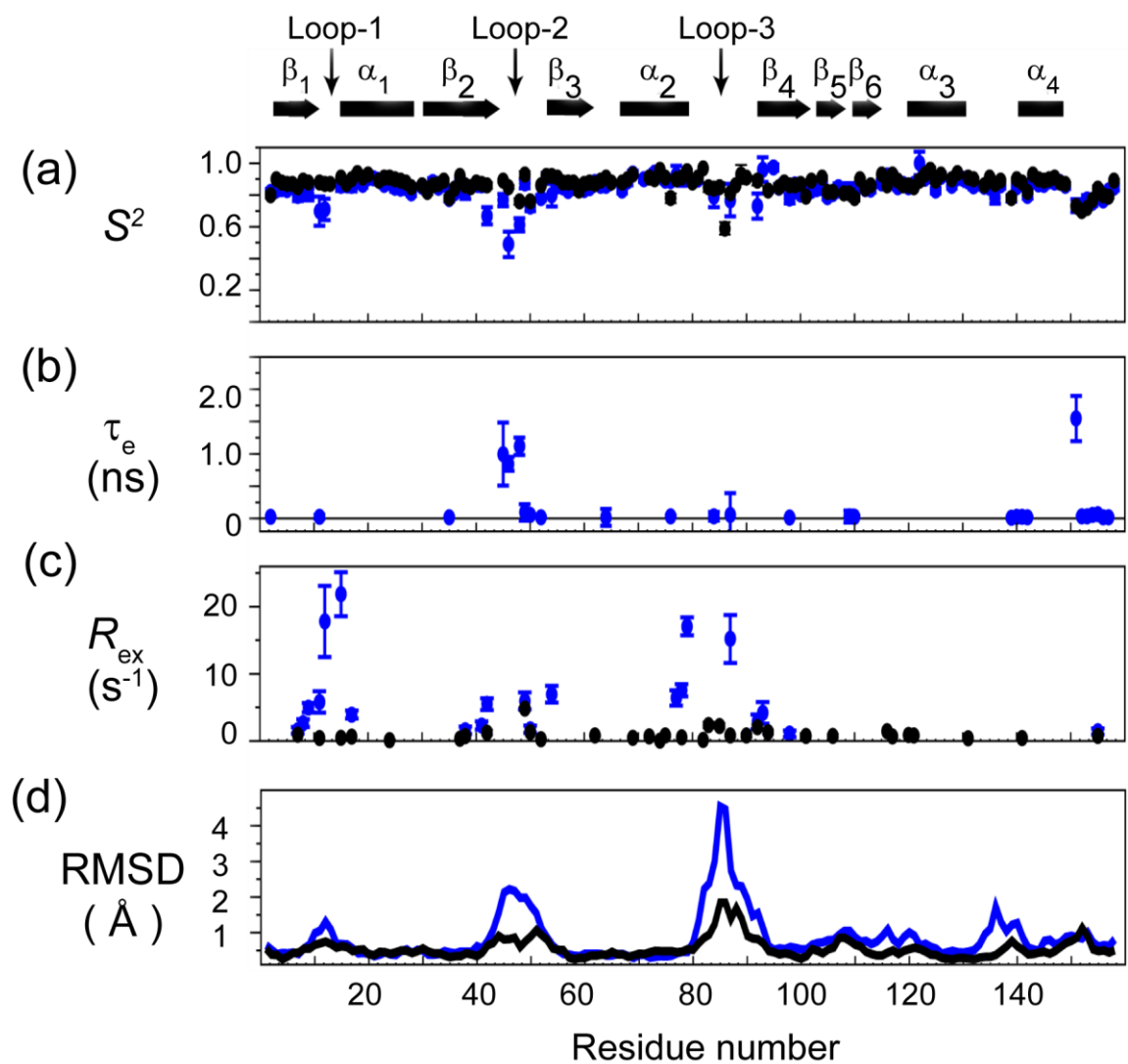


Figure 2.7. Comparison of the model-free parameters versus residue number between the binary complex HPPK•MgAMPCPP (blue) and the ternary complex HPPK•MgAMPCPP•DMHP (black) at 0.5 mM protein concentration. The secondary structure is represented at the top of the figure. (a) Squared order parameter S^2 ; (b) correlation time (τ_e) of internal motion; (c) chemical exchange contribution R_{ex} ; and (d) backbone rms deviations.

Figure 2.8

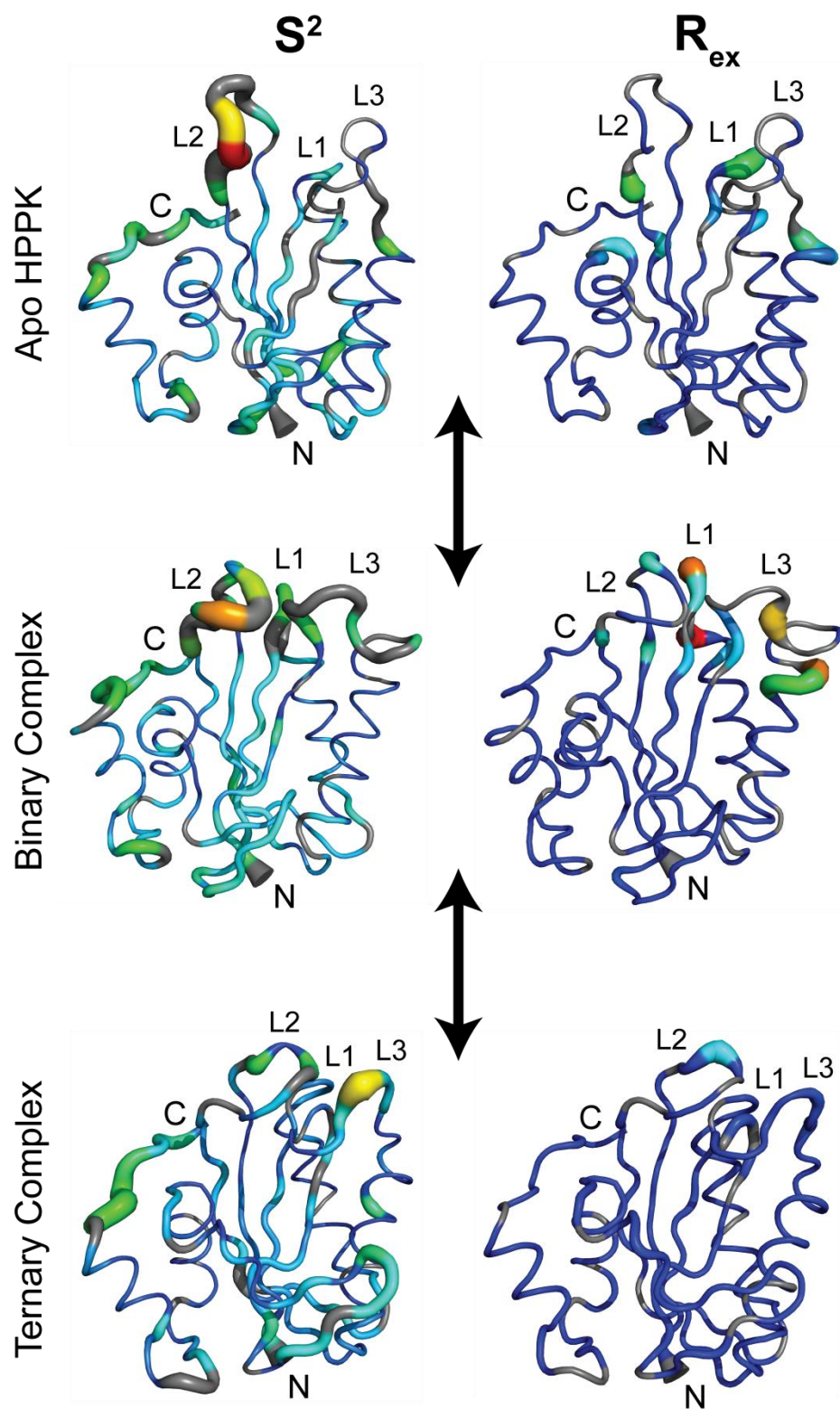


Figure 2.8 (cont'd)

Figure 2.8. Mapping of the dynamic parameters S^2 (left panel) and R_{ex} (right panel) to the structures of the apo HPPK (top), the binary complex (HPPK•MgAMPCPP, middle), and the ternary complex (HPPK•MgAMPCPP•DMHP, bottom). The S^2 values are scaled by a color gradient with the smallest value in red and the largest value in blue. The R_{ex} values are scaled by a color gradient with the smallest value in cyan and the largest value in red, and residues without an R_{ex} term are colored in blue. Residues without the measured dynamic parameters, i.e., 12 proline residues and other residues with missing or overlapping NH cross-peaks, are in grey. The figure was made with PyMOL (45).

References

REFERENCES

- 1 Bermingham, A., and Derrick, J. P. (2002) The folic acid biosynthesis pathway in bacteria: Evaluation of potential for antibacterial drug discovery, *Bioessays* 24, 637-648.
- 2 Blakley, R. L., and Benkovic, S. J. (1984) in *Folates and pterins*, John Wiley & Sons, New York.
- 3 Benkovic, S. J., and Hammes-Schiffer, S. (2003) A perspective on enzyme catalysis, *Science* 301, 1196-1202.
- 4 Boehr, D. D., Dyson, H. J., and Wright, P. E. (2006) An NMR perspective on enzyme dynamics, *Chem. Rev.* 106, 3055-3079.
- 5 Hammes, G. G. (2002) Multiple conformational changes in enzyme catalysis, *Biochemistry* 41, 8221-8228.
- 6 Hammes-Schiffer, S., and Benkovic, S. J. (2006) Relating protein motion to catalysis, *Annu. Rev. Biochem.* 75, 519-541.
- 7 Henzler-Wildman, K., and Kern, D. (2007) Dynamic personalities of proteins, *Nature* 450, 964-972.
- 8 Olsson, M. H. M., Parson, W. W., and Warshel, A. (2006) Dynamical contributions to enzyme catalysis: Critical tests of a popular hypothesis, *Chem. Rev.* 106, 1737-1756.
- 9 Saen-Oon, S., Ghanem, M., Schramm, V. L., and Schwartz, S. D. (2008) Remote mutations and active site dynamics correlate with catalytic properties of purine nucleoside phosphorylase, *Biophys. J.* 94, 4078-4088.
- 10 Xiao, B., Shi, G. B., Chen, X., Yan, H., and Ji, X. (1999) Crystal structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, a potential target for the development of novel antimicrobial agents, *Structure* 7, 489-496.
- 11 Stammers, D. K., Achari, A., Somers, D. O., Bryant, P. K., Rosemond, J., Scott, D. L., and Champness, J. N. (1999) 2.0 Å x-ray structure of the ternary complex of 7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase from *Escherichia coli* with atp and a substrate analogue, *FEBS Lett.* 456, 49-53.

- 12 Shi, G., Gao, J., and Yan, H. (1999) ^1H , ^{13}C and ^{15}N resonance assignments of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase and its complex with mgamppcp, *J. Biomol. NMR* 14, 189-190.
- 13 Shi, G., Gong, Y., Savchenko, A., Zeikus, J. G., Xiao, B., Ji, X., and Yan, H. (2000) Dissecting the nucleotide binding properties of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase with fluorescent 3'(2)'-o-anthraniloyladenine 5'-triphosphate, *Biochim. Biophys. Acta* 1478, 289-299.
- 14 Blaszczyk, J., Shi, G., Yan, H., and Ji, X. (2000) Catalytic center assembly of hppk as revealed by the crystal structure of a ternary complex at 1.25 Å resolution, *Structure* 8, 1049-1058.
- 15 Bermingham, A., Bottomley, J. R., Primrose, W. U., and Derrick, J. P. (2000) Equilibrium and kinetic studies of substrate binding to 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from *Escherichia coli*, *J. Biol. Chem.* 275, 17962-17967.
- 16 Xiao, B., Shi, G., Gao, J., Blaszczyk, J., Liu, Q., Ji, X., and Yan, H. (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.
- 17 Shi, G., Blaszczyk, J., Ji, X., and Yan, H. (2001) Bisubstrate analogue inhibitors of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: Synthesis and biochemical and crystallographic studies, *J. Med. Chem.* 44, 1364-1371.
- 18 Li, Y., Gong, Y., Shi, G., Blaszczyk, J., Ji, X., and Yan, H. (2002) Chemical transformation is not rate-limiting in the reaction catalyzed by *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, *Biochemistry* 41, 8777-8783.
- 19 Blaszczyk, J., Li, Y., Shi, G., Yan, H., and Ji, X. (2003) Dynamic roles of arginine residues 82 and 92 of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: Crystallographic studies, *Biochemistry* 42, 1573-1580.
- 20 Li, Y., Wu, Y., Blaszczyk, J., Ji, X., and Yan, H. (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.

- 21 Blaszczyk, J., Shi, G., Li, Y., Yan, H., and Ji, X. (2004) Reaction trajectory of pyrophosphoryl transfer catalyzed by 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, *Structure* 12, 467-475.
- 22 Blaszczyk, J., Li, Y., Wu, Y., Shi, G., Ji, X., and Yan, H. (2004) Essential roles of a dynamic loop in the catalysis of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, *Biochemistry* 43, 1469-1477.
- 23 Li, Y., Blaszczyk, J., Wu, Y., Shi, G., Ji, X., and Yan, H. (2005) Is the critical role of loop 3 of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase in catalysis due to loop 3 residues arginine-84 and tryptophan-89? Site-directed mutagenesis, biochemical, and crystallographic studies, *Biochemistry* 44, 8590-8599.
- 24 Li, G., Felczak, K., Shi, G., and Yan, H. (2006) Mechanism of the conformational transitions in 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as revealed by NMR spectroscopy, *Biochemistry* 45, 12573-12581.
- 25 Marley, J., Lu, M., and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins, *J. Biomol. NMR* 20, 71-75.
- 26 Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Prog. Nucl. Magn. Reson. Spectrosc.* 34, 93-158.
- 27 Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) Nmrpipe: A multidimensional spectral processing system based on unix pipes, *J. Biomol. NMR* 6, 277-93.
- 28 Johnson, B. A., and Blevins, R. A. (1994) Nmrview: A computer program for the visualization and analysis of NMR data, *J. Biomol. NMR* 4, 603-614.
- 29 Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed src homology 2 domain studied by ¹⁵N NMR relaxation, *Biochemistry* 33, 5984-6003.

- 30 Markley, J. L., Horsley, W. J., and Klein, M. P. (1971) Spin-lattice relaxation measurements in slowly relaxing complex spectra, *J. Chem. Phys.* 55, 3604-3605.
- 31 Fushman, D., Cahill, S., and Cowburn, D. (1997) The main-chain dynamics of the dynamin pleckstrin homology (ph) domain in solution: Analysis of N-15 relaxation with monomer/dimer equilibration, *J. Mol. Biol.* 266, 173-194.
- 32 Koradi, R., Billeter, M., and Wüthrich, K. (1996) Molmol: A program for display and analysis of macromolecular structures, *J. Mol. Graph.* 14, 51-55.
- 33 Walker, O., Varadan, R., and Fushman, D. (2004) Efficient and accurate determination of the overall rotational diffusion tensor of a molecule from N-15 relaxation data using computer program rotdif, *J. Magn. Reson.* 168, 336-345.
- 34 Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data, *J. Biomol. NMR* 16, 23-28.
- 35 de la Torre, J. G., Huertas, M. L., and Carrasco, B. (2000) Hydronmr: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations, *J. Magn. Reson.* 147, 138-146.
- 36 Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results, *J. Am. Chem. Soc.* 104, 4559-4570.
- 37 Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, *J. Am. Chem. Soc.* 104, 4546-4559.
- 38 Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) Deviations from the simple two-parameter model-free approach to the interpretation of N-15 nuclear magnetic-relaxation of proteins, *J. Am. Chem. Soc.* 112, 4989-4991.
- 39 Richey, D. P., and Brown, G. M. (1969) The biosynthesis of folic acid. IX. Purification and properties of the enzymes required for the formation of dihydropteroic acid, *J. Biol. Chem.* 244, 1582-1592.
- 40 Talarico, T. L., Dev, I. K., Dallas, W. S., Ferone, R., and Ray, P. H. (1991) Purification and partial characterization of 7,8-dihydro-6-hydroxymethylpterin-

pyrophosphokinase and 7,8-dihydropteroate synthase from *Escherichia coli* MC4100, *J. Bacteriol.* 173, 7029-32.

- 41 Talarico, T. L., Ray, P. H., Dev, I. K., Merrill, B. M., and Dallas, W. S. (1992) Cloning, sequence analysis, and overexpression of *Escherichia coli* *folK*, the gene coding for 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase, *J. Bacteriol.* 174, 5971-5977.
- 42 Schurr, J. M., Babcock, H. P., and Fujimoto, B. S. (1994) A test of the model-free formulas - effects of anisotropic rotational diffusion and dimerization, *J Magn. Reson. B* 105, 211-224.
- 43 Åkerud, T., Thulin, E., Van Etten, R. L., and Akke, M. (2002) Intramolecular dynamics of low molecular weight protein tyrosine phosphatase in monomer-dimer equilibrium studied by NMR: A model for changes in dynamics upon target binding, *J. Mol. Biol.* 322, 137-152.
- 44 Palmer, A. G., Kroenke, C. D., and Loria, J. P. (2001) Nuclear magnetic resonance methods for quantifying microsecond- to-millisecond motions in biological macromolecules, *Methods Enzymol.* 339, 204-238.
- 45 DeLano, W. L. (2002), DeLano Scientific, Palo Alto, CA, USA.

CHAPTER 3

Nuclear Magnetic Resonance and Molecular Dynamics

Studies of the N10A and Q50A HPPK

Introduction

Structural studies of *E. coli* HPPK show that it goes through dramatic conformational changes in three loops during its catalytic cycle (Figures 1.6 and 3.1) (1-3). Several lines of evidence indicate that the conformational changes in HPPK are important for its catalysis. First, kinetic studies have shown that the product release is the rate-limiting step in HPPK-catalyzed reaction (4). For the product to be released, loops 2 and 3 must move away to unseal the active center (5). Second, the conformational changes are necessary to bring catalytic residues R82 and R92 into the active center for catalysis (5, 6). Third, the essential roles of loop 3 in catalysis are further supported by a study of the loop 3 deletion mutation (7). Compared with the WT HPPK, the mutant enzyme has a 100-fold higher K_d for HP. The rate-limiting step shifts to the chemical step and the rate constant for the chemical step decreases by a factor of 1.1×10^5 . The crystal structure of the mutant enzyme shows that the catalytic center is not fully assembled and sealed. Fourth, site-directed mutagenesis, biochemical and X-ray crystallographic studies have shown that the critical functional role of loop 3 is not due to the interactions of R84 or W89 with the substrates (8). In sum, these results indicate that the loops rather than individual residues play important functional roles.

The three flexible loops are coupled together in the HPPK substrate ternary complex through a hydrogen bond network (2). As shown in Figure 3.2, this network includes N10 in loop 1, Q50 and P47 in loop 2, and W89, P91 and R92 in loop 3. Central

to the hydrogen bond network are N10 and Q50, whose side-chain amides are involved in 6 of the 7 hydrogen bonds. This network is not present in the ligand-free (apo) HPPK. Structural studies show that the active center of HPPK is assembled and sealed by the coupling of the three loops via the hydrogen bond network upon the binding of both substrates.

To determine the roles of the hydrogen bond network in the assembly of the active center and catalysis, N10 and Q50 have been replaced with alanine individually. Kinetic studies of the Q50A and N10A HPPK mutant enzymes have been completed and the results are listed in Table 3.1. The binding affinity does not change significantly for the first substrate, ATP, with K_d increased 1.4 fold over WT for the N10A HPPK and 1.5 fold for Q50A HPPK. The K_d for the second substrate, HP, increased 12 fold and 230 fold for the Q50A and N10A HPPK, respectively. The binding studies of WT HPPK have shown that it has very weak binding affinity towards HP without MgATP binding first, indicating certain conformational changes are required for efficient HP binding. The equilibrium constant for the chemical step is similar between WT and Q50A mutant. However, both the forward (k_3) and the reverse (k_{-3}) rate constants are significantly smaller for the Q50A mutant. Product release is still the rate limiting step for the Q50A HPPK catalyzed reaction while the rate limiting step is the chemical step for the N10A HPPK. Without the hydrogen bond network, it seems that the catalytically efficient

complex cannot be properly formed, thus leads to a decreased binding affinity of substrate and a lower catalytic efficiency.

The structures of the two mutant enzymes have been determined by X-ray crystallography using AMPCPP (an ATP analogue) and HP (personal communication with Jaroslaw Blaszczyk). However, none of the crystal structures contains the substrate HP. For the N10A HPPK, the HP binding site is occupied by an acetate ion. For the Q50A HPPK, the HP binding site is occupied by an acetate ion or a pterin derivative with equal probabilities. The conformations of the three loops in both structures are similar to those in the crystal structure of the HPPK•MgADP. In particular, loop 3 is in a super-open conformation, very different from the closed conformation in the ternary substrate complex HPPK•HP•MgAMPCPP.

Since the crystallization failed to get the ternary complex for either N10A or Q50 HPPK, we turned to an NMR approach to assess the structural effects of both mutations. Furthermore, we investigated the dynamic properties of the two mutant enzymes by ^{15}N - NMR relaxation studies and MD simulations to assess the effects of the mutations on the dynamic properties of HPPK.

Materials and Methods

Sample Preparation. Uniformly ^{15}N -labeled *E. coli* HPPK mutants (N10A, Q50A) were purified according to the protocol developed in our lab (9). The NMR sample of the apo HPPK was prepared by dissolving lyophilized protein powder into 20

mM sodium phosphate at pH 7.4 with 5% D₂O/95% H₂O. The NMR sample of the HPPK ternary complex was prepared by adding a small amount of (~0.5 mg) HDMP into an HPPK ·AMPCPP ·Mg²⁺ solution and an overnight incubation at 4 °C overnight. Undissolved HDMP was removed by centrifugation. The NMR samples contained 0.5 mM HPPK protein, 5 mM AMPCPP, ~2 mM HDMP, and 20 mM MgCl₂. Assuming that the affinity of HDMP for HPPK is similar to that of HP, calculations based on the dissociation constants of HP suggest that HPPK ternary complex is the dominant form (>99% for Q50A HPPK and >97% for N10A HPPK).

Spectral Assignment. The resonance assignment for the Q50A HPPK was obtained by analysis of the 3D ¹⁵N-dispersed NOESY spectrum and comparison to the NMR spectra of the WT HPPK. For the N10A mutant enzyme, HNCACB, CBCA(CO)NH and ¹⁵N - dispersed NOESY spectra were used for the resonance assignment.

¹⁵N NMR Relaxation Experiments. NMR relaxation experiments were carried out at 25 °C on a Varian Unity INOVA 600 MHz NMR spectrometer equipped with 5mm triple-resonance probes and z-axis pulsed field gradients. The pulse sequences developed by Farrow *et al* (10) were used for measuring of R_1 , R_2 and ¹H-¹⁵N heteronuclear steady-state NOEs. For the R_1 measurement, nine different relaxation delay times were used: 11.1, 44.4, 166.5, 388.5, 666, 943.5, 110.0, 1443.0 and 1998 ms. The second time point was repeated for error estimation. Nine different relaxation delay times

were used for R_2 measurement: 0, 17.4, 34.8, 52.2, 69.6, 87.0, 104.4, 121.9 and 139.3 ms. The third time point was repeated for error estimation. A recycle delay of 2.5 s was used for both R_1 and R_2 measurements. Two experiments were performed for the measurement of ^1H - ^{15}N steady-state NOE: one with 3 s proton pre-saturation following a 5 s recycle delay and the other without pre-saturation but a 8 s total recycle delay instead. NOEs were derived from the ratio of peak intensities of these two spectra. Duplicate spectra were acquired for error estimation. All the spectra were recorded with 1944 complex data points in the proton dimension and 100 in the nitrogen dimension. The number of transients was 16 for the R_1 , R_2 measurements and 64 for the ^1H - ^{15}N steady-state NOE measurements.

The relaxation data were processed with NMRPipe (11) and analyzed with NMRView (12). A Lorentz-to-Gauss window function was applied to all spectra. Zero-filling was used to make 4096 total points in the proton dimension and 1024 in the nitrogen dimension. The peak intensities were extracted using the rate analysis subroutine of NMRview. The exponential decay curves were fitted to a two parameter exponential equation using the program CURVEFIT, available from the website of Palmer's group. The HetNOE subroutine of NMRView was used to calculate the ^1H - ^{15}N steady-state NOEs.

Rotational Diffusion and Model-Free Analysis. Rotational diffusion was analyzed using the R_2/R_1 ratio of residues chosen by following criteria. First, the residue

is from a well-organized secondary structure. Second, the NOE is no less than 0.65. Third, the R_2/R_1 ratio is within one standard deviation of the mean value. The mean structure of the NMR ensembles of the HPPK ternary complex, HPPK•MgAMPCPP•DMHP (PDB code 2F63) (13), was taken as the representative structure. Protons were added using tleap module in Amber10. Tensor2 (14) was used to optimize the rotational diffusion tensor against the experimental data. HydroNMR (15) was used to predict the rotational diffusion tensor based on the protein structure.

The microdynamic parameters under the Lipari-Szabo formalism (16-18) were extracted using the TENSOR2 program (14). Briefly, the ^{15}N relaxation data were interpreted in terms of motion of the N-H bond. Five models of increasing complexity were tested, including model 1 (S^2), model 2 (S^2, τ_e), model 3 (S^2, R_{ex}), model 4 ($S^2, \tau_e, R_{\text{ex}}$), and model 5 (S^2, τ_e, S_f^2), where S^2 is the squared order parameter, τ_e is the correlation time describing the internal motion, S_f^2 is the second order parameter describing the fast internal motion, and R_{ex} is the chemical exchange contribution to R_2 . *F*-test was then used to select the model that satisfies the data with the lowest number of parameters. Standard errors in the model-free parameters were evaluated by 100 Monte Carlo simulations.

$^3\text{J}_{\text{HNHA}}$ (J-coupling constant between the amide proton and CA proton)

Measurement. The standard HNHA pulse sequence (19) was used to collect the spectrum. The spectra were processed following the aforementioned protocol. Peak

position and intensity were located using NMRview Peakpick macro. $^3J_{\text{HNHA}}$ constants were estimated using the equation:

$$S_{\text{cross}}/S_{\text{diag}} = -\tan^2(2\pi J_{\text{HN}}\xi)$$

where S_{cross} and S_{diag} are the cross peak and the diagonal peak intensity, respectively, and 2ξ represents the active time of $^3J_{\text{HNHA}}$ in the pulse sequence.

Molecular Dynamics Simulation. MD simulations were performed using the program Amber 10 with the AMBER03 force field. Two sets of 50 ns simulation data were acquired for the Q50A HPPK, and four sets of 50 ns simulation data and two sets of 20 ns simulation data were acquired for the N10A HPPK. The X-ray crystal structure of the *E coli* HPPK ternary complex (PDB ID: 1f9y) was used as the starting point for MD simulations. This complex contains Mg, HP and AMPCPP. The WT substrate ternary complex was generated by replacing AMPCPP with ATP. Q50 and N10 were substituted with alanine virtually to create the mutant HPPK ternary complexes. The standard parameters of ATP in AMBER were used during simulations. The force field parameters for HP were derived according to the following protocol. First, the HP coordinate was obtained from the X-ray crystal structure and hydrogen atoms were added manually. Then the structure was optimized and atomic charges calculated at the HF/6-31G* using the Gaussian03 program. The force field parameters were obtained using the antechamber module in Amber 10. The interactions with histidine residues were carefully examined

and the protonation at the ϵ position was used. The default protonation states under neutral pH for all other residues were used without further consideration.

The fully assembled HPPK substrate ternary complex was neutralized with sodium ions and solvated by a periodic box of TIP3P water molecules that extended at least 12 Å from the protein atoms. The solvated complex was subjected to 1000 steps of steepest descent energy minimization followed by 1500 steps conjugated gradient minimization, where the solute atoms, including both the protein and ligands, were fixed by a harmonic potential with a force constant of 50.0 kcal/Å. The MD dynamics simulation was carried out using the PMEMD module in Amber 10. The simulations were initially run at a constant volume for 20 ps to heat up the system from 0 K to 300 K, then at constant pressure for 20 ps to allow the adjustment of solvent density. The SHAKE algorithm (20) was used to constrain all bond lengths involving hydrogen atoms, permitting a 2 fs time step. A non-bonded pair list cut off of 12.0 Å was used, and the non-bonded pair list was updated every 25 steps. The particle mesh Ewald method (21) was used to evaluate the contribution of long-range electrostatic interactions. The temperature was controlled by Langevin thermostat (22) with a collision frequency of 5 ps⁻¹. This temperature controlled method is significantly more efficient at equilibrating the system temperature than the Berendsen temperature-coupling scheme (23) that was the recommended method for older versions of AMBER. The coordinates were saved every 2 ps.

The first 20 ns of the 50 ns trajectories were considered equilibration time and the rest were used for further analysis, including hydrogen bond, RMSF analysis and clustering. The shorter simulation was also analyzed and the results were similar to the longer one if not mentioned. The PTRAJ module in AMBER tool 1.4 was for data analysis. VMD 1.8.6 (24) was used to visualize the structure and trajectory.

Order Parameter Calculation. MD simulation and NMR spin relaxation spectroscopy are complementary tools to investigate the dynamics of bio-molecules. Comparison of order parameters from experiments and MD simulation is increasingly used to validate the conformational sampling of MD simulations. Traditionally, order parameters were calculated using the autocorrelation function approach where the global tumbling was removed by superimposing MD trajectory to the reference structure. Recently, the isotropic reorientational eigenmode dynamics (IRED) approach (25) was introduced to calculate the order parameters based on MD simulation. It has two advantages over the autocorrelation time approach: first, there is no need to separate the overall tumbling from the internal motion. Second, the snapshots generated from the MD trajectory are treated analytically to yield an isotropic ensemble. Order parameters were calculated using both methods and no significant difference was found. Only the order parameters obtained from the IRED approach were reported. It was found that the order parameters computed over the trajectory longer than the overall correlation tumbling time could include motions that would not be reflected in the experimental order parameter

values (26), thus the order parameters were calculated using a 5 ns time window based on our trial results. The final order parameters were the average values from these calculations.

Hydrogen Bond Analysis. The hydrogen bond analysis was carried out using the PTRAJ module in Amber 10. The criteria for hydrogen bond analysis were the distance between hydrogen donor and acceptor less than 3.5 Å and the angle formed by the donor, hydrogen and acceptor between 120 ° and 180 °. Only hydrogen bonds with occupancies larger than 5% were reported.

Results

NMR Chemical Shift Analysis. Chemical shift is very sensitive to local chemical environments. The ^{15}N - ^1H chemical shift changes between WT and mutant enzymes can be used as an indicator of their structural differences. Figure 3.3 shows ^1H - ^{15}N HSQC spectra of the apo Q50A and the Q50A ternary complex, each overlaid with their corresponding WT HPPK spectra. Overall, the paired spectra show a good overlap, which indicates that the perturbations of the Q50A substitution are small for both the apo form and the ternary complex.

Figure 3.4 shows the overlay of the spectra of N10A HPPK with the corresponding ones of the WT HPPK. The spectrum of the apo form matches well with the WT spectrum as in the case of Q50A HPPK. However, the spectrum for the ternary complexes changes significantly. There are some minor peaks in the N10A ternary

complex spectrum and the results from ZZ-exchange experiments (32) have shown that these peaks are not due to slow exchange.

The chemical shift changes (Figure 3.5) were calculated using the following equation:

$$\Delta\text{NH} = \sqrt{\left(\frac{\Delta\text{N}}{10}\right)^2 + (\Delta\text{H})^2}$$

where ΔH , ΔN , and ΔNH are ^1H , ^{15}N , and weighted chemical shift changes, respectively.

For the apo HPPK, the plotted chemical shifts are mainly for the residues in the core region, because the resonances of most of residues in the three loops are not observed.

For the ternary complex, in contrast, resonances are observed for almost all the residues including those in the three regions. For most residues in the core region, the chemical shift difference between the WT HPPK and the mutant enzymes is less than 0.1 ppm for both the apo form and the ternary complex. In contrast, all three loops show some large chemical shift changes in the ternary complexes of the mutant enzymes. The results indicate that the conformations of the three loops are significantly different between the WT HPPK and the mutant enzymes.

$^3\text{J}_{\text{HNHA}}$ Measurement and Dihedral Angle Analysis. The homonuclear three-bond $^3\text{J}_{\text{HNHA}}$ coupling constant is related to the backbone dihedral angle ϕ in the protein and can be described by the Karplus equation:

$$J = A \cos^2 \phi + B \cos \phi + C$$

Therefore, $^3J_{\text{HNHA}}$ values were measured for the WT HPPK, N10A and the Q50A HPPK substrate ternary complexes to assess the effects of the mutations on the conformation of HPPK. AMPCPP and HDMP were used to mimic substrates ATP and HP, respectively. $^3J_{\text{HNHA}}$ was measured for 110, 111 and 93 residues of the WT, Q50A and N10A protein complexes, respectively, with other $^3J_{\text{HNHA}}$ values not obtained due to weak or missing peaks. Figure 3.6 shows the correlation of $^3J_{\text{HNHA}}$ between Q50A and WT. It is clear that there is high correlation between these two and the linear fitting gives a slope of 0.98 ± 0.02 , which is a clear indication of high structural similarity. The correlation of $^3J_{\text{HNHA}}$ between N10A and WT HPPK were also plotted in Figure 3.6 and a slope of 0.95 ± 0.02 was obtained. These results are consistent with those of the chemical shift analysis, both indicating that the conformations of the mutant enzymes are very similar to that of the WT enzyme in the core region.

With respect to the three loops, the analysis was focused on the ternary complex form. $^3J_{\text{HNHA}}$ could be measured for 14 residues in the three loops for the WT ternary complex, 6 residues for the N10A ternary complex, and 14 residues for the Q50A residues. The measured $^3J_{\text{HNHA}}$ coupling constants for the loop residues in all three HPPK complexes are listed in table 3.3. Of the 12 loop residues measured for the WT complex, 12 were also measured for the Q50A complex but only four measured for the N10A complex. Of the 12 residues measured for both the WT and the Q50A complex, two residues have significantly different $^3J_{\text{HNHA}}$ values (>1 Hz) and both are located in

loop 2 (D52 and Y53). Of the four residues measured for both the WT and the N10A complex, three residues have significantly different $^3J_{\text{HNHA}}$ values, one in loop 1 (L11) and two in loop 2 (D49 and Q50). The different $^3J_{\text{HNHA}}$ values indicate that these residues adopt different ϕ dihedral angles, suggesting that the conformations of loops 1 and 2 of the N10A complex and loop 2 of the Q50A complex are significantly different from those of the WT complex.

Measured ^{15}N Relaxation Parameters. ^{15}N relaxation data could be obtained for 138 and 135 residues from 146 non-proline residues for Q50A and Q10A HPPK ternary complexes, respectively. Figure 3.7 summarizes the relaxation for the Q50A and the N10A ternary complex along with the data for WT HPPK ternary complex for comparison. The mean R_1 , R_2 and NOE value for the Q50A ternary complex are $1.30 \pm 0.05 \text{ s}^{-1}$, $12.38 \pm 1.35 \text{ s}^{-1}$ and 0.80 ± 0.07 , respectively. They are $1.12 \pm 0.04 \text{ s}^{-1}$, $13.06 \pm 0.53 \text{ s}^{-1}$ and 0.75 ± 0.07 , for the N10A HPPK and $1.32 \pm 0.055 \text{ s}^{-1}$, $11.51 \pm 0.94 \text{ s}^{-1}$ and 0.82 ± 0.07 for WT HPPK, respectively

The averaged R_2/R_1 ratio of the selected residues based on the criteria as described in the methods was used to estimate the isotropic correlation time of the protein. The estimated total correlation time is $9.29 \pm 0.01 \text{ ns}$ and $10.29 \pm 0.01 \text{ ns}$ for Q50A and N10A HPPK ternary complex, respectively. The predicted correlation time based on WT HPPK ternary complex structure is 10.75 ns . The smaller correlation time indicates that the monomer form is dominant in the solution. The ratio of the estimated principal

components is: 1: 0.903:0.977 and 1:0.85:0.90 for Q50A and the N10A HPPK, respectively. They are similar to the ratio calculated using the mean NMR structure of the ternary complex (1:0.86:0.92).

Model-free Analysis. The results of the model-free analysis were plotted in Figure 3.8. Model-free parameters could be extracted for 129 residues from both mutant enzyme ternary complexes. For the Q50A ternary complex, the mean S^2 value is 0.90 ± 0.04 . 12 residues are one standard deviation of the mean value (residues 2, 30, 46, 50, 83, 86, 87, 96, 100, 105, 106, 108, 122, 142, 151 and 156). Among them, residue 46 and 50 are in loop 2, residue 84, 86 and 87 are in loop 3, residue 100, 105, 106 and 108 are in the region close to the β -sheet 5 and 6, and residue 151 and 156 are in the C-terminal region. This pattern is similar to that of WT HPPK although there are some differences for specific residues.

An R_{ex} term is required for 43 residues to get good fitting for the Q50A ternary complex, indicating that Q50A ternary complex is quite dynamic on the μ s to ms timescale, because the R_{ex} term describes the dynamics on this timescale. Of these 43 residues, 13 (residues 15, 46, 53, 82, 83, 85, 87, 88, 89, 90, 92, 93, and 94) have significant R_{ex} (>1.5 Hz). For the WT ternary complex, this number is 4. The increase in number of residues with R_{ex} is an indication that Q50A HPPK is more dynamic on the μ s-ms timescale compared to WT protein. Among them, 10 residues are in loop 3 and 2 residues are in loop 2.

For the N10A HPPK ternary complex, the mean S^2 value is 0.88 ± 0.03 , similar to those of WT and Q50A HPPK. There are 31 residues with S^2 one standard deviation below the mean. Among these residues, 2 are near loop 1, 6 are in or near loop 2, 4 residues are in loop 3, 6 are in the loop connecting β -strand 5 and 6, and 8 are in the C-terminal region. This pattern is similar to those of the WT and Q50A HPPK ternary complexes.

An R_{ex} term is required for 53 residues in the N10A ternary complex to get good fitting. Among them, 20 have a significant R_{ex} (>1.5 Hz). This number is 13 and 4 for the Q50A and the WT HPPK ternary complex, respectively. The increased number of residues with an R_{ex} term indicates that the N10A complex is more dynamic on the μ s-ms timescale compared to WT. Among them, 9 are in loop 3, 4 are in loop 2 and 4 are in loop 1.

Compared with WT HPPK ternary complex, both N10A and Q50A HPPK ternary complexes have more residues with significant R_{ex} values, suggesting that both N10A and Q50A ternary complexes are more flexible than the WT counterpart. It is clear that for both mutants, the residues with significant R_{ex} are mainly located in the catalytic regions. The R_{ex} values of N10A HPPK are much larger compared to those of the Q50A protein. There is no significant difference between these two mutant proteins in terms of the rest of the protein, which are mainly α -helices and β -sheets. This result is consistent

with the X-ray crystal structure of HPPK which shows that the core structure is very stable to either binding or mutation.

MD Simulation Analysis and Order Parameter Comparison. MD simulation can provide more detailed information about dynamic changes that cannot be obtained from a NMR experiment. Figure 3.9 shows the root mean square fluctuation (RMSF) plots calculated based on the MD trajectories of the WT and the mutant enzyme ternary complexes. The RMSF patterns are similar. The mean RMSF are 0.695, 0.704 and 0.631 Å for the WT, the Q50A and the N10A ternary complex, respectively. Surprisingly, the mean RMSF of the N10A complex is slightly smaller than that of the WT complex. This difference in mean RMSF is mainly due to lower RMSFs in some C-terminal regions, particularly residues 115-120. It should be noted that the RMSFs of the three loops important for catalysis are similar between the mutant and the WT HPPK complex. Interestingly, loop 2 of the Q50A complex has significantly higher RMSFs but residue 86 in loop 3 has a significantly lower RMSF. However, loop 3 of the Q50A complex remains flexible as it has high RMSFs.

The backbone ^{15}N - ^1H order parameters (S^2) determined from the MD simulations and the model-free analysis of the NMR relaxation data are plotted in Figure 3.9A and 3.9B for the Q50A and the N10A ternary complex, respectively. The order parameters calculated from the MD simulation data and measured by the model-free analysis of the NMR relaxation data are similar. The average order parameters from the

MD simulation data are 0.850 and 0.845 for the Q50A and N10A complex, respectively, comparable to the averages of the experimentally measured values of 0.90 and 0.88 for the Q50A and N10A complex. It is well known that the order parameters calculated from MD simulation tend to be smaller than experimental values in loops and other flexible regions, which is the cause of the lower average calculated order parameters.

Hydrogen Bond Analysis. Hydrogen bond analysis was carried out on the MD simulation data using the PTRAJ module of the program AmberTools (version 1.4) and the results are summarized in Table 3.2. The hydrogen bond network connecting the three loops together in the WT HPPK ternary complex is well conserved during the MD simulation. There are a total of five hydrogen bonds between the loops, one between N10 and Q50, three between N10 and residues P51 and R92, and one between Q50 and W89, all with high percentages of occurrence. Substitution of N10 with alanine abolishes all hydrogen bonds involving the side-chain amide of N10, three of the five hydrogen bonds that connect the three loops, two between loops 1 and 2 and one between loops 1 and 3. No new hydrogen bonds are observed, but the three loops are still connected by two hydrogen bonds, one between loops 1 and 3 and the other between loops 2 and 3. Substitution of Q50 with alanine removes all hydrogen bonds involving the side-chain amide of Q50, including two of the five hydrogen bonds between the three loops, and no new hydrogen bonds are observed. The mutation eliminates the sole hydrogen bond between loops 2 and 3, but loops 1 and 3 are still connected by one hydrogen bond and

loops 1 and 3 by two hydrogen bonds. There are five hydrogen bonds between residues of loop 2 in the WT HPPK ternary complex. Three of them are formed between Q50 and P47/Q48/Y53 and two formed between Q48 and P44/G46. All of them are well conserved during the MD simulation of the WT ternary complex. Among these hydrogen bonds, only two of them involve the side-chain amide of Q50 and the rest are expected to be conserved in the Q50A ternary complex. However, none of them is conserved during MD simulation of the Q50A complex, with the largest occurrence at only ~12%. The loss of these hydrogen bonds might also contribute to the increased fluctuation of loop 2. In contrast, all of these hydrogen bonds are conserved in the MD simulation of the N10A complex.

Discussion

The active center of HPPK is fully assembled and some of the important catalytic residues move into the active center only after both substrates bind to the enzyme forming a ternary substrate complex and loops 2 and 3 move in to seal the active center. The NMR studies, both the chemical shift mapping and the J-coupling constant analysis, indicate that both N10 and Q50 are important for the assembling of the active center, because the core structures of both mutant enzymes are nearly the same as that of the WT enzyme but the conformations of the three loops are significantly different between the mutant enzymes and the WT enzyme. In support of this interpretation, the chemical shifts of the core residues are very similar between the N10A and the WT HPPK and

essentially the same between the Q50A and the WT HPPK (Figures 3.3-3.5). The $^3J_{\text{HNHA}}$ constants of the mutant enzymes are correlated well with those of the WT enzyme (Figure 3.6). The results are consistent with those of the crystallographic analysis, indicating that both mutations do not perturb the core structure of the enzyme. The NMR data cannot tell the conformations of the three loops in the apo form, as most residues of the three loops cannot be observed in this form of the enzyme. In the ternary complex form, the chemical shifts of the three loops are significantly different between the mutant enzymes and the WT enzyme, particularly between the N10A mutant and the WT enzyme. There are also significant differences in the J-coupling constants of the three loops between the mutant enzymes and the WT enzyme. Taken together, the data indicate that the three loops of the mutant enzyme ternary complexes assume conformations different from those of the WT enzyme.

Both mutations significantly affect the dynamic properties of HPPK. Motions on the ps-ns timescale are similar between the mutant enzymes and the WT enzyme based on the order parameters measured by NMR and calculated from the MD simulation data. However, both mutant enzymes, particularly the three loops, are more mobile on the μ s-ms timescale based on two lines of evidence. First, both mutant enzymes have more missing or weak peaks, which are most likely due to conformational exchange on the μ s-ms timescale. Second, significantly more residues require an R_{ex} term for the model-free analysis of the mutant enzyme ternary complexes to account for conformational exchange

on the μ s-ms timescale, all of which are located in the three loop regions. The MD simulations are of limited usage in this regard, as the simulation times are far from reaching the μ s-ms timescale, because of technological limitations.

The effects on the conformations and dynamic properties caused by the N10A mutation are more severe than those caused by the Q50A mutation based on the NMR studies, including the extent of chemical shift perturbations, the number of missing and weak peaks, and the number residues requiring an R_{ex} term for the model-free analysis. This is consistent with the roles of the two residues in the hydrogen bond network that connects the three loops. The side-chain amide of N10 is involved in three of the five hydrogen bonds in the hydrogen bond network and that of Q50 only two. Consequently, the N10A mutation causes a more severe weakening or disruption of the hydrogen bond network and the coupling of three loops than the Q50A mutation and therefore more severe effects on the conformations and dynamical properties of the three loops. Most likely the three loops are in equilibrium between active and inactive conformations. In the WT enzyme, the active conformation is stabilized by the hydrogen bond network anchored by N10 and Q50 after the formation of the ternary substrate complex. The predominant conformation is active for catalysis. In the mutant enzymes, because of the weakening or disruption of the hydrogen bond network that connect the three loops, the equilibrium shifts to an inactive conformation or inactive conformations. The equilibrium shift is more dramatic in the N10A mutant enzyme than in the Q50A mutant enzyme as

revealed by the NMR data, causing more significant changes in the thermodynamic and kinetic parameters of the N10A mutant enzyme (Table 3.1). Since either mutation causes a significant decrease in the affinity of MgATP for the enzyme, the loop coupling is not important for the binding of the first substrate. Because both mutations cause a significant decrease both in the affinity of HP for the enzyme and in the rate constant for the chemical step, the loop coupling is important for the binding of the second substrate and the full assembling of the active center and catalysis. These roles of the loop coupling are consistent with the results of a previous mutagenesis study removing a significant part of loop 3 as described in the Introduction section.

In conclusion, both N10 and Q50 are important for connecting the three catalytic loops and the loop coupling is important for the binding of the second substrate HP and the full assembling and stabilization of the active center and catalysis.

Table 3.1. Thermodynamic and kinetic parameters of WT, Q50A and N10A HPPK

	$K_d(\text{MgATP}) (\mu\text{M})$	$K_d(\text{HP}) (\mu\text{M})$	$k_3 (\text{s}^{-1})$	$k_{-3} (\text{s}^{-1})$	$k_4 (\text{s}^{-1})$
WT	2.6 ± 0.06	0.17 ± 0.01	16	20	1.8
Q50A	2.0 ± 0.2	2.1 ± 0.2	2.3	3.3	2.2
N10A	3.7 ± 0.2	39 ± 2	0.072	*	*

*: not applicable

Table 3.2. Hydrogen bonds between the three catalytic loops observed during the MD simulations

	WT		Q50A		N10A		Distance in WT HPPK X-ray structure (Å)
Hydrogen bond	Occurrence (%)	Distance (Å)	Occurrence (%)	Distance (Å)	Occurrence (%)	Distance (Å)	
N10-OD1... R92-NH	99.42	3.01 ±0.14	97.6	3.01 ±0.15			2.87
P51-O...N10-NH ₂	99.36	2.98 ±0.16	96.28	2.97 ±0.16			2.9
Q50-OE1...N10-NH ₂	91.1	3.06 ±0.18					2.91
R92-O...N10-NH	89.23	3.22 ±0.15	82.93	3.23 ±0.15	86.64	3.08 ±0.14	3.01
P47-O...Q50-NH ₂	98.84	2.96 ±0.15			99.06	2.96 ±0.15	2.86
W89-O...50-NH ₂	97.29	2.99 ±0.17			69.2	3.04 ±0.18	2.82
Q48-O...50-NH	80.34	2.99 ±0.16	7.4	3.15 ±0.19	73.91	2.98 ±0.17	3.15
Q50-OE1...53-OH	78.01	3.05 ±0.23			97.66	2.87 ±0.19	2.81
P47-O...Q50-NH ₂	98.84	2.96 ±0.15			99.06	2.96 ±0.15	2.86
52-O...N10-NH ₂			1.24	3.00 ±0.16			
Q48-NH ₂ ...P44O	70.86	2.95 ±0.17	11.80	2.97 ±0.17	42.14	2.96 ±0.16	
Q48-NH ₂ ...G46O	67.31	3.01 ±0.18	11.56	3.01 ±0.18	41.75	2.98 ±0.17	
P47-O...Q48-NH ₂	11.93	3.19 ±0.18			5.27	3.21 ±0.18	

Table 3.3 $^3J_{\text{HNHA}}$ coupling constants (Hz) of loop residues of the WT, the Q50A, and the N10A HPPK ternary complex

	residue number	$^3J_{\text{HNHA}}$ wtHPPK	$^3J_{\text{HNHA}}$ Q50A HPPK	$^3J_{\text{HNHA}}$ N10A HPPK
loop 1	9		7.82	
	10	7.98	8.39	9.20
	11	5.95	8.34	
	12			
	13			
loop 2	44	8.53	7.59	
	45			
	46			
	47			5.88
	48	6.49	6.75	7.59
	49	5.00	5.80	7.43
	50			
	51	3.13	1.96	
	52	9.22	8.09	
	53	8.35	8.81	
loop 3	82	8.17	8.23	
	83			
	84		4.00	
	85	5.23	5.56	
	86	5.07		
	87			6.15
	88			
	89	5.48		
	90			
	91	5.71	4.99	5.91
	92		2.79	
	93	6.96	7.04	

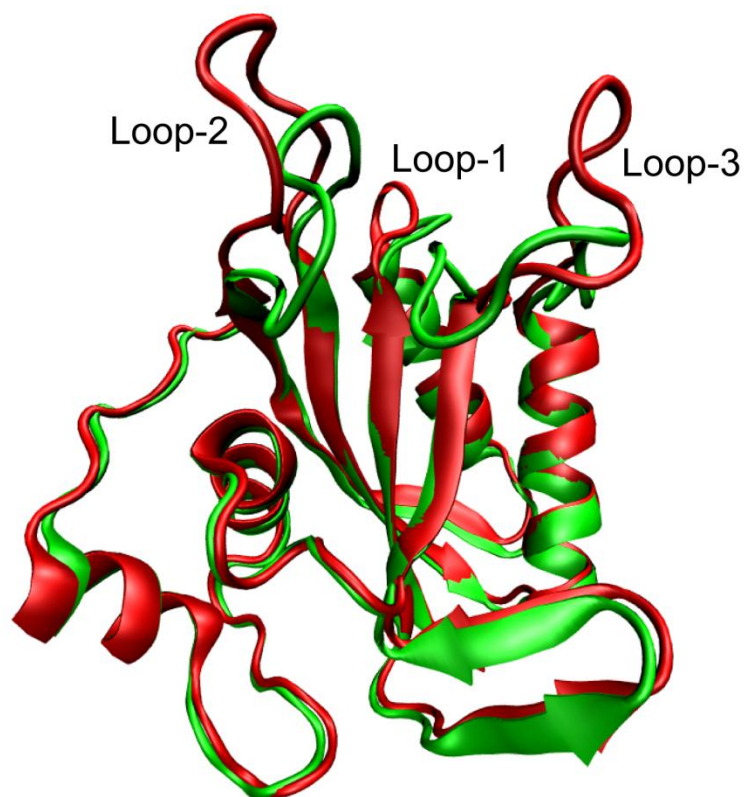


Figure 3.1. The crystal structure of the apo HPPK (PDB ID: 1HKA) and representative structure of the HPPK substrate ternary complex (PDB ID: 1Q0N). The apo HPPK is shown in red color and the ternary complex in green color. Only the protein backbone is shown for simplicity.

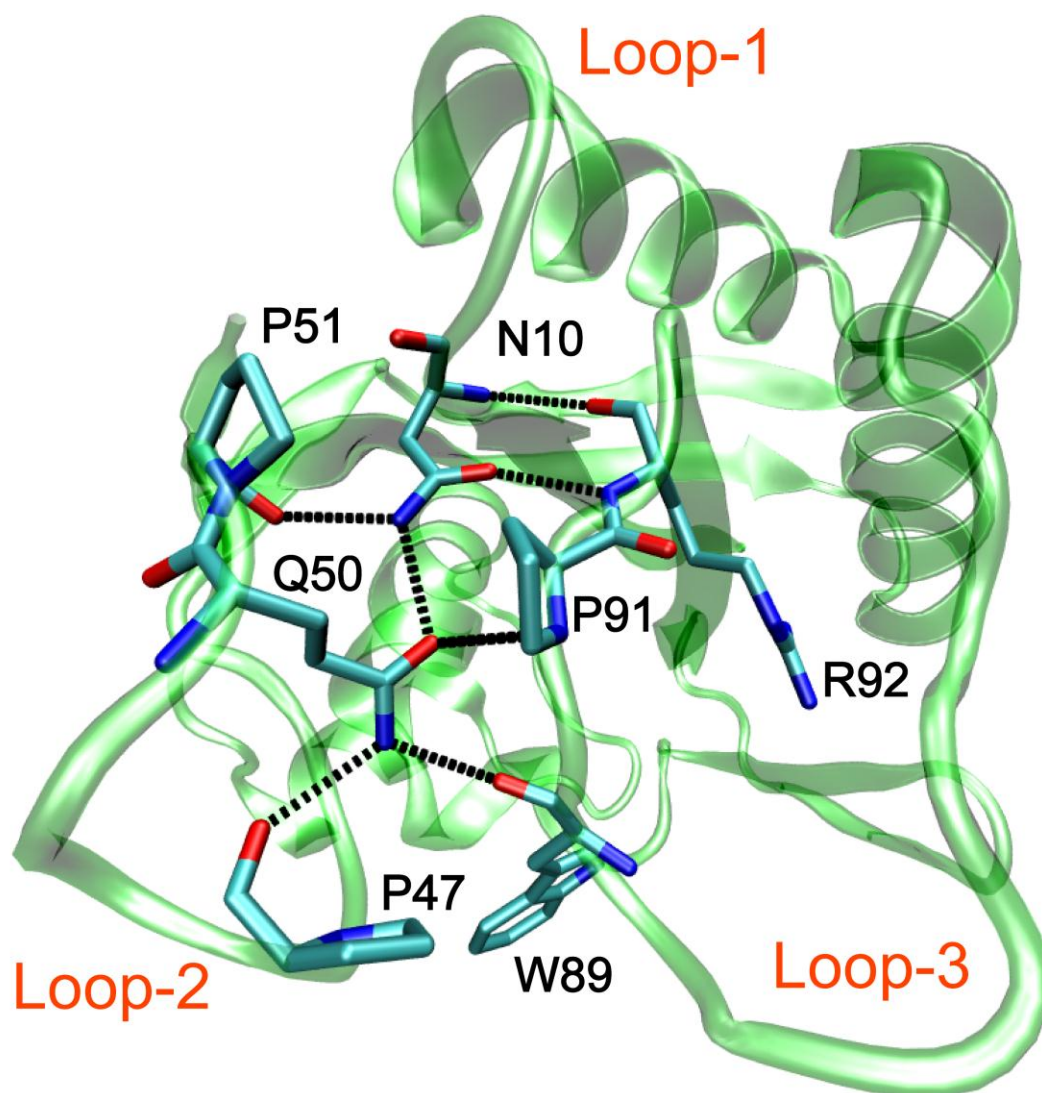


Figure 3.2. Schematic illustration of the coupling of three flexible loops through a hydrogen bond network in the HPPK ternary complex. This hydrogen bond network involves N10 from loop 1, P47, and Q50 and P51 from loop 2, and W89, P91, and R92 from loop 3. This network is not observed in apo HPPK. The hydrogen bonds are represented by dashed lines. Key residues are represented using bond model.

Figure 3.3

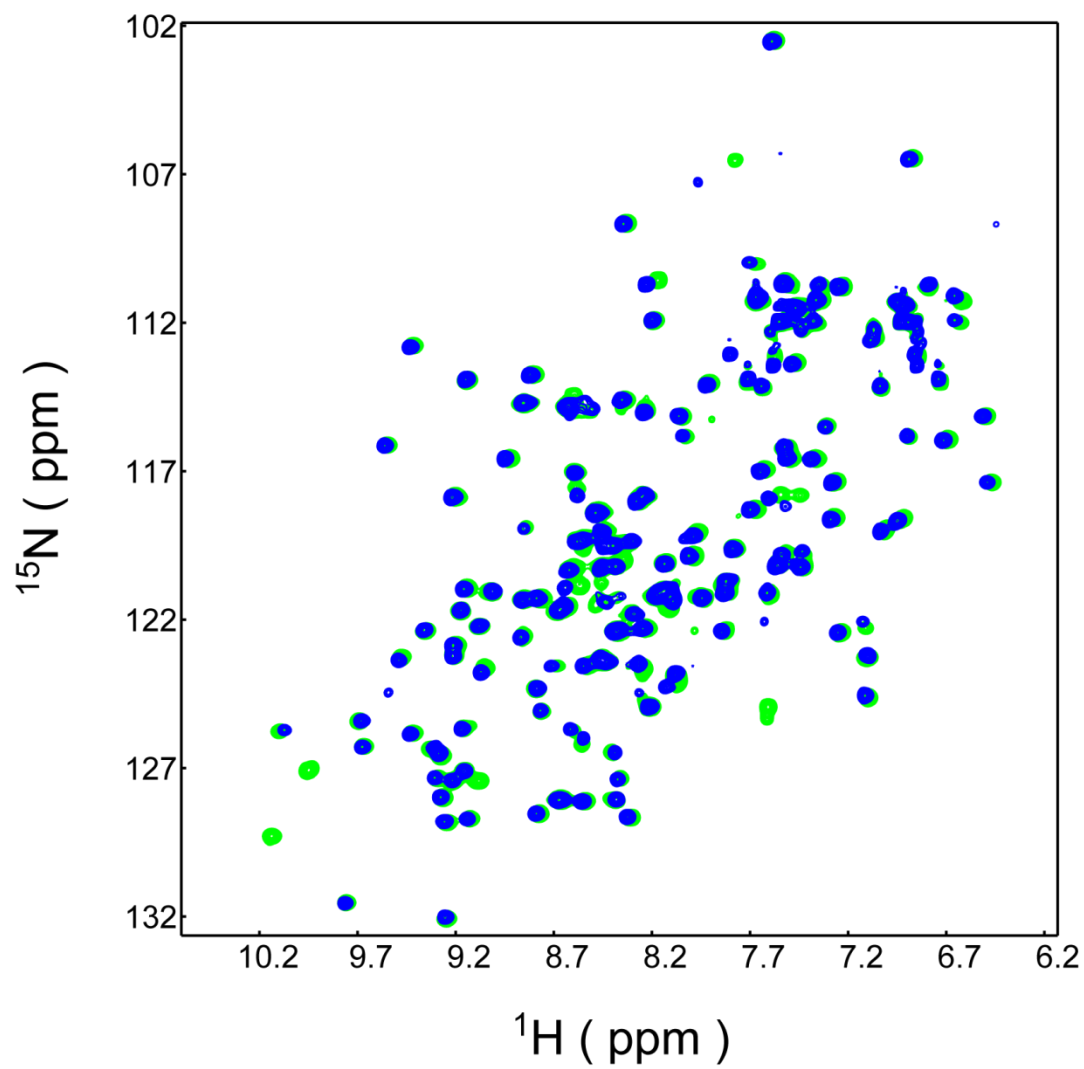


Figure 3.3 (cont'd)

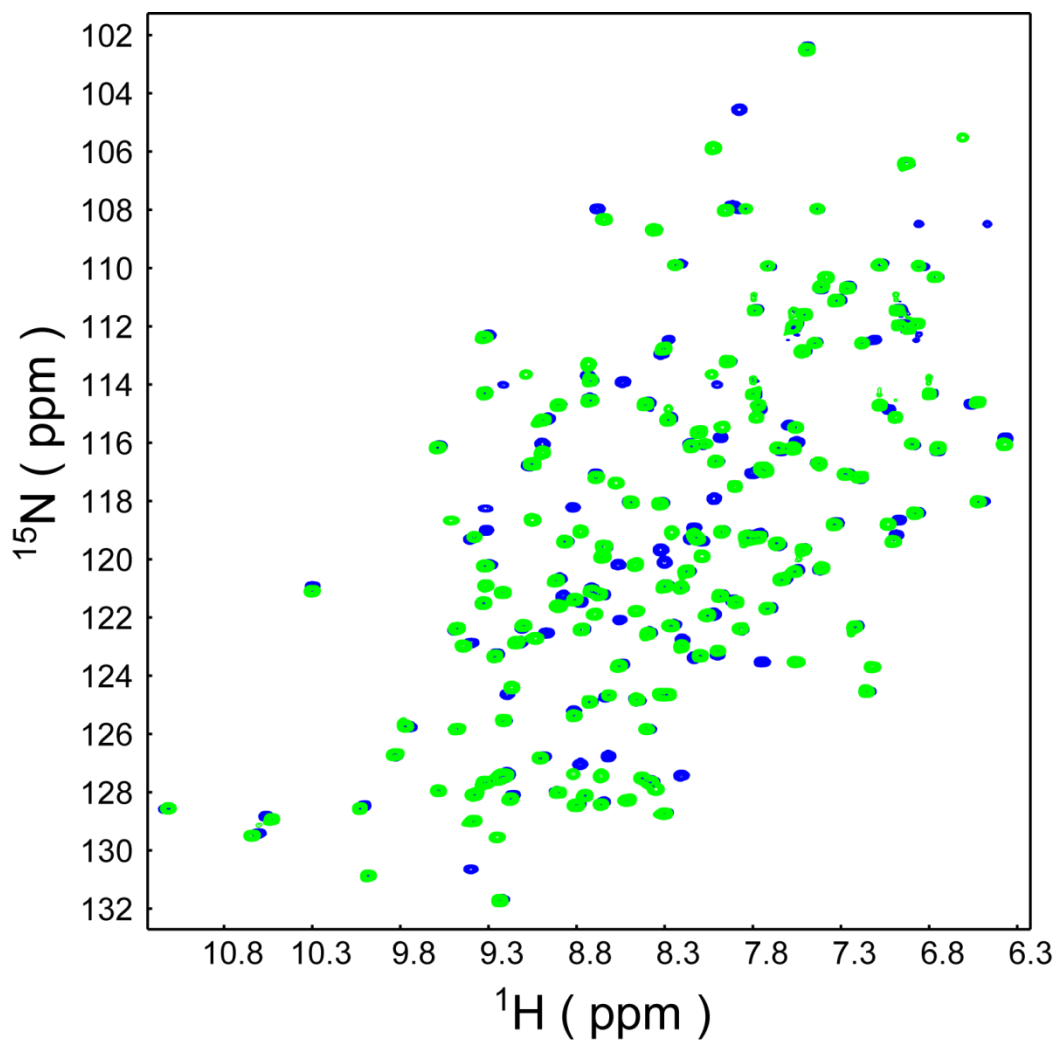


Figure 3.3. Overlay of the ^1H - ^{15}N HSQC spectra of the wild-type and Q50A HPPK mutant enzymes. Upper panel: Apo forms; lower panel: HPPK•AMPCPP•HDMP ternary complexes. Blue and green colors represent WT and Q50A HPPK mutant enzymes, respectively.

Figure 3.4

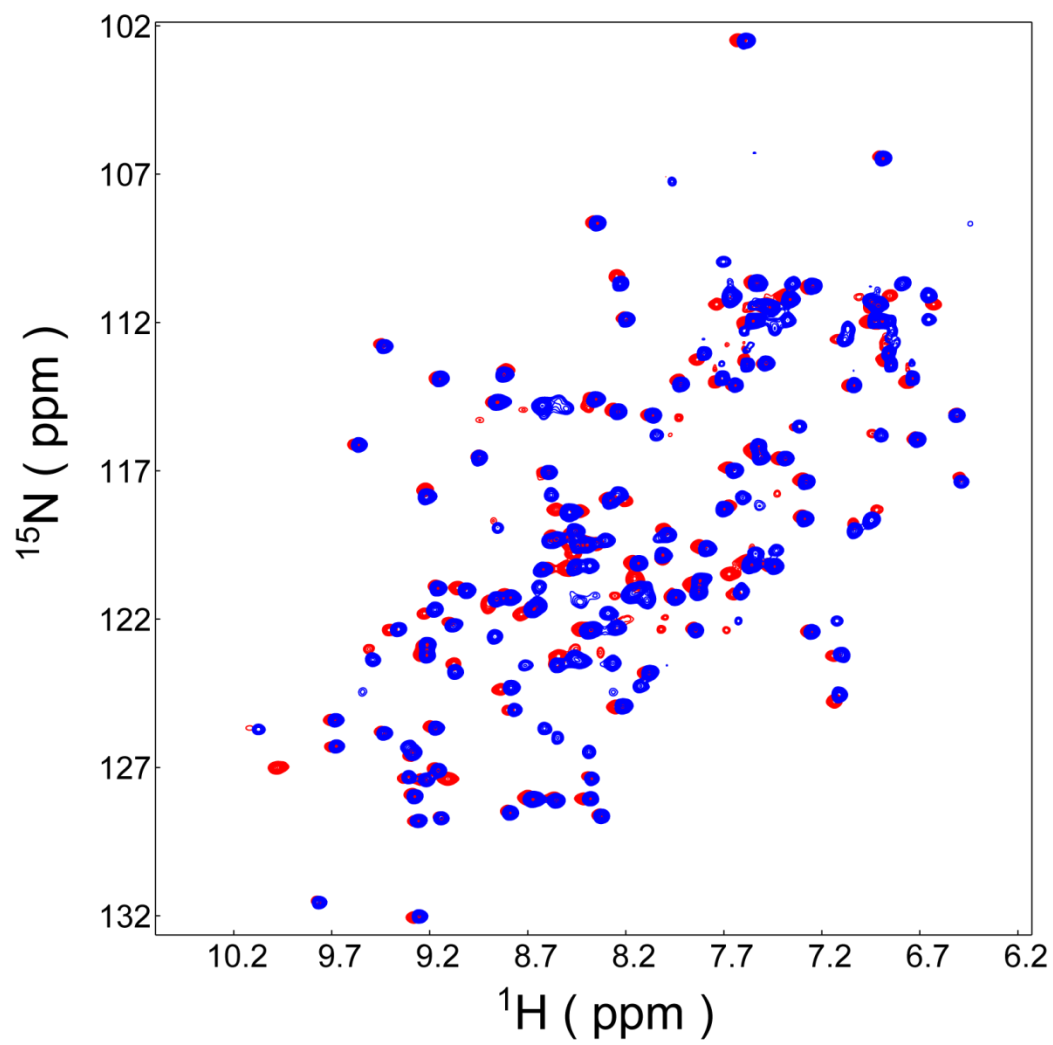


Figure 3.4 (cont'd)

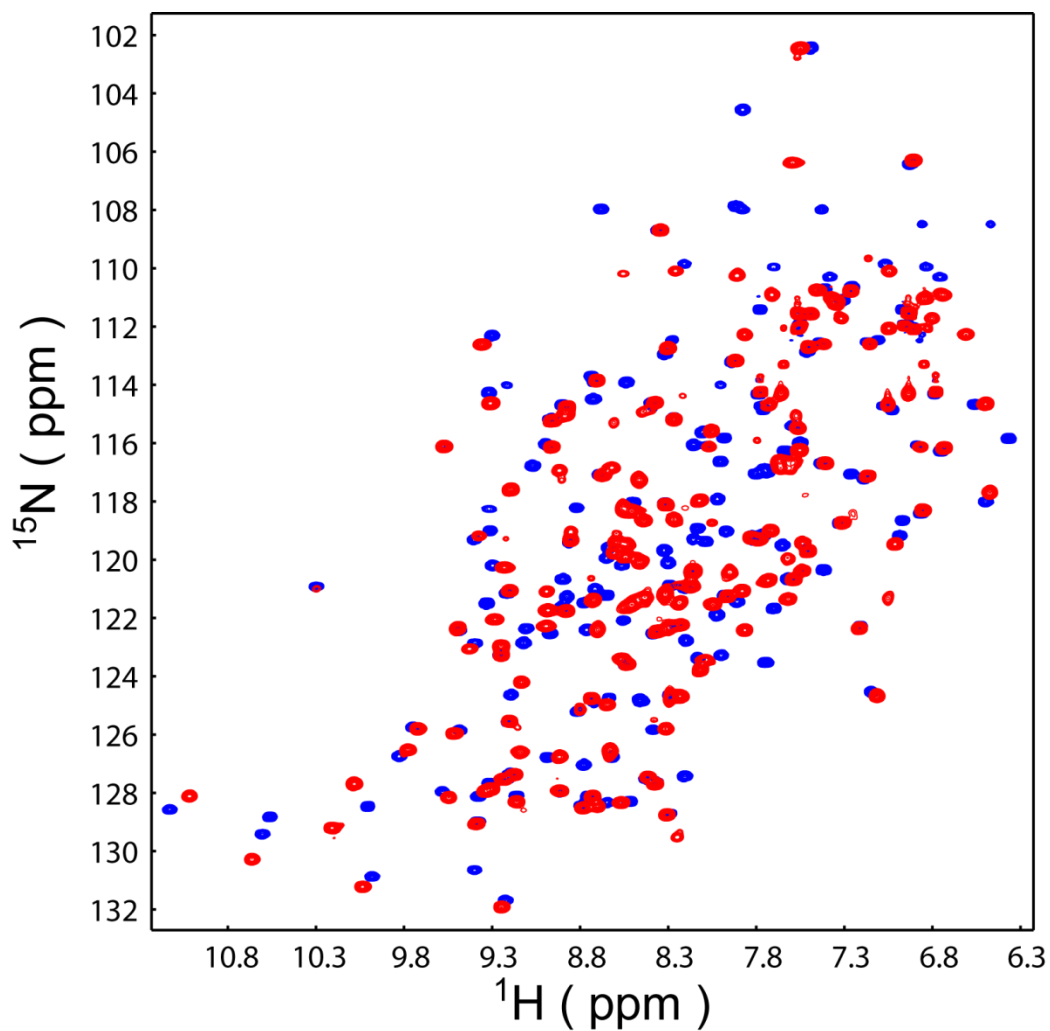


Figure 3.4. Overlay of the ^1H - ^{15}N HSQC spectra of the wild-type and N10A HPPK mutant enzymes. Upper panel: Apo forms; lower panel: HPPK•AMPCPP•HDMP ternary complexes. The blue and red colors represent WT and N10A HPPK mutant enzymes, respectively.

Figure 3.5

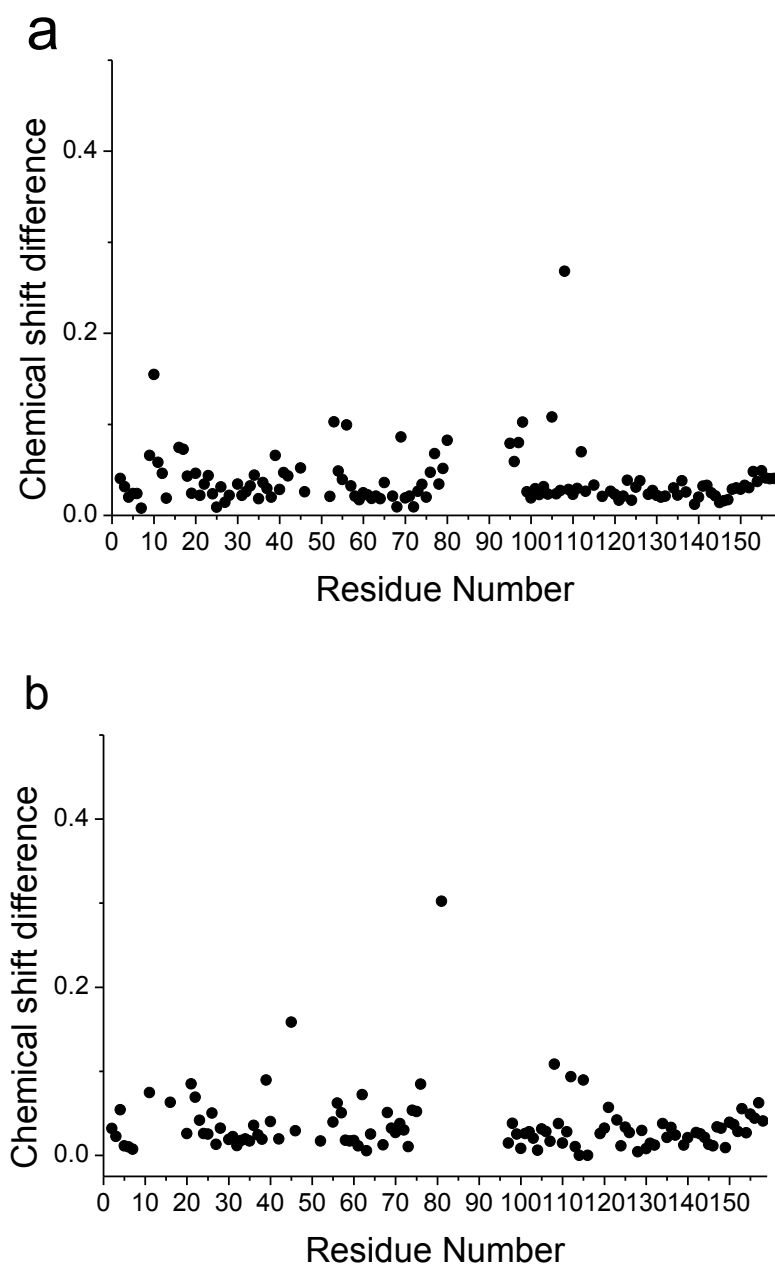


Figure 3.5 (cont'd)

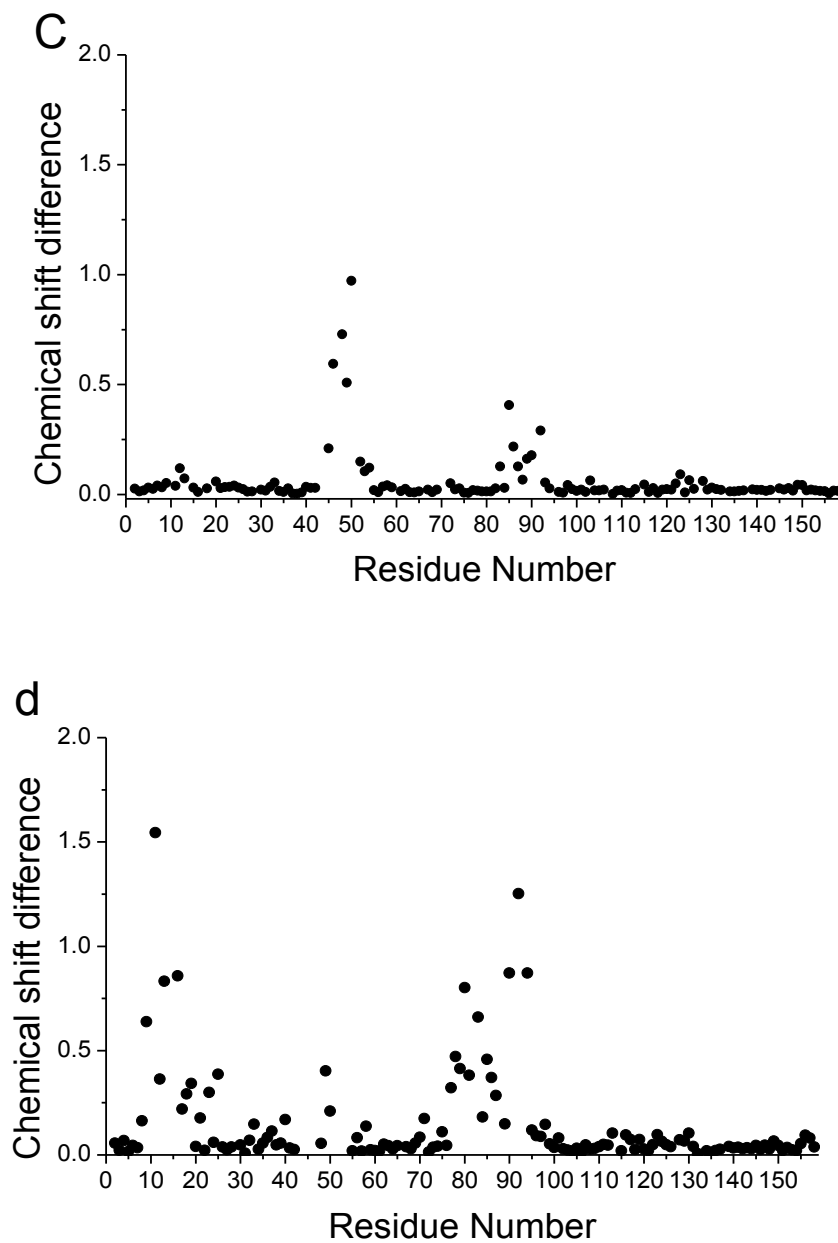


Figure 3.5. Chemical shift mapping for HPPK mutant enzymes. a: Apo Q50A HPPK; b: Apo N10A HPPK; c: Q50A HPPK ternary complex; d: N10A HPPK ternary complex

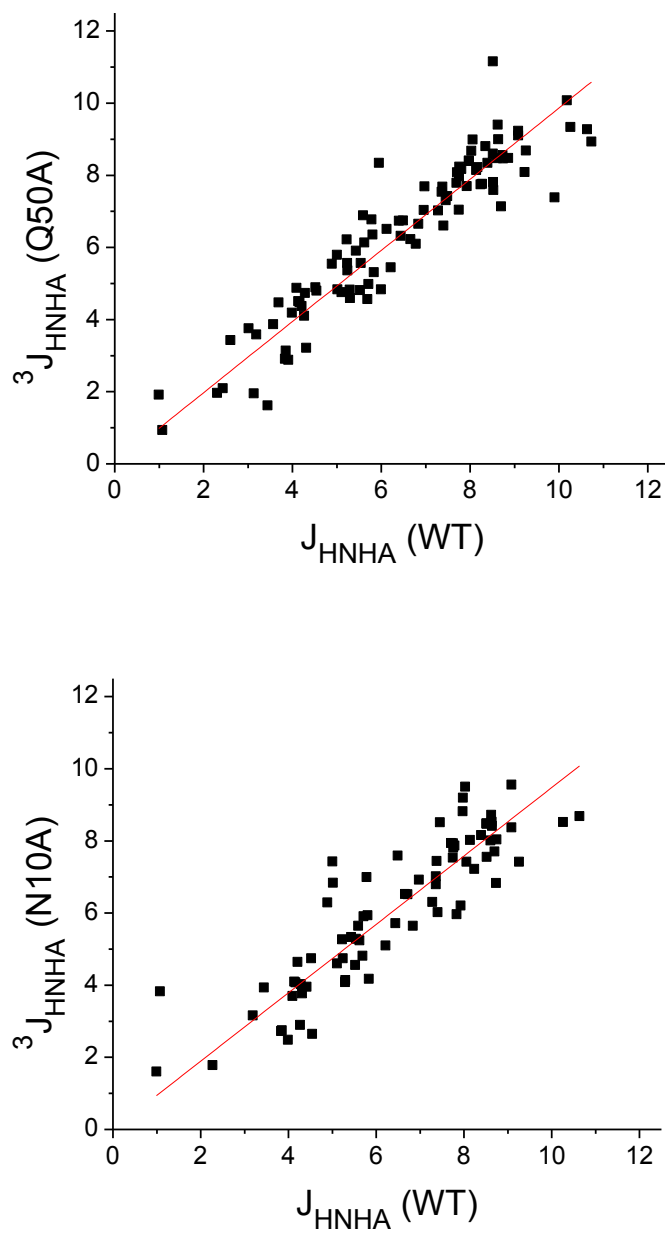


Figure 3.6. Correlation of $^3J_{\text{HNHA}}$ coupling constants between mutant and WT HPPK ternary complexes. Upper panel: Q50A HPPK; lower panel: N10A HPPK

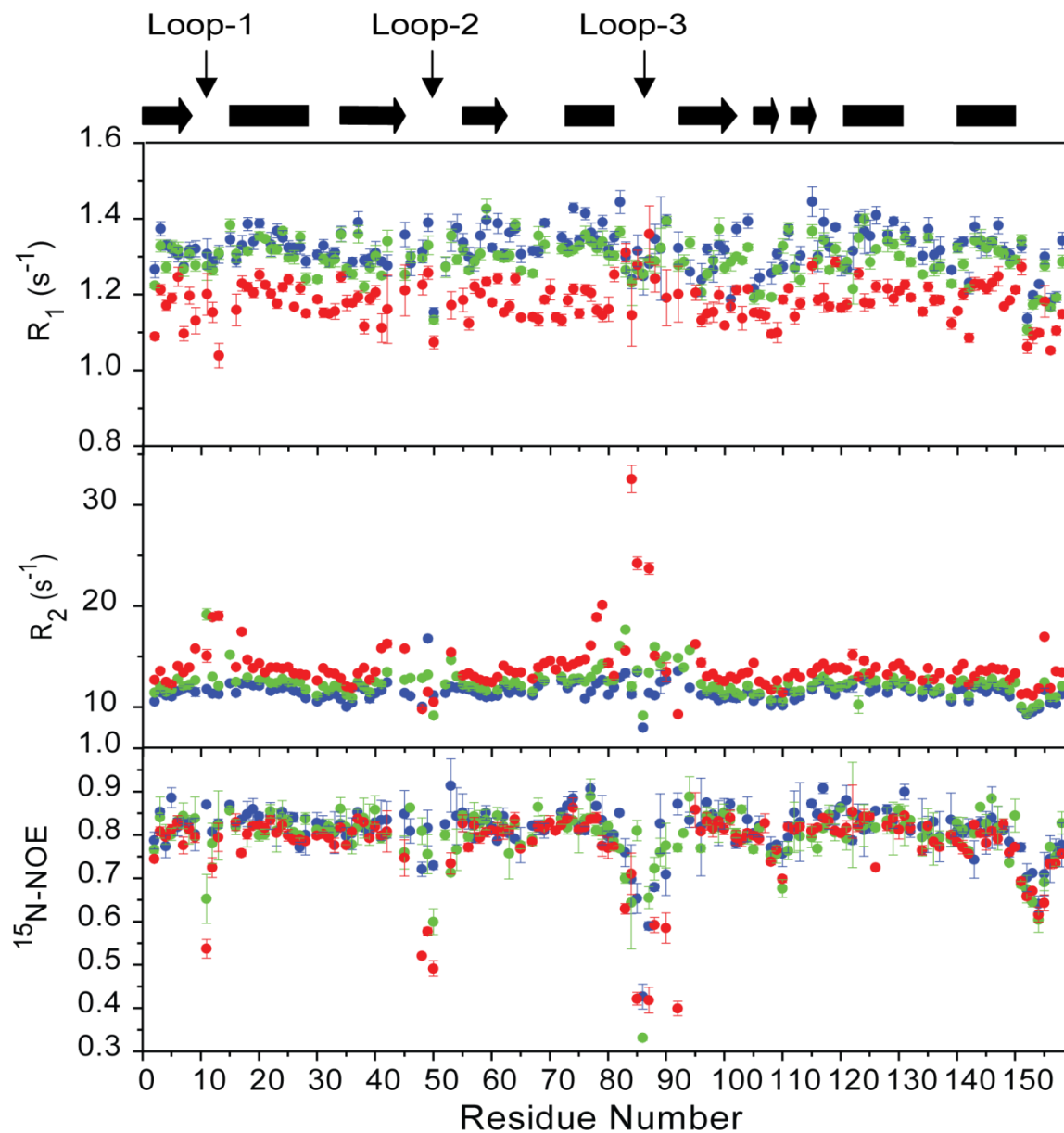


FIGURE 3.7. Comparison of ^{15}N relaxation parameters versus residue number between WT, Q50A, and N10A HPPK ternary complexes. The secondary structure elements are drawn at the top of the figure. Blue, red, and green colors represent WT, N10A, and Q50A HPPK enzymes, respectively.

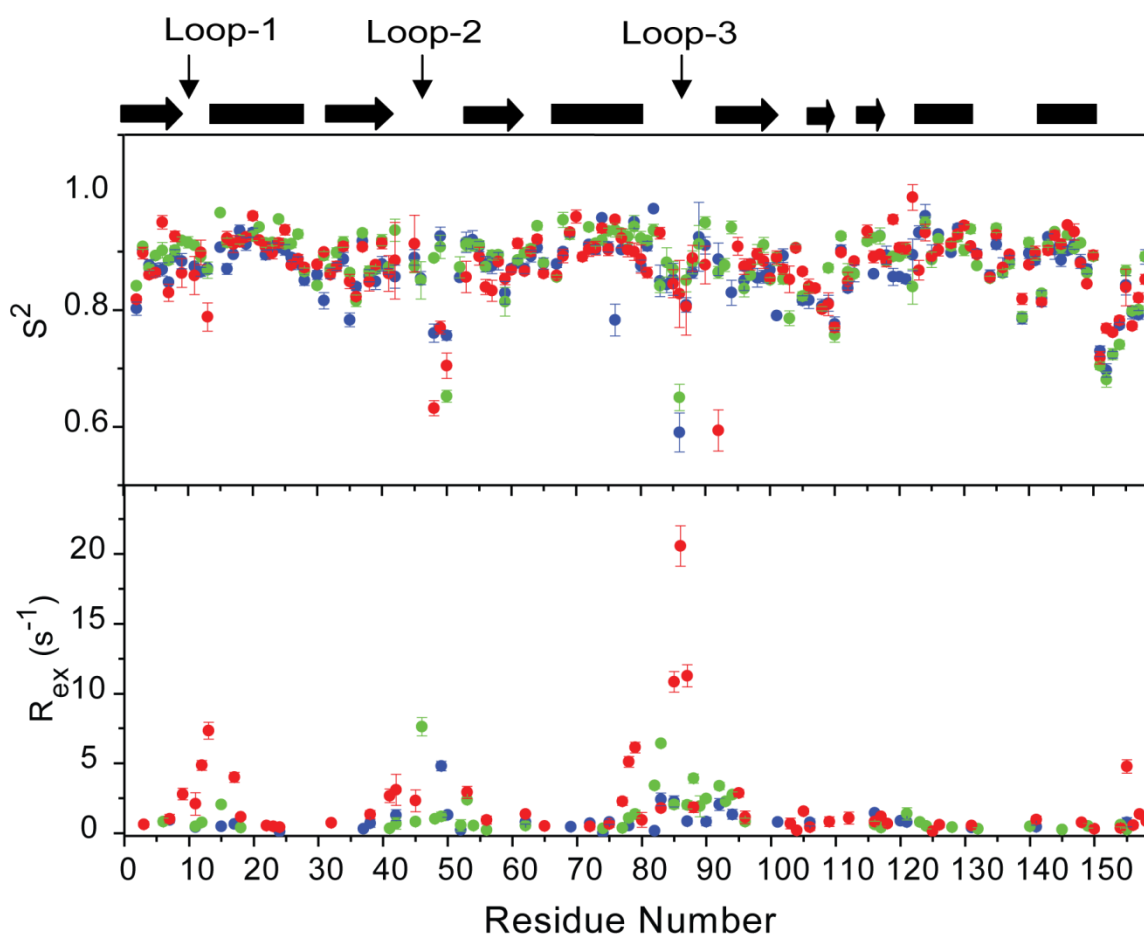


FIGURE 3.8. Comparison of the model-free parameters versus residue number between WT, Q50A, and the N10A HPPK ternary complexes. Blue, red, and green colors represent WT, N10A, and Q50A HPPK, respectively.

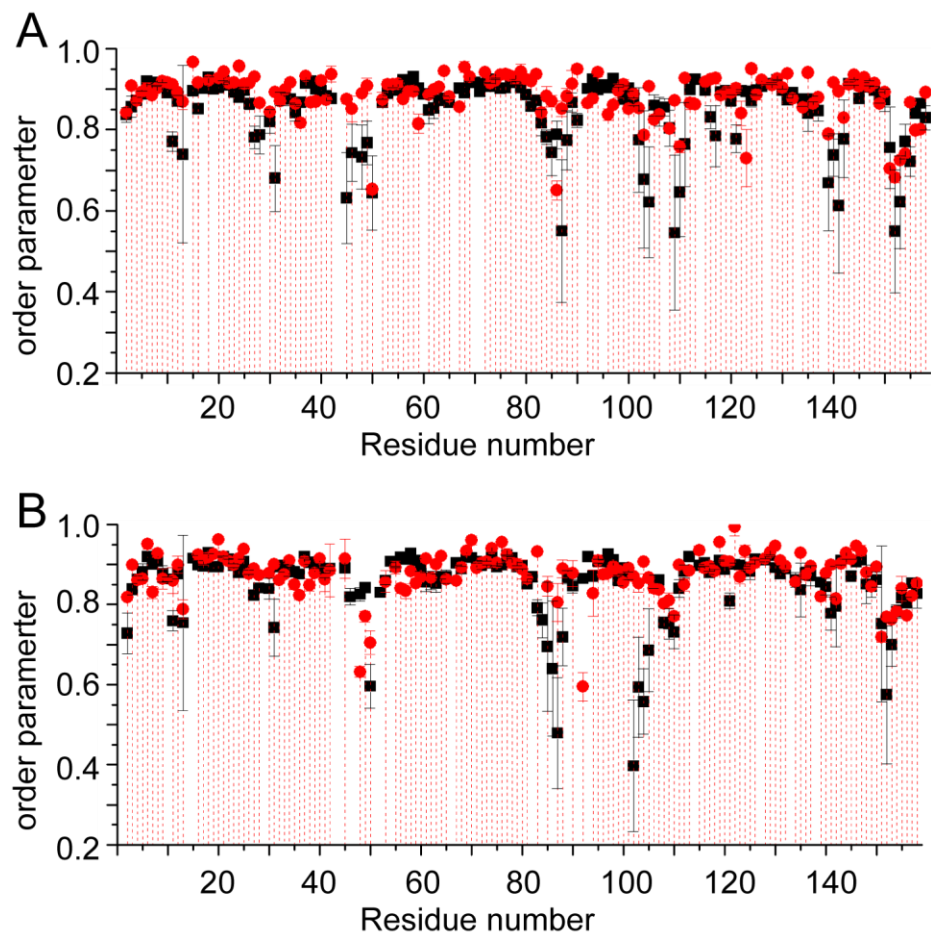


Figure 3.9. Correlation of the order parameter (S^2) from NMR spectroscopy and MD simulation. S^2 values from MD (black square) and NMR (red circle) are plotted along the amino acid sequences of Q50A (A) and N10A (B) HPPK enzymes.

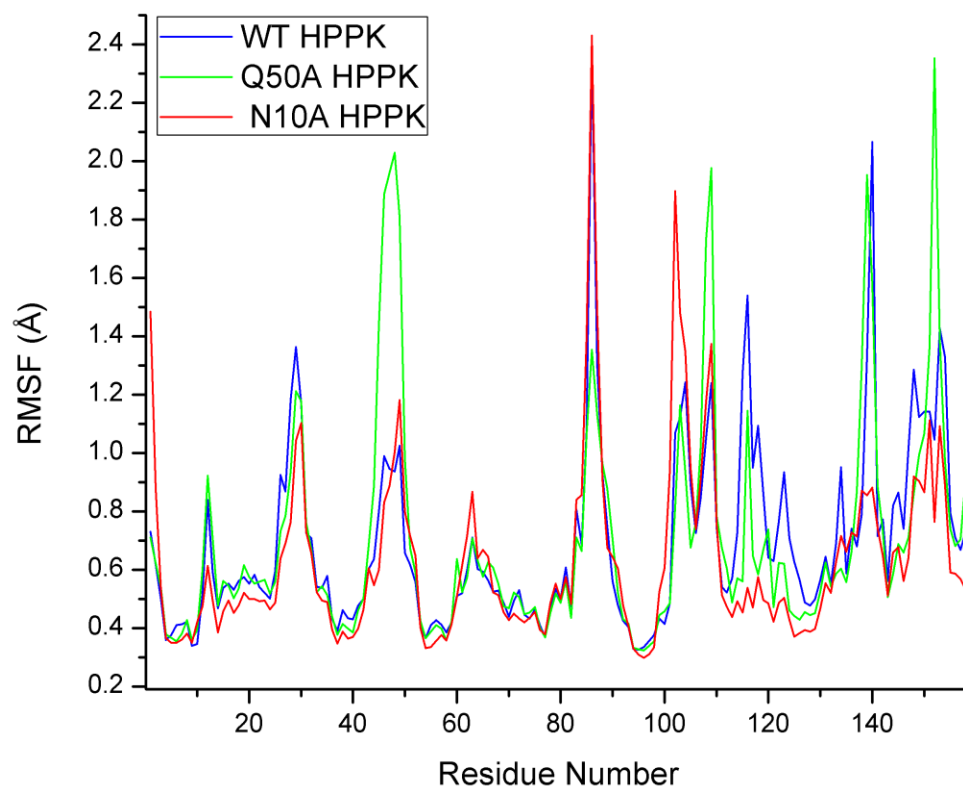


Figure 3.10. Comparison of RMSFs of WT (blue), N10A (red) and Q50A (green) HPPK enzymes.

References

REFERENCES

1. Xiao, B., Shi, G. B., Gao, J. H., Blaszczyk, J., Liu, Q., Ji, X. H., and Yan, H. G. (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.
2. Blaszczyk, J., Shi, G., Yan, H., and Ji, X. (2000) Catalytic center assembly of HPPK as revealed by the crystal structure of a ternary complex at 1.25 angstrom resolution. *Structure* 8, 1049-1058.
3. Blaszczyk, J., Shi, G. B., Li, Y., Yan, H. G., and Ji, X. H. (2004) Reaction trajectory of pyrophosphoryl transfer catalyzed by 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Structure* 12, 467-475.
4. Li, Y., Gong, Y. C., Shi, G. B., Blaszczyk, J., Ji, X. H., and Yan, H. G. (2002) Chemical transformation is not rate-limiting in the reaction catalyzed by *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Biochemistry* 41, 8777-8783.
5. Blaszczyk, J., Li, Y., Shi, G. B., Yan, H. G., and Ji, X. H. (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. Li, Y., Wu, Y., Blaszczyk, J., Ji, X. H., and Yan, H. G. (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.
7. Blaszczyk, J., Li, Y., Wu, Y., Shi, G. B., Ji, X. H., and Yan, H. G. (2004) Essential roles of a dynamic in the catalysis of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Biochemistry* 43, 1469-1477.
8. Li, Y., Blaszczyk, J., Wu, Y., Shi, G. B., Ji, X. H., and Yan, H. G. (2005) Is the critical role of loop 3 of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase in catalysis due to loop 3 residues arginine-84 and tryptophan-89? Site-directed mutagenesis, biochemical, and crystallographic studies. *Biochemistry* 44, 8590-8599.

9. Shi, G. B., Gao, J. H., and Yan, H. G. (1999) H-1, C-13 and N-15 resonance assignments of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase and its complex with MgAMPPCP. *J. Biomol. NMR* 14, 189-190.
10. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Formankay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology-2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* 33, 5984-6003.
11. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe - a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277-293.
12. Johnson, B. A., and Blevins, R. A. (1994) NMRView: a computer-program for the visualization and analysis of NMR data. *J. Biomol. NMR* 4, 603-614.
13. Li, G. Y., Felczak, K., Shi, G. B., and Yan, H. G. (2006) Mechanism of the conformational transitions in 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as revealed by NMR spectroscopy. *Biochemistry* 45, 12573-12581.
14. Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* 16, 23-28.
15. de la Torre, J. G., Huertas, M. L., and Carrasco, B. (2000) HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* 147, 138-146.
16. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* 112, 4989-4991.
17. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules .1. Theory and range of validity. *J. Am. Chem. Soc.* 104, 4546-4559.
18. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules .2. Analysis of experimental results. *J. Am. Chem. Soc.* 104, 4559-4570.

19. Vuister, G. W., and Bax, A. (1993) Quantitative J correlation - a new approach for measuring homonuclear 3-bond J(H(N)H(Alpha)) coupling-constants in N-15 enriched proteins. *J. Am. Chem. Soc.* 115, 7772-7777.
20. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-integration of cartesian equations of motion of a system with constraints - molecular dynamics of N-Alkanes. *J. Comput. Phys.* 23, 327-341.
21. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh ewald - an N.Log(N) method for ewald eums in large systems. *J. Chem. Phys.* 98, 10089-10092.
22. Adelman, S. A., and Doll, J. D. (1976) Generalized langevin equation approach for atom-solid-surface scattering - general formulation for classical scattering off harmonic solids. *J. Chem. Phys.* 64, 2375-2388.
23. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684-3690.
24. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph.* 14, 33-38.
25. Prompers, J. J., and Bruschweiler, R. (2002) General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J. Am. Chem. Soc.* 124, 4522-4534.
26. Koller, A. N., Schwalbe, H., and Gohlke, H. (2008) Starting structure dependence of NMR order parameters derived from MD simulations: Implications for judging force-field quality. *Biophys. J.* 95, L4-L6.
27. Farrow, N. A., Zhang, O. W., Formankay, J. D., and Kay, L. E. (1994) A heteronuclear correlation experiment for simultaneous determination of N-15 longitudinal decay and chemical-exchange rates of systems in slow equilibrium. *J. Biomol. NMR* 4, 727-734.

Chapter 4

Insight into *E. coli* HPPK Substrate Binding Mechanism and Conformational Transition Pathway: Molecular Dynamics Simulation Studies

Introduction

Structural studies show that HPPK undergoes dramatic conformational changes in its catalytic loops through its catalytic cycle. The conformational changes are important for catalysis as shown by combined biochemical and structural studies. However, how the conformational transitions occur is largely unknown. Because high-resolution structures are available for many stages of its catalytic cycle and the protein is small and monomeric, computational approaches have been used to investigate the dynamic properties of the enzyme. Yang et al. studied the loop conformational dynamics of HPPK in the ligand-free (apo) form and its binary complex with MgATP by standard MD simulations and local-enhanced sampling (1). They found that the crystallographic B-factors underestimate the loop dynamics considerably and the open conformation of loop 3 in the binary complex is accessible to the apo HPPK, which supports that HPPK uses a selected-fit mechanism instead of the induced-fit mechanism for ATP binding. Su and Cukier studied the conformational dynamics of the HPPK•MgATP complex using restraint MD simulation and ATP binding to HPPK using reweight MD simulation (2). It was found that the HPPK•MgATP binary complex can sample a great span of conformations from apo-like to more ligand-bound-like conformation, which supports the hypothesis that the ligand binding by HPPK follows a selected-fit mechanism. They also studied the conformation of the HP binding site in the HPPK•MgATP binary complex (derived by removal of HP from the HPPK substrate ternary complex) using a Hamiltonian Replica Exchange method (HREM) (3). It was shown that there are several different conformations of the HP binding site based on cluster analysis of the key residues for HP binding. This result indicates that the HP binding may also follow a selected-fit mechanism.

This chapter describes a more systemic computational study of the conformational dynamics of HPPK in the first three stages of its catalytic cycle. The conformational sampling of the side-chains of core active-site residues was investigated by standard MD simulations of the apo HPPK, the binary substrate complex, and the ternary substrate complex. The potential pathways for the conformational transitions from the apo HPPK to the binary substrate complex to the ternary substrate complex were obtained by targeted MD simulations of the apo HPPK and the binary substrate complex.

Methods

Molecular Dynamics Simulations. The crystal structures of the apo HPPK (PDB ID: 1HKA), the binary complex HPPK•MgAMPCPP, (unpublished), and the ternary complex HPPK•AMPCPP•HP (PDB ID: 1Q0N) were used as the starting coordinates for the MD simulations of the apo HPPK, the binary substrate complex HPPK•MgATP, the ternary substrate complex HPPK•MgATP•HP, respectively. The unpublished crystal structure of the binary complex HPPK•MgAMPCPP was chosen as the starting coordinate for the MD simulation of the binary substrate complex HPPK•MgATP, because MgAMPCPP is a better MgATP analogue for HPPK. The protein, substrate (and substrate analogues) and Mg^{2+} were kept in the coordinate files. Water molecules coordinated with Mg^{2+} were also kept: four in the HPPK•MgATP complex and three in the HPPK•MgATP•HP complex. All others were deleted from the coordinate files. AMPCPP was virtually substituted to the native substrate ATP. The simulations were carried and analyzed in the same way as in Chapter 3.

Principal Component Analysis (PCA). PCA was used to determine the large-scale concerted motion direction in each form of HPPK. PCA is a widely used method to simplify the study of the multidimensional data by reducing the dimensionality. PCA

decompose the total variance of the atom RMS fluctuations over an MD trajectory into a small set of modes that have a relative large contribution to this variance and a large set of modes with a small contribution. This method has been shown to be very effective in analyzing concerted motions in proteins (4).

The involvement coefficient, by definition, is the dot product between the unit vector of a PC and the unit vector of a probe direction (5). The involvement coefficient v_i of the i th PC is defined as:

$$v_i = \frac{|R_i \cdot R_{diff}|}{|R_i| \cdot |R_{diff}|} = \cos \theta$$

in which R_i is the vector of the i th PC, and R_{diff} is the vector of the difference between two structures, where θ is the angle between R_i and R_{diff} . The cumulative involvement coefficient μ_n of the first n PCs is computed as

$$\mu_n = \sum_{i=1}^n \mu_i^2$$

The involvement coefficient can be used as an indicator of the degree of overlap between a PC and a probe direction. In our case, the probe direction is defined as the conformation difference vector between two different forms of HPPK. The involvement coefficient of the first PC of the apo HPPK can describe how much the largest motion direction of the apo HPPK overlap with the conformational change direction from the apo to the binary complex. The cumulative coefficient quantifies how much low-indexed PCs overlap with the conformation difference.

Targeted MD simulations. The conformational transitions during substrate binding were studied by targeted MD simulations (7), which enable simulation of large conformational change from the initial to final state by using an additional energy functional term that decreases the RMSD between those two states. The targeted MD

implemented in AMBER 10 (6) was used to calculate the trajectories. The additional energy term in AMBER 10 has the following form:

$$E_t = 0.5 * f * n * (\text{RMSD}_t - \text{RMSD}_{\text{target}})^2 \quad (1.1)$$

where f is the force constant for targeted MD, n is the number of atoms that were used to calculate the RMSD, RMSD_t is the RMSD at time t and $\text{RMSD}_{\text{target}}$ is the targeted RMSD. Since the masks for fitting and RMSD calculation are separately defined in AMBER 10, it is possible to fit to one part of the structure but calculate the RMSD (and thus the restraint force) for another part of the structure.

The conformation transitions of HPPK through the formation of the ternary substrate complex were simulated using a two-stage targeted MD simulation: from the apo HPPK to the binary substrate complex HPPK•MgATP to the ternary substrate complex HPPK•MgATP•HP. Two model complexes were prepared: HPPK^{apo}•ATP (HPPK•MgATP complex with apo HPPK conformation) and HPPK^{bin}•MgATP•HP complex (HPPK•MgATP•HP complex with HPPK•MgATP conformation). The former was prepared by placing ATP and Mg²⁺ into the apo HPPK structure and the latter by placing HP in the HPPK•MgATP structure by superimposition of the corresponding structures described earlier. The preparation for the targeted MD run was comprised of three rounds of energy minimization and short NVT and NPT equilibration periods followed by 2 ns of regular MD simulation. Three snapshots from the regular MD simulations were randomly chosen as the reference structure for targeted MD simulations. Three 3 ns targeted MD simulations were run using a force constant of 15 kcal/mol/Å⁻², based on our trial results. The RMSD was reduced gradually during the 3 ns simulation. The core structure residues (1-8, 15-42, 55-79, 94-158) were used for fitting and the

whole protein was used for RMSD calculation. The same protocol was applied to the targeted MD simulations of the second stage.

Results

Stability of the MD Simulations. The MD simulations are stable based on the time evolution of RMSD (Figure 4.2). Equilibrium is reached based on RMSD after 9, 9 and 5 ns for the apo, HPPK•ATP, and HPPK•ATP•HP complexes, respectively. The trajectories after reaching equilibrium were used for further analysis. Surprisingly, there is a jump in the RMSD of the apo HPPK at ~20 ns but the RMSD goes back to around the equilibrium value in ~2 ns.

The Side-chain Conformation Sampled in the MD Simulations. There are five residues interacting with HP and 12 residues interacting with ATP as revealed by X-ray crystallography. Among them, 11 residues interact with substrates through their side-chains. These residues are Y53, N55, and F123 for HP binding, and R82, R84, R88, W89, R92, H115, Y116, and R121 for ATP binding. Among them, R82, R84, R88, W89, and R92 are located in catalytic loop 3 that goes through most dramatic conformational changes during substrate binding. These residues have to be positioned close to ATP through conformational changes before the re-orientation of their side-chains for ATP binding. As discussed before, standard MD simulation is not suitable for studying this kind of conformational change. Thus, the side-chains of these residues are not considered in this study. This study focuses on residues Y53, N55, H115, Y116, R121, and F123. The results show that these side-chain motions are quite restricted as expected in the ternary substrate complex (Figure 4.2). In the apo form and the binary substrate complex, these side-chains can also sample the conformations of the ternary substrate complex. Interestingly, the side-chains of Y53 and N55 samples more conformational space in the binary substrate complex than in the apo form, indicating that the binding of MgATP

makes these side-chains more mobile. This enhanced mobility may be caused by the motion of loop 2. As shown by NMR studies, binding of MgAMPCPP, an MgATP analogue, makes loop 2 more mobile.

Principal Component Analysis. PCA analysis was used to study large-scale concerted motions in different substrate binding states of HPPK, and to determine if these motions correlate with the conformational transitions between different states. To this end, involvement coefficients and cumulative involvement coefficients were calculated using equations 4.1 and 4.2, respectively. The former is a measure of the relevance of a PC to the conformational transition, and the latter a measure of the contributions from a set of PCs.

Figure 4.4 shows the individual and cumulative involvement coefficients of the first 50 PCs of the apo HPPK and the HPPK•MgATP complex. The results show that none of PCs are highly relevant to the conformational transitions. The cumulative involvement coefficient of the first 50 PCs is only 0.5 for apo HPPK and only 0.36 the HPPK•MgATP complex. Apparently the PCA analysis failed to identify motions relevant to the conformational transitions.

Targeted MD Simulations. Because the PCA analysis failed to identify motions relevant to the conformational transitions, targeted MD simulations were adopted to investigate the conformational transitions. After carefully inspecting the trajectories of the targeted MD simulations, several parameters were defined and monitored as the conformational transitions progress (Figures 4.4 and 4.5).

Conformational Transition from the Apo HPPK to the HPPK•MgATP Complex. The conformational transition of loop 2 can be described by the pseudo-dihedral angle CA55-CA43-CA48-CA51 (Figure 4.4 A). Its progression towards the conformation of the binary complex is nearly monotonic. The loop reaches the

conformation of the binary complex in ~ 1 ns and is stabilized after 2 ns. The conformation of loop 3 can be described by the pseudo-dihedral angles CA97-CA81-CA86-CA89 (Figure 4.4 B) and CA89-CA88-CA87-CA76 (Figure 4.4C). Its progression towards the conformation of the binary complex is not straightforward and involves a twist motion. The twist motion of loop 3 makes it possible to position R92 into the ATP binding mode. During the simulation, the side-chain of R92 first forms a hydrogen bond with the backbone of E77 at 1.2 ns, then shifts the H-bond partner to the backbone of R82 until the loop adopts a helix-like conformation, which breaks the H-bond and position R92 for ATP binding.

Conformational Transition from HPPK•MgATP to HPPK•MgATP•HP

Complex. For the first 1 ns, the loop 2 moves closer to the active site in a kind of hinge motion. However, it is hard to find a single dihedral angle to describe this process. Instead, the distance between CA46 and CA95 residues is used for this purpose. As shown in Figure 4.5A, this distance correlates well with loop 2 movement. From ~ 500 ps, loop 3 starts to move closer to the active site, which can be characterized by the pseudo-dihedral angle CA86-CA88-CA97-CA72 (Figure 4.5B). From ~ 750 ps, the small helix inside loop 3 starts to untwist, which can be characterized by a pseudo-dihedral angle CA86-CA85-CA88-CA89 (Figure 4.5C).

Hydrogen Bond Analysis. The model complexes (both HPPK^{apo}-ATP and HPPK^{bin}-ATP•HP) were created each by superimposition of two corresponding crystal structures. It is found that the substrates can be docked into the active center without steric hindrance.

Figure 4.7 shows the superposition of the active centers of the apo HPPK, the binary complex, and the HPPK ternary complex. In the structure of the apo HPPK, the core residues T112, L98, R121, H115 and Y116 are in a good position to bind ATP.

However, the loop-3 residues that interact with MgATP, R82, R84, and R92, are far away from the active center. In the structure of the binary substrate complex HPPK•ATP, residues Q55, F123 and Y53 in the HPPK•ATP binary complex are in a position to bind HP. However, the loop 2 residues that interact with HP are away from the active center. In both cases, the active center is not blocked and is accessible to the substrates.

The protein-substrate interactions were further examined after 2 ns MD simulations. For HPPK^{apo}•MgATP complex, the conformations of the three catalytic loops remain similar to those of the apo HPPK and the catalytically important residues R82 and R92 are still far away from MgATP, although the interactions of the nearby residues with MgATP were further optimized. Figure 4.8 shows the final snapshot of the complex in the MD simulation. The R92 side-chain forms hydrogen bonds with E77 and G81, which might keep R92 from getting into active site.

The conformational change of R92 was monitored during the targeted MD simulation of the first stage, from HPPK^{apo} to HPPK^{bin}. As it interacts with E77 and G81 in the starting structure and with ATP in the final state, the conformational change was monitored through tracking the distances between them. The optimization of the interactions between D95 and the magnesium ion was also monitored through their distance. The results are shown in Figure 4.9. The interactions are maintained for the first 750 ps as loop 2 and loop 3 go through the conformational transition. During this stage, loop 2 and loop 3 moves closer to each other. After that, the R92 move towards ATP till 2 ns as indicated by the increased distance between R92 and E77/G81 and reduced distance between R92 and ATP. The interaction is optimized in the remaining 1 ns. Thus, the loop 3 conformational change occurs before R92 starts to move into the active site.

The interactions between HP and HPPK in the $\text{HPPK}^{\text{bin}} \cdot \text{MgATP} \cdot \text{HP}$ complex are optimized after 2 ns of the standard MD simulation except for L45, as shown in Figure 4.8. Similar to the $\text{HPPK}^{\text{apo}} \cdot \text{ATP}$ complex, the overall conformation is similar to the starting structure, the $\text{HPPK} \cdot \text{MgATP}$ binary complex, and no dramatic conformational change was observed in the loops. R82 keeps the conformation as in the binary complex. W89, which interacts with ATP through a hydrogen bond and forms part of the binding pockets of HP, is also in a conformation similar to that in the binary complex.

The positioning of these residues is monitored during the targeted MD simulation of the $\text{HPPK}^{\text{bin}} \cdot \text{MgATP} \cdot \text{HP}$ complex. Figure 4.10 shows the positioning process as measured by the distances between L45 and HP, W89 and ATP, and R82 and N74. At the beginning, R82 points outside the ATP binding pocket. From around 1.2 ns, it starts to move in the right direction to get closer to the active site and reaches the position in the $\text{HPPK} \cdot \text{MgATP} \cdot \text{HP}$ complex after ~2 ns. W89 is positioned to interact with ATP after R82 moves to the right direction. The carbonyl oxygen of L45 forms a hydrogen bond with HP. It moves into the active site very quickly. However, no stable interaction is formed till ~2 ns, as the distance fluctuates from 3-4 Å.

Discussion

The three catalytic loops, particularly loops 2 and 3, of HPPK undergo dramatic conformational changes through its catalytic cycle. In this study, the conformational transitions from the apo form to the binary substrate complex to the ternary substrate complexes are investigated by MD simulations. The results show that the conformational transitions are complex and cannot be described by some dominant modes of motion. None of the first 50 PCs obtained by PCA analysis of the trajectories of the standard MD simulations are highly relevant to the conformational transitions, as their involvement

coefficients are all small and the cumulative involvement coefficients of the first 50 PCs are also small. This is consistent with the result of the analysis of the apo HPPK by Su and Cukier using an elastic normal mode analysis (2), which found a mode of closing motion of the loops but its direction does not correlate with the motions needed for the formation of the ternary complex. The targeted MD simulations show that motions of loop 3 are more complex than those of loop 2 in the conformational transitions, which is consistent with the complex conformational changes revealed by structural studies by X-ray crystallography and NMR. It appears that loop 2 can reach the target conformations more easily than loop 3, but it is stabilized only after loop 3 reaches its target conformations. The guanidinium groups of R82 and R92, two most important residues for HPPK catalysis, cannot move into the active center without the conformational transitions of loop 3, which is in agreement with the deletion mutagenesis study of this loop.

The conformational changes of the loops are not required for the access of the substrates to the active center or the interactions of the core active site residues with the substrates. Both MgATP and HP can be placed in their binding sites in the apo HPPK without steric hindrance. The core active site residues in the apo form can sample the conformations of the binary and the ternary substrate complex (Figure 4.3), in support of a selected-fit mechanism. Biochemical studies show that HPPK follows an ordered mechanism with MgATP binding to the enzymes first. HP can bind to HPPK in the absence of MgATP, but its affinity is low, in the millimolar range. The binding of MgATP enhances the affinity of HP for HPPK by more than three orders of magnitude. The dramatic enhancement of HP affinity for HPPK is probably due to the conformational change of loop 2 that is stabilized by loop 3. The conformation of loop 3 is in turn stabilized by the interactions of R82 and R92 with MgATP.



Figure 4.1. The structures of the apo form, binary complex, and ternary complex of HPPK.

Green, the apo HPPK (PDB ID: 1HKA); purple, HPPK•MgAMPCPP complex

(unpublished structure); red, HPPK•MgAMPCPP•HP complex (PDB ID: 1Q0N).

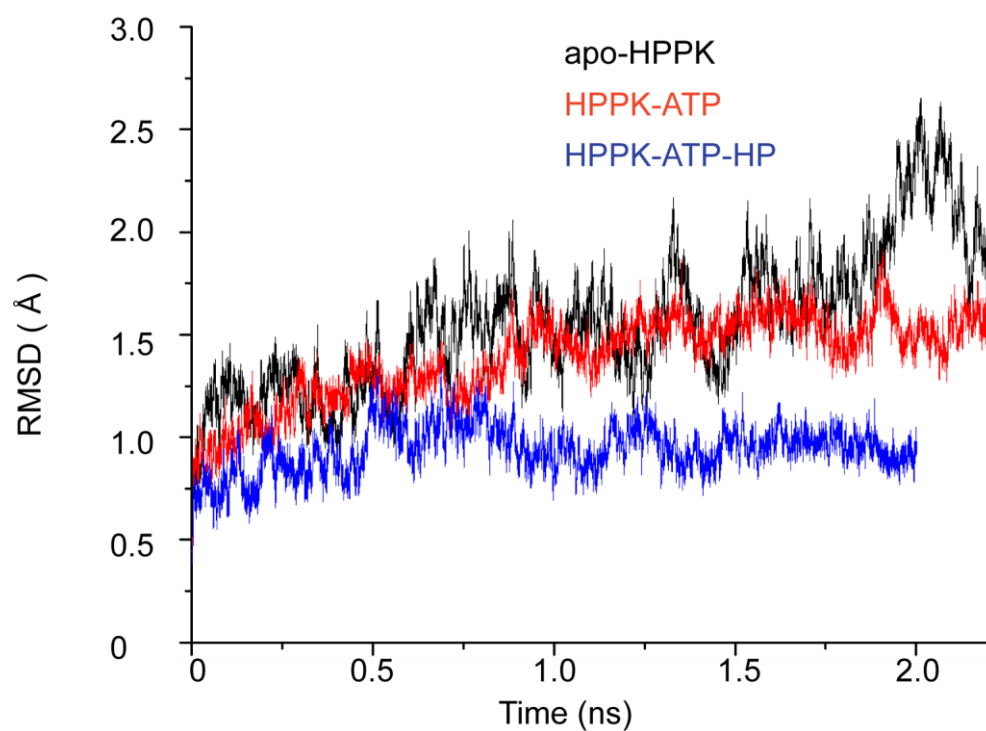


Figure 4.2: C α RMSDs of the apo HPPK (black), the HPPK•MgATP complex (red), and the HPPK•MgATP•HP complex (blue).

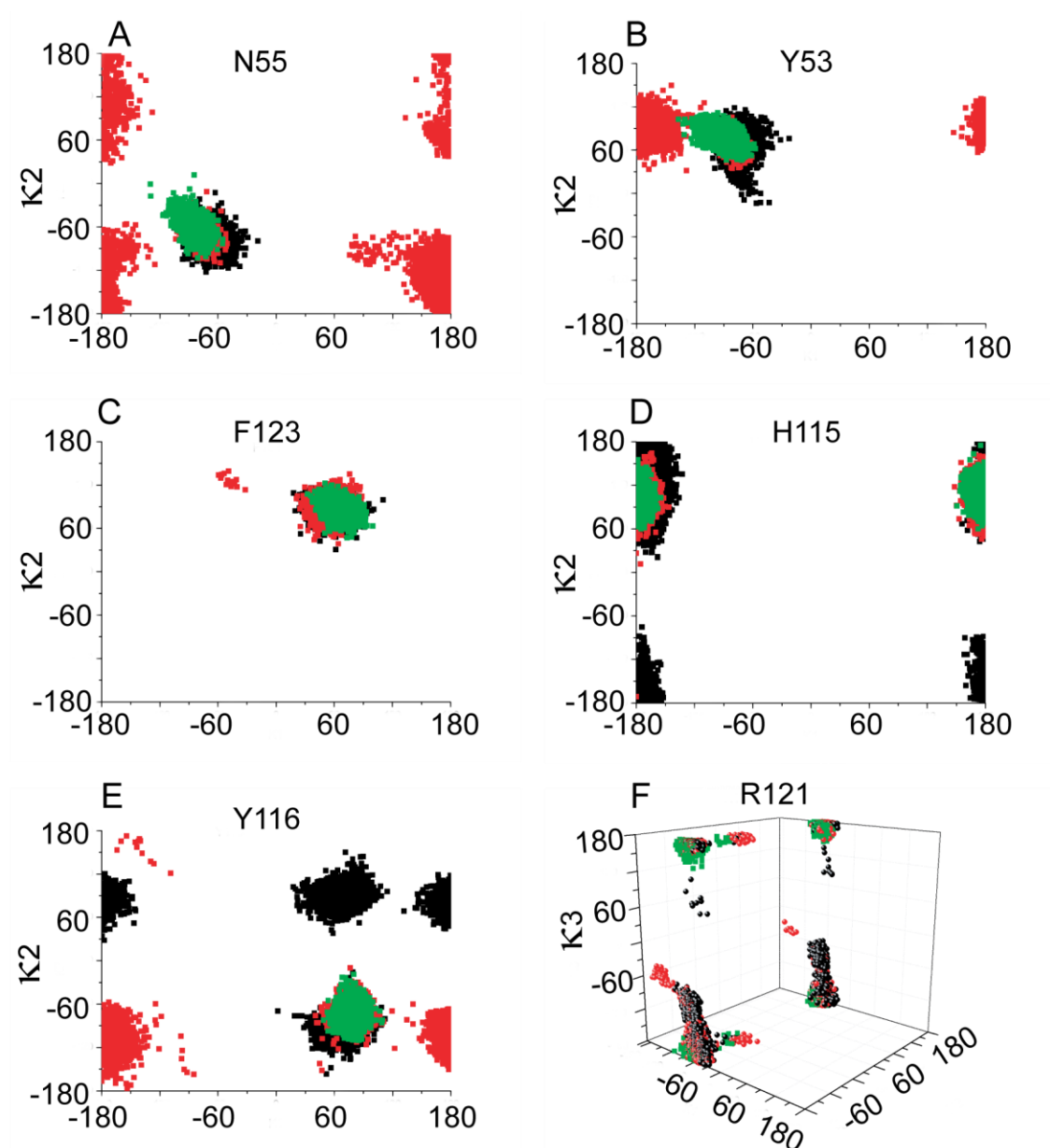


Figure 4.3. Side-chain conformations sampled during the MD simulations of the apo HPPK (black), HPPK•MgATP complex (red) and HPPK•MgATP•HP complex (green). The side-chain conformation was defined by side-chain dihedral angles χ_1 , χ_2 and χ_3 . See text for the detailed definition. (χ_1, χ_2, χ_3) plot was used to represent the arginine side-chain conformation and the (χ_1, χ_2) plot was used for all the other residues. The panels A, B, C, D, E and F are for residue N55, Y53, F123, H115, Y116 and R121, respectively

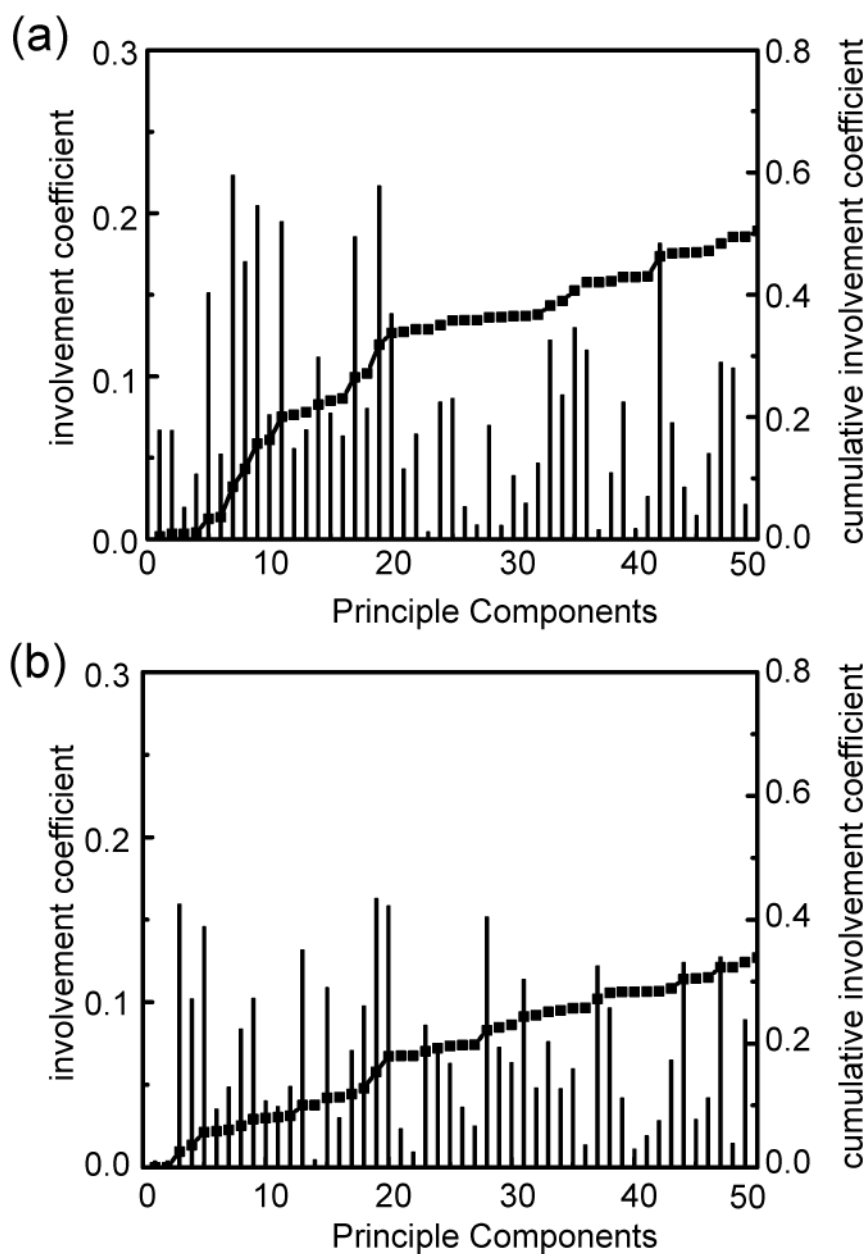


Figure 4.4: The involvement coefficients (thin bar; left axis) and the cumulative involvement coefficients (thick line, right y-axis) of the first 50 PCs. The involvement coefficients are calculated by projecting the PCs of the MD simulation of the apo HPPK onto the difference vector between the apo and the binary complex (a) and projecting the PCs of the MD simulation of the binary complex onto the difference vector between the binary and ternary complexes (b).

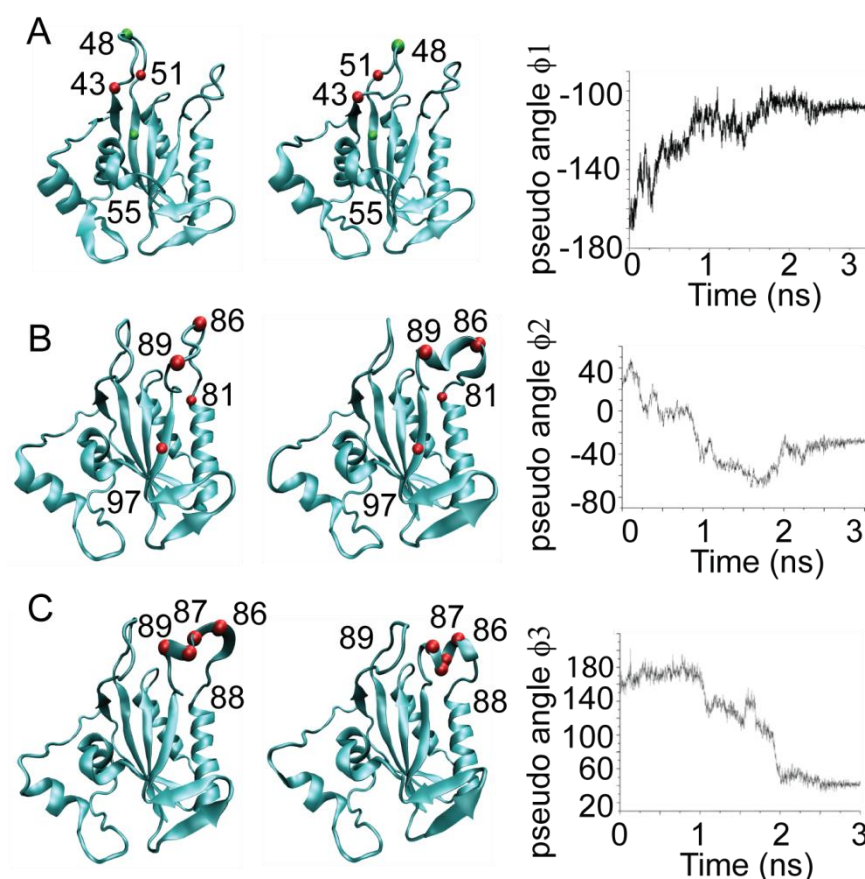


Figure 4.5. The characteristic motions and reaction coordinates from the apo HPPK to HPPK•MgATP complex. In each row, the left panel represents the initial conformation, the middle panel the target conformation, and the right panel the time evolution of the chosen reaction coordinates along the TMD trajectory. The reaction coordinates are pseudo dihedral angles CA55-CA43-CA48-CA51 (4A), CA97-CA81-CA86-CA89 (4B), and CA89-CA88-CA87-CA86 (4C). The involved atoms are shown in balls in the cartoon drawings.

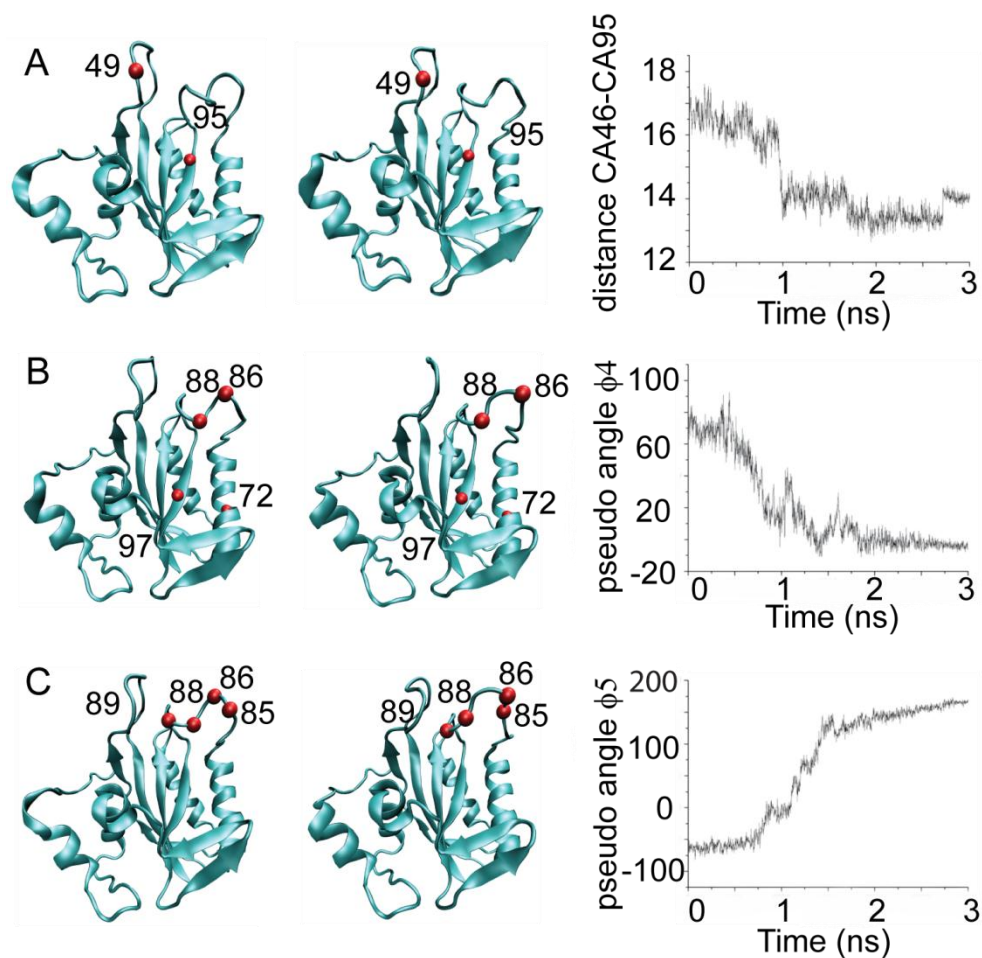


Figure 4.6. The characteristic motions and reaction coordinates from HPPK•MgATP to HPPK•MgATP•HP complex. In each row, the left panel represents the initial conformation, the middle panel the target conformation, and the right panel the time evolution of the chosen reaction coordinates along the TMD trajectory. The reaction coordinates are the distance between CA46 to CA95 (5A) and the pseudo dihedral angle CA86-CA88-CA97-CA72 (5B), and CA86-CA85-CA88-CA89 (5C). The involved atoms are shown in balls in the cartoon drawings.

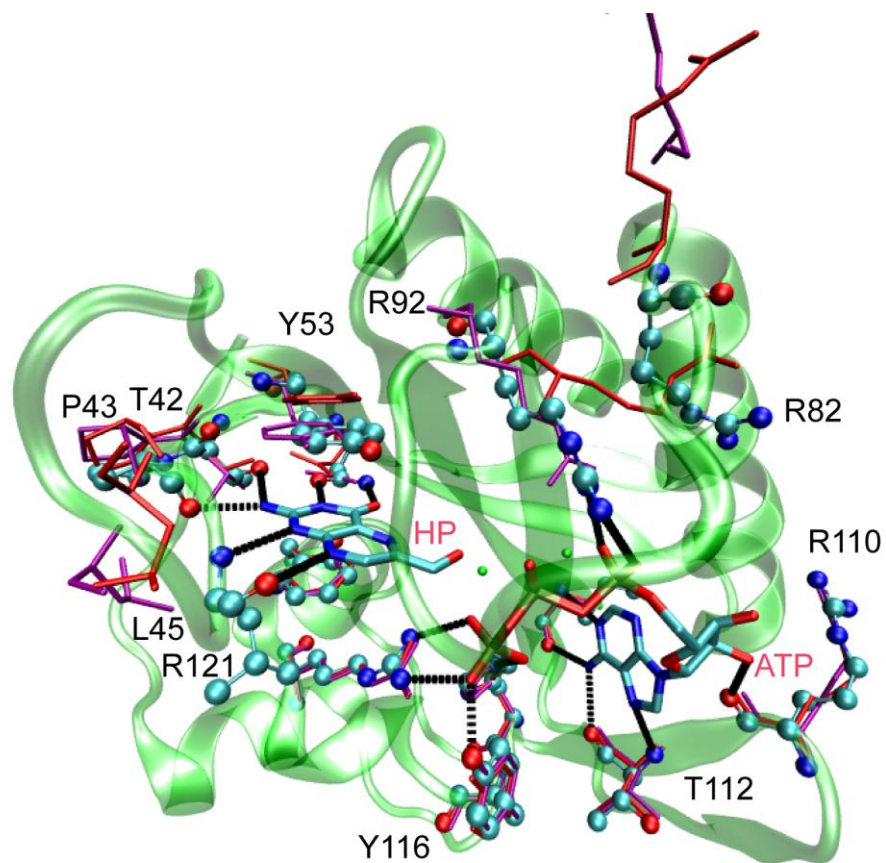


Figure 4.7. The conformations of the active site residues in the crystal structures of the apo HPPK, HPPK•MgATP and HPPK•MgATP•HP complexes. Hydrogen bonds are shown as black dashed lines. The active site residues in HPPK•MgATP•HP complex are shown in ball and stick model. The red and purple licorice model is for apo HPPK and HPPK•MgATP complex, respectively. HPPK in the ternary complex is shown in the green transparent cartoon.

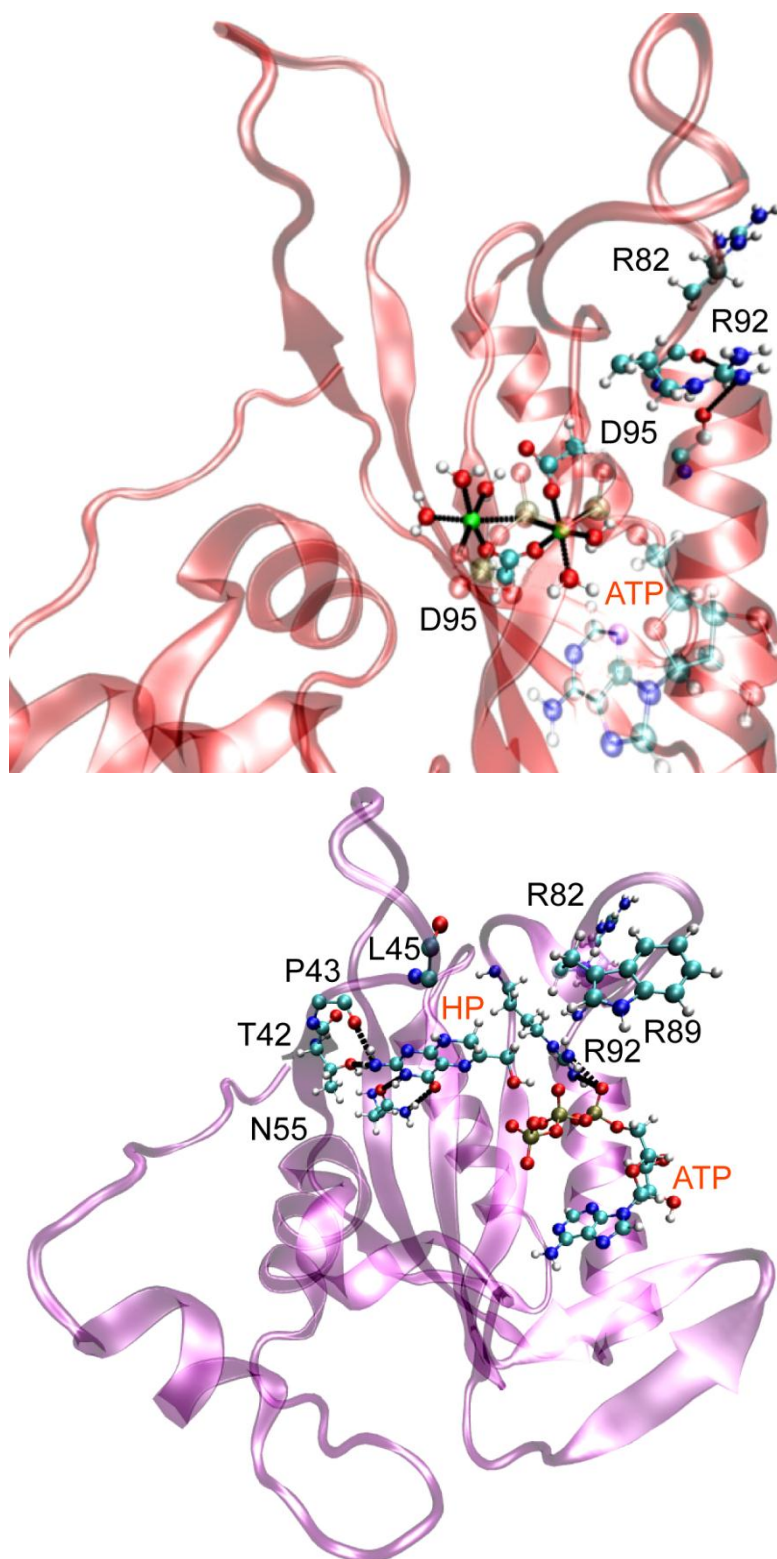


Figure 4.8. Snapshots of the HPPK^{apo}•ATP complex (upper panel) and the HPPK^{bin}•ATP•HP complex (lower panel) after 2 ns of MD simulation. The protein is shown in cartoon and the selected residues and ligands are shown in ball and stick. The hydrogen bonds and ion coordination are shown in black dashed lines.

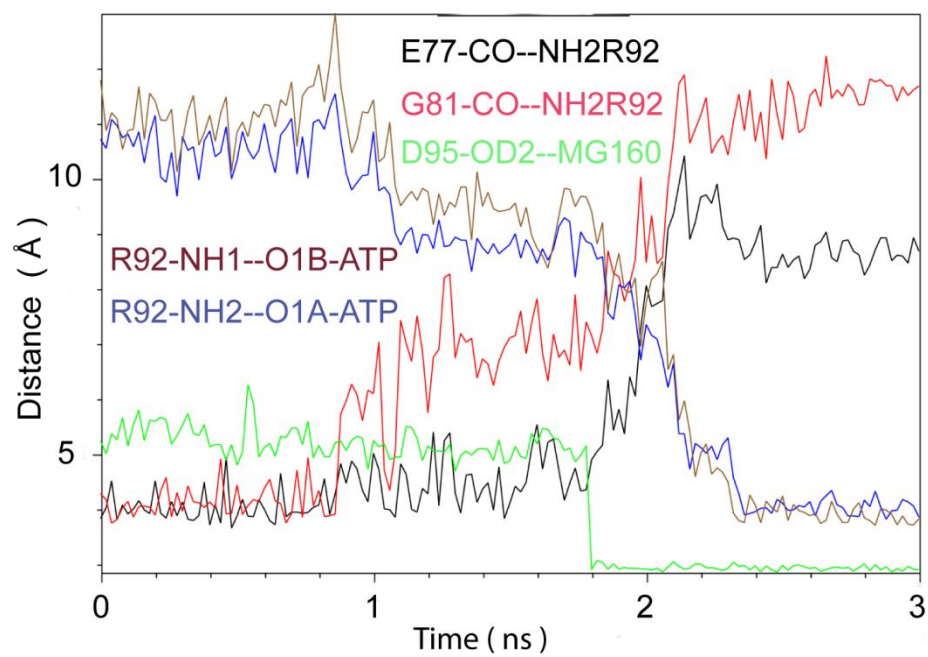


Figure 4.9. The positioning of R92 and D95 in the HPPK•MgATP binary complex as shown in the targeted MD simulation. The distances of six pair of atoms are plotted.

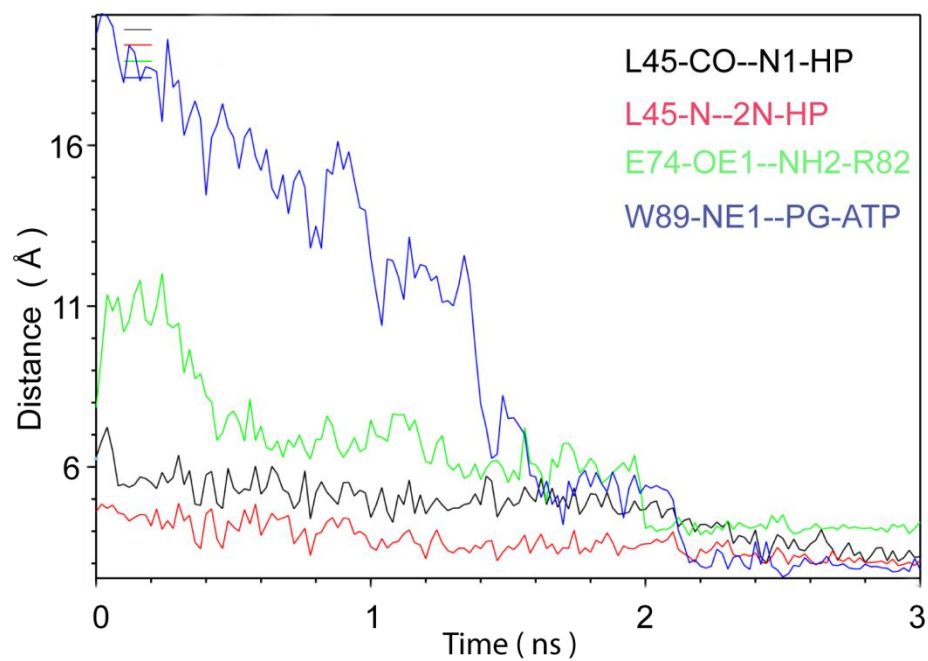


Figure 4.10. The positioning of L45, W89 and R82 in the HPPK•MgATP•MgHP complex as shown in the targeted MD simulation. Four interatomic distances were used as indicators as shown in the figure and were plotted along the simulation time.

References

REFERENCES

- 1 Yang, R., Lee, M. C., Yan, H. G., and Duan, Y. (2005) Loop conformation and dynamics of the *Escherichia coli* HPPK apo-enzyme and its binary complex with MgATP. *Biophys. J.* 89, 95-106.

- 2 Su, L., and Cukier, R. I. (2009) An enhanced molecular dynamics study of HPPK•ATP conformation space exploration and ATP binding to HPPK. *J. Phys. Chem. A* 113, 2025-2035.

- 3 Su, L., and Cukier, R. I. (2009) Hamiltonian replica exchange method study of *Escherichia coli* and *Yersinia pestis* HPPK. *J. Phys. Chem. B* 113, 16197-16208.

- 4 Ma, J. (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes, *Structure* 13, 373–380.

- 5 Ma, J., and Karplus, M. (1997) Ligand-induced conformational changes in ras p21: a normal mode and energy minimization analysis, *J. Mol. Biol.* 274, 114-131.

- 6 Schlitter, J., Engels, M., and Kruger, P. (1994) Targeted molecular-dynamics - a new approach for searching pathways of conformational transitions. *J. Mol. Graph.* 12, 84-89.

- 7 D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo,, M. Crowley, R. C. W., W. Zhang, K.M. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I., Kolossv áry, K. F. W., F. Paesani, J. Vanicek, X.Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke,, L. Yang, C. T., J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin,, and Kollman, a. P. A. (2008) AMBER 10, University of California, San Francisco.

Chapter 5

Conclusion and Perspective

HPPK is one of the key enzymes in the folate biosynthetic pathway, a potential for development of antimicrobial agents. Its catalytic mechanism and structure have been extensively studied. HPPK follows an ordered kinetic mechanism and product release is the rate limiting step of the HPPK-catalyzed reaction. The enzyme goes through dramatic conformational changes during its catalytic cycle, and the active center is assembled only after both substrates bind to the enzyme. The conformational changes are important for HPPK catalysis, but the molecular details of the conformational transitions are largely unknown. In this thesis work, we have investigated the conformational dynamics of HPPK from the apo form to the binary substrate complex and then to the ternary substrate complex, the first half of the catalytic cycle of HPPK, by a combination of experimental and computational studies.

¹⁵N-relaxation experiments were first carried out to characterize the dynamic properties on the first half of the HPPK catalytic cycle. Model-free analysis of the ¹⁵N NMR relaxation data together with the weak or missing signals revealed that the three catalytic loops are mobile on a wide range of timescales in the absence of ligand. The binding of the nucleotide to HPPK does not cause major changes in the dynamic properties of the enzyme. The internal mobility of HPPK is not reduced and is even moderately increased in the binary complex, particularly in the catalytic loops. However, the internal mobility of the catalytic loops is significantly reduced upon the formation of the ternary complex, but some mobility remains.

Our molecular dynamics simulation studies showed that HPPK uses the selected-fit mechanism for binding both substrates, because the side chains of key residues in the core region sample the conformations of different substrate binding forms of HPPK. The

PCA analysis of the 20 ns simulation showed that none of the principal components with large fluctuations are significantly correlated with the conformational transitions from the apo form to the ternary substrate complex. Targeted MD was used to explore the potential pathways of the conformational transition. It was found that loop 3 undergoes more complex motions than loop 2 in the conformational transitions and loop 2 is stabilized only after loop 3 reaches its target conformations.

Two mutated HPPK proteins were used to study the functional roles of HPPK dynamics: N10A and Q50A. Both N10 and Q50 are key residues in the hydrogen bond network that connects the three catalytic loops (loop coupling) but neither has direct interaction with the substrates or products. The structural and dynamic perturbations were studied by NMR. Chemical shift analysis showed that the loop conformations of both mutated enzymes are significantly different from those of the WT HPPK, as indicated by the large chemical shift changes in the ternary complexes of the mutated enzymes. The core region is essentially unperturbed, as the chemical shift differences between the WT HPPK and the mutated enzymes are less than 0.1 ppm for both the apo form and the ternary complex. This conclusion is supported by the J-coupling analysis. The high correlation between ternary complexes of either mutant enzyme and WT HPPK indicates high structural similarity. The significant differences in the measured $^3J_{\text{HNHA}}$ values of the loop regions suggest that the conformations of loops 1 and 2 of the N10A complex and loop 2 of the Q50A complex are significantly different from those of the WT complex.

^{15}N relaxation studies showed that both N10A and Q50A HPPK ternary complexes have more residues with significant R_{ex} values, suggesting that both N10A

and Q50A ternary complexes are more flexible than the WT counterpart. The effects on the conformations and dynamic properties caused by the N10A mutation are more severe than those caused by the Q50A mutation based on the extent of chemical shift perturbations, the number of missing and weak peaks, and the number residues requiring an R_{ex} term for the model-free analysis of the ^{15}N relaxation data. In combination with the results of the biochemical and crystallographic studies, the NMR data indicate that the loop coupling is important for the binding of the second substrate HP and the full assembling and stabilization of the active center and catalysis.

While my studies have provided some important insights to the conformational dynamics of HPPK, additional studies are needed for a more complete understanding. Below I propose several further studies of the conformational dynamics of HPPK by NMR.

The first study is to investigate the dynamics of the second half of the catalytic cycle, namely, product release. As product release is rate-limiting for HPPK-catalyzed reaction, it is important to understand the dynamic properties of this process. This will provide a complete view of the dynamic properties of HPPK through its catalytic cycle. ^{15}N -relaxation method can be used for this purpose.

The second study is to use ^{13}C probes for NMR dynamics studies. The information obtained from ^{15}N -based method is limited to the protein backbone and the side chains of tryptophan, arginine and asparagine and glutamine residues. Since substrate binding and product release of HPPK involves extensive hydrophobic

interactions with protein side chains, ^{13}C NMR relaxation studies may provide more detailed information about the interactions between HPPK and its substrates/ligands.

The third study is to use paramagnetic relaxation enhancement (PRE) method to study the loop conformational ensemble. PRE is widely used to study protein structure and dynamics. Compared to the short distance restraint of less than 6 Å provided by NOE measurement, PRE provides long range distance restraints in the range of 10-35 Å. With this information, it is possible to detect and visualize lowly populated states of protein. HPPK is suitable for PRE measurement. One of the problems with ^{15}N relaxation studies on HPPK is that the peaks of some key residues, such as some loop 2 and loop 3 residues are missing due to either solvent exchange or intrinsic flexibility of those regions, which makes it impossible to obtain the dynamic information of those residues from NMR relaxation studies. One way to use PRE to study the conformational dynamics is to place paramagnetic probes on the loops (2 and 3) that undergo dramatic conformational changes during the catalytic cycle. Since the largest displacement of loop residues is ~18 Å as revealed by X-ray crystallography, the dynamic information of these loops can be readily obtained by using the rigid core residue as a reference. Since there is not native cysteine residue in HPPK, the labeling position can be precisely controlled by introducing cysteine mutation.