

MSU
4
2007

**LIBRARY
Michigan State
University**

This is to certify that the
dissertation entitled

**THE DYNAMICS OF ENZYMATIC REACTION:
COMPUTATIONAL AND EXPERIMENTAL STUDIES**

presented by

LISHAN YAO

has been accepted towards fulfillment
of the requirements for the

Ph.D

degree in

Chemistry



Major Professor's Signature

5/2/06

Date

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

**THE DYNAMICS OF ENZYMATIC REACTIONS:
COMPUTATIONAL AND EXPERIMENTAL STUDIES**

By

Lishan Yao

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Chemistry
and Department of Biochemistry and Molecular Biology

2006

ABSTRACT

THE DYNAMICS OF ENZYMATIC REACTIONS: COMPUTATIONAL AND EXPERIMENTAL STUDIES

By

Lishan Yao

The roles of protein dynamics in enzymatic catalysis have been investigated by a combination of experimental and computational approaches. The enzyme systems under the studies include yeast cytosine deaminase (yCD), *Bacillus subtilis* guanine deaminase (bGD), and *Staphylococcus aureus* dihydroneopterin aldolase (SaDHNA).

yCD catalyzes the hydrolytic deamination of cytosine to uracil. The enzyme is of great biomedical interest, because it can also catalyze the deamination of the prodrug 5-fluorocytosine to form the anticancer drug 5-fluorouracil. A multidisciplinary approach has been employed to elucidate the catalytic mechanism of yCD. Computational studies, including quantum chemical calculations and molecular dynamics (MD) simulations, permit the suggestion of a complete reaction path for the deamination reaction at atomic resolution. The reaction path involves the formation of two tetrahedral intermediates. Abstraction and addition of protons are critical for facilitating the reaction, and the carboxyl group of glutamate-64 at the catalytic center plays a critical role in shuttling the protons. One product release path was identified that involves the loop containing phenylalanine 114 (F114) and the C-terminal helix of the protein. Nuclear magnetic resonance spectroscopy (NMR) has been used to study the catalytic mechanism and dynamics of the enzyme. A series of NMR studies have been carried out to study the binding of the anticancer drug 5-fluorouracil to the enzyme. Together with transient kinetic analysis, the NMR studies show that product release is rate-limiting in the

activation of the prodrug 5-fluorocytosine by yCD. NMR studies also indicate that the binding of an intermediate state analog rigidifies the F114 loop and C-terminal helix, while the binding of product maintains their flexibility on the second to minute timescale. Flexibility changes of the F114 loop on the picosecond to nanosecond timescale were found based on both NMR relaxation and MD simulation studies.

Guanine deaminase catalyzes the hydrolytic deamination of guanine and plays an important role in the regulation of guanine nucleotide pool. A complete reaction path of the bGD-catalyzed reaction has been proposed by using a combination of a multilayer quantum mechanical method and an MD simulation method. Two residues, Glutamate 55 and Aspartate 114, play important roles in the proton shuttling in the bGD-catalyzed reaction. The dynamical properties of the active site have been elucidated by the MD simulations.

Dihydroneopterin aldolase catalyzes the conversion of dihydroneopterin to 6-hydroxymethyl-7,8-dihydropterin (HP) in the biosynthesis of folate. The dynamic properties of SaDHNA and its complex with the reaction product HP have been investigated by MD simulations. Flexible regions of the apo protein surrounding the active site have been identified and, upon binding HP, show that the active site is rigidified. Specific residues important to product binding and the catalytic mechanism of DHNA are associated with these flexible regions, and their interactions with HP account for most of the binding energy. An exit path for HP has been found and the barrier to its release estimated.

Dedicated to itself

ACKNOWLEDGMENTS

Firstly, I would like to thank my advisors Dr. Honggao Yan and Dr. Robert I. Cukier for providing me the opportunity to study enzyme system computationally and experimentally. Their guidance has been benefiting me and their enthusiasm for research will influence me further in my future.

Secondly, I would like to thank the members of my guidance committee Dr. Jim Geiger, Dr. James Harrison and Dr. Michael Garavito for their advice and critical review of my thesis.

Thirdly, I would like to say thanks to Dr. Aizhuo Liu and Guangyu Li for their help in NMR experiments and Dr. Stepan Sklenak and Dr. Yuxiang Bu for their help in quantum mechanical calculations. Their kindness and patience has made me learn things faster and easier.

Fourthly, I would like to thank Dr. Yue Li and Yan Wu for teaching me molecular biology and protein purification techniques. I would also like to express my gratitude to members of Dr. Cukier's lab Dr. Steve Seibold, Hongfeng Lou and Li Su and members of Dr. Yan's lab Yi Wang, Dr. Jaroslaw Blaszczyk, Jifeng Wang and Zhengwei Lu. It is a great pleasure to work with them.

TABLE OF CONTENT

LIST OF FIGURES	ix
LIST OF TABLES	xvi
Chapter 1	1
Introduction.....	1
Part I. Theoretical Background.....	3
Section 1. NMR Relaxation.....	3
Section 2. Molecular Dynamics Simulation.....	22
Section 3. ONIOM: A Multilayered Integrated Quantum Mechanics Method	28
Part II Studied Proteins	31
Section 1. Yeast Cytosine Deaminase	31
Section 2. Dihydroneopterin Aldolase	32
Section 3. Guanine Deaminase.....	33
Reference	35
Appendices.....	39
Chapter 2.....	40
Product Release Is Rate-Limiting in the Activation of the Prodrug 5-Fluorocytosine by Yeast Cytosine Deaminase	40
Experimental procedures	42
Results.....	48
Discussion	54
Appendices.....	61
Chapter 3.....	68
Yeast Cytosine Deaminase Dynamics Study by NMR and Molecular Dynamics Simulation.....	68
Introduction.....	68
Materials and Methods.....	71
Results.....	75
Discussion	84
References.....	87
Appendices.....	90
Chapter 4.....	98
The Dynamics Changes of YCD along Reaction Cycle Revealed by HD exchange ..	98
Introduction.....	98
Materials and Methods.....	101
Results.....	103
Summary	114
Reference	117
Appendices.....	119

Chapter 5	127
A Molecular Dynamics Exploration of the Catalytic Mechanism of Yeast Cytosine Deaminase	127
Introduction	127
Methods	129
Results and Discussion	137
Concluding remarks	147
Reference	150
Appendices	152
Chapter 6	166
Product Release Study of Yeast Cytosine Deaminase by a Combination Method of ONIOM and Molecular Dynamics Simulation	166
Introduction	166
Method	167
Results and discussion	172
Conclusion	175
Reference	177
Appendices	178
Chapter 7	182
A Molecular Dynamics Study of The Ligand Release Path in Yeast Cytosine Deaminase	182
Introduction	182
Methods	185
Results and Discussion	188
Conclusions	199
Reference	201
Appendices	202
Chapter 8	211
Reaction Mechanism of Guanine Deaminase: An ONIOM and Molecular Dynamics Simulation Study	211
Introduction	211
Methods	213
Results	216
Discussion	228
Conclusion	230
References	231
Appendices	233
Chapter 9	240
Mechanism of Dihydroneopterin Aldolase: A Molecular Dynamics Study of the Apo Enzyme and Its Product Complex	240
Introduction	240
Methods	243

Results and Discussion247

Conclusions.....261

References.....266

Appendices.....268

LIST OF FIGURES

Figure 1-1.Periodic boundary condition box. The box in the center is box P (primary) ..	39
Figure 1-2 Aldolase reaction catalyzed by DHNA. A represents acid, B represents base.	39
Figure 2-1 Stopped-flow analysis of the activation of the prodrug 5FC by yCD. (A) The time course of the reaction as monitored by OD ₂₉₆ . The reaction mixture contained 10 μ M yCD and 250 μ M 5FC. The solid line was obtained by nonlinear least square fit to an equation with an exponential and a linear term. (B) The linear dependence of the rate constant of the exponential term on the concentration of 5FC.	63
Figure 2-2 Quench-flow analysis of the activation of the prodrug 5FC by yCD. The solid lines were obtained by global fitting using the numerical analysis program DYNAFIT.	64
Figure 2-3 ¹ H- ¹⁵ N HSQC spectra of yCD. Sequential assignments are indicated with residue numbers. (A) 1.5 mM yCD; (B) 1.5 mM yCD + 12 mM 5FU.....	65
Figure 2-4 ¹⁹ F NMR spectra of free (unbound) 5FU and the yCD-bound 5FU. The smaller peaks near the strong peak of free 5FU in the upper spectrum are ¹³ C satellite peaks.	66
Figure 2-5 ¹⁹ F NMR spectra of yCD labeled with 5-fluorotryptophan. (A) 1.5 mM 5-fluorotryptophan-labeled yCD; (B) 1.5 mM 5-fluorotryptophan-labeled yCD + 24.7 mM 5FU.	66
Figure 2-6 NMR saturation transfer analysis of the binding of 5FU to the 5-fluorotryptophan-labeled yCD. The NMR sample contained 1.5 mM 5-fluorotryptophan-labeled yCD and 24.7 mM 5FU. The data in panels A and B were obtained by irradiating the ¹⁹ F NMR nuclei of the bound and free yCD, respectively. The solid lines were obtained by nonlinear least square fit of the data as described in the “Experimental Procedures” section.....	67
Figure 3-1 ¹⁵ N relaxation data of yCD in the (a). apo, (b). inhibitor 5FPy bound form measured at 600 MHz and plotted vs residue.	91
Figure 3-2 Model-free dynamics results for the (a) apo form (b) inhibitor 5FPy bound form yCD from NMR ¹⁵ N relaxation data.	92
Figure 3-3 RMSF and order parameter of the (a), apo form. (b), inhibitor 2Py bound form. (c). product bound form. The x-ray B-factors of apo and 2Py bound form were converted and plotted in red.....	93

Figure 3-4 X-ray structure of yCD complexed with 2Py. Three regions 111-117, 150-158 and 72*-81* (from adjacent protomer) are highlighted in gold. Three residues F114, W152 and I156 which blocks the active site were also shown.....	94
Figure 3-5 τ_e value calculated from MD simulation for the apo form, inhibitor 2Py complex and product complex.....	94
Figure 3-6 two sidechain dihedral angles defined in MD data analysis for (a). F114, (b). W152, (c). I156.....	95
Figure 3-7 F114 sidechain dihedral angle ω_1 and ω_2 of the (a). apo form, (b). inhibitor 2Py bound form, (c). product uracil bound form in 10 ns MD simulation.....	95
Figure 3-8 The sidechain dihedral angle ω_1 and ω_2 of W156 in apo form.....	96
Figure 3-9 I156 sidechain dihedral angle ω_1 and ω_2 of the (a). apo form, (b). inhibitor 2Py bound form, (c). product uracil bound form in 10 ns MD simulation.....	96
Figure 3-10 The sidechain conformations of F114, W152 and I156.....	97
Figure 3-11 RMSF vs S2 plot of the (a) α helix and β sheet residues, (b) other region residues.....	97
Figure 4-1 The deamination of cytosine/5FC to uracil/5FU through a tetrahedral intermediate.....	120
Figure 4-2 The chemical shift difference (a). in the apo and 5FPy bound form. (b). in the apo and 5FU bound form.....	120
Figure 4-3 The colored tube plot of backbone CAs based on chemical shift difference (a). in apo and 5FPy bound form. (b). in apo and 5FU bound form. The difference increases with the color changing from blue to red. The residues with no assignment were painted white.....	121
Figure 4-4 Exchange time for (a) apo, (b) 5FPy complex, (c) 5FU complex. \blacktriangle represents no exchanges (exchange time longer than ~2000 minutes (apo) or ~4000 minutes (5FPy and 5FU complex)), Δ represents fast exchanges (exchange time shorter than ~5 minutes).....	122
Figure 4-5 Number of water molecules around NH within 3.5Å from N atom in MD simulation. \blacktriangle represents no exchanges residues (exchange time longer than ~2000 minutes).....	123
Figure 4-6 (a). Percentages of H-bond formation between NH and other residues. (b). Percentages of H-bond formation between NH and solvent water during 1.5 ns MD simulation.....	123

Figure 4-7 X-ray structure of yCD complexed with 2Py. Regions with significant increase of HD exchange time were highlighted in gold. 124

Figure 4-8 Summary of HD exchange data of yCD in (a). apo, (b). 5FPy complex and (c) 5FU complex forms at pH 7. Residues are colored according to the logarithm of protection factor: blue, residues without exchanges during experimental time ($\log P \geq 7.0$); red, residues with fast exchanges ($\log P \leq 2.0$); orange to iceblue, residues with slow exchanges ($7.0 > \log P > 2.0$). The crystal structure of apo form subunit one was used in the picture..... 124

Figure 4-9 The ratio between the exchange time of (a). 5FPy and apo yCD. (b). 5FU and apo yCD. Δ represents the binding of ligand changes the residue exchange time from no exchange to slow exchange or from slow exchange to fast exchange or from no exchange to fast exchange. \blacktriangle represents the reverse change..... 125

Figure 4-10 The exchange time ratio in apo form at pH 7.16 and 8.05..... 125

Figure 4-11 The exchange time ratio in 5FPy bound form at (a). pH 7.10 and 7.62, (b). pH 7.62 and 8.50..... 126

Figure 4-12 The exchange time ratio in 5FU bound form at pH 7.10 and 7.50. 126

Figure 5-1. (a) Ribbon representation of the crystal structure of yCD drawn according to the coordinate of the 1.14 Å crystal structure.⁴ The figure was prepared with Molscript²⁶ and Raster3D.^{27,28} (b) Schematic drawing of the coordination sphere of the catalytic zinc ion and polar interactions between yCD and a reaction intermediate analog as revealed by X-ray crystallography.^{4,5} The distances (Å) between the zinc ion and its ligands and between the heavy atoms involved in hydrogen bonds are obtained from one subunit (A) of the 1.14 Å crystal structure of yCD with the reaction intermediate analog complex. 159

Figure 5-2 Schematic representation of the path of the yCD-catalyzed reaction as revealed by our quantum chemical calculations⁸ and molecular dynamic simulations (this work): 1. water/cytosine active site. 2. hydroxide/Glu64H cytosine complex. 3 hydroxide/Glu64 cytosineH complex. 4. intermediate I complex. 5. intermediate II complex. 6. uracil/ammonia complex. 7. uracil/water complex..... 160

Figure 5-3 RMSDs of CA (residues 15-158) for subunit one and subunit two compared with the corresponding crystal structures. The first 500 ps were treated as an equilibration stage, while the last 1500 ps were used to do the data analysis. Left panel: free form, right panel: intermediate analog complex. 161

Figure 5-4 (a) RMSDs from the crystal structure of the C-terminal helix (left panel) and (b) the Phe114 loop (right panel) in the free yCD simulation, after equilibration..... 161

Figure 5-5 RMSFs of CAs in the yCD simulations compared with the crystal structure B-factors. (a) Left panel: free yCD simulation. (b) Right panel: yCD intermediate analog complex..... 162

Figure 5-6 Schematic representation and MD snapshots of the yCD active site. Top panel: free form. Bottom panel: intermediate analog complex form.....	163
Figure 5-7 Changes in the CB-CG-CD-OE1 dihedral angle of Glu64. (a). Free yCD MD simulation. (b). yCD water/cytosine model (see 1 in Figure 2). (c) yCD intermediate I complex (see 4 in Figure 2).	164
Figure 5-8 Potential of mean force for rotation around the CB-CG-CD-OE2 dihedral angle of Glu64H.....	165
Figure 6-1 The flow chart of MD ONIOM combination used in Zn-O bond cleavage process.....	178
Figure 6-2 The ONIOM energy changes along the Zn-O4 distance scan.....	178
Figure 6-3 The rearrangement of the active site during MD simulation after the Zn-O4 bond cleavage.....	179
Figure 6-4 The average forces between Zn and O4. The bars in the plot represent the errors.	179
Figure 6-5 The potential of mean force along the Zn-O4 bond cleavage.	180
Figure 6-6 The three states defined based on PMF during Zn-O4 bond cleavage.	180
Figure 6-7 The distance change between Zn-OE1 during the cleavage of Zn-O4.	181
Figure 7-1 a. 5-fluro-uracil (5FU) b. 5-fluro-cytosine (5FC)	203
Figure 7-2 CA RMSFs of: (a). protomer 1. (b). protomer 2 at 300 K (black) and 320 K (green). The crystal B-factors were converted to RMSFs (red). The difference (blue) is the RMSF difference of the 300 K and 320 K data.	203
Figure 7-3 CA RMSF plots of the first 5 PCA modes at 300 K and 320 K and their difference: (a) protomer 1. (b) protomer 2.....	204
Figure 7-4 (a) Schematic graph of 50 trajectory snapshots of CA atoms projected out of the first 5 PCA motion modes at 300 K and 320 K. The F114 loop and C-terminal helix were labeled to give a better view. Water molecules can diffuse into the active site at 320 K through the path in between F114 loop and C-terminal helix, but not at 300 K. (b) One snapshot of apo MD simulation at 320 K. Water molecules diffuse into the active site through the triangle mouth defined by C91, F114 and I156. (c) 3D plots of distances d1 (C91 CA-F114 CZ), d2 (C91 CA-I156 CD) and d3 (F114 CZ-I156) at 300 K (red) and 320 K (black).	205
Figure 7-5 (a). Total CA RMSD versus time obtained by comparing the MD snapshots with the crystal structure. (b). The average restraint force between cytosine and the	

protein in each window. (c). Hydrogen bond lifetimes between cytosine and the protein during the cytosine pushing along path 1.	206
Figure 7-6 (a). CA RMSFs in 2 ns regular MD simulation. (b). RMSF difference by comparing the first stage of cytosine pushing with the regular simulation along path 1. (c). RMSF difference by comparing the second stage with the regular simulation.....	207
Figure 7-7 Schematic graph of 50 trajectory snapshots of CA atoms projected out of the first 5 PCA motion modes: (a). Regular simulation. (b). Pushing simulation stage 1. (c). Pushing simulation stage 2.....	208
Figure 7-8 The cytosine release paths from the pushing simulation.	208
Figure 7-9 Mass weighted RMSD plot of F114 and I156 along the pushing path of cytosine (path 1).....	209
Figure 7-10 (a). RMSD of all CA atoms compared with the crystal structure. (b). The average force calculated for the cytosine pushing along path 2. (c). The CA fluctuation differences between the regular simulation and the cytosine pushing.....	210
Figure 8-1 The deamination reaction catalyzed by GD.	235
Figure 8-2 Active site interaction of the guanine bound complex from the ONIOM calculation.	235
Figure 8-3 Reaction mechanism proposed for deamination of guanine to xanthine catalyzed by GD. All the species are labeled with numbers (see text).	236
Figure 8-4 Zn-O bond cleavage.	237
Figure 8-5 The proton transfer from Zn bound water to Asp114 through a water bridge.	237
Figure 8-6 Alternative mechanism with protonated Asp114.....	238
Figure 8-7 The tautomerization of guanine	239
Figure 9-1 CA atom RMSDs from the simulations and corresponding crystal structure B-factors. (a) Apo form. (b) HP bound form.	272
Figure 9-2 Comparison of the RMSFs derived from the average protomer RMSF and the octamer RMSF (see text for the difference in these RMSFs). (a) Apo form. (b) HP bound form.	272
Figure 9-3 A representative snapshot of two adjacent subunits of apo DHNA from the MD simulation. HP is placed for the identification of the active site. One subunit is colored in green, and the other in red. The regions corresponding to large RMSF	

differences are highlighted in gold color. For clarity, only one set of the regions is highlighted. 273

Figure 9-4 (a) RMSF differences between the apo DHNA and HP complex simulations. The differences based on the average protomer and octamer RMSF methods are displayed with solid and dashed lines, respectively. (b) RMSF differences between the apo DHNA and HP complex simulations excluding protomer 6, based on the octamer RMSF method. 273

Figure 9-5 Time evolution of the ψ dihedral angle of Asp 50 and the ϕ of Thr51 in the eight subunits displayed sequentially as one trajectory. 274

Figure 9-6 Principal Component Analysis eigenvalues of the first 20 modes in the apo and HP complex simulations. The total fluctuation over all the modes is 96.6 \AA^2 for the apo and 67.2 \AA^2 for the complex simulation. 274

Figure 9-7 The CA atom RMSF projections onto the first principal component eigenvector for the apo and HP bound forms. The enhanced fluctuation regions of apo versus HP complex are evident and similar to those found as displayed in Figure 4. 275

Figure 9-8 The CA distances: d_1 (Tyr*54 to Ile105), d_2 (Tyr*54 to Ala69) and d_3 (Ala69 to Ile105) parametric on the trajectory for the apo (blue) and HP bound (red) forms. There is a major and minor component for the apo form. The HP bound form distance fluctuations are smaller and fairly well confined to a part of the apo major state volume. 275

Figure 9-9 Active site interactions based on the MD simulations of apo enzyme (a) and HP complex (b). Residues in one protomer are distinguished from the other protomer by “*”s. 276

Figure 9-10 (a) Electrostatic and (b) van der Waals interaction energies between HP and all residues within 5 \AA of HP. The energies are averages over the 2 ns simulation and the protomers. 277

Figure 9-11 CA RMSF comparison between the push and regular simulations, indicating the small disturbance in the protein fluctuations from the exit of HP for the chosen force constant and pulling rate. 277

Figure 9-12 CA RMSD of the 50 ps averaged structure in each window relative to the average structure in the HP complex regular simulation. 278

Figure 9-13 Total CA fluctuation of DHNA along the exit trajectory of HP showing a range of protein fluctuations that span the bound to free form results. 278

Figure 9-14 CA RMSD of the flexible regions of the 50 ps averaged structure in each window relative to the average fragments in the HP complex regular simulation. (a) Residues 15–25. (b) Residues 45*–55*. (c) Residues 68–74. (d) Residues 100–110. 279

Figure 9-15 CA fluctuations of flexible regions along the exit path of HP. (a) Residues 15–25. (b) Residues 45*–55*. (c) Residues 68–74. (d) Residues 100–110. (e) sum of (a), (b), (c), and (d). 280

Figure 9-16 The interaction energies between HP and its surroundings over the exit pathway. (a) Electrostatic interactions between HP, and Glu22, Glu74 and Lys100. (The solvation energy of the starting point (–14.5 kcal/mol) was set to zero, to give a better view for this figure). (b) Overall interactions between HP and residues 15–25, 45*–55*, 68–74, 100–110, solvent and the other parts of the protein. 281

Figure 9-17 The intramolecular energetic and entropic contributions of HP, and its interaction energy with the environment and solvent, along the exit pathway. 281

LIST OF TABLES

Table 2-1 Kinetic Constants for the Activation of the Prodrug 5FC by yCD ^a	61
Table 3-1 The RMSF of center of mass of various helixes in the apo form.....	90
Table 4-1 The statistics of HD exchange times for the apo, 5FPy and 5FU complex in various pHs.	119
Table 5-1 Force constants for the Zn complex.	152
Table 5-2 Charges (<i>e</i> units) of Zn complex used in the yCD simulations.....	153
Table 5-3 RMSD for free and intermediate analog complex forms	154
Table 5-4 Hydrogen bonds in yCD free and intermediate analog complex forms	155
Table 5-5 Hydrogen bonds in yCD water/cytosine simulation.....	156
Table 5-6 Hydrogen bonds in yCD hydroxide/Glu64H cytosine simulation	156
Table 5-7 Hydrogen bonds in yCD hydroxide/Glu64 cytosineH simulation	157
Table 5-8 Hydrogen bonds in yCD intermediate I/II simulations	157
Table 5-9 Hydrogen bonds in yCD uracil/ammonia and uracil/water simulations.	158
Table 5-10 Chemical mutation free energies between water and ammonia.	158
Table 7-1 Total CA fluctuations and contributions from first 5 modes at 300 K and 320 K.....	202
Table 8-1 Comparison of ONIOM optimized structure with crystal structure of the imidazole inhibitor complex	233
Table 8-2 Hydrogen bonds in GD substrate bound form.....	233
Table 8-3 Hydrogen bonds in complex 8 bound form.....	234
Table 8-4 Hydrogen bonds in complex 9 bound form.....	234
Table 9-1 Active Site Interactions in Apo DHNA.....	268
Table 9-2 Hydrogen Bonds between HP and Its Surroundings in HP complex ^(a)	269
Table 9-3 Active Site Interactions in HP Complex	270

Table 9-4 Hydrogen bonds between Asp [*] 46, Asn71, and Leu72 in HP Complex	270
Table 9-5 Hydrogen Bonds with Glu24 and Ile25 in HP Complex	271
Table 9-6 RMSF of Arg118 Backbone N Atom in the Regular and Pushing MD Simulations	271

Chapter 1

Introduction

It has been long realized that the static structure of an enzyme alone is insufficient to explain how it precisely works. Proteins are dynamic molecules that often undergo conformational changes while performing their specific functions, such as an enzyme reaction or ligand binding. The dynamic properties intrinsic to a protein structure may provide information on the location and the energetics of the conformational change process, and are thus the focus of many biophysical studies. The focus of my thesis work is to study the roles of protein dynamics in enzymatic catalysis using both nuclear magnetic resonance (NMR) spectroscopy and molecular dynamics (MD) simulation methods. NMR is a very powerful technique for both biomolecular structure determinations and dynamics and thermodynamics of biological macromolecules. On the other hand, MD can provide details concerning individual atom motions as a function of time, which can be used to calculate thermodynamic parameters by applying statistical mechanics. MD and NMR can confirm each other and are complimentary. By combining these two techniques, one can obtain richer information about protein dynamics.

One drawback of MD simulation due to commonly used force fields is that electrons are not treated explicitly and chemical bonds are restrained by harmonic potentials. No chemical bond cleavage and formation can be studied directly by conventional MD methods. A Quantum mechanics (QM) method must be used for the study of the chemical steps of an enzymatic reaction, but in practice the size of biological systems prohibits the use of high-level QM methods on the whole system. A multilayered QM method has been developed to address this problem. Together with NMR and the computational methods, one can study enzymatic reaction mechanisms, dynamics and their relationship.

Part I. Theoretical Background

Section 1. NMR Relaxation

NMR can be used to monitor the dynamic behavior of a protein at a multitude of specific sites. Moreover, protein movements on a broad range of timescales can be monitored using various types of NMR experiments — nuclear spin relaxation rate measurements report the internal motions on fast (sub nanoseconds) and slow (microseconds to milliseconds) timescales as well as the overall rotational diffusion of the molecule (5–50 nanoseconds), whereas rates of magnetization transfer among protons with different chemical shifts and proton exchange report movements of protein domains on the very slow timescales (milliseconds to days). These features make NMR a unique and powerful tool in studying protein dynamics related to protein functions (Kay 1998; Ishima and Torchia 2000; Palmer 2001; Palmer et al. 2001; Stone 2001; Akke 2002).

The relaxation rate constants depend on the spectral density functions that quantify the frequency dependence of stochastic motions modulating the dipolar, chemical shift anisotropy, or quadrupolar interactions. The relaxation data are interpreted in terms of overall rotational diffusion of the molecule and intramolecular dynamics at specific atomic sites. Some brief illustrations about relaxation mechanisms in macromolecules are discussed below.

1). ps-ns timescale relaxation

In the semiclassical theory of spin relaxation, the Hamiltonian for the spin system is written as the sum of a deterministic quantum-mechanical Hamiltonian that acts only on the spin system, H_0 , and a stochastic Hamiltonian, $H_I(t)$, that couples the spin system to the lattice (Abragam 1961; Cavanagh et. al. 1996):

$$H(t) = H_0 + H_I(t) \quad (1)$$

The Hamiltonian $H_I(t)$ is regarded as time-dependent perturbations acting on the main time-independent Hamiltonian, H_0 . The Liouville equation of the density operator is:

$$d\sigma(t)/dt = -i[H(t), \sigma(t)] \quad (2)$$

In the interaction representation with

$$\sigma^* = e^{iH_0 t} \sigma e^{-iH_0 t}, H_I^*(t) = e^{iH_0 t} H_I(t) e^{-iH_0 t}$$

the equation (2) becomes

$$d\sigma^*(t)/dt = -i[H^*(t), \sigma^*(t)]$$

which can be solved by using time-dependent perturbation theory approximated to the second order

$$d\sigma^*(t)/dt = -i[H^*(t), \sigma^*(0)] - \int_0^t d\tau [H_I^*(t), [H_I^*(t-\tau), \sigma^*(0)]] \quad (3)$$

Since $H^*(t)$ varies randomly in time, we shall perform an average of the lattice ensembles. It will be assumed further that (Abragam 1961):

(a). It is permissible to neglect the correlation between $H^*(t)$ and $\sigma^*(0)$ in the average of (3) and average them separately. So the first term of equation (3) is zero because the average of $H^*(t)$ is zero (random perturbation). The justification of this assumption is as

following: The density function $\sigma^*(0)$ depends on the behavior of $H^*(t')$ ($t' \leq 0$) (equation (3)), and since there is a correlation between $H^*(t)$ taken at different times, in principle there is a correlation between $H^*(t)$ ($t > 0$) and $\sigma^*(0)$. But if $t \gg \tau_c$ (τ_c is the correlation time of $H^*(t)$), this correlation can be neglected and therefore average $H^*(t)$ and $\sigma^*(0)$ separately.

(b). It is permissible to replace $\sigma^*(0)$ by $\sigma^*(t)$ on the right hand side of equation (3).

Assuming at $t = 0$ the system is in equilibrium, $\sigma^*(0)$ can be expressed as

$$\sigma^*(0) = \sigma(0) = \exp(-\hbar H_0 / kT) / \text{tr}\{\exp(-\hbar H_0 / kT)\}$$

$\exp(-\hbar H_0 / kT)$ can be approximated by

$$\exp(-\hbar H_0 / kT) \approx E - \hbar H_0 / kT$$

at room temperature. E is the identity operator. For example, for ^{15}N atom with spin angular momentum $1/2$ at 300 K and 14.09 T magnetic field (600 MHz for proton resonance), $\langle \alpha | \hbar H_0 / kT | \alpha \rangle$ is 3.04×10^{-5} very small compared to the expectation of E

which is 1. Since E commutes with any operator, equation (3) only changes the

$-\hbar H_0 / kT$ part of $\sigma^*(0)$, which is a small perturbation. Therefore $\sigma^*(0)$ can be replaced

by $\sigma^*(t)$.

(c). It is permissible to extend the upper limit of integral (3) to $+\infty$. Since $t \gg \tau_c$, the integral from t to $+\infty$ is negligible.

(d). It is permissible to neglect all higher-order terms on the right-hand side of (3).

With these assumptions and omitting the bar on $\sigma^*(t)$ which stands for the average density matrix, equation (3) becomes

$$d\sigma^*(t)/dt = - \int_0^\infty d\tau \overline{[H_1^*(t), [H_1^*(t-\tau), \sigma^*(t)]]} \quad (4)$$

The solution of (4) is E which means all the states will have same probability corresponding to the infinite temperature situation. One more term σ_0 the equilibrium density operator needs to be introduced into (4) to ensure that the spin system relaxes to equilibrium, which modifies (4) to

$$d\sigma^*(t)/dt = - \int_0^\infty d\tau \overline{[H_1^*(t), [H_1^*(t-\tau), \sigma^*(t) - \sigma_0]]} \quad (5)$$

This equation can be obtained more rigorously by treating the lattice quantum mechanically (Abragam 1961). Then σ^* is transformed back to σ which gives

$$d\sigma(t)/dt = -i[H_0, \sigma] - \frac{1}{2} \int_{-\infty}^\infty d\tau \overline{[H_1(t), [e^{-iH_0\tau} H_1(t-\tau) e^{iH_0\tau}, \sigma(t) - \sigma_0]]}$$

For any physical observable, Q , the expectation value of Q can be denoted as:

$$d\langle Q \rangle / dt = \frac{d}{dt} Tr\{Q\sigma(t)\} = Tr\left\{Q \frac{d\sigma(t)}{dt}\right\} \quad (6)$$

The longitudinal and transverse of relaxation times, T_1 and T_2 can be calculated by replacing Q by I_z and I^+ (or I^-). In order to proceed, the nature of the perturbation Hamiltonian H_1 must be considered. In a real system, a very large number of physical interactions give rise to stochastic Hamiltonians capable of mediating spin relaxation. For spin 1/2 nuclei in diamagnetic biological macromolecules (such as backbone amide N),

the dominant relaxation mechanisms are the magnetic dipolar coupling with amide proton and chemical shift anisotropic (CSA) mechanisms (Cavanagh et. al. 1996).

a). Dipolar relaxation (Fischer et al. 1998)

Considering two-spin system I and S

$$\frac{d}{dt} \begin{bmatrix} \Delta \langle I_z(t) \rangle \\ \Delta \langle S_z(t) \rangle \end{bmatrix} = - \begin{bmatrix} \rho_I & \sigma_{IS} \\ \sigma_{IS} & \rho_S \end{bmatrix} \begin{bmatrix} \Delta \langle I_z(t) \rangle \\ \Delta \langle S_z(t) \rangle \end{bmatrix} \quad (7)$$

in which

$$\Delta \langle I_z(t) \rangle = \langle I_z(t) \rangle - \langle I_z \rangle_{eq}$$

$$\Delta \langle S_z(t) \rangle = \langle S_z(t) \rangle - \langle S_z \rangle_{eq}$$

$$\rho_I = d^2 \{ 6j(\omega_I) + 2j(\omega_I - \omega_S) + 12j(\omega_I + \omega_S) \}$$

$$\rho_S = d^2 \{ 6j(\omega_I) + 2j(\omega_I - \omega_S) + 12j(\omega_I + \omega_S) \}$$

$$\sigma_{IS} = d^2 \{ -2j(\omega_I - \omega_S) + 12j(\omega_I + \omega_S) \}$$

$$\text{with } d^2 = \frac{1}{8} \left(\frac{\mu_0}{4\pi} \right)^2 \left(\frac{\gamma_I \gamma_S \hbar}{r^3} \right)^2$$

where μ_0 is the permeability of free space, γ is gyromagnetic ratio, r is the distance between two dipoles, $j(\omega)$ is the spectral density evaluated at frequency ω . The longitudinal relaxation time T_I is the reciprocal of the longitudinal relaxation rate, ρ . The cross term is the Nuclear Overhauser Effect which allows for transfer of information between two nuclei. A similar expression can be obtained for I^+ , S^+

$$\frac{d}{dt} \begin{bmatrix} \langle I^+(t) \rangle \\ \langle S^+(t) \rangle \end{bmatrix} = - \begin{bmatrix} i\omega_I + \frac{1}{T_2^I} & 0 \\ 0 & i\omega_S + \frac{1}{T_2^S} \end{bmatrix} \begin{bmatrix} \langle I^+(t) \rangle \\ \langle S^+(t) \rangle \end{bmatrix} \quad (8)$$

where

$$\frac{1}{T_2^I} = d^2 \{4j(0) + 3j(\omega_I) + j(\omega_I - \omega_S) + 6j(\omega_I + \omega_S) + 6j(\omega_S)\}$$

$$\frac{1}{T_2^S} = d^2 \{4j(0) + 3j(\omega_S) + j(\omega_I - \omega_S) + 6j(\omega_I + \omega_S) + 6j(\omega_I)\}$$

T_2 is the transverse relaxation time.

b). Chemical shift anisotropic relaxation

$$\frac{1}{T_1^I} = c^2 \{6j(\omega_I)\}$$

$$\frac{1}{T_2^I} = c^2 \{4j(0) + 3j(\omega_I)\}$$

where $c^2 = \frac{\omega_I^2 \Delta\sigma^2}{18} \left(1 + \Delta\eta^2/3\right)$ with $\Delta\sigma = \sigma_{11} - \frac{\sigma_{22} + \sigma_{33}}{2}$, $\Delta\eta = \frac{\sigma_{22} - \sigma_{33}}{\sigma_{11} - \sigma_{iso}}$, and

$$\sigma_{iso} = \frac{\sigma_{11} + \sigma_{22} + \sigma_{33}}{3} \text{ with } \sigma_{11} \geq \sigma_{22} \geq \sigma_{33}. \sigma_{ii} \text{ is the diagonal part of the chemical}$$

shift tensor. By combining the dipolar interaction mechanism and the chemical shift

anisotropy mechanism, T_1 , T_2 and NOE for I spin are given by:

$$\frac{1}{T_1} = d^2 \{6j(\omega_I) + 2j(\omega_I - \omega_S) + 12j(\omega_I + \omega_S)\} + c^2 \{6j(\omega_I)\}$$

$$\frac{1}{T_2} = d^2 \{4j(0) + 3j(\omega_I) + j(\omega_I - \omega_S) + 6j(\omega_I + \omega_S) + 6j(\omega_S)\} + c^2 \{4j(0) + 3j(\omega_I)\} + R_{ex}$$

$$NOE = 1 + (\gamma_I / \gamma_S) d^2 \{-2j(\omega_I - \omega_S) + 12j(\omega_I + \omega_S)\} \times T_1$$

(9)

R_{ex} is micro-millisecond timescale motion induced by conformation exchange or chemical exchange. This motion only contributes to T_2 relaxation but not T_1 or NOE (discussed below).

In order to extract dynamics information from T_1 , T_2 and NOE , the explicit form of $j(\omega)$ has to be derived.

Defining

$$J(\omega) = \int_{-\infty}^{+\infty} G(\tau) e^{-i\omega\tau} d\tau$$

$G(\tau)$ is the correlation function. $J(\omega)$ is the spectral density of the interaction $G(\tau)$.

$j(\omega)$ is the real part of $J(\omega)$. For example, if

$$G(\tau) = \frac{1}{5} \exp[-\tau / \tau_m]$$

$j(\omega)$ will be

$$j(\omega) = \frac{2}{5} \frac{\tau_m}{(1 + \omega^2 \tau_m^2)}$$

where τ_m is the correlation time. But for a large system like a protein, with a distribution of relaxation times, the description of the correlation function becomes much more complicated. By assuming the overall rotation of the protein is isotropic, the most

commonly used spectral density function given by the Lipari-Szabo model free formalism (Lipari and Szabo 1982; Lipari and Szabo 1982):

$$j(\omega) = \frac{2}{5} \left[\frac{S^2 \tau_m}{1 + (\omega \tau_m)^2} + \frac{(1 - S^2) \tau_c}{1 + (\omega \tau_c)^2} \right] \quad (10)$$

The corresponding correlation function is

$$G(\tau) = \frac{1}{5} \exp[-\tau / \tau_m] \left\{ S^2 + (1 - S^2) \exp(-\tau / \tau_e) \right\} \quad (11)$$

in which $\tau_c^{-1} = \tau_m^{-1} + \tau_e^{-1}$, S^2 is the square of the generalized order parameter that characterizes the amplitude of the intra-molecular motion in a molecular reference frame, τ_m is the correlation time for global motion and τ_e is the effective correlation time for internal motions. The first part of (11) corresponds to the global rotation and the second part is the internal motion. The order parameter is defined by

$$S^2 = \sum_{q=-2}^2 \left| Y_2^q(\theta, \phi) \right|^2$$

in which $Y_2^q(\theta, \phi)$ is a modified second-order spherical harmonic function, defined by

$$Y_2^q(\theta, \phi) = \begin{cases} (3 \cos^2 \theta - 1) / 2 & q = 0 \\ \sqrt{3/2} \sin \theta \cos \theta e^{i\phi} & q = 1 \\ \sqrt{3/8} \sin^2 \theta e^{i2\phi} & q = 2 \end{cases}$$

$$Y_2^{-q}(\theta, \phi) = Y_2^{q*}(\theta, \phi)$$

The order parameter satisfies the inequality $0 \leq S^2 \leq 1$, in which lower values indicate larger amplitudes of angular motions. The model-free spectral density function was extended by Clore and co-workers (Clore et al. 1990):

$$j(\omega) = \frac{2}{5} \left[\frac{S^2 \tau_m}{1 + (\omega \tau_m)^2} + \frac{(S_f^2 - S^2) \tau_c}{1 + (\omega \tau_c)^2} \right] \quad (12)$$

by introducing a second exponential in the correlation function with

$$G(\tau) = \frac{1}{5} \exp[-\tau / \tau_m] \left\{ S^2 + (1 - S_f^2) \exp(-\tau / \tau_f) + (S_f^2 - S^2) \exp(-\tau / \tau_s) \right\} \quad (13)$$

τ_f and τ_s are correlation time for fast and slow internal motions. S_f^2 is the order parameter for the fast timescale internal motions. So far, we have introduced six parameters including R_{ex} , τ_f , τ_m , τ_e , S^2 , S_f^2 . More parameters can be introduced by adding more exponentials into equation (13). However, there are only three experimental observables (T_1 , T_2 and NOE) we can measure. It is impossible to get all six parameters unless more experiments are performed in different magnetic fields. On the other hand, some of these parameters might be redundant to describe dynamics of one specific spin system. In practice, 5 models are commonly used to describe the motion (Mandel et al. 1995).

Model 1: S^2

Model 2: S^2 and τ_e

Model 3: S^2 and R_{ex}

Model 4: S^2 , τ_e and R_{ex}

Model 5: S^2 , S_f^2 and τ_e

For those parameters not included in one specific model, they are set to 0 (except S_f^2

which is set to 1.0). The χ^2 test is used to justify whether one model is sufficient to

describe the motion and F hypothesis test is used to judge whether the experimental data can be better described by introducing more parameters (Mandel et al. 1995). A new study shows that the F test prefers simple models (such as model 1, 2) which underestimate τ_e and/or R_{ex} (d'Auvergne and Gooley 2003; Chen et al. 2004). New methods, the Bayesian Information Criteria (BIC) and Akaike's Information Criteria (AIC), are also used to compare different models (d'Auvergne and Gooley 2003; Chen et al. 2004; Ding et al. 2005).

Before the model-free analysis, global correlation time τ_m has to be determined. By choosing rigid residues essentially with $\tau_e = 0$, $R_{ex} = 0$ and $S^2 = 1$, the ratio of T_2 and T_1 is:

$$\frac{T_1}{T_2} = \frac{d^2 \{4j(0) + 3j(\omega_I) + j(\omega_I - \omega_S) + 6j(\omega_I + \omega_S) + 6j(\omega_S)\} + c^2 \{4j(0) + 3j(\omega_I)\}}{d^2 \{6j(\omega_I) + 2j(\omega_I - \omega_S) + 12j(\omega_I + \omega_S)\} + c^2 \{6j(\omega_I)\}}$$

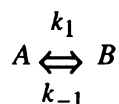
in which

$$j(\omega) = \frac{2}{5} \left[\frac{\tau_m}{1 + (\omega\tau_m)^2} \right]$$

Thus the ratio is only a function of τ_m , which can be obtained by using a non-linear least square fitting method. After the evaluation of τ_m , model-free analysis can be performed for internal motions. Model-free analysis has become a common method to extract dynamics information out of relaxation parameters, though other methods can also be used (Peng and Wagner 1992). But the order parameter generated by the model-free method gives a direct measurement of the angular flexibility of individual spin system (e.g. backbone NH).

2). Microsecond to second relaxation

If a chemical or conformational exchange process alters the magnetic environment of atoms in a molecule, either by transferring the atom to a new chemical context or by altering the surrounding milieu, then the resulting time dependence of the resonance frequencies of the nuclear spins contributes to dephasing of transverse coherences and transfer of magnetization between sites. These phenomena are referred to as chemical or conformational exchange. Consider (Palmer et al. 2001)



where A, B are two magnetically distinct environments. The chemical exchange rate constant, k_{ex} , is given by

$$k_{ex} = k_1 + k_{-1} = k_1 / p_B = k_{-1} / p_A$$

$$\text{where } \begin{cases} p_A = [A]/([A] + [B]) \\ p_B = 1 - p_A \end{cases}$$

Two situations will be discussed: (a) Only one signal is observed which means k_{ex} are larger than the chemical shift difference between A and B. (b) A and B are observable as separated signals. The first case is referred as fast exchange while the second case is referred as slow exchange.

a). Fast exchanges

Since in the fast exchanges only one signal is observed, an effective Hamiltonian can be introduced as (Wennerst.H 1972):

$$H(t) = f_A(t)H_A(t) + f_B(t)H_B(t)$$

$$\begin{cases} f_A = 1 \\ f_B = 0 \end{cases} \text{ when spin stays in state A. } \begin{cases} f_A = 0 \\ f_B = 1 \end{cases} \text{ when spin stays in state B. No transition}$$

state is assumed. By using the same procedure described in the first section, T_1 , T_2 and NOE contributions from the exchange can be derived:

$$\begin{aligned} \frac{1}{T_1} &= p_A / T_{1A} + p_B / T_{1B} \\ \frac{1}{T_2} &= p_A / T_{2A} + p_B / T_{2B} + p_A p_B (\omega_A - \omega_B)^2 / k_{ex} \\ NOE &= p_A NOE_A + p_B NOE_B \end{aligned} \quad (14)$$

ω_A, ω_B are resonance frequencies for A, B. $T_{1A}, T_{1B} (T_{2A}, T_{2B})$ are longitudinal

(transverse) relaxation time for A, B. The assumptions $k_{ex} \gg \frac{1}{T_{1A}}, \frac{1}{T_{1B}}, \frac{1}{T_{2A}}, \frac{1}{T_{2B}}$ and

$k_{ex} \ll \frac{1}{\tau_{mA}}, \frac{1}{\tau_{mB}}$ are used for the derivation, which means the conformational

exchange rate has to be much faster than relaxation rates but much slower than the molecular global tumbling rate. $k_{ex} \gg \omega_A - \omega_B$ is also implicitly used to guarantee only one signal is observed. Eq. (14) shows that for residues with multiple conformations, T_1 , T_2 and NOE are more complicated. But if A and B have quite similar relaxation rates and the NOE of one form is dominant, we can simply add a $R_{ex} (R_{ex} = p_A p_B (\omega_A - \omega_B)^2 / k_{ex})$ to the T_2 term in eq. (9).

b). Slow exchanges

In the slow exchange, with the only condition $k_{ex} \ll \frac{1}{\tau_{mA}}, \frac{1}{\tau_{mB}}$, a more general

equation for longitudinal relaxation can be derived by modified Bloch equations

(McConnell 1958; Palmer et al. 2001):

$$\frac{d}{dt} \begin{bmatrix} \Delta \langle M_A^Z(t) \rangle \\ \Delta \langle M_B^Z(t) \rangle \end{bmatrix} = - \begin{bmatrix} \frac{1}{T_{1A}} + p_B k_{ex} & -p_A k_{ex} \\ -p_B k_{ex} & \frac{1}{T_{1B}} + p_A k_{ex} \end{bmatrix} \begin{bmatrix} \Delta \langle M_A^Z(t) \rangle \\ \Delta \langle M_B^Z(t) \rangle \end{bmatrix} \quad (15)$$

in which

$$\begin{aligned} \Delta \langle M_A^Z(t) \rangle &= \langle M_A^Z(t) \rangle - \langle M_A^Z \rangle_0 \\ \Delta \langle M_B^Z(t) \rangle &= \langle M_B^Z(t) \rangle - \langle M_B^Z \rangle_0 \end{aligned}$$

$\langle M_A^Z \rangle_0, \langle M_B^Z \rangle_0$ denote the equilibrium expectation value of M_A^Z and M_B^Z .

The solution of this equation is:

$$\begin{bmatrix} \Delta \langle M_A^Z(t) \rangle \\ \Delta \langle M_B^Z(t) \rangle \end{bmatrix} = \begin{bmatrix} a_{AA}(t) & a_{AB}(t) \\ a_{BA}(t) & a_{BB}(t) \end{bmatrix} \begin{bmatrix} \Delta \langle M_A^Z(0) \rangle \\ \Delta \langle M_B^Z(0) \rangle \end{bmatrix} \quad (16)$$

where

$$\begin{aligned}
a_{AA}(t) &= \frac{1}{2} \left(1 - \frac{\frac{1}{T_{1A}} - \frac{1}{T_{1B}} + k_{ex}(p_B - p_A)}{\lambda_+ - \lambda_-} \right) \exp(-\lambda_- t) \\
&+ \frac{1}{2} \left(1 + \frac{\frac{1}{T_{1A}} - \frac{1}{T_{1B}} + k_{ex}(p_B - p_A)}{\lambda_+ - \lambda_-} \right) \exp(-\lambda_+ t) \\
a_{BB}(t) &= \frac{1}{2} \left(1 + \frac{\frac{1}{T_{1A}} - \frac{1}{T_{1B}} + k_{ex}(p_B - p_A)}{\lambda_+ - \lambda_-} \right) \exp(-\lambda_- t) \\
&+ \frac{1}{2} \left(1 - \frac{\frac{1}{T_{1A}} - \frac{1}{T_{1B}} + k_{ex}(p_B - p_A)}{\lambda_+ - \lambda_-} \right) \exp(-\lambda_+ t) \\
a_{AB}(t) &= \frac{k_{ex} p_A}{\lambda_+ - \lambda_-} [\exp(-\lambda_- t) - \exp(-\lambda_+ t)] \\
a_{BA}(t) &= \frac{k_{ex} p_B}{\lambda_+ - \lambda_-} [\exp(-\lambda_- t) - \exp(-\lambda_+ t)]
\end{aligned}$$

and

$$\lambda_{\pm} = \frac{1}{2} \left\{ \frac{1}{T_{1A}} + \frac{1}{T_{1B}} + k_{ex} \pm \left[\left(\frac{1}{T_{1A}} - \frac{1}{T_{1B}} + k_{ex}(p_B - p_A) \right)^2 + 4 p_A p_B k_{ex}^2 \right]^{1/2} \right\}$$

T_{1A}, T_{1B} are the longitudinal relaxation times for A, B. So for both A and B, the relaxation is no longer a single exponential decay. Actually this formula is valid not only in slow exchange but also in fast exchanges. This complicated relaxation formula can be reduced to equation (14) in the fast exchange limit. Assuming $k_{ex} \gg \omega_A - \omega_B$, only one signal is observed with the intensity $\langle M_A^Z(t) \rangle + \langle M_B^Z(t) \rangle$. If $k_{ex} \gg \frac{1}{T_{1A}}, \frac{1}{T_{1B}}$, it can

be shown that the exponential term with λ_+ disappears while the rate λ_- can be approximately by (Taylor expansion):

$$\lambda_- \approx \frac{P_A}{T_{1A}} + \frac{P_B}{T_{1B}} - P_A P_B \frac{\left(\frac{1}{T_{1A}} - \frac{1}{T_{1B}}\right)^2}{k_{ex}} \approx \frac{P_A}{T_{1A}} + \frac{P_B}{T_{1B}} \quad (17)$$

A typical protein molecule has T_1 in seconds for amide nitrogen, for example yCD (35 KDa) has $T_1 \sim 1.3$ s. If k_{ex} is about 100 s^{-1} or more, it is safe to omit the quadratic form in λ_- , which will give the same result as shown in (14).

Modified Bloch equations can also be used for transverse relaxation rate calculation:

$$\frac{d}{dt} \begin{bmatrix} \langle M_A^+(t) \rangle \\ \langle M_B^+(t) \rangle \end{bmatrix} = - \begin{bmatrix} i\omega_A + \frac{1}{T_{2A}} + p_B k_{ex} & -p_A k_{ex} \\ -p_B k_{ex} & i\omega_B + \frac{1}{T_{2B}} + p_A k_{ex} \end{bmatrix} \begin{bmatrix} \langle M_A^+(t) \rangle \\ \langle M_B^+(t) \rangle \end{bmatrix} \quad (18)$$

The solution of these equations is:

$$\begin{bmatrix} \langle M_A^+(t) \rangle \\ \langle M_B^+(t) \rangle \end{bmatrix} = \begin{bmatrix} a_{AA}(t) & a_{AB}(t) \\ a_{BA}(t) & a_{BB}(t) \end{bmatrix} \begin{bmatrix} \langle M_A^+(0) \rangle \\ \langle M_B^+(0) \rangle \end{bmatrix} \quad (19)$$

in which

$$a_{AA}(t) = \frac{1}{2} \left(1 - \frac{-i\Delta\omega + \frac{1}{T_{2A}} - \frac{1}{T_{2B}} + k_{ex}(p_B - p_A)}{\lambda_+ - \lambda_-} \right) \exp(-\lambda_- t) \\ + \frac{1}{2} \left(1 + \frac{-i\Delta\omega + \frac{1}{T_{2A}} - \frac{1}{T_{2B}} + k_{ex}(p_B - p_A)}{\lambda_+ - \lambda_-} \right) \exp(-\lambda_+ t)$$

$$a_{BB}(t) = \frac{1}{2} \left(1 + \frac{-i\Delta\omega + \frac{1}{T_{2A}} - \frac{1}{T_{2B}} + k_{ex}(p_B - p_A)}{\lambda_+ - \lambda_-} \right) \exp(-\lambda_- t) \\ + \frac{1}{2} \left(1 - \frac{-i\Delta\omega + \frac{1}{T_{2A}} - \frac{1}{T_{2B}} + k_{ex}(p_B - p_A)}{\lambda_+ - \lambda_-} \right) \exp(-\lambda_+ t)$$

$$a_{AB}(t) = \frac{k_{ex} p_A}{\lambda_+ - \lambda_-} [\exp(-\lambda_- t) - \exp(-\lambda_+ t)]$$

$$a_{BA}(t) = \frac{k_{ex} p_B}{\lambda_+ - \lambda_-} [\exp(-\lambda_- t) - \exp(-\lambda_+ t)]$$

and

$$\lambda_{\pm} = \frac{1}{2} \left\{ -i\omega_A - i\omega_B + \frac{1}{T_{2A}} + \frac{1}{T_{2B}} + k_{ex} \pm \left[\left(-i\Delta\omega + \frac{1}{T_{2A}} - \frac{1}{T_{2B}} + k_{ex}(p_B - p_A) \right)^2 + 4p_A p_B k_{ex}^2 \right]^{1/2} \right\}$$

$$\Delta\omega = |\omega_A - \omega_B|$$

Just as the longitudinal relaxation, it appears that the transverse relaxation includes two exponential decays. But it can be reduced to a single exponential decay under the fast exchange limit, with the same procedure used to simplify λ_{\pm} as in longitudinal

relaxation. With the conditions $k_{ex} \gg \frac{1}{T_{2A}}, \frac{1}{T_{2B}}$ and $k_{ex} \gg \omega_A - \omega_B$, the coefficient

of λ_+ approaches 0 and λ_- can be approximated by:

$$\begin{aligned}
\lambda_- &\approx -i(P_A\omega_A + P_B\omega_B) + \frac{P_A}{T_{2A}} + \frac{P_B}{T_{2B}} \\
&+ P_AP_B \frac{(\omega_A - \omega_B)^2 - \left(\frac{1}{T_{2A}} - \frac{1}{T_{2B}}\right)^2 - 2i(\omega_A - \omega_B)\left(\frac{1}{T_{2A}} - \frac{1}{T_{2B}}\right)}{k_{ex}} \quad (20) \\
&\approx -i(P_A\omega_A + P_B\omega_B) + \frac{P_A}{T_{2A}} + \frac{P_B}{T_{2B}} + P_AP_B \frac{(\omega_A - \omega_B)^2}{k_{ex}}
\end{aligned}$$

where $-i(P_A\omega_A + P_B\omega_B)$ is the Larmor precession frequency which is zero when measuring the relaxation on resonance, or can be removed by using Carr-Purcell-Meiboom-Gill (CPMG) pulses (Luz and Meiboom 1963). Details will be discussed in the next section. One condition $|\omega_A - \omega_B| > \left|\frac{1}{T_{2A}} - \frac{1}{T_{2B}}\right|$ was used for the second step approximation of (20). This assumption can be justified as following: as shown in (9) and (12), the largest term contributing to $\frac{1}{T_2}$ is $j(0)$ which is only proportional to τ_m the global tumbling time (assuming the internal motion has a shorter timescale.), suggesting $\frac{1}{T_2}$ is not sensitive to the local environmental changes. In yCD $\frac{1}{T_2}$ of the amide nitrogen is about 25 s^{-1} , the $\frac{1}{T_2}$ difference between different conformations should be smaller than this value. On the other hand, for a exchange rate k_{ex} about 1000 s^{-1} , assuming conformation A and B have same populations and T_2 , $|\omega_A - \omega_B|$ has to be larger than 71 rad/s to contribute 5% of $\frac{1}{T_2}$ which is the typical detection limit in experiments (experimental uncertainty). So the conformational changes of the spin

system with $|\omega_A - \omega_B| < \left| \frac{1}{T_{2A}} - \frac{1}{T_{2B}} \right|$ can hardly be detected. Based on the conditions used to derive (20), we can refine the valid range of (20) used for conformational exchange study in a protein such as yCD. The lower limit of k_{ex} is the transverse relaxation rate $\sim 25 \text{ s}^{-1}$ and the upper limit is the global tumbling rate $\sim 55 \times 10^6 \text{ s}^{-1}$. So the valid k_{ex} should be in between $\sim 250 \text{ s}^{-1}$ and $\sim 5.5 \times 10^6 \text{ s}^{-1}$, primarily on a millisecond to microsecond timescale. Many biological processes occur with time constants within this range. If we are only interested in the motion on this timescale, CPMG and spin locked techniques $T_{1\rho}$ can be used to extract k_{ex} based on equations derived similar to (20).

c. Transverse relaxation during CPMG sequences

A single spin-echo period, $\tau/2 - 180^\circ - \tau/2$ can be described as a 180° pulse (in x or y direction) in between two $\tau/2$ delays. The 180° pulse changes M^+ to M^- (and M^- to M^+), while during $\tau/2$ delays, M^+ (A and B) relaxes with transition matrix C consisted of the element $a_{ij}(\tau)$ from equation (19) and M^- relaxes with C^* (complex conjugate of C). The relaxation in 180° pulse period is neglected because the pulse length is usually much shorter than $\tau/2$. After $2n$ spin echo periods, $M^+(2n\tau)$ is given by:

$$M^+(2n\tau) = (CC^*CC^*)^n M^+(0) \quad (21)$$

The general solution of equation (21), a combination of two exponentials, is rather complicated (Carver and Richards 1972; Jen 1978; Davis et al. 1994). If $P_A \gg P_B$, a one-exponential decay can be approximated for the solution with the relaxation rate constant of A is (Ishima and Torchia 1999):

$$\frac{1}{T_2}(\tau) = \frac{1}{T_2}(\tau \rightarrow 0) + \left(p_A p_B \Delta\omega^2 k_{ex} \right) / \left[k_{ex}^2 + \left(p_A^2 \Delta\omega^4 + 144 / \tau^4 \right)^{1/2} \right]$$

In the fast exchange limit, the solution is approximated by single exponential with the decay rate (Luz and Meiboom 1963):

$$\frac{1}{T_2}(\tau) = \frac{1}{T_2}(\tau \rightarrow 0) + \left(p_A p_B \Delta\omega^2 / k_{ex} \right) \left[1 - 2 \tanh(k_{ex} \tau / 2) / k_{ex} \tau \right]$$

where $\frac{1}{T_2}(\tau \rightarrow 0) = p_A / T_{2A} + p_B / T_{2B}$. In practice, two τ values can be selected

τ_1, τ_2 to measure transverse relaxation, then

$$\frac{1}{T_2}(\tau_1) - \frac{1}{T_2}(\tau_2) = 2 \left(p_A p_B \Delta\omega^2 / k_{ex} \right) \left\{ \left[\tanh(k_{ex} \tau_1 / 2) / k_{ex} \tau_1 \right] - \left[\tanh(k_{ex} \tau_2 / 2) / k_{ex} \tau_2 \right] \right\}$$

By comparing the rates at these two delays, fast exchanges can be identified. In most cases of practical interest, CPMG experiments in proteins will be applicable to chemical exchange processes with values of $k_{ex} < 1 \times 10^4 \text{ s}^{-1}$ as a result of experimental constraints on the minimum value of τ (Palmer et al. 2001). Sample heating at high pulsing rates is a major limitation. In addition, the delay time in between two 180° pulses has to be much longer than the pulses length; otherwise the assumption in equation (21) may not be valid. (e.g. The 180° pulse length of ^{15}N is $\sim 80 \mu\text{s}$ in Chemistry department Varian 600 MHz machine, the safe τ minimum should be $\sim 0.5 \text{ ms}$). The up-limit of τ should be less than half of T_2 ($\sim 40 \text{ ms}$ for yCD) in order to give reasonable sensitivity and enough data points for exponential fitting before the signal decays to zero. The function $\tanh(x)/x$ changes dramatically when x is small, for example, x changes from 0.5 to 5 corresponding $\tanh(x)/x$ changes from 0.92 to 0.2. Therefore, practically CPMG experiments are sensitive to $k_{ex} \sim 10^3 \text{ s}^{-1}$ provided that $\Delta\omega$ is reasonably big.

In a summary, protein dynamics in ps-ns and μ s-ms characterized by heteronuclear spin relaxation NMR spectroscopy is discussed above. The relaxation rate was derived by using time dependant perturbation method. For backbone NH spin system, this perturbation mainly comes from NH dipolar-dipolar interaction and nitrogen chemical shift anisotropy effect. The motion of protein that causes the perturbation time dependant is separated to global tumbling motion and local fluctuations of individual spin sites. Model-free method was described, with the assumption of spin vector local time correlation function as exponential forms. The order parameter was introduced which describes the flexibility of individual spin vector.

Unlike the ps-ns protein dynamics, motion on the μ s-ms scale is primarily caused by protein conformational exchanges. The relaxation rate for this timescale motion can be derived from the modified Bloch equation. The μ s-ms motion only changes T_2 but not T_1 for the fast exchanges. The reason can be simply illustrated as following. The exchanges switch the Hamiltonian between two conformations, which consists of the major term time independent H_0 and minor term time dependent $H(t)$ in equation (1). Because H_0 commutes with I_z but not I^+ or I^- , the major term contributes to T_2 but not T_1 relaxation. Therefore the influence of exchange over T_1 relaxation is rather small.

Section 2. Molecular Dynamics Simulation

MD simulation can provide a great detail concerning individual particle motions as a function of time (Karplus and McCammon 2002). Thus, it can be used to address specific questions about the properties of a model system, often more easily than

experiments on the actual system. For many aspects of biomolecular function, it is these details that are of interest. Of course, experiments play an essential role in validating the simulation methodology. Another advantage of simulations is that users have complete control of the potential function. So, the users can transmute the potential from that representing one system to another during a simulation in the calculation of free energy differences.

MD simulates matter at the atomic level. Given an intermolecular potential, and an initial configuration for the molecules, it will provide molecular configurations for all time by solving Newton's equations of motion (Lanczos 1966).

$$m_i a_i = f_i \quad a_i = \ddot{r}_i \text{ (dot = time derivative)}$$

$$f_i = -\nabla_i V \quad \nabla_i \text{ means gradient with respect to } r_i.$$

V is the intermolecular potential. We will assume that it is pairwise additive.

$$V = \frac{1}{2} \sum_{\substack{i,j \\ i \neq j}} V_{ij}$$

The above second order in time equations may be written as

$$\dot{r}_i = p_i / m \quad 6N \text{ 1st order in time differential equations}$$

$$\dot{p}_i = f_i$$

A lot of force-field development deals with the intramolecular part of the potential. To develop it, one fits quantum chemistry data on: vibrational frequencies, dipole moments, derivatives of the energy and etc. Typical intramolecular force fields will have terms as (Pearlman et al. 1995):

$$E = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_r (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} (1 + \cos[n\phi - \gamma])$$

$$+ \sum_{i>j}^{atom} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{i>j}^{atom} \frac{q_i q_j}{\epsilon r_{ij}}$$

Nature deals with ~ Avogadro's number of particles. On a computer, ~ 10,000 - 50,000 is the current practical limit for conventional computers. For parallel machines with special codes, this is changing toward very large numbers of molecules (1,000,000). For many purposes, this is not necessary. Many simulations were run with 256 molecules, if you want properties of a small molecule solvent. If you want to do a small protein, say cytochrome c – with about 100 residues, and properly solvated for, need about 6,000 H₂O molecules. The surface of the box becomes very important; e.g., for 1000 = 10³ molecules in a cubic box, 488 molecules appear on the cube faces! This drove the introduction of *Periodic Boundary Conditions* (PBC, Figure 1-1). You just need to store the particle coordinates in box P (not the other boxes). When a particle moves out of box P, it is replaced in the list of coordinates by the one that moves into the box. Simulating a periodically replicated system is better than a box in vacuum but introduces problems with long range interactions. Modern MD codes such as AMBER (Pearlman et al. 1995) use special Ewald-based methods (Essmann et al. 1995) to deal with these issues.

For one atom, the *j*th, the vector position at time *t* is *r_j(t)*. Over a time interval *T*, RMS average deviation is defined

$$\langle r_j(t) \rangle^T = \frac{1}{T} \int_0^T r_j(t) dt$$

Consider the fluctuations in position with respect to this time evolution. The natural definition is to construct the matrix

$$\begin{bmatrix} \langle dx_j(t) dx_j(t) \rangle^T & \langle dx_j(t) dy_j(t) \rangle^T & \langle dx_j(t) dz_j(t) \rangle^T \\ \langle dy_j(t) dx_j(t) \rangle^T & \langle dy_j(t) dy_j(t) \rangle^T & \langle dy_j(t) dz_j(t) \rangle^T \\ \langle dz_j(t) dx_j(t) \rangle^T & \langle dz_j(t) dy_j(t) \rangle^T & \langle dz_j(t) dz_j(t) \rangle^T \end{bmatrix}$$

where $dx_j(t) = x_j(t) - \langle x_j(t) \rangle^T$ etc.

So that $\langle dx_j(t) dx_j(t) \rangle^T = \langle x_j^2(t) \rangle^T - \langle x_j(t) \rangle^T \langle x_j(t) \rangle^T$

This matrix is a lot of information to store/analyze, so it is more usual to take its trace (sum the diagonal elements) to obtain

$$rms_j = \langle dr_j(t) \bullet dr_j(t) \rangle^T = \langle r_j(t) \bullet r_j(t) \rangle^T - \langle r_j(t) \rangle^T \bullet \langle r_j(t) \rangle^T$$

This gives the time-averaged fluctuation of j th atom's position relative to its time averaged position.

If we want to do the same on e.g. a residue basis, then define

$$rms_{resk} = \sqrt{\frac{\sum_{j \in resk} \langle dr_j(t) \bullet dr_j(t) \rangle^T}{\sum_{j \in resk} 1}}$$

Similarly the deviation from X-ray structure is defined as:

$$dev = xray_{resk} = \sqrt{\frac{\sum_{j \in resk} (\langle r_j(t) \rangle^T - r_j(0)) \cdot (\langle r_j(t) \rangle^T - r_j(0))}{\sum_{j \in resk} 1}}$$

where $r_j(0) = r_j(X - ray)$.

Adding up all the deviations from X-ray time averaged over some time interval and plotting as a function of time from initial time gives some indication of how long it takes to equilibrate the protein/solvent system from the protein point of view.

Sometimes we want a thermostated system, e.g. NVT canonical system. To fix the temperature, we may scale the velocities. The average kinetic energy is

$$\langle E_k \rangle = \frac{3}{2} N k_B T$$

Equipartition says

$$\left\langle \frac{1}{2} m v_{ix}^2 \right\rangle = k_B T / 2$$

where $\langle \rangle$ means average over the Maxwellian $P(v_{ix})$ for each particle i . So, evaluate the “instantaneous temperature”

$$\frac{\sum_i \frac{1}{2} m_i v_i^2}{3/2 N} = k_B T$$

If T is not T_{ref} , the desired temperature, multiply all the v_i 's by the same factor to make $T = T_{ref}$. If this is done every MD step, then you will maintain the target temperature, T_{ref} . More sophisticated methods of temperature control can be used (Berendsen et al. 1984).

Thermodynamic Perturbation Methods.

The connection between the Helmholtz free energy $A(N, V, T)$ (one can do the same for the Gibbs free energy $G(N, P, T)$) and the potential energy function $U(R^N)$ used in e.g. MD is

$$A(N, V, T) = -k_B T / (N! \Lambda^{3N}) \ln \int dR^N e^{-U(R^N) / k_B T}$$

N is number of particles in the system, V is volume, T is temperature, R^N denotes the coordinates of each atom, k_B is the Boltzman constant, Λ is $(2\pi m k_B T)^{-1/2}$. Consider another system with potential energy $U^*(R^N)$. Then

$$\begin{aligned} A^*(N, V, T) - A(N, V, T) &= -k_B T \ln \left[\frac{\int dR^N e^{-(U^*(R^N) - U(R^N)) / k_B T} e^{-U(R^N) / k_B T}}{\int dR^N e^{-U(R^N) / k_B T}} \right] \\ &= -k_B T \ln \left\langle e^{-(U^*(R^N) - U(R^N)) / k_B T} \right\rangle_U \end{aligned}$$

The last line defines an ensemble average, done by e.g. doing a long MD trajectory, where the configurations are generated by doing the MD with the $U(R^N)$ potential.

If we want to find out binding free energy for a ligand L with a protein P , we can define $A(N, V, T)$ as the free energy for the binding state, and $A^*(N, V, T)$ as the free energy for the dissociated state. Obviously, the free energy difference is the binding free energy between ligand L and protein P . But in MD, we can't just simulate these two states, since MD can only give us a small conformational space in real time simulation for a protein, which means the average we get from above equation may not be the true binding free energy. Instead if we define some intermediate state,

$$U_{\lambda}(R^N) = \begin{cases} U(R^N) & \lambda = 0 \\ f(U(R^N), U^*(R^N), \lambda) & \lambda \in (0,1) \\ U^*(R^N) & \lambda = 1 \end{cases}$$

Then

$$A^*(N, V, T) - A(N, V, T) = \int_0^1 \left\langle \frac{\partial f}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

This equation could be integrated to give us binding free energy, as long as the path we choose is reversible.

Section 3. ONIOM: A Multilayered Integrated Quantum Mechanics

Method

The standard high level *ab initio* quantum mechanics (QM) method is well known to be computationally expensive, which make it extremely difficult to study enzymatic reaction. Here we introduce a ONIOM method developed by Keiji Morokuma and his colleagues (Svensson et al. 1996; Dapprich et al. 1999; Vreven and Morokuma 2000; Vreven et al. 2001; Torrent et al. 2002; Morokuma 2003; Vreven et al. 2003), a onion like multi-level method, combining different leverls of quantum chemical methods as well as molecular mechanics method. The concept of ONIOM method is rather simple. We will illustrate it by using two-layer ONIOM, while extending to more layers is straightfoward. Assuming a protein with well defined active site (for example from X-ray structure), in order to study the reaction mechanism we have to treat the whole system with resonably good QM method (such as DFT, Density Functional Theory), which is impossible with the current computer resources. Instead we can use ONIOM concept to separate the system (named the real system) into two parts: 1. the active site which we are

primarily interested in (named the model system), treated with high level QM method (DFT etc.). 2. the rest of protein, which influences the reaction mainly through long range electrostatics, treated with low level QM or MM method (AM1 or force field etc.). Three calculations are carried out: the high level calculation of the model system, the low level calculation of the real system and the low level calculation of the model system. The total energy of the system will be:

$$E(\text{ONIOM,real}) = E(\text{high,model}) + E(\text{low,real}) - E(\text{low,model})$$

This treatment inevitably introduces an error if considering the high level calculation of the whole system as the “correct” method. But if the error is a constant for two different structures, their relative energy difference will be evaluated correctly by ONIOM method. In this case, if the change of ligand (e.g. from reactant to intermediate state) doesn’t cause a drastic change beyond the active site, the ONIOM method should give a reasonably accurate result. The two most important questions when using the ONIOM are how to choose different layers and how to select the methods to describe these layers (Morokuma 2003). The answer to the first question is to determine which parts of the system are more important and which parts are less important. In enzyme-catalyzed reaction, the reactant as well as residues that are directly involved in the bond cleavage and formation are most important and the residues away from the reactant might be less important. The former can be selected as the high level system, while the latter can be selected as the low level system. The high level QM method used in ONIOM depends on the system and problems one wants to study, and practically also depends on the size of the system. Once the high level is selected, two structures along the reaction can be used

to calculate the energy difference by treating the real system in high level. Then, the low level methods will be judged by the reproducibility of energy difference.

The implementation of ONIOM is straightforward if there is no covalent bond in between low level layer and the high level layer. The force used to drive the structure to the minimum is just (Dapprich et al. 1999):

$$\nabla E(\text{ONIOM, real}) = \nabla E(\text{high, model}) + \nabla E(\text{low, real}) - \nabla E(\text{low, model})$$

And the Hessian matrix is the second derivative of ONIOM energy, which provides the information about normal vibrational frequencies:

$$\nabla^2 E(\text{ONIOM, real}) = \nabla^2 E(\text{high, model}) + \nabla^2 E(\text{low, real}) - \nabla^2 E(\text{low, model})$$

If there are covalent bonds between low level layer and high level layer, link atoms are used to saturate the model layer. But the link atoms are constrained by the atom pairs that form covalent bonds. In this case, Jacobian matrix J is introduced to project the forces on the link atoms to the corresponding atom pairs (Dapprich et al. 1999):

$$\nabla E(\text{ONIOM, real}) = \nabla E(\text{high, model}) + \nabla E(\text{low, real}) \times J - \nabla E(\text{low, model}) \times J$$

And the second derivative is:

$$\nabla^2 E(\text{ONIOM, real}) = \nabla^2 E(\text{high, model}) + J^T \times \nabla E(\text{low, real}) \times J - J^T \times \nabla E(\text{low, model}) \times J$$

where J^T is the transpose of J .

Part II Studied Proteins

Section 1. Yeast Cytosine Deaminase

Yeast cytosine deaminase (yCD), a zinc metalloenzyme, catalyzes the hydrolytic deamination of cytosine to uracil. yCD is of great biomedical interest, because it also catalyzes the deamination of the prodrug 5-fluorocytosine (5FC), which is one of the most widely used prodrugs for gene-directed enzyme prodrug therapy (GDEPT) for the treatment of cancer (Greco and Dachs 2001). The challenge in cancer therapy is to kill tumor cells without damaging normal cells. GDEPT meets the challenge by activating a prodrug in the tumor, thereby minimizing damage to normal tissues (Aghi et al. 1998; Greco and Dachs 2001).

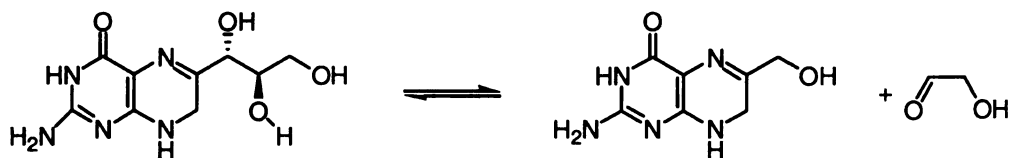
The structure of yCD has been recently determined at high resolution with (Ireton et al. 2003; Ko et al. 2003) and without (Ireton et al. 2003) the inhibitor 2-pyrimidinone, a transition state analogue. yCD is a homodimeric enzyme, and the structure of each monomer consists of a central β -sheet flanked by two α -helices on one side and four α -helices on the other. The homodimeric protein contains two active centers, and each active center is composed of residues within a single subunit and a catalytic zinc ion coordinated with a histidine (His62), two cysteines (Cys91 and Cys94), and a water molecule in the substrate-free enzyme. The Zn-bound water serves as a nucleophile in the deamination reaction. The bound inhibitor is completely buried, being covered by a lid composed of Phe114 from the loop between β 4 and α D, and Trp152, and Ile156, both from the C-terminal helix.

The catalytic apparatus of yCD, including the catalytic zinc, its coordinated residues, and the proton shuttle, is very similar to that of *E. coli* cytidine deaminase (Ireton et al. 2003; Ko et al. 2003), which has been extensively studied (Schramm and Bagdassarian 1999; Snider et al. 2002; Snider et al. 2002) and is a paradigm for understanding enzymatic catalysis. Crystal structures have been determined for the complexes of the enzyme with 3-deazacytidine (Xiang et al. 1996), 3,4-dihydrozebularine (Xiang et al. 1995), 3,4-hydrated 2-pyrimidinone riboside (Xiang et al. 1995), and the product uridine (Xiang et al. 1997), all at 2.30 Å resolution except the structure of the 3,4-dihydrozebularine complex, which is at 2.20 Å resolution. These crystal structures represent different states of the catalytic cycle and have provided important insights into the catalytic mechanism of the enzyme.

On the basis of the similarity of yCD and CDA active site structures and early studies of the mechanism of CDA catalysis, an analogous reaction mechanism was proposed for yCD (Ireton et al. 2003). Our recent quantum chemical study (Sklenak et al. 2004) using the ONIOM (B3LYP:PM3) method has revealed a complete path for the deamination reaction catalyzed by yCD.

Section 2. Dihydroneopterin Aldolase

DHNA catalyzes the conversion of 7, 8-dihydroneopterin (DHNP) to 6-hydroxymethyl-7, 8-dihydropterin (HP) in the *de novo* biosynthesis of folic acid from guanosine triphosphate, which is present in microbes but not humans.



It is a good target for antibacterial chemotherapy just like dihydropteroate synthase and dihydrofolate reductase (Dale et al. 1997). *S. aureus* DHNA structures have been published for the apo and HP bound forms (Hennig et al. 1998). DHNA is an octamer in nature with 121 amino acids for each monomer (Hennig et al. 1998). The octamer is a hollow cylinder formed by a 'head to head' assembly of two such tetrameric rings. The active site sits in between two adjacent subunits of the tetramer. The bottom of the active site has a glutamate (Glu74) that serves as an "anchor" for binding and there are residues lining the active site that stabilize HP. A general acid and base are needed for the aldolase reaction (Figure 1-2).

Section 3. Guanine Deaminase

Guanine deaminase (GD), a Zn metalloenzyme, catalyzes the hydrolytic deamination of guanine into xanthine and therefore plays an important role in nucleotide metabolism. The crystal structure of GD from *Bacillus subtilis* (bGD) complexed with imidazole was solved by Liaw (Liaw et al. 2004) and his colleagues at a resolution of 1.17 Å. bGD forms a homodimer with 156 residues in each monomer. The overall structure of each monomer consists of a center five-stranded β -sheet sandwiched by six helices (Liaw et al. 2004). The homodimer GD contains two active sites, including one Zn atom in each active site. The Zn atom is coordinated with His53, Cys83, Cys86

and one water molecule. It is interesting that the active site is buried by a swapped C-terminal tail from an adjacent subunit, the first domain-swapped structure in the Cytidine Deaminase superfamily (Liaw et al. 2004). It was proposed that this tail not only seals the active site but also is used to recognize specific substrates (Liaw et al. 2004). The bGD catalyzed reaction is believed to proceed through a tetrahedral transition state while Glu55 acts as the proton shuttle. But the detailed mechanism is still unknown.

The thesis is arranged as follows: 1. Experimental studies of yCD are included in Chapters 2-4. 2. Computational studies of yCD are presented in Chapters 5-7. 3. A computational study of GD is given in Chapter 8. 4. A computational study of DHNA is included in Chapter 9. Chapter 2 describes kinetics studies of the activation of 5FC and an NMR study of the product release process. Chapter 3 describes the yCD dynamics study (on the ps-ns time scale) by using NMR relaxation measurements and MD simulations. Chapter 4 describes the yCD dynamics study (on the sec-min time scale) by using the NMR HD exchange method. Chapter 5 investigates by MD simulation the yCD catalyzed reaction mechanism. Chapter 6 describes the ONIOM and MD study of 5FU release from yCD. Chapter 7 studies the ligand release path from yCD by using the MD simulation method. Chapter 8 describes the GD catalytic mechanism by using the MD and ONIOM methods. Chapter 9 describes DHNA dynamical properties in the apo and product complex by the use of MD simulations.

Reference

Abraham, A. (1961). The principles of nuclear magnetism.

Aghi, M., C. M. Kramm, et al. (1998). "Synergistic anticancer effects of ganciclovir/thymidine kinase and 5-fluorocytosine/cytosine deaminase gene therapies." J Natl Cancer Inst 90(5): 370-80.

Akke, M. (2002). "NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis." Current Opinion in Structural Biology 12(5): 642-647.

Berendsen, H. J. C., J. P. M. Postma, et al. (1984). "Molecular-Dynamics with Coupling to an External Bath." Journal of Chemical Physics 81(8): 3684-3690.

Carver, J. P. and R. E. Richards (1972). "General 2-Site Solution for Chemical Exchange Produced Dependence of T2 Upon Carr-Purcell Pulse Separation." Journal of Magnetic Resonance 6(1): 89-&.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G. and Nicholas, J.S. (1996). Protein NMR Spectroscopy.

Chen, J. H., C. L. Brooks, et al. (2004). "Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation." Journal of Biomolecular Nmr 29(3): 243-257.

Clore, G. M., A. Szabo, et al. (1990). "Deviations from the Simple 2-Parameter Model-Free Approach to the Interpretation of N-15 Nuclear Magnetic-Relaxation of Proteins." Journal of the American Chemical Society 112(12): 4989-4991.

Dale, G. E., C. Broger, et al. (1997). "A single amino acid substitution in Staphylococcus aureus dihydrofolate reductase determines trimethoprim resistance." Journal of Molecular Biology 266(1): 23-30.

Dapprich, S., I. Komaromi, et al. (1999). "A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives." Journal of Molecular Structure-Theochem 462: 1-21.

d'Auvergne, E. J. and P. R. Gooley (2003). "The use of model selection in the model-free analysis of protein dynamics." Journal of Biomolecular Nmr 25(1): 25-39.

Davis, D. G., M. E. Perlman, et al. (1994). "Direct Measurements of the Dissociation-Rate Constant for Inhibitor-Enzyme Complexes Via the T-1-Rho and T-2 (Cpmg) Methods." Journal of Magnetic Resonance Series B 104(3): 266-275.

- Ding, Z. F., G. Lee, et al. (2005). "PhosphoThr peptide binding globally rigidifies much of the FHA domain from Arabidopsis receptor kinase-associated protein phosphatase." Biochemistry 44(30): 10119-10134.
- Essmann, U., L. Perera, et al. (1995). "A Smooth Particle Mesh Ewald Method." Journal of Chemical Physics 103(19): 8577-8593.
- Fischer, M. W. F., A. Majumdar, et al. (1998). "Protein NMR relaxation: theory, applications and outlook." Progress in Nuclear Magnetic Resonance Spectroscopy 33(4): 207-272.
- Greco, O. and G. U. Dachs (2001). "Gene directed enzyme/prodrug therapy of cancer: Historical appraisal and future perspectives." J. Cell. Physiol. 187(1): 22-36.
- Hennig, M., A. D'Arcy, et al. (1998). "Crystal structure and reaction mechanism of 7,8-dihydroneopterin aldolase from Staphylococcus aureus." Nature Structural Biology 5(5): 357-362.
- Ireton, G. C., M. E. Black, et al. (2003). "The 1.14 Å crystal structure of yeast cytosine deaminase evolution of nucleotide salvage enzymes and implications for genetic chemotherapy." Structure 11: 961-972.
- Ishima, R. and D. A. Torchia (1999). "Estimating the time scale of chemical exchange of proteins from measurements of transverse relaxation rates in solution." Journal of Biomolecular Nmr 14(4): 369-372.
- Ishima, R. and D. A. Torchia (2000). "Protein dynamics from NMR." Nature Structural Biology 7(9): 740-743.
- Jen, J. (1978). "Chemical Exchange and Nmr T2 Relaxation - Multisite Case." Journal of Magnetic Resonance 30(1): 111-128.
- Karplus, M. and J. A. McCammon (2002). "Molecular dynamics simulations of biomolecules." Nature Structural Biology 9(9): 646-652.
- Kay, L. E. (1998). "Protein dynamics from NMR." Nature Structural Biology 5: 513-517.
- Ko, T.-P., J.-J. Lin, et al. (2003). "Crystal structure of yeast cytosine deaminase. Insights into enzyme mechanism and evolution." J. Biol. Chem. 278: 19111-19117.
- Lanczos, M. (1966). "The variational principle of mechanics."
- Liaw, S. H., Y. J. Chang, et al. (2004). "Crystal structure of Bacillus subtilis guanine deaminase - The first domain-swapped structure in the cytidine deaminase superfamily." Journal of Biological Chemistry 279(34): 35479-35485.

Lipari, G. and A. Szabo (1982). "Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity." Journal of the American Chemical Society 104(17): 4546-4559.

Lipari, G. and A. Szabo (1982). "Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results." Journal of the American Chemical Society 104(17): 4559-4570.

Luz, Z. and S. Meiboom (1963). "Nuclear Magnetic Resonance Study of Protolysis of Trimethylammonium Ion in Aqueous Solution - Order of Reaction with Respect to Solvent." Journal of Chemical Physics 39(2): 366-&.

Mandel, A. M., M. Akke, et al. (1995). "Backbone Dynamics of Escherichia-Coli Ribonuclease Hi - Correlations with Structure and Function in an Active Enzyme." Journal of Molecular Biology 246(1): 144-163.

McConnell, H. M. (1958). "Reaction Rates by Nuclear Magnetic Resonance." Journal of Chemical Physics 28(3): 430-431.

Morokuma, K. (2003). "ONIOM and its applications to material chemistry and catalyses." Bulletin of the Korean Chemical Society 24(6): 797-801.

Palmer, A. G. (2001). "NMR probes of molecular dynamics: Overview and comparison with other techniques." Annual Review of Biophysics and Biomolecular Structure 30: 129-155.

Palmer, A. G., C. D. Kroenke, et al. (2001). Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. Nuclear Magnetic Resonance of Biological Macromolecules, Pt B. 339: 204-238.

Pearlman, D. A., D. A. Case, et al. (1995). "Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules." Computer Physics Communications 91(1-3): 1-41.

Peng, J. W. and G. Wagner (1992). "Mapping of Spectral Density-Functions Using Heteronuclear Nmr Relaxation Measurements." Journal of Magnetic Resonance 98(2): 308-332.

Schramm, V. L. and C. K. Bagdassarian (1999). Deamination of nucleosides and nucleotides and related reactions. Enzymes, Enzyme Mechanisms, Proteins, and Aspects of NO Chemistry. C. D. Poulter. New York, Amsterdam. 5: 71-100.

Sklenak, S., L. S. Yao, et al. (2004). "Catalytic mechanism of yeast cytosine deaminase: An ONIOM computational study." Journal of the American Chemical Society 126(45): 14879-14889.

- Snider, M. J., D. Lazarevic, et al. (2002). "Catalysis by entropic effects: The action of cytidine deaminase on 5,6-dihydrocytidine." Biochemistry 41(12): 3925-3930.
- Snider, M. J., L. Reinhardt, et al. (2002). "N-15 kinetic isotope effects on uncatalyzed and enzymatic deamination of cytosine." Biochemistry 41(1): 415-421.
- Stone, M. J. (2001). "NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding." Accounts of Chemical Research 34(5): 379-388.
- Svensson, M., S. Humbel, et al. (1996). "ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition." Journal of Physical Chemistry 100(50): 19357-19363.
- Torrent, M., T. Vreven, et al. (2002). "Effects of the protein environment on the structure and energetics of active sites of metalloenzymes. ONIOM study of methane monooxygenase and ribonucleotide reductase." Journal of the American Chemical Society 124(2): 192-193.
- Vreven, T., B. Mennucci, et al. (2001). "The ONIOM-PCM method: Combining the hybrid molecular orbital method and the polarizable continuum model for solvation. Application to the geometry and properties of a merocyanine in solution." Journal of Chemical Physics 115(1): 62-72.
- Vreven, T. and K. Morokuma (2000). "On the application of the IMOMO (integrated molecular orbital plus molecular orbital) method." Journal of Computational Chemistry 21(16): 1419-1432.
- Vreven, T., K. Morokuma, et al. (2003). "Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints." Journal of Computational Chemistry 24(6): 760-769.
- Wennerst.H (1972). "Nuclear Magnetic-Relaxation Induced by Chemical Exchange." Molecular Physics 24(1): 69-&.
- Xiang, S., S. A. Short, et al. (1995). "Transition-state selectivity for a single hydroxyl group during catalysis by cytidine deaminase." Biochemistry 34(14): 4516-23.
- Xiang, S., S. A. Short, et al. (1996). "Cytidine deaminase complexed to 3-deazacytidine: a "valence buffer" in zinc enzyme catalysis." Biochemistry 35(5): 1335-41.
- Xiang, S., S. A. Short, et al. (1997). "The structure of the cytidine deaminase-product complex provides evidence for efficient proton transfer and ground-state destabilization." Biochemistry 36(16): 4768-74.

Appendices

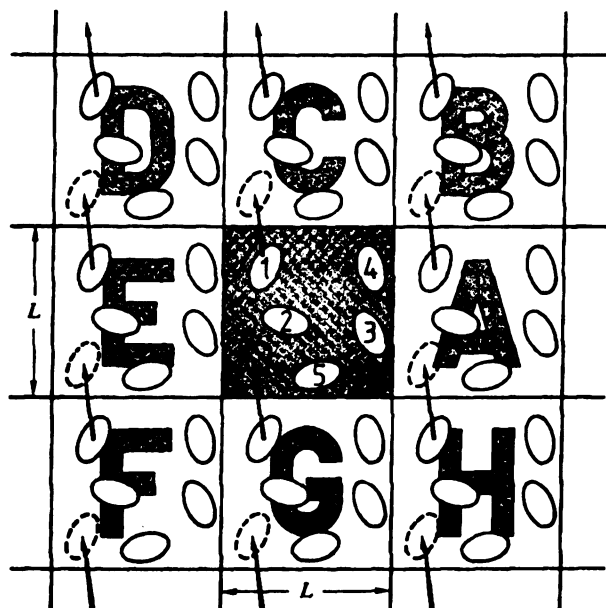


Figure 1-1. Periodic boundary condition box. The box in the center is box P (primary)

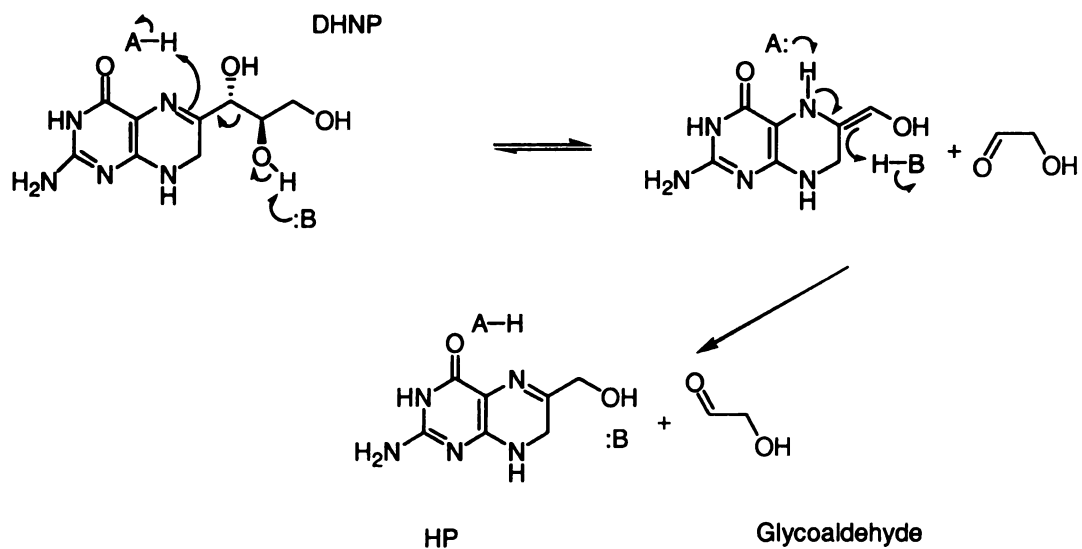


Figure 1-2 Aldolase reaction catalyzed by DHNA. A represents acid, B represents base.

Chapter 2

Product Release Is Rate-Limiting in the Activation of the Prodrug 5-Fluorocytosine by Yeast Cytosine Deaminase

Yeast cytosine deaminase (yCD), a zinc metalloenzyme, catalyzes the hydrolytic deamination of cytosine to uracil (Scheme 1A). yCD is of great biomedical interest because it also catalyzes the deamination of the prodrug 5-fluorocytosine (5FC, Scheme 1B), which is one of the most widely used prodrugs for gene-directed enzyme prodrug therapy (GDEPT) for the treatment of cancer (1, 2). The challenge in cancer therapy is to kill tumor cells without damaging normal cells. GDEPT meets the challenge by activating a prodrug in the tumor, thereby minimizing damage to normal tissues (1, 2). In cytosine deaminase-based GDEPT, the prodrug 5FC is converted to 5-fluorouracil (5FU) by the enzyme. 5FU is an anticancer drug used to treat breast, colon, rectal, stomach and pancreatic cancers, and is the drug of choice for treating colorectal carcinoma. However, the drug has high gastrointestinal and hematological toxicities. In contrast, the prodrug 5FC is fairly nontoxic to human, because of the lack of the CD activity in human cells. By producing 5FU in the tumor, the CD/5FC system minimizes the undesired toxic effects of 5FU.

The structure of yCD has been recently determined at high resolution both in the apo form (3) and in complex with the inhibitor 2-pyrimidinone, a reaction intermediate analog (3, 4). The enzyme is a homodimeric enzyme, and the structure of each monomer

consists of a central β -sheet flanked by two α -helices on one side and three α -helices on the other. The homodimeric enzyme contains two active centers, and each active center is composed of residues within a single subunit and a catalytic zinc ion coordinated with a histidine (His62), two cysteines (Cys91 and Cys94), and a water molecule in the substrate-free enzyme, the latter serving as a nucleophile in the deamination reaction. The inhibitor is bound in a hydrated adduct (4-[R]-hydroxyl-3,4-dihydropyrimidine), and the 4-hydroxyl group is coordinated with the catalytic zinc. The bound inhibitor is completely buried, being covered by a lid composed of Phe114 from the loop between β 4 and α D, and Trp152, and Ile156, both from the C-terminal helix. It has been suggested that the C-terminal helix may serve as a “gate” controlling the access to the active center and therefore both substrate binding and product release (4). Surprisingly, the structure of apo yCD is essentially the same as that of the inhibitor complex with an RMSD of 0.23 Å for the backbone atoms between the two structures. The active center in the apo enzyme is also covered by the same cluster of hydrophobic residues and appears inaccessible to the substrate (3). The crystal structure of the apo enzyme also contains a second zinc ion in the substrate binding pocket. However, the non-catalytic zinc is coordinated with water molecules only and has no direct interactions with the protein. The effects of the non-catalytic zinc on the protein structure are not clear. The yCD-catalyzed reaction is believed to proceed via a tetrahedral intermediate with a conserved glutamate (Glu64) serving as a proton shuttle (3-5).

The catalytic apparatus of yCD, including the catalytic zinc, its coordinated residues, and the proton shuttle, is very similar to that of *E. coli* cytidine deaminase (3, 4), which has been extensively studied (6-8) and is a paradigm for understanding

enzymatic catalysis. yCD has emerged as another excellent model system for studying this class of enzymatic reactions and the role of conformational dynamics in enzymatic catalysis, because high resolution structures have been determined (1.14 Å and 1.60 Å for the two structures of the reaction intermediate analog complex (3, 4) and 1.43 Å for the structure without the reaction intermediate analog (3)). Furthermore, yCD is only about half the size of *E. coli* cytidine deaminase and is amenable to high resolution NMR analysis, which is being carried out currently in our laboratory.

We are interested in elucidating how yCD catalyzes the activation of the prodrug 5FC to the anticancer drug 5FU and the role of conformational dynamics in the catalysis. In this paper, we show by a combination of transient kinetic and NMR studies that product release is rate limiting in the activation of the prodrug 5FC by yCD and may involve multiple steps.

Experimental procedures

Materials. All chemicals and biochemicals were from commercial sources. Cytosine, 5-fluorocytosine, and 5-fluorouracil were purchased from Sigma. [6-³H]-5-fluorocytosine was purchased from Moravsek. [¹⁵N]-NH₄Cl was purchased from Isotec. Restriction enzymes and T4 ligase were purchased from New England Biolabs. *Pfu* DNA polymerase and the pET-17b vector were purchased from Stratagene and Novagen, respectively.

Cloning. The yCD gene was amplified by PCR from yeast genomic DNA. The primers for PCR were: 5'-GGG ATC CAT ATG GCA AGC AAG TGG GAT CAG-3' (forward primer with a *Nde* I site) and 5'-GGA ATT CTA CTC ACC AAT ATC TTC AAA CC-3' (reverse primer with an *EcoR* I site). The PCR product was digested with

Nde I and *Eco*R I restriction enzymes and ligated with the vector pET-17b digested with the same restriction enzymes. The ligation mixture was transformed into the *E. coli* strain DH5 α . The correct coding sequence of the cloned yCD gene was verified by DNA sequencing. The DNA of the expression construct (pET17b-yCD) was then transformed into the *E. coli* strain BL21(DE3)pLysS.

Site-Directed Mutagenesis. The mutant W10H was made by PCR-based site directed mutagenesis. The forward and reverse primers for the mutagenesis were 5'-CATATGGCAAGC AAGCACGATCAGAAGGGTATGGACATTGCC-3' and 5'-GGCAATGTCCATACCCTT CTGATCGTGCTTGCTTGCCATATG-3. The mutation was verified by DNA sequencing. The entire gene was sequenced in order to confirm no unintended mutation.

Expression and Purification. Five ml of LB medium containing 100 μ g/ml of ampicillin and 20 μ g/ml chloramphenicol was inoculated with a fresh colony of the expression strain BL21(DE3)pLysS containing pET17b-yCD. The culture was grown overnight at 33 °C with vigorous shaking (~200 rpm). The overnight culture was first checked for the expression of yCD by SDS-PAGE and used for seeding a large scale growth. The expression of yCD was induced by the addition of IPTG to a final concentration of 0.5 mM when the OD₆₀₀ of the culture reached 1.2. The culture was then cooled down to room temperature and grown for six more hours. The *E. coli* cells were harvested by centrifugation, washed once with buffer A (50 mM potassium phosphate, pH 7.9), and kept at -20 °C until use.

The frozen bacterial paste was thawed at room temperature and suspended in 100 ml of pre-cooled buffer A. The cells were disrupted by a French press. The resulting

lysate was centrifuged for 20 min at ~27,000 g and 4 °C. Polyethyleneimine was added to the pooled supernatant to a final concentration of 0.1%. The solution was centrifuged immediately after mixing at 15,000 g for 30 min. Ammonium sulfate was added in small portions to the supernatant under constant stirring to 40% saturation. After another hour of stirring, the solution was centrifuged at ~27,000 g for 20 min. The supernatant was loaded to a phenyl Sepharose column equilibrated with 40% saturation ammonium sulfate in buffer B (50 mM potassium phosphate, pH 7.0). The column was washed with the equilibrium solution until the OD₂₈₀ of the effluent was <0.05 and eluted with a linear ammonium sulfate gradient (40-0% saturation) in buffer B. Fractions containing yCD were identified by SDS-PAGE and concentrated to ~15 ml by an Amicon concentrator using an YM10 membrane. The protein solution was then applied to a Sephadex G-75 column equilibrated with 20 mM Tris-HCl, pH 7.5. The column was developed with the same buffer. Fractions from the gel filtration column were monitored by OD₂₈₀ and SDS-PAGE. Pure yCD fractions were pooled and concentrated to 10-20 ml. The concentrated yCD was dialyzed against 2 mM potassium phosphate buffer (pH 7.0) and lyophilized.

¹⁵N-Labeling. For ¹⁵N-labeling of yCD, the expression strain BL21(DE3)pLysS containing pET17b-yCD was grown in a M9 medium with ¹⁵NH₄Cl as the sole nitrogen source. The cells were grown first at 33 °C and were induced by 0.5 mM IPTG when the culture reached an OD₆₀₀ of 1.2. The culture was further incubated at room temperature for 10 h. The ¹⁵N-labeled yCD was purified by the same procedure as for the unlabeled protein.

5-Fluorotryptophan Labeling. To produce 5-fluorotryptophan labeled yCD, 1 L of M9 medium containing 100 µg/ml of ampicillin and 20 µg/ml chloramphenicol was

inoculated with 5 ml culture of the expression strain BL21(DE3)pLysS containing pET17b-yCD and grown overnight at 33 °C with vigorous shaking. The *E. coli* cells were centrifuged down and re-suspended in 3 L of M9 medium containing 100 µg/ml of ampicillin and 20 µg/ml chloramphenicol and grown until OD₆₀₀ reached 1.0. 5-Fluorotryptophan was then added to a final concentration of 50 mg/L. The culture was incubated further at 33 °C until OD₆₀₀ reached 1.4 and then induced with 0.5 mM IPTG (final concentration). The cells were harvested by centrifugation after 4 h of incubation at room temperature. The labeled protein was purified and the percentage of 5-fluorotryptophan labeling was determined by mass spectrometry.

pH Indicator Assay. For measuring the steady-state kinetic parameters of 5FC, the reaction mixture contained 1 mM 5FC, 0.09 mM cresol red, and 8.1-12 µg/ml yCD in 100 mM bicine, 100 mM NaCl, pH 7.5. The reaction was initiated by addition of the enzyme at 25 °C and followed by monitoring the absorption of the pH indicator cresol red at 572 nm. The values of the steady state kinetic parameters were estimated by the numerical analysis of the time course of the reaction using the program DYNAFIT (9). The steady-state kinetic parameters of cytosine were determined by initial velocity analysis. The reaction mixture contained 0.09 mM cresol red, ~0.6 µg/ml yCD, 0.2, 0.5, 0.8, 1.2, 2.4, 4.8, or 9.6 mM cytosine. The initial rates were analyzed according to the standard Michaelis-Menten equation.

Stopped-Flow Analysis. Stopped-flow experiments were performed in an Applied Photophysics SX.18MV-R stopped-flow spectrofluorometer at 25 °C. One syringe contained yCD, and the other contained the substrate 5FC. Both yCD and 5FC were dissolved in 20 mM sodium phosphate, 150 mM NaCl, pH 7.3. The absorption at 296 nm

was monitored. The data were analyzed by nonlinear least square fit to an exponential equation using the program Origin (OriginLab).

Quench-Flow Analysis. Quench-flow experiments were carried out with a KinTek RQF-3 rapid quench-flow instrument at 25 °C. All reaction components were dissolved in 20 mM phosphate, 150 mM NaCl, pH 7.3. One syringe contained yCD, and the other contained the substrate 5FC. A trace amount of [6-³H]-5FC was used to follow the reactions. The reactions were quenched with 0.5 M acetic acid, 10 mM 5FC, and 10 mM 5FU. The substrate 5FC and product 5FU were separated by TLC on a RP-18 F_{254s} aluminium sheet (Merck) developed with acetonitrile and acetic acid at a 50:1 ratio. After the TLC plate was air dried for 10 min, 5FC and 5FU were spotted under UV light. The fluorescent spots were cut out and soaked in 1 ml of the developing solution in scintillation vials and shaken for 30 min at room temperature. Radioactivities were measured with a Beckman LS6500 scintillation counter after the addition of 7.5 ml scintillation fluid to each scintillation vial. In pre-steady state experiments, the reaction mixture contained 20 μM yCD and various concentrations of 5FC. In single-turnover experiments, the reaction mixture contained 300 μM yCD and 15 μM 5FU. All concentrations were those after mixing. The presteady-state and single-turnover data were first analyzed by nonlinear least square fit to appropriate single exponential equations as previously described (10). The amplitudes and rate constants were then used to set the initial values for fitting the data to the complete mechanism by numerical analysis using the program DYNAFIT (9).

¹H-¹⁵N HSQC NMR Spectroscopy. NMR samples were prepared by dissolving ¹⁵N-labeled yCD in 100 mM potassium phosphate, 100 μM NaN₃, pH 7.0 made in H₂O/²H₂O

(13/1). DSS (20 μ M) was used as an internal standard for chemical shift calibration. The initial protein concentration was 1.5 mM in protomers. The NMR samples containing 12 and 20 mM 5FU were made by adding aliquots of a concentrated 5FU solution. The NMR sample containing 75 mM 5FU was made by adding 5FU powder. The sensitivity-enhanced ^1H - ^{15}N HSQC spectra were acquired at 25 $^{\circ}\text{C}$ on a Varian INOVA 600 spectrometer. The spectrum widths were 9476 and 2400 Hz for ^1H and ^{15}N dimensions, respectively. 160 (t_1 , ^{15}N dimension) \times 1946 (t_2 , ^1H dimension) complex data points were recorded for each spectrum. The number of transients was 32 for each FID with a 1.5 s delay between transients. The NMR data were processed with the program NMRPipe (11).

^{19}F NMR Spectroscopy. ^{19}F NMR experiments were performed on a Varian INOVA 300 or 600 MHz NMR spectrometer. The NMR samples were made with the same buffer as that for the ^1H - ^{15}N HSQC NMR experiments. The NMR spectra were acquired with a spectral width of 5000 Hz, 5000 data points, and 1024 transients for each spectrum with a 2 s delay between transients. The NMR data were processed with the program VNMR (Varian Associates). The ^{19}F chemical shifts were referenced to $\text{CF}_3\text{C}_6\text{H}_5$.

Saturation transfer experiments were carried out as described by Lian and Roberts (12). The NMR sample for the saturation transfer experiments contained 1.5 mM yCD labeled with 5-fluorotryptophan and 24.7 mM 5FU. Two complementary saturation transfer experiments were performed, one saturating a ^{19}F NMR signal of free yCD and the other saturating a ^{19}F NMR signal of 5FU-bound yCD. The signals were saturated with a low power gated decoupling pulse. The pulse power was set to -2 dB on the basis of the results of preliminary tests. Eight spectra were acquired for each saturation transfer

experiment with saturation times of 0, 0.04, 0.06, 0.08, 0.1, 0.12, 0.2, and 0.4 s. The control experiment was performed with the saturation frequency set at one end of the spectrum. The intensity of the monitored peak is described by the following equation.

$$I/I_0 = (k/\lambda)e^{-\lambda t} + R/\lambda$$

where I is the peak intensity of the monitored species when the other species is saturated, I_0 is the peak intensity of the monitored species in the control experiment, k is the rate constant for the conversion of the monitored species to the saturated species, R is the longitudinal relaxation rate constant for the monitored species, and $\lambda = k + R$. The data were fitted by nonlinear least squares regression to an exponential equation using the program Origin.

$$I_{rel} = Ae^{-\lambda t} + c$$

where I_{rel} is the peak intensity of the monitored species relative to that of the control experiment. When the NMR signal of the bound yCD is saturated, the product of A and λ provides the value for the dissociation rate constant. When the NMR signal of the free yCD is saturated, the product of A and λ is the rate constant for the conversion of the free yCD to the bound yCD. The concentrations of the free enzyme, the bound enzyme, and the free ligand can be calculated from the standard equilibrium relationship. The association rate constant can be then obtained because the rate constant for the conversion of the free yCD to the bound yCD is the product of the association rate constant and the concentration of free 5FU at equilibrium.

Results

Steady State Kinetic Analysis. CD is traditionally assayed by measuring UV absorbance changes (13). Although the method is convenient for routine measurement of CD activity, it is not accurate for measuring kinetic constants, because both the substrate cytosine and the product uracil have high molar extinction coefficients and similar maximum absorbance wavelengths. For example, for an accurate measurement of the K_m value of cytosine (1.1 mM), the substrate concentration must go up to at least 5 mM (~5 times K_m), preferably 10 mM (~10 times K_m). The absorbance for 5 mM cytosine is 30.5 at the maximum absorbance wavelength (267 nm) and 6.1 at the wavelength (286 nm) normally used for measuring CD activity. To overcome this problem, we developed a pH indicator assay. Although the use of pH indicators to monitor the progress of enzymatic reactions has a long history, the technique has not been used for measuring the kinetics of the CD-catalyzed reaction. The pH indicator assay took advantage of the high pK_a (9.25) of ammonia so that essentially all ammonia generated by the CD-catalyzed reaction is converted to ammonium at the assay pH (7.5). The resultant pH change causes a change in the absorbance of the pH indicator at 572 nm. The absorbance change is proportional to the ammonia generated by the CD-catalyzed reaction if the pK_a of the pH indicator is equal to that of the buffer. Cresol red was used as the pH indicator because its pK_a (8.3) matches with that of bicine (8.35) used for making the assay buffer. Because kinetic parameters change with pH, the pH changes were kept within 0.05 unit. CO₂ absorption effects were not an issue, because the reaction rates with the amount of the enzyme in the assay were at least an order of magnitude higher than the basal rates obtained without the enzyme (control). Therefore, the CO₂ absorption effects could be safely ignored. We also tested the phenol red (pK_a = 7.81) and POPSO (pK_a = 7.82) combination, and the kinetic

constants obtained by the two systems were the same. We chose the higher pK_a system because it gave a higher signal with the same small magnitude of pH change. The buffering capacity was compensated by increasing the concentration of the buffering component. The results of the kinetic measurement by the pH indicator assay are summarized in Table 1. The results showed that yCD has a slightly higher catalytic efficiency (k_{cat}/K_m) for the prodrug 5FC than for the natural substrate cytosine. The catalytic efficiency of yCD for 5FC is ~10-fold higher than that of the *E.coli* enzyme (14), in agreement with the report that yCD is a better enzyme for CD/5FC-based GDEPT (15).

Transient Kinetic Analysis. In order to determine the rate constants for the individual steps of the prodrug activation, we performed the transient kinetic analysis of the reaction. We first performed stopped-flow spectrophotometric analysis of the reaction. As shown in Figure 2-1A, there was a burst increase followed by a steady decrease in absorbance. A similar phenomenon was observed for *E. coli* cytosine deaminase (14). The burst increase in absorbance was attributed to the change in the environment of the substrate upon binding to the enzyme, the decrease to the conversion of the substrate to the products. The absorbance changes could be phenomenologically described by an exponential and a linear term. The rate constant of the exponential term is linearly dependent on the concentrations of the substrate (Figure 2-1B). The apparent association and dissociation rate constants were estimated to be $0.21 \mu M^{-1}s^{-1}$ and $33 s^{-1}$, respectively, which were probably the low limits because the substrate complex was rapidly converted to the product complex as shown by a quench-flow analysis.

Because the extinction coefficient of the bound 5FC is slightly higher than that of the free 5FC, it was difficult to estimate how much 5FU was formed at the initial phase of the reaction. We then performed several stopped-flow experiments with a pH indicator. It turned out that the data obtained in the first 50 ms was very noisy and unusable. Quench-flow experiments were used instead to obtain definitive kinetic information for the prodrug activation. The results are shown in Figure 2-2. The burst experiments (the top three lines in Figure 2-2) clearly indicated that product release is the rate-limiting step in the activation of the prodrug. A single turnover experiment in which the enzyme concentration was larger than that of the substrate was also performed (the bottom line in Figure 2-2). The data from both burst and single turnover experiments were analyzed by global fitting according to the following minimal kinetic mechanism (Scheme 2) using the numerical analysis program DYNAFIT (9, 10). The initial values for the global numerical analysis were obtained by nonlinear least squares fit of the data to appropriate exponential equations.

The rate constants obtained by the global numerical analysis are summarized in Table 1. The rate constant for the forward reaction is eight times that of product release and more than ten times of k_{cat} . The K_m and k_{cat} values calculated from the individual rate constants were very similar to those determined by steady-state kinetic measurements, indicating that the individual rate constants determined by the transient kinetic analysis are consistent with the steady-state kinetic parameters. The slow product release is probably due to the slow dissociation of 5FU and was confirmed by the NMR analysis of the binding of 5FU to yCD.

NMR Analysis. To confirm that the product release is rate-limiting in the prodrug activation, several NMR experiments were performed. First, we acquired ^1H - ^{15}N HSQC spectra of ^{15}N -labeled yCD in the absence of 5FU and in the presence of 12, 20, and 80 mM 5FU. As shown in Figure 2-3B, many residues had two sets of cross-peaks under unsaturated conditions, indicating that 5FU was in slow exchange with the complex of yCD and 5FU on NMR time scale and yCD was not saturated with 5FU. Only one set of HSQC cross-peaks were obtained when 5FU reached ~80 mM (the maximum solubility), indicating that yCD was predominately in the bound form at this concentration of 5FU (data not shown). The sequential resonance assignment of yCD in complex with the reaction intermediate analog 5-fluoro-2-pyrimidinone has been achieved by 3D double and triple resonance NMR experiment, including ^{15}N -edited 3D NOESY, HNCA, HN(CO)CA, HN(CA)CB, HN(COCA)CB, HNCO, HN(CA)CO, HCCH-TOCSY, C(CO)NH-TOCSY, and H(CCO)NH-TOCSY data (16), using single, double, and triple labeled yCD samples (Yao et al., unpublished). Because most of the cross peaks in the HSQC spectra are well resolved, many cross peaks of unliganded yCD (Figure 2-3A) and its complex with 5FU (Figure 2-3B) could be assigned by comparison with the assigned ^1H - ^{15}N HSQC spectrum of yCD in complex with the reaction intermediate analog. Most of the cross-peaks that moved upon the binding of 5FU belong to the residues in the vicinity of the active site. The K_d of the binary product complex was estimated to be ~20 mM, because about equal amounts of yCD was in the free form and the bound form at 20 mM 5FU. Second, we acquired ^{19}F NMR spectra of free 5FU and in complex with yCD (Figure 2-4). The formation of the yCD complex shifted the ^{19}F NMR signal of 5FU by ~4.5 ppm. The results clearly indicated that the free 5FU is in slow exchange with the

yCD and 5FU complex on the ^{19}F NMR timescale. Third, we acquired ^{19}F NMR spectra of yCD labeled with 5-fluorotryptophan in the absence and presence of 5FU (Figure 2-5). The degree of the 5-fluorotryptophan labeling was estimated to be ~90% on the basis of the ^1H - ^{15}N HSQC spectra of single (^{15}N) yCD and double labeled (^{15}N and ^{19}F) yCD proteins. The k_{cat} and K_{m} values of the ^{19}F -labeled yCD were $15 \pm 0.4 \text{ s}^{-1}$ and $80 \pm 6 \text{ }\mu\text{M}$, respectively, indicating that the 5-fluorotryptophan labeling has no significant effects on the kinetic properties of yCD. The ^{19}F -labeled yCD showed two ^{19}F NMR peaks, and one was much sharper than the other. The binding of 5FU shifted the broader NMR peak by ~2 ppm, but caused no significant change in the position of the narrower peak (Figure 2-5). The wild-type yCD has two tryptophan residues, Trp10 at the N-terminus (away from the active site) and Trp152 at the active site. The two ^{19}F NMR peaks were assigned on the basis of the comparison with the ^{19}F NMR spectrum of the yCD mutant W10H, which showed only one peak from Trp152 (data not shown). Surprisingly, the peak shifted belonged to Trp10 at the N-terminus rather than Trp152 at the active site. The result again indicated that 5FU is in slow exchange with its complex with the enzyme on the NMR time scale. Furthermore, the intensities of the free yCD and the bound yCD were about equal in the presence of 24.7 mM 5FU, again indicating that the K_{d} value for the binary complex is ~20 mM. Fourth, we determined the exchange rate constants by NMR saturation transfer experiments. The results are shown in Figure 2-6. The results were analyzed using a simple one-step binding model. The dissociation rate constant was determined by saturating the ^{19}F NMR signal of the bound yCD, which was 13 s^{-1} . The rate for the formation of the complex, a product of the association rate constant and the concentration of free 5FU at equilibrium, was determined by saturating the ^{19}F NMR

signal of the free yCD, which was 14 s^{-1} . The concentration of the free 5FU was calculated to be 24 mM using the standard equilibrium relationship. The association rate constant therefore was $0.6\text{ mM}^{-1}\text{s}^{-1}$, and the K_d was 22 mM. The results indicated that 5FU is in slow exchange with its yCD complex but has a rather high K_d .

Discussion

yCD is of great biomedical interest because it catalyzes the deamination of the prodrug 5FC to form the anticancer drug 5FU. By a combination of transient kinetic and NMR studies, we have clearly shown that the product release is a rate-limiting step in the activation of the prodrug 5FC by yCD. The transient kinetic studies showed that the rate constant of the chemical step for the forward reaction (250 s^{-1}) is ~8 times that of the product release (31 s^{-1}) and ~15 times k_{cat} (17 s^{-1}). The transient kinetic results are consistent with those of the steady state kinetic analysis in the sense that the k_{cat} and K_m values calculated from the rate constants determined by the transient kinetic analysis are in close agreement with those measured by the steady state kinetic analysis. The steady-state kinetic parameters, however, are insufficient for the description of the kinetics of the prodrug activation. The yCD-catalyzed activation of the prodrug 5FC generates two products, 5FU and ammonia. The four NMR experiments described in the results section clearly indicated that the release of 5FU is rate limiting in the activation of the prodrug 5FC by yCD.

There are two possible causes for the slow release of 5FU by yCD. One involves the breaking of the coordination between 5FU and the catalytic zinc, and the other the opening of the lid that covers the active site. When 5FC is converted to 5FU at the active site of yCD, the oxygen at position 4 of 5FU is coordinated with the catalytic zinc. Using

the ONIOM methodology (17-24), our recent computational study of the deamination of cytosine showed that energetically, the cleavage of the O4-Zn bond is rather difficult either in the presence or in the absence of ammonia (5). The computational study suggests that uracil is liberated from the zinc by an oxygen exchange mechanism that involves the formation of a *gem*-diol intermediate from the Zn bound uracil and a water molecule, the C4-O^{Zn} cleavage, and the regeneration of the Zn-coordinated water. The rate determining step in the oxygen exchange is the formation of the *gem*-diol intermediate, which is also the rate determining step for the overall yCD-catalyzed deamination reaction.

5FU is also likely to be completely buried at the active site of yCD. The structure of yCD in complex with the reaction intermediate analog 2-pyrimidinone has been recently determined at high resolutions (3, 4). The bound reaction intermediate analog is completely buried, being covered by a lid composed of Phe114, Trp152, and Ile156, the latter two of which are from the C-terminal helix. It is possible that the opening of the lid may also determine the rate of product release in yCD. Furthermore, the structure of apo yCD is essentially the same as that of the reaction intermediate complex (3). The active center is also covered by the same cluster of hydrophobic residues in the apo enzyme and appears inaccessible to the substrate (3). It appears that the opening of the lid is required not only for product release but also substrate binding. Thus, the release or binding of 5FU may involve multiple steps as illustrated in Scheme 3. where E^o and E^c represent the open and closed conformations of the unliganded enzyme, respectively, E^c-5FU is a complex in which 5FU is coordinated with the catalytic zinc and yCD is in a closed conformation, E^c.5FU is a complex in which 5FU is not coordinated with the zinc and the

enzyme is still in a closed conformation, and $E^o.5FU$ is a complex in which 5FU is not coordinated with the zinc but the enzyme is in an open conformation.

The NMR results are interesting, because the results not only identified that the release of 5FU is a slow step in the deamination of 5FC but also showed that 5FU has a low affinity for yCD (high K_d). In general, slow exchange means tight binding and a low K_d . One explanation for the high K_d is that yCD exists in two conformations, a closed conformation with an inaccessible active center and an open conformation with an accessible active center, as illustrated in Scheme 3. 5FU binds to the open conformation only. The equilibrium is in favor of the closed conformation, which is most populated and was crystallized (3). The apparent K_d is the reciprocal of the overall equilibrium constant, which is a product of the equilibrium constant for the conformational transition of the unliganded yCD and the equilibrium constant for the binding of 5FU to the open conformation. An equilibrium constant in favor of the closed conformation will raise the apparent K_d by a factor of the reciprocal of the equilibrium constant. The unusually low apparent association rate constant may also be attributed to the multiple step nature of the binding of 5FU to yCD.

yCD belongs to the CDA family of purine/pyrimidine deaminases (3, 4). In addition to yCD and other fungal cytosine deaminases, members of the family of enzymes include cytidine deaminases, guanine deaminases, and riboflavin biosynthesis enzymes. These enzymes have similar structures with a zinc-containing catalytic apparatus (25). The structures of the family of enzymes reported to date are all in a closed conformation. Because these enzymes all have a closed conformation and most likely follow a similar chemical mechanism, product release may also be rate-limiting in the

reactions catalyzed by other members of this family of enzymes. However, a lack of a viscosity dependence on the k_{cat} of *E. coli* cytidine deaminase suggests that product release may not be rate-limiting in the reaction catalyzed by the enzyme (26). On the other hand, full ^{15}N kinetic isotope effects are manifested in the reaction catalyzed by mutants with significantly reduced catalytic efficiencies but not the wild-type cytidine deaminase (7), indicating that there is another step influencing the rate besides the slow $\text{C}^4\text{-N}^4$ bond cleavage in the deamination of cytidine.

In conclusion, the transient kinetic and NMR data together clearly showed that the release of 5FU is rate-limiting in the activation of the prodrug 5FC by yCD and may involve multiple steps, the interconversion of yCD between a closed and an open conformation and the cleavage or formation of the coordination between 5FU and the catalytic zinc of yCD.

References

- (1) Aghi, M., Hochberg, F., and Breakefield, X. O. (2000) Prodrug activation enzymes in cancer gene therapy. *J. Gene Med.* 2, 148-164.
- (2) Greco, O., and Dachs, G. U. (2001) Gene directed enzyme/prodrug therapy of cancer: Historical appraisal and future prospectives. *J. Cell. Physiol.* 187, 22-36.
- (3) Ireton, G. C., Black, M. E., and Stoddard, B. L. (2003) The 1.14 Å crystal structure of yeast cytosine deaminase evolution of nucleotide salvage enzymes and implications for genetic chemotherapy. *Structure* 11, 961-972.
- (4) Ko, T.-P., Lin, J.-J., Hu, C.-Y., Hsu, Y.-H., Wang, A. H.-J., and Liaw, S.-H. (2003) Crystal structure of yeast cytosine deaminase. Insights into enzyme mechanism and evolution. *J. Biol. Chem.* 278, 19111-19117.
- (5) Sklenak, S., Yao, L. S., Cukier, R. I., and Yan, H. G. (2004) Catalytic mechanism of yeast cytosine deaminase: An ONIOM computational study. *J. Am. Chem. Soc.* 126, 14879-14889.
- (6) Schramm, V. L., and Bagdassarian, C. K. (1999) Deamination of nucleosides and nucleotides and related reactions, in *Enzymes, Enzyme Mechanisms, Proteins, and Aspects of NO Chemistry* (Poulter, C. D., Ed.) pp 71-100, Amsterdam, New York.
- (7) Snider, M. J., Reinhardt, L., Wolfenden, R., and Cleland, W. W. (2002) ¹⁵N kinetic isotope effects on uncatalyzed and enzymatic deamination of cytidine. *Biochemistry* 41, 415-421.
- (8) Snider, M. J., Lazarevic, D., and Wolfenden, R. (2002) Catalysis by entropic effects: The action of cytidine deaminase on 5,6-dihydrocytidine. *Biochemistry* 41, 3925-3930.
- (9) Kuzmic, P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: Application to HIV proteinase. *Anal. Biochem.* 237, 260-273.
- (10) 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.
- (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, 277-93.

- (12) Lian, L. Y., and Roberts, G. C. K. (1993) Effects of chemical exchange on NMR spectra, in *NMR of Macromolecules* (Roberts, G. C. K., Ed.), IRL Press, New York.
- (13) Ipata, P. L., Marmocchi, F., Magni, G., and Felicioli, R. (1971) Baker's yeast cytosine deaminase. Some properties and allosteric inhibition by nucleosides and nucleotides. *Biochemistry* 10, 4270-4276.
- (14) Porter, D. J. T. (2000) *Escherichia coli* cytosine deaminase: the kinetics and thermodynamics for binding of cytosine to the apoenzyme and the Zn^{2+} holoenzyme are similar. *Biochim. Biophys. Acta* 1476, 239-252.
- (15) Kievit, E., Bershad, E., Ng, E., Sethna, P., Dev, I., Lawrence, T. S., and Rehemtulla, A. (1999) Superiority of yeast over bacterial cytosine deaminase for enzyme/prodrug gene therapy in colon cancer xenografts. *Cancer Res.* 59, 1417-1421.
- (16) Gardner, K. H., and Kay, L. E. (1998) The use of ^2H , ^{13}C , ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* 27, 357-406.
- (17) Dapprich, S., Komaromi, I., Byun, K. S., Morokuma, K., and Frisch, M. J. (1999) A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct.* 462, 1-21.
- (18) Humbel, S., Sieber, S., and Morokuma, K. (1996) The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for n-butane conformation and $\text{S}(\text{N})2$ reaction: $\text{RCl}+\text{Cl}$. *J. Chem. Phys.* 105, 1959-1967.
- (19) Kuno, M., Hannongbua, S., and Morokuma, K. (2003) Theoretical investigation on nevirapine and HIV-1 reverse transcriptase binding site interaction, based on ONIOM method. *Chem. Phys. Lett.* 380, 456-463.
- (20) Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S., and Morokuma, K. (1996) ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and $\text{Pt}(\text{P}(\text{t-Bu})(3))(2)+\text{H}_2$ oxidative addition. *J. Phys. Chem.* 100, 19357-19363.
- (21) Svensson, M., Humbel, S., and Morokuma, K. (1996) Energetics using the single point IMOMO (integrated molecular orbital plus molecular orbital) calculations: Choices of computational levels and model system. *J. Chem. Phys.* 105, 3654-3661.
- (22) Vreven, T., Mennucci, B., da Silva, C. O., Morokuma, K., and Tomasi, J. (2001) The ONIOM-PCM method: Combining the hybrid molecular orbital method and

- the polarizable continuum model for solvation. Application to the geometry and properties of a merocyanine in solution. *J. Chem. Phys.* **115**, 62-72.
- (23) Vreven, T., and Morokuma, K. (2000) On the application of the IMOMO (integrated molecular orbital plus molecular orbital) method. *J. Comput. Chem.* **21**, 1419-1432.
- (24) Vreven, T., Morokuma, K., Farkas, O., Schlegel, H. B., and Frisch, M. J. (2003) Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. *J. Comput. Chem.* **24**, 760-769.
- (25) Liaw, S. H., Chang, Y. J., Lai, C. T., Chang, H. C., and Chang, G. G. (2004) Crystal structure of *Bacillus subtilis* guanine deaminase - The first domain-swapped structure in the cytidine deaminase superfamily. *J. Biol. Chem.* **279**, 35479-35485.
- (26) Snider, M. J., Gaunitz, S., Ridgway, C., Short, S. A., and Wolfenden, R. (2000) Temperature effects on the catalytic efficiency, rate enhancement, and transition state affinity of cytidine deaminase, and the thermodynamic consequences for catalysis of removing a substrate "anchor". *Biochemistry* **39**, 9746-9753.

Appendices

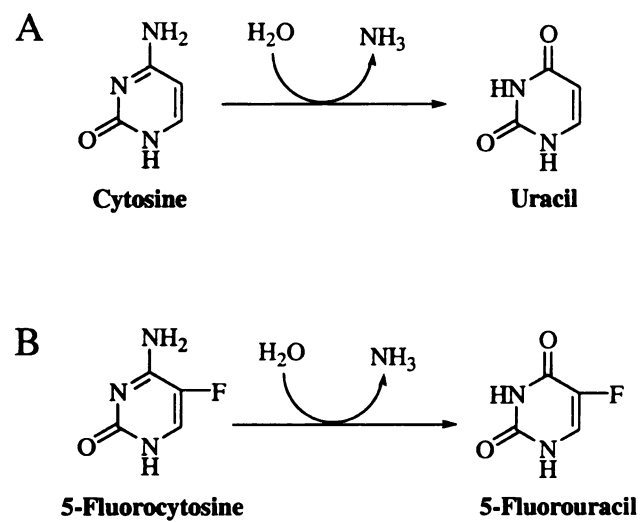
Table 2-1 Kinetic Constants for the Activation of the Prodrug 5FC by yCD^a

Steady State Kinetics						
5-Fluorocytosine			Cytosine			
K_m (mM)	k_{cat} (s ⁻¹)	k_{cat}/K_m (M ⁻¹ s ⁻¹)	K_m (mM)	k_{cat} (s ⁻¹)	k_{cat}/K_m (M ⁻¹ s ⁻¹)	
0.16±0.01	17±0.4	1.1×10 ⁵	1.1±0.08	91±4	8.2×10 ⁴	
Transient Kinetics						
k_1 (μM ⁻¹ s ⁻¹) (s ⁻¹)	k_{-1} (s ⁻¹)	k_2 (s ⁻¹)	k_{-2} (s ⁻¹)	k_3 (s ⁻¹)	K_m (mM)	k_{cat}
0.48±0.04	93±20	250±30	91±20	31±2	0.11 ^b	21 ^b

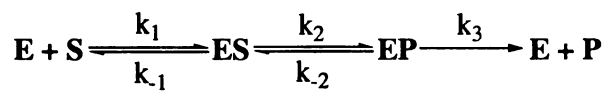
^aThe steady-state kinetic parameter for cytosine are presented for comparison.

^bThe K_m and k_{cat} were calculated from the individual rate constants.

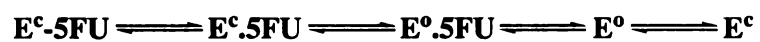
Scheme 1



Scheme 2



Scheme 3



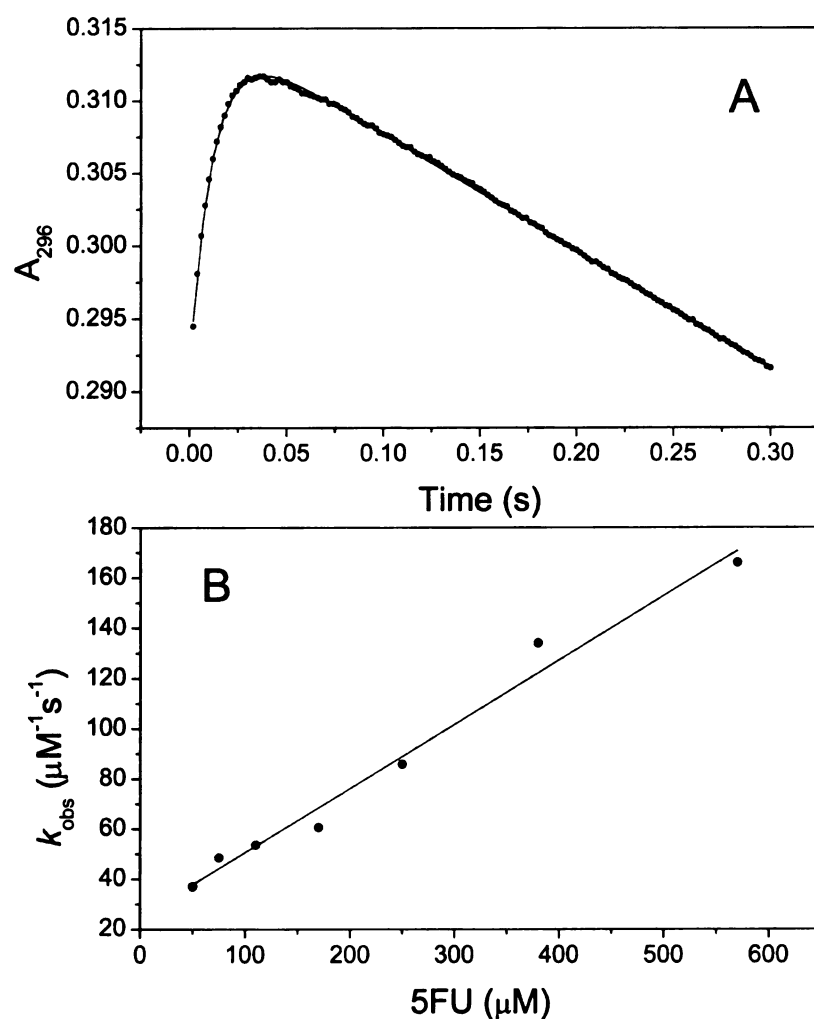


Figure 2-1 Stopped-flow analysis of the activation of the prodrug 5FC by yCD. (A) The time course of the reaction as monitored by OD_{296} . The reaction mixture contained 10 μM yCD and 250 μM 5FC. The solid line was obtained by nonlinear least square fit to an equation with an exponential and a linear term. (B) The linear dependence of the rate constant of the exponential term on the concentration of 5FC.

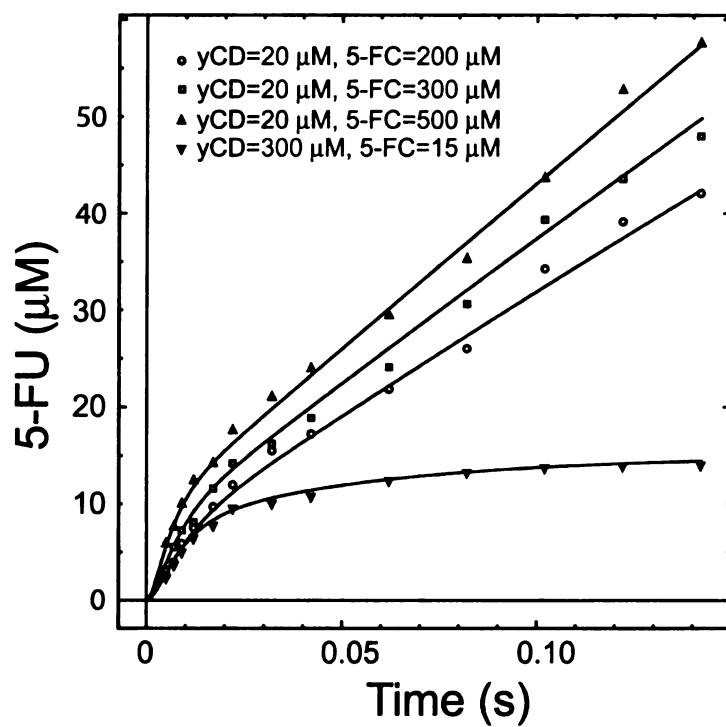


Figure 2-2 Quench-flow analysis of the activation of the prodrug 5FC by yCD. The solid lines were obtained by global fitting using the numerical analysis program DYNAFIT.

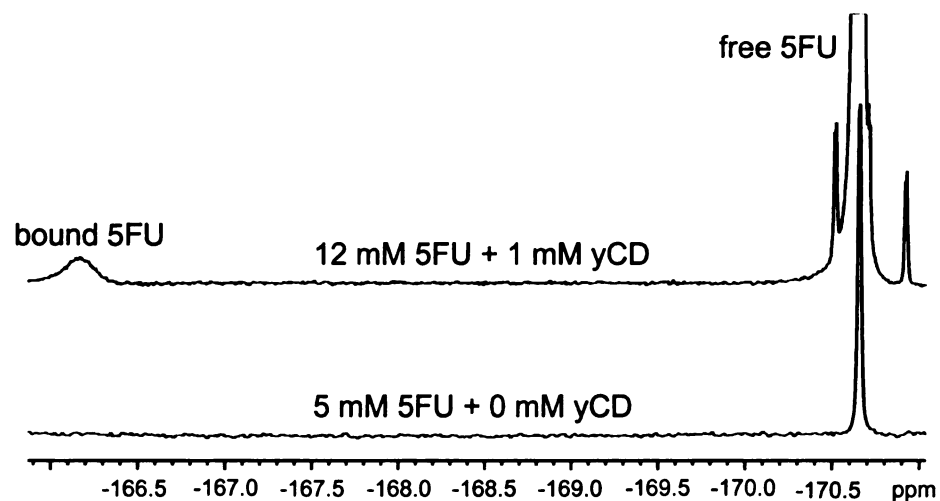


Figure 2-4 ^{19}F NMR spectra of free (unbound) 5FU and the yCD-bound 5FU. The smaller peaks near the strong peak of free 5FU in the upper spectrum are ^{13}C satellite peaks.

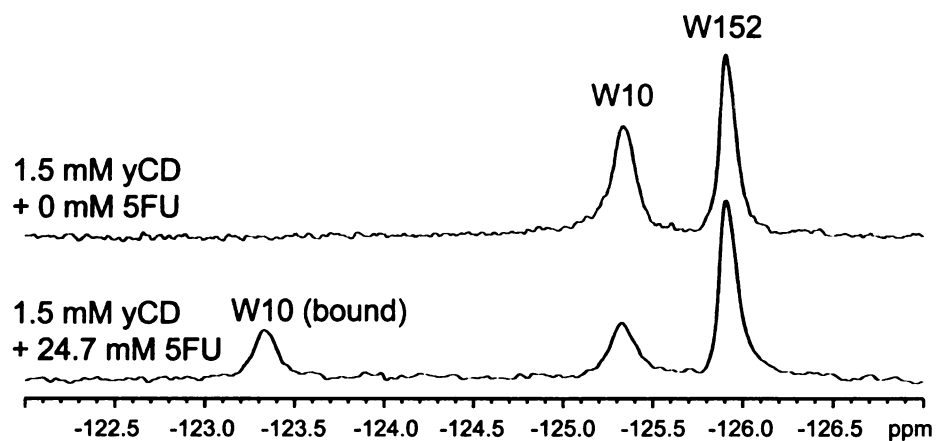


Figure 2-5 ^{19}F NMR spectra of yCD labeled with 5-fluorotryptophan. (A) 1.5 mM 5-fluorotryptophan-labeled yCD; (B) 1.5 mM 5-fluorotryptophan-labeled yCD + 24.7 mM 5FU.

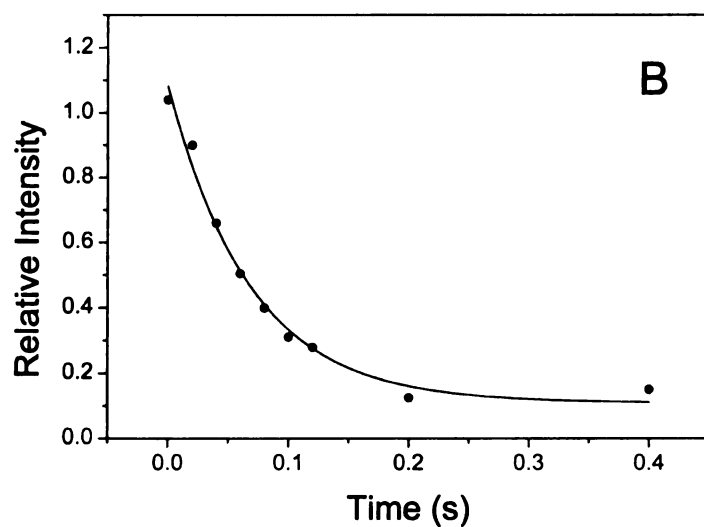
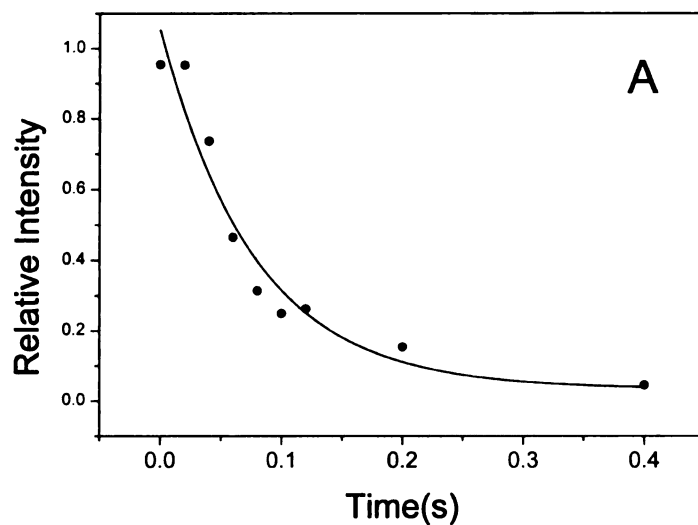


Figure 2-6 NMR saturation transfer analysis of the binding of 5FU to the 5-fluorotryptophan-labeled yCD. The NMR sample contained 1.5 mM 5-fluorotryptophan-labeled yCD and 24.7 mM 5FU. The data in panels A and B were obtained by irradiating the ^{19}F NMR nuclei of the bound and free yCD, respectively. The solid lines were obtained by nonlinear least square fit of the data as described in the “Experimental Procedures” section.

Chapter 3

Yeast Cytosine Deaminase Dynamics Study by NMR and Molecular Dynamics Simulation

Introduction

Yeast cytosine deaminase (yCD) catalyzes the deamination of cytosine to uracil. It can also catalyze the deamination of the prodrug 5-fluorocytosine (5-FC) to the anticancer drug 5-fluorouracil (5-FU). Thus, a potential system for gene-directed enzyme prodrug therapy combines yCD and the prodrug 5-FC. X-ray structures of yCD in the apo form (Ireton et al. 2003) and in complex with the potent inhibitor 2-pyrimidinone (2Py) (Ireton et al. 2003; Ko et al. 2003), are available. yCD is a homodimer with one active site in each protomer. The yCD protomer contains a center β sheet ($\beta 1 \sim \beta 5$, parallel to each other) with two α helices ($\alpha 1$ and $\alpha 5$) on one side and four α helices ($\alpha 2$, $\alpha 3$, $\alpha 4$ and $\alpha 6$) on the other side. Each active center contains a single catalytic zinc ion that is tetrahedrally coordinated by a histidine (His62), two cysteines (Cys91 and Cys94), and a water molecule in the substrate-free enzyme or the inhibitor in the complex. Surprisingly, the structure of apo yCD is essentially the same as that of the transition state analog complex with the active site completely buried with the average root mean square deviation (RMSD) 0.23 Å for all backbone atoms.

The immediate question is whether the dynamics are the same, though the two structures are almost identical. In this paper, we explore the dynamic property of the apo and inhibitor bound form by using an NMR backbone relaxation method (Abragam 1961;

Fischer et al. 1998; Ishima and Torchia 2000; Palmer 2001), which primarily describes the motion on the picosecond (ps) to nanosecond (ns) timescale. The inhibitor we use is 5-fluoro-2-pyrimidinone (5FPy) which is a transition state analog for the reaction from 5FC to 5FU. Model-free analysis was used to extract the order parameter S^2 (Lipari and Szabo 1982; Lipari and Szabo 1982; Mandel et al. 1995) describing the motion of each backbone NH vector.

MD simulation can provide a great detail concerning individual atom motions as a function of time (Karplus and McCammon 2002). Previous MD simulations for the apo form, and reactant cytosine, various intermediate and product uracil complexes (Yao et al. 2005) all showed a buried active site during the ~2.0 ns simulations, with the overall protein quite rigid. In this study we performed three 10 ns MD simulations for apo form, inhibitor (2Py) bound form and product (uracil) bound form to see the possible dynamics difference. Since these MD simulations also describe the motion on a ps to ns timescale, the computational results can provide complementary information to NMR experimental data about protein dynamics on this timescale. The order parameter and internal motion correlation time were calculated for each amide NH vector. Though direct comparison of order parameters from MD and model-free analysis is not straightforward because of the limitations of both methods (Korzhnev et al. 2001; Case 2002), the order parameter changes of yCD due to the inhibitor binding should be comparable.

As is well known, the order parameter is defined by (Lipari and Szabo 1982)

$$S^2 = \sum_{q=-2}^2 \left| Y_2^q(\theta, \phi) \right|^2 \quad (1)$$

in which $Y_2^q(\theta, \phi)$ is a modified second-order spherical harmonic function, defined by

$$Y_2^q(\theta, \phi) = \begin{cases} (3\cos^2\theta - 1)/2 & q = 0 \\ \sqrt{3/2} \sin\theta \cos\theta e^{i\phi} & q = 1 \\ \sqrt{3/8} \sin^2\theta e^{i2\phi} & q = 2 \end{cases} \quad (2)$$

$$Y_2^{-q}(\theta, \phi) = Y_2^{q*}(\theta, \phi)$$

Essentially the order parameter S^2 describes the fluctuation of the angular coordinates of the NH vector. A small order parameter indicates a large angular fluctuation, from a more flexible NH vector, and usually arises from a larger translational motion of the nitrogen atom. The translational fluctuations of the nitrogen can be characterized by the root mean square fluctuation (RMSF) of N. But, since the lower bound of the order parameter is 0 and the upper bound of the RMSF could be several angstroms on the ps-ns timescale, the question is whether there is a quantitative correlation between RMSF and order parameter in this wide range. Also, one may question whether the correlation is different in different secondary structure regions. In addition, how sensitive is the order parameter to protein motion? More importantly, does it follow that a residue with a smaller order parameter is really the residue with larger flexibility? These questions will be addressed in this study.

One advantage of an MD simulation is that the motion of each atom in the simulation system is determined. In the apo and 2Py bound form X-ray structures (Ireton et al. 2003; Ko et al. 2003) three residues, F114, W152 and I156, block the active site. It has been proposed that one possible product release path is in between the F114 loop (111-117) and the C-terminal helix (150-158), where two residues, F114 and I156, are identified that must move away in order to let the ligand out (Yao et al. 2006). The side-chain dynamics of F114, W152 and I156 should be interesting in this regard, and will be discussed in the Results Section.

Materials and Methods

Construction, expression and purification of ^{15}N labeled yCD were described as before (Yao et al. 2005). An NMR sample of apo (5FPy bound) form was 1.5 mM ^{15}N labeled yCD (with 20 mM 5FPy) in 93% H_2O /7% D_2O 100 mM phosphate buffer. NMR spectroscopy was performed on a 600 MHz Varian Inova spectrometer at 25 °C. The pulse sequences for R_1 , R_2 , NOE measurements are based on sensitivity-enhanced gradient-based HSQC experiment (Farrow et al. 1994). For R_1 , R_2 measurements, a recycle delay of 2.0 s was used between transients. For the NOE measurements, a recycle delay of 8 s was used for the non-NOE reference, and 4 s was used for the NOE measurement. A train of 120° hard pulses at 5 ms intervals with a sinc soft pulse inserted was used for 4 s proton saturation, with minimum perturbation of water magnetization found during saturation. The R_1 (R_2) experiments were performed using 16 (32) transients. 64 transients were used for NOE measurements. For R_1 measurements, 9 spectra were recorded using delays 11.1, 55.5, 133.2, 233.1, 377.4, 555.0, 888.0, 1332.0, 1998.0 milliseconds (ms). For R_2 measurement, 11.77, 23.53, 35.30, 47.07, 58.84, 70.60, 82.37, 109.8, 137.3 ms were used for the delay.

The spectral widths were 9476 and 2040 Hz for ^1H and ^{15}N dimensions, respectively; 128 (t_1 , ^{15}N dimension) and 1946 (t_2 , ^1H dimension) complex data points were recorded for each spectrum. The NMR data were processed with NMRPipe (Delaglio et al. 1995). The peak heights were measured by using NMRView (Johnson and Blevins 1994). The R_1 , R_2 relaxation rates were obtained by fitting the peak intensity to a two-parameter exponential decay function using Curvefit (A.G. Palmer III, Columbia University), and uncertainties in fitted parameters were estimated using Jackknife

simulations (Eisenmann et al. 2004). The average error for R_1 and R_2 is 2.6% and 3.6% for the 5FPy bound form and 4.5% and 4.7% for the apo form. It was proposed that the fit of peak volume gives more accurate results than peak height because the search routine of peak height to find the maximum in a peak overestimates intensity, especially for weak peaks. They usually come with a longer delay time which could underestimate the R_1 and R_2 rates in the exponential fitting (Viles et al. 2001). But, the disadvantage of the peak volume method is that it is difficult to quantify the volume of overlapped peaks. For well-separated peaks in the 5FPy bound form, peak volume was also used for R_1 , R_2 determinations in the 5FPy bound form. The average difference for R_1 and R_2 is 1.6% and 3.2%, respectively, which is within the error of R_1 and R_2 in peak height fitting. The peak height fitting should give a sufficiently accurate rate; thus, it was used for data analysis.

The NOE enhancement was calculated as a ratio between the cross-peak intensities with and without proton saturation. The errors in NOE values were propagated from the uncertainty of peak height based on RMS noise in the spectra (Nicholson et al. 1992). The average percentage error of NOE is 5.0% for the 5FPy bound form and 5.0% for the apo form.

Model-free calculations were performed using the ModelFree program provided by Prof. Arthur G. Palmer (Palmer et al. 1991). The global tumbling time and anisotropy of the rotational diffusion tensor were obtained from R_2/R_1 ratios using the program TENSOR2 (Dosset et al. 2000). Residues with NOEs larger than 0.6 or

$$\left\{ \left(\langle T_2 \rangle - T_{2,n} \right) / \langle T_2 \rangle \right\} - \left\{ \left(\langle T_1 \rangle - T_{1,n} \right) / \langle T_1 \rangle \right\} > 1.5SD \quad (3)$$

were filtered out. And then residues with

$$\left| \left(\langle R_2 / R_1 \rangle - R_{2,n} / R_{1,n} \right) \right| > 1.5SD \quad (4)$$

were also filtered out. The subscript n is for residue n and SD stands for standard deviation. The calculated principle components of the diffusion tensor, Dz : Dy : Dx, are 1.14 : 1.07 : 1.00 for the 5FPy bound form and 1.09 : 1.05 : 1.00 for the apo form, based on the relaxation data. So, the motional anisotropy of yCD is very small and can be neglected (Korzhnev et al. 2001), Therefore, isotropic global motion was adopted in the dynamics analysis. The global tumbling time is 18.8 ns and 19.5 ns for the 5FPy bound and apo forms, respectively.

Five models with various combinations of model-free parameters are used in the model-free analysis (Mandel et al. 1995). They are model 1 $\{S^2\}$, model 2 $\{S^2, \tau_e\}$, model 3 $\{S^2, R_{ex}\}$, model 4 $\{S^2, \tau_e, R_{ex}\}$ and model 5 $\{S_f^2, S_s^2, \tau_e\}$. Akaike's information criterion (AIC) was used as the model selection method (Akaike 1973; d'Auvergne and Gooley 2003). AIC is computed as $\chi^2 + 2k$, where χ^2 is the fitting error and k is the number of parameters in the selected model (d'Auvergne and Gooley 2003; Chen et al. 2004). All five models can be compared simultaneously by using AIC (Chen et al. 2004; Ding et al. 2005), and the model with lowest AIC is selected. Since AIC doesn't provide information on the quality of fit, 500 Monte Carlo simulations are used to determine whether the selected model is sufficient to describe the data for model 1, 2 and 3. If $\chi^2 < \chi^2(0.1)$ where 0.1 is the pre-chosen critical value, the model was considered to be sufficient. For model 4 and 5, only χ^2 equal to 0 was considered to be sufficient. In the model-free analysis, τ_e was allowed to range up to 5 ns in models 2, 4 and 5.

A reliable estimate of the overall rotation correlation times is very important for the internal motion. The equations (3) and (4) are supposed to filter out the residues with conformational exchanges on micro-millisecond timescale and/or large τ_e , which gives a good initial guess of the global tumbling time. The global tumbling time was optimized during the model-free analysis of local dynamics by using an iterative scheme. The tumbling time was scanned from -0.5 ns to +0.5 ns around the initial value with 0.1 ns each step, when the model-free analysis was performed. The summation of AIC over all spin systems in each step was compared and the lowest one was selected, which gives the best global tumbling time. The final global tumbling time is 18.6 ns and 19.6 ns for the 5FPy bound and apo forms, respectively, quite similar to the initial values, indicating that equations (3) and (4) are effective in removing residues with large τ_e and/or R_{ex} for yCD. Three residues, 77, 154, 156, cannot be fitted in the apo form; two residues, 38 and 85, cannot be fitted in the 5FPy bound form.

MD simulation

To explore the dynamic property of yCD in the apo and bound forms, 10 ns MD simulations were performed for apo form, inhibitor (2Py) complex and product uracil complex. In the simulation, the inhibitor used was 2Py, slightly different from 5FPy in experiment. The only difference is the proton of the fifth position of the pyrimidine ring is substituted by fluorine in 5FPy. There is no hydrogen bond donor close to fluorine; the substitution of fluorine by proton in the simulation should not introduce much effect in terms of the dynamical properties of the protein. The Zn complex and ligand charge parameterization have been described previously (Yao et al. 2005), along with the details of the simulation protocol. All three simulations are performed with constant temperature

300 K and constant volume. The coordinates were saved every 2 ps. The first 2 ns of simulation time were treated as the equilibration period and not included in the data analysis.

According to the original model-free formalism (Lipari and Szabo 1982), the internal dynamics of an amide NH vector can be characterized with the correlation function,

$$C(t) = \langle \mu_0 \bullet \mu_t \rangle = S^2 + (1 - S^2) e^{-t/\tau_e} \quad (5)$$

where μ is the NH vector, and parameterized with the order parameter $S^2 = C(\infty)$ and the internal motion time constant τ_e , defined by

$$\tau_e = \frac{1}{C(0) - C(\infty)} \int_0^{\infty} (C(t) - C(\infty)) dt \quad (6)$$

The correlation function was calculated only up to 4 ns to ensure adequate statistics. S^2 values were obtained by averaging $C(t)$ over the last 500 ps (Chen et al. 2004); τ_e values were obtained based on equation (6). Since yCD is a homodimer, an estimate of the error for S^2 and τ_e can be computed from the differences between the two protomers.

Results

¹⁵N relaxation in Apo and 5FPy bound form

¹⁵N NMR relaxation parameters of the yCD apo and 5FPy bound forms were determined at 298 K at 600 MHz (Figures 3-1, 3-2). 123 residue peaks in the apo and 136 residue peaks in the bound form were resolved well enough for quantitative analysis. Backbone NOEs are sensitive to motion in the picosecond to nanosecond time range, with lower NOEs indicating greater motions on this timescale. The average NOE for the apo form is 0.79±0.06, slightly lower than that for the 5FPy bound form 0.81±0.05. The

theoretical NOE value at 600 MHz is 0.83 for a completely rigid isotropic protein with global tumbling time of 19 ns. The high NOE values indicate that the protein is quite rigid in both apo and 5FPy bound forms. Though the average NOEs are similar, some regions show differences between the apo and 5FPy bound form. The NOEs of K115 and S116 from the F114 loop, which covers the active site, are both 0.66 in the apo form, significantly smaller than the 0.84 and 0.73 found in the bound form; The average NOE of the C-terminal helix, 150~158, another region covering the active site, is 0.70 ± 0.09 in the apo form, smaller than the 0.82 ± 0.04 found in the 5FPy bound form. It suggests that these two regions undergo motions on the ps to ns timescale in the apo form, but not in the 5FPy bound form.

Model-free dynamics from ^{15}N relaxation

The average S^2 is 0.93 ± 0.05 in the apo form, comparable to 0.91 ± 0.05 in the 5FPy bound form, indicating that the overall protein is quite rigid in both forms. Most of the residues in both forms have S^2 values larger than 0.8 (Figure 3-2). In some regions, the apo form S^2 values appear to be smaller than for the bound form, but the difference is rather small. For example, the average S^2 of the F114 loop (111-117) is 0.87 ± 0.06 in the apo form and 0.90 ± 0.04 in bound form. The C-terminal helix shows no apparent differences for S^2 , but larger τ_e 's were observed in the apo form, consistent with the NOE data (Figure 3-2). There are 41 residues with τ_e larger than 0, among which 18 residues have τ_e larger than 100 ps in the apo form, compared with 37 residues with τ_e larger than 0, where 6 residues have τ_e larger than 100 ps. in the 5FPy bound form. Therefore there are more residues in the apo form undergoing motion in picoseconds. It appears that more residues in the 5FPy bound form have conformational exchange on the

μ s-ms timescale than in the apo form. 17 residues have R_{ex} with the average 3.96 s^{-1} in the apo form, compared to 49 residues with R_{ex} average 2.84 s^{-1} in the 5FPy bound form. Though more residues in 5FPy bound form require an R_{ex} term in the model-free analysis, only 6 residues have R_{ex} greater than 4.0 s^{-1} comparable to 8 residues in the apo form.

Protein dynamics from MD simulation

It appears that yCD is quite stable in the 10 ns simulation with CA root mean square deviation (RMSD) from the crystal structure less than 1.5 \AA in apo, inhibitor 2Py bound and product uracil bound forms (data not shown). The root mean square fluctuations (RMSF) of backbone CAs were calculated for the apo form, as shown in Figure 3-3a. Most of the flexible residues are from regions between α helix and β sheet. It is interesting that $\alpha 4$, $\alpha 5$ and $\alpha 6$ are more flexible than other α helices and the β sheet, which will be discussed later. The pattern of RMSFs is similar to the x-ray B-factors. But residues 72~81 and 111~117 are significantly more flexible in the MD simulation than indicated by the corresponding x-ray B-factors (Figure 3-3a). The flexibility of these two regions is consistent with experimental S^2 values in the apo form (Figure 3-2a). The former region, a small helix between $\alpha 2$ and $\beta 3$, is in contact with the C-terminal helix (150-158) in the adjacent protomer, and R73 forms a salt bridge with E154 (Figure 3-4). The later region, the F114 loop between $\beta 4$ and $\alpha 4$ covering the active site, is quite rigid based on the x-ray B-factors, but rather flexible in the MD simulation. The upper panel of Figure 3-3a shows the order parameters calculated from the apo form simulation, which qualitatively correlates well with the RMSFs. For most of the residues, S^2 is well defined though, for some residues with low S^2 values, the error is quite large. Seven residues, including 59, 74, 75, 81, 82, 116, and 117 have S^2 errors larger than 0.1, which are from

flexible regions, with the exception of 82. All these residues also have quite large τ_e values (~ 1000 ps) with large errors as shown in Figure 3-5. Simulations of 10 ns may not be long enough to determine the order parameter of these residues, which are likely to have longer timescale motion. The average τ_e is 1240 ± 224 ps for residues 72~81, 1139 ± 539 ps for residues 111~117, significantly larger than the average for the whole protein of 396 ± 439 ps. So, the flexibility increase of residues 72~81 and 111~117 in the MD simulation, compared with the X-ray B-factors, comes from the ns timescale motion.

The binding of 2Py doesn't change the protein overall flexibility as shown in Figure 3-3b. It appears that residues 72~81 and 111~117 are slightly more rigid than in the apo form, except for K115, which seems to be rigid in one protomer but flexible in the other. The average RMSF for the former is 0.72 \AA in the 2Py bound form, which is slightly smaller than the 0.85 \AA in the apo form. The average RMSF for the latter is 0.92 \AA in the bound form compared with 1.21 \AA in the apo form. The rigidity increase in these two regions can be confirmed by the calculated order parameters S^2 and τ_e (Figure 3-3b, 5). The average order parameter of residues 72~81 (111~117) is 0.78 (0.64) in the 2Py bound form, which is larger than the 0.65 (0.49) found in the apo form. The average τ_e 's of residues 72~81 and 111~117 are 449 ps and 561 ps in the 2Py bound form, about 2 times smaller than for the apo form. The experimental order parameter also implies a flexibility decrease of the F114 loop (111~117), but the difference in 72~81 is rather small (except 74: S^2 0.79 ± 0.03 in apo, 0.83 ± 0.03 in bound form) when comparing the bound and apo forms (Figure 3-2).

The RMSF of the C-terminal helix displays a sharp decrease in D155 due to the hydrogen bond interaction between the carboxyl group of D155 and 2Py, while in the apo

form no such pattern was observed (Figure 3-3). The order parameter of D155 is 0.86 ± 0.01 in the 5FPy bound form, slightly larger than the 0.80 ± 0.03 in the apo form, while τ_e is $\sim 260 \pm 160$ ps in the 2Py bound form, about 3 times smaller than that in the apo form ($\sim 750 \pm 537$ ps). In both forms, I156, G157 and E158 are quite flexible.

The RMSFs of the uracil bound form is shown in Figure 3-3c, which is similar to the apo and 2Py bound form. Residues 72~81 and the F114 loop (111~117) are also slightly rigidified compared to the apo form. The former has an average RMSF of 0.69 \AA , S^2 of 0.78 and τ_e of 641 ps, while the latter has average RMSF of 0.80 \AA , S^2 of 0.64 and τ_e of 813 ps.

The side-chain dynamics of F114, W152 and I156

It has been observed that the yCD active site is blocked by the side-chains of F114, W152 and I156 in two crystal structures (Ireton et al. 2003; Ko et al. 2003). The dynamics of these three side-chains should be interesting because they may contribute to ligand binding or release. Two side-chain dihedral angles, ω_1 and ω_2 , were defined to study the side-chain motion as shown in Figure 3-6. ω_1 describes the rotation of the side-chain around the $C\alpha-C\beta$ axis, and ω_2 describes the rotation of the side-chain around the $C\beta-C\gamma$ axis. Figure 3-7a shows the ω_1 , ω_2 plot along MD trajectory of the apo form. It appears that there are two conformations, one major conformation with $\omega_1 \sim 190^\circ$ and $\omega_2 \sim 250^\circ$, and one minor conformation with $\omega_1 \sim 210^\circ$ and $\omega_2 \sim 70^\circ$. The 180° difference between the two ω_2 corresponds to a phenol ring flip, which gives the same structure since the phenol ring is axially symmetric. The ω_1 average is $187^\circ \pm 15.6^\circ$, consistent with the corresponding X-ray structure angle $174^\circ \pm 1.4^\circ$ (average over two protomers). Even though the fluctuation appears to be small, ω_1 can span approximately 80° , from 150° to

230° (Figure 3-7a). The ω_2 value of F114 in the apo yCD X-ray structure is $82^\circ \pm 2.9^\circ$ falling into the sampling region of MD simulation. After the binding of the inhibitor, the flip of the phenol ring is more frequent because the two groups of dots (ω_1 and ω_2 snapshot values) are more evenly distributed than in the apo form (Figure 3-7b). The ω_1 average is $184^\circ \pm 13.3^\circ$, close to the apo form with slightly smaller fluctuation. This value is also consistent with the angle from the inhibitor bound form the X-ray structure of $174^\circ \pm 1.4^\circ$. While the binding of the inhibitor seems not to change the conformational states, but only the distribution, the binding of the product introduces a new state with $\omega_1 \sim 290^\circ$ and $\omega_2 \sim 275^\circ$ (Figure 3-7c). The new state shares a similar probability to be visited as the other two states found in the MD simulation. So, the product bound form appears to explore the largest dihedral space of the F114 side-chain.

Figure 3-8 shows the W152 ω_1 , ω_2 dihedral angles of the apo form yCD. Surprisingly, there are also two states; one with $\omega_1 \sim 280^\circ$, $\omega_2 \sim 70^\circ$, another with $\omega_1 \sim 250^\circ$, $\omega_2 \sim 175^\circ$. It is interesting that the X-ray structure W152 values $\omega_1 \sim 165^\circ$, $\omega_2 \sim 235^\circ$ are not included in these two states, but further investigation suggests that the X-ray conformation is quite similar to the second one. The binding of the inhibitor or the product doesn't change the shape of the two regions in Figure 3-8 (data not shown) suggesting that the accessible dihedral space of W152 is similar in all three forms.

The side-chain ω_1 , ω_2 plot for I156 is shown in Figure 3-9, with two major conformations captured in the apo form MD simulation. The first conformation has $\omega_1 \sim 60^\circ$, $\omega_2 \sim 75^\circ$; the second one has $\omega_1 \sim 60^\circ$, $\omega_2 \sim 175^\circ$. The corresponding X-ray structure dihedral angles are $\omega_1 58.4^\circ$, $\omega_2 175.1^\circ$, which fall into the second conformation. The difference between the two conformations mainly comes from ω_2 ,

which is the rotation of C_δ along the C_β - $C_{\gamma 1}$ (there are two C_γ in I156) axis. The binding of the inhibitor introduces the third state with $\omega 1 \sim 60^\circ$, $\omega 2 \sim 280^\circ$, a new rotamer around the C_β - C_γ axis. The binding of product shows a similar pattern of side-chain dihedral plot as the inhibitor bound form, with the third state slightly more sparse.

In summary, it appears that the side-chains of F114, W152 and I156 are quite flexible in the apo, inhibitor bound and product bound forms. The flip of the F114 phenol ring is observed in all three forms, and the binding of the product introduces a new state with $\omega 1 \sim 290^\circ$ and $\omega 2 \sim 275^\circ$ (Figure 3-7c). There are two conformations of the W152 side-chain in all three forms, two conformations of I156 in the apo form, and three conformations of I156 in the inhibitor and product bound form. The representative side-chain conformations of F114, W152 and I156 are plotted in Figure 3-10, where the backbone yCD coordinates are taken from the inhibitor bound form X-ray structure.

The correlation between MD S^2 and RMSF

1). The α helix and β sheet regions

As shown in Figure 3-3, the order parameter correlates well with the RMSF in the apo, 2Py and uracil bound forms, qualitatively. Larger RMSF residues usually have smaller S^2 values. But, since the minimum of S^2 is zero while the maximum of a RMSF can be quite large, this correlation may not be linear for residues with small S^2 . In order to see what the relationship between S^2 and RMSF is, they were plotted in x and y-axis together as shown in Figure 3-11. All α helix and β sheet residues are plotted in Figure 3-11a, while the rest of the protein residues are plotted in Figure 3-11b (156, 157, 158 were included in Figure 3-11b since they don't form a strict α helix in the x-ray structure). All three forms, apo, 2Py and uracil complex, were included in Figure 3-11, and the order

parameters with error larger than 0.1 were removed. As shown in Figure 3-11a, high S^2 usually corresponds to small RMSF, but residues with low S^2 have quite scattered RMSFs. And, it is interesting that α helix residues have more flexibility than β sheet ones for the same order parameter. For example, for a residue with S^2 0.85, the RMSF of an α helix residue can be 0.5 Å to 1.0 Å while a β sheet residue only varies from 0.3 Å to 0.5 Å. Since a β sheet forms the core of yCD, it is not a surprise that the β sheet might be more rigid than the α helix. Some α helices seem to be more flexible than other α helices and the β sheet regions, based on RMSFs, such as $\alpha 4$, $\alpha 5$, and $\alpha 6$ (Figure 3-3). The average RMSFs of $\alpha 4$, $\alpha 5$, $\alpha 6$ (not including 156, 157 and 158) are 0.72 Å compared with 0.41 Å for the other secondary structure regions. But the order parameters are quite similar, with the average 0.85 for the former and 0.87 for the latter. As is well known, an order parameter describes the angular fluctuation of the backbone NH vector, while the backbone nitrogen RMSF describes the translational motion of an N atom that is essentially the same as the backbone CA atom RMSF in yCD (data not shown). If a group of residues undergoes a translational motion without rotating the NH vector, the CA RMSF will be large but the order parameter will be 1.0. In this case, order parameter underestimates the motion of residues, which seems to be the case in $\alpha 4$, $\alpha 5$, $\alpha 6$ for apo, 2Py bound and uracil bound forms. In order to identify whether this is true, the overall translational motion of helices was investigated for the apo form. The center of mass for the backbone heavy atoms (CA, CO, N, O) of each helix was calculated. The fluctuation of the center of mass was calculated as shown in Table 8-1. The RMSF of $\alpha 1$, $\alpha 2$, $\alpha 3$ is ~ 0.3 Å while the RMSF of $\alpha 4$, $\alpha 5$, $\alpha 6$ is two times larger ~ 0.6 Å. It proves that the

fluctuation difference between $\alpha 1$, $\alpha 2$, $\alpha 3$ and $\alpha 4$, $\alpha 5$, $\alpha 6$ comes from the overall motion instead of local motion of individual residues in helices.

Two approximate boundaries of the α helix and β sheet can be drawn in the RMSF- S^2 plot. The absolute values of the slopes for the two boundaries are 0.91 Å and 3.2 Å in the β sheet and 1.7 Å and 9.2 Å in the α helix. The smaller slope in the β sheet region suggests that generally the order parameter is more sensitive to the motion in the β sheet but less sensitive in an α helix.

2). The other regions

Figure 3-11b shows the RMSF S^2 plot for regions other than the α helices and β sheet. A cone like graph with rigid residues on the lower right side and flexible residues on the upper left side can be seen, similar to the α helix and β sheet regions. There is no one-to-one relationship between S^2 and RMSF evident. For example, a residue with S^2 0.80 can have an RMSF range from 0.4 Å to 1.4 Å. On the other hand, a residue with RMSF 0.8 Å can have an order parameter span from 0.5 to 0.85. As discussed above, S^2 and RMSF describe different kinds of motion. Considering a system composed only of one free NH vector in space, the translational motion of N is completely independent of the rotation of the vector, and that means the RMSF and S^2 are independent. But, in protein system, due to the sp^2 hybridization of the backbone amide N, CA-CO-N-H has to maintain a trigonal planar geometry, which means that the angular coordinate of the NH vector depends on the orientation of the peptide plane (For the simplicity of the argument, the bending of NH away from trigonal planar geometry is not considered). The orientation of one peptide plane is usually restrained by its position in a structure. The relationship between orientation and position might be very complicated and dependent

on the local environment as well as the whole protein structure. Larger position fluctuations will give more possibilities for different orientation fluctuations, which is consistent with the cone-like plot of order parameter and RMSF for yCD. The absolute value of the slope of the two boundaries is 1.3 Å and 5.7 Å, respectively. The large range in between these two boundaries will make the interpretation of an order parameter complicated. For example, the residue with S^2 0.8 could be more flexible than the residue with S^2 0.5, as shown in Figure 3-11b. But, on the other hand, as discussed above, residues 72~81 and 111~117 show an increase of rigidity based on both RMSF and S^2 values, due to the binding of 2Py. So, for the same region in different forms (apo, bound etc.), the flexibility change can be accurately detected by the order parameter.

Discussion

The crystal structures of yCD in the apo and inhibitor 2Py bound forms are essentially the same, with the active site completely buried. Our NMR relaxation study shows that both forms are quite rigid. The F114 loop, which covers the active site, appears to be more flexible in the apo form, based on order parameters. The C-terminal helix undergoes more ps-ns timescale motion in the apo than in the bound form based on NOE experiments. The flexibility change is consistent with our HD exchange results that showed that the F114 loop and C-terminal helix are more flexible in the apo form, even though these two kinds of experiments study motions on completely different timescales (manuscript in preparation). The relaxation measurement mainly studies the motion in the ps-ns time scale, while an HD exchange measurement studies motion on a second to minute timescale.

The backbone flexibility change in the F114 loop was also confirmed by an MD simulation study. The average RMSF over CA atoms in this region is about 0.3 Å greater in the apo than in the inhibitor bound form. The analysis of the order parameter and correlation time τ_c from simulation provides a consistent conclusion. The average S^2 of backbone NH vectors in the F114 loop decreases ~0.15 while τ_c decreases about 2 fold. Another region, residues 72~81, (a loop) that interacts with the C-terminal helix from the adjacent protomer also shows a decrease of flexibility due to the binding of the inhibitor. The binding of the product uracil displays similar backbone flexibility changes in these two regions.

It has been proposed that the motion of the F114 loop and C-terminal helix is important for ligand release. In the X-ray structure of yCD, three residue side-chains F114, W152 and I156 block the active site. It appears that these side-chains are quite flexible in the apo, 2Py bound and uracil bound forms. The flip of the phenol ring of F114 was observed in all three forms. A new conformation appears after the binding of uracil but not 2Py. The large span for ω_1 of ~80° in the apo form can also rotate the F114 side-chain away from the crystal conformation. There are two conformations of the W152 side-chain in all three forms, and two (three) conformations of I156 in apo (bound) form. Together with the motion of the backbone of the F114 loop and the C-terminal helix, the ligand release path in between these two regions seems to be feasible.

Backbone order parameter S^2 has been widely used in NMR experiments to describe the flexibility of residues. But, since S^2 is related to the angular flexibility of the NH vector instead of a direct reflection of the RMSF of backbone N (or CA), how S^2 and RMSF are correlated is of great interest. It appears that, qualitatively, they correlate quite

well, with smaller S^2 usually corresponding to larger RMSF. But there is no quantitative correlation between them. For example, the average RMSF of $\alpha 4$, $\alpha 5$, $\alpha 6$ is about double that of the other secondary structure regions, but the average order parameter is the same. Further investigation suggests that $\alpha 4$, $\alpha 5$, $\alpha 6$ undergo two times larger center of mass translation motion than the other helices. The order parameter fails to describe the translation motion of the whole helix, since an individual NH vector doesn't change its orientation in this process. It is also interesting that the order parameter seems to be more sensitive to the motion in the β sheet than in an α helix. In the regions other than the α helix and β sheet, the plot of RMSF versus S^2 has a cone shape, with the tip corresponding to large order parameter and small RMSF. A residue with small S^2 can have quite different RMSF values, while a residue with large RMSF can have quite different S^2 . On the other hand, in the different forms of yCD that we studied, the increase of RMSF corresponds to the decrease of order parameter, such as for the F114 loop. So, the order parameter appears to be more accurate to describe the backbone flexibility difference of the same region in different protein forms, but less accurate to describe the flexibility of the different regions in the same protein form.

References

- Abragam, A. (1961). The principles of nuclear magnetism.
- Akaike, H. (1973). "In Information Theory and an Extension of the Maximum Likelihood Principle." Proceedings of the 2nd International Symposium on Information Theory: 267-281.
- Case, D. A. (2002). "Molecular dynamics and NMR spin relaxation in proteins." Accounts of Chemical Research **35**(6): 325-331.
- Chen, J. H., C. L. Brooks, et al. (2004). "Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation." Journal of Biomolecular Nmr **29**(3): 243-257.
- d'Auvergne, E. J. and P. R. Gooley (2003). "The use of model selection in the model-free analysis of protein dynamics." Journal of Biomolecular Nmr **25**(1): 25-39.
- Delaglio, F., S. Grzesiek, et al. (1995). "Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes." Journal of Biomolecular Nmr **6**(3): 277-293.
- Ding, Z. F., G. Lee, et al. (2005). "PhosphoThr peptide binding globally rigidifies much of the FHA domain from Arabidopsis receptor kinase-associated protein phosphatase." Biochemistry **44**(30): 10119-10134.
- Dosset, P., J. C. Hus, et al. (2000). "Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data." Journal of Biomolecular Nmr **16**(1): 23-28.
- Eisenmann, A., P. Neudecker, et al. (2004). "Treatment of peak intensity uncertainties in NMR relaxation data analysis can lead to severe artifacts." Monatshefte Fur Chemie **135**(9): 1089-1099.
- Farrow, N. A., R. Muhandiram, et al. (1994). "Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology-2 Domain Studied by N-15 Nmr Relaxation." Biochemistry **33**(19): 5984-6003.
- Fischer, M. W. F., A. Majumdar, et al. (1998). "Protein NMR relaxation: theory, applications and outlook." Progress in Nuclear Magnetic Resonance Spectroscopy **33**(4): 207-272.
- Ireton, G. C., M. E. Black, et al. (2003). "The 1.14 Å crystal structure of yeast cytosine deaminase evolution of nucleotide salvage enzymes and implications for genetic chemotherapy." Structure **11**: 961-972.

- Ireton, G. C., M. E. Black, et al. (2003). "The 1.14 angstrom crystal structure of yeast cytosine deaminase: Evolution of nucleotide salvage enzymes and implications for genetic chemotherapy." Structure **11**(8): 961-972.
- Ishima, R. and D. A. Torchia (2000). "Protein dynamics from NMR." Nature Structural Biology **7**(9): 740-743.
- Johnson, B. A. and R. A. Blevins (1994). "Nmr View - a Computer-Program for the Visualization and Analysis of Nmr Data." Journal of Biomolecular Nmr **4**(5): 603-614.
- Karplus, M. and J. A. McCammon (2002). "Molecular dynamics simulations of biomolecules." Nature Structural Biology **9**(9): 646-652.
- Ko, T. P., J. J. Lin, et al. (2003). "Crystal structure of yeast cytosine deaminase - Insights into enzyme mechanism and evolution." Journal of Biological Chemistry **278**(21): 19111-19117.
- Ko, T.-P., J.-J. Lin, et al. (2003). "Crystal structure of yeast cytosine deaminase. Insights into enzyme mechanism and evolution." J. Biol. Chem. **278**: 19111-19117.
- Korzhnev, D. M., M. Billeter, et al. (2001). "NMR studies of Brownian tumbling and internal motions in proteins." Progress in Nuclear Magnetic Resonance Spectroscopy **38**(3): 197-266.
- Lipari, G. and A. Szabo (1982). "Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity." Journal of the American Chemical Society **104**(17): 4546-4559.
- Lipari, G. and A. Szabo (1982). "Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results." Journal of the American Chemical Society **104**(17): 4559-4570.
- Mandel, A. M., M. Akke, et al. (1995). "Backbone Dynamics of Escherichia-Coli Ribonuclease Hi - Correlations with Structure and Function in an Active Enzyme." Journal of Molecular Biology **246**(1): 144-163.
- Nicholson, L. K., L. E. Kay, et al. (1992). "Dynamics of Methyl-Groups in Proteins as Studied by Proton-Detected C-13 Nmr-Spectroscopy - Application to the Leucine Residues of Staphylococcal Nuclease." Biochemistry **31**(23): 5253-5263.
- Palmer, A. G. (2001). "NMR probes of molecular dynamics: Overview and comparison with other techniques." Annual Review of Biophysics and Biomolecular Structure **30**: 129-155.
- Palmer, A. G., M. Rance, et al. (1991). "Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance C-12 Heteronuclear Nmr-Spectroscopy." Journal of the American Chemical Society **113**(12): 4371-4380.

Viles, J. H., B. M. Duggan, et al. (2001). "Potential bias in NMR relaxation data introduced by peak intensity analysis and curve fitting methods." Journal of Biomolecular Nmr **21**(1): 1-9.

Yao, L. S., Y. Li, et al. (2005). "Product release is rate-limiting in the activation of the prodrug 5-fluorocytosine by yeast cytosine deaminase." Biochemistry **44**(15): 5940-5947.

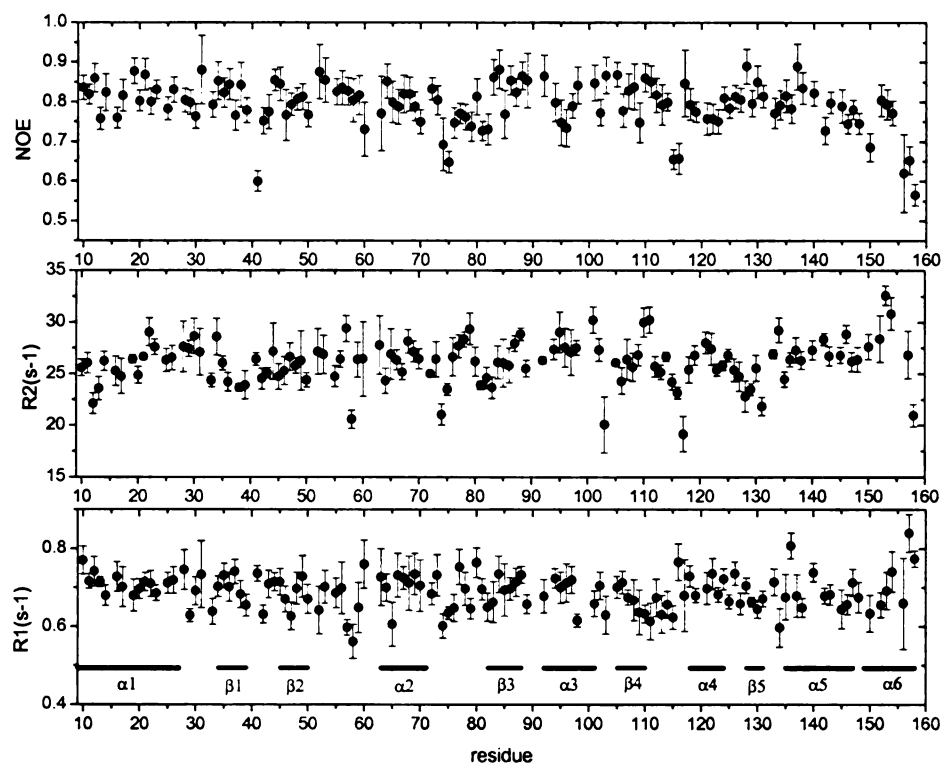
Yao, L. S., S. Sklenak, et al. (2005). "A molecular dynamics exploration of the catalytic mechanism of yeast cytosine deaminase." Journal of Physical Chemistry B **109**(15): 7500-7510.

Yao, L. S., H. G. Yan, et al. (2006). "A molecular dynamics study of the ligand release path in yeast Cytosine Deaminase (submitted)."

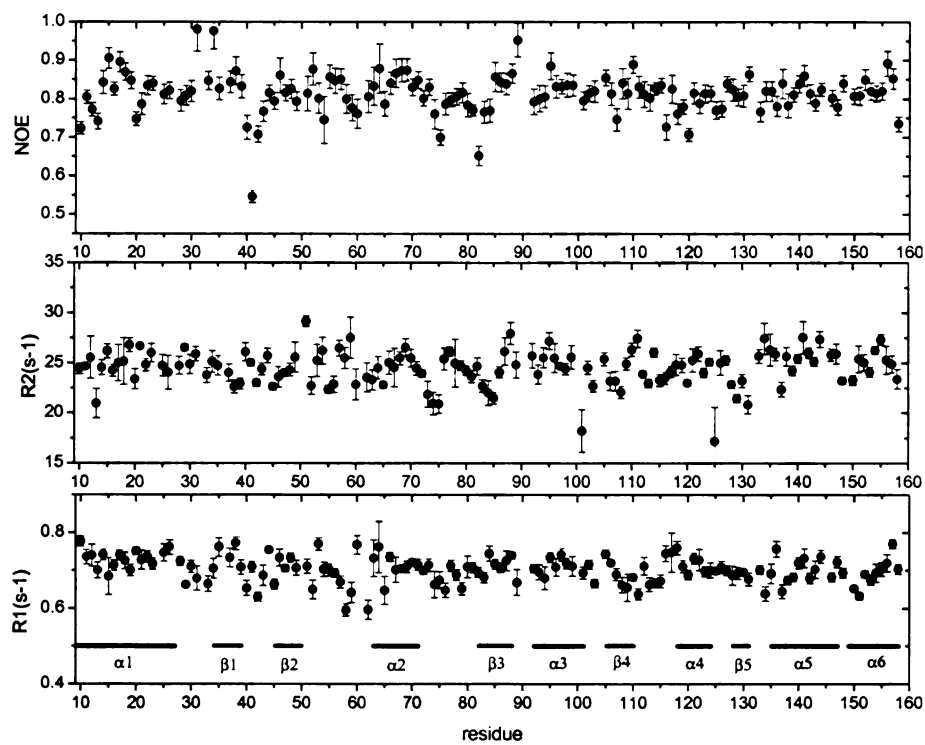
Appendices

Table 3-1 The RMSF of center of mass of various helixes in the apo form

	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$	$\alpha 5$	$\alpha 6$
RMSF(Å)	0.28±0.01	0.30±0.01	0.27±0	0.64±0.02	0.57±0.05	0.59±0.03

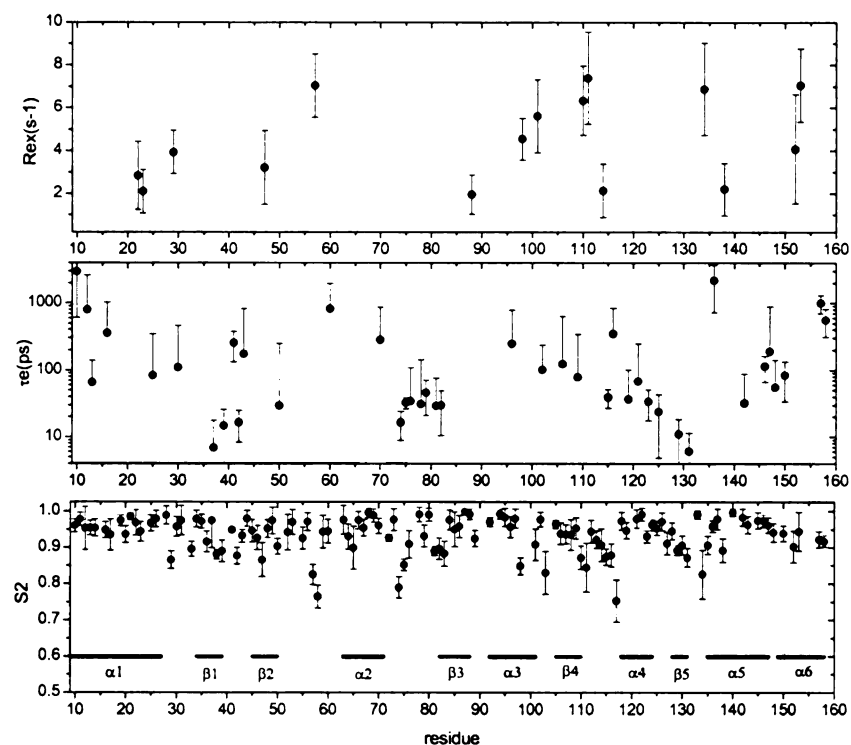


(a)

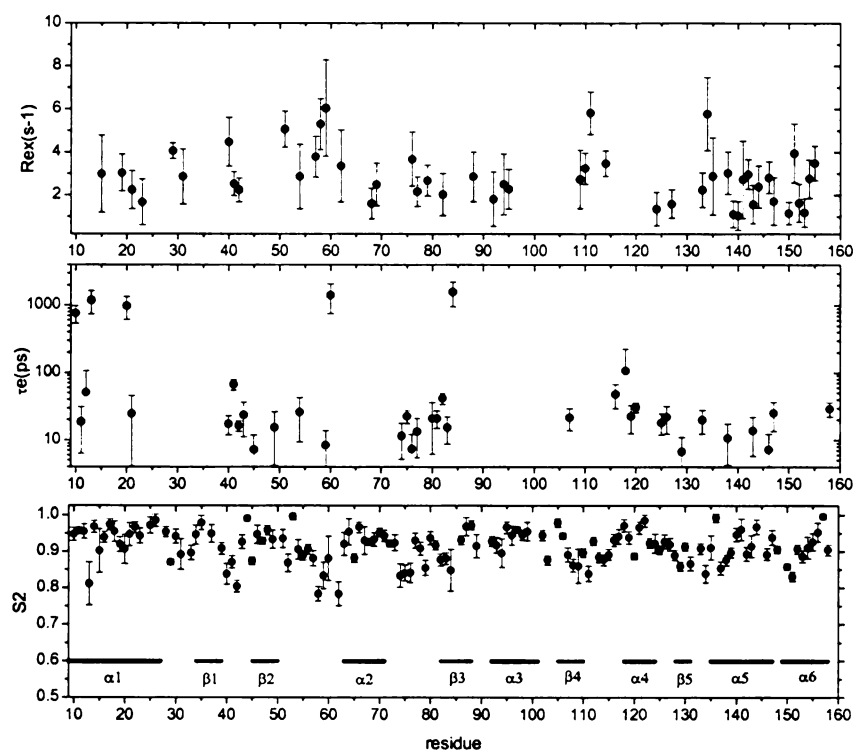


(b)

Figure 3-1 ^{15}N relaxation data of yCD in the (a). apo, (b). inhibitor 5FPy bound form measured at 600 MHz and plotted vs residue.



(a)



(b)

Figure 3-2 Model-free dynamics results for the (a) apo form (b) inhibitor 5FPy bound form yCD from NMR ^{15}N relaxation data.

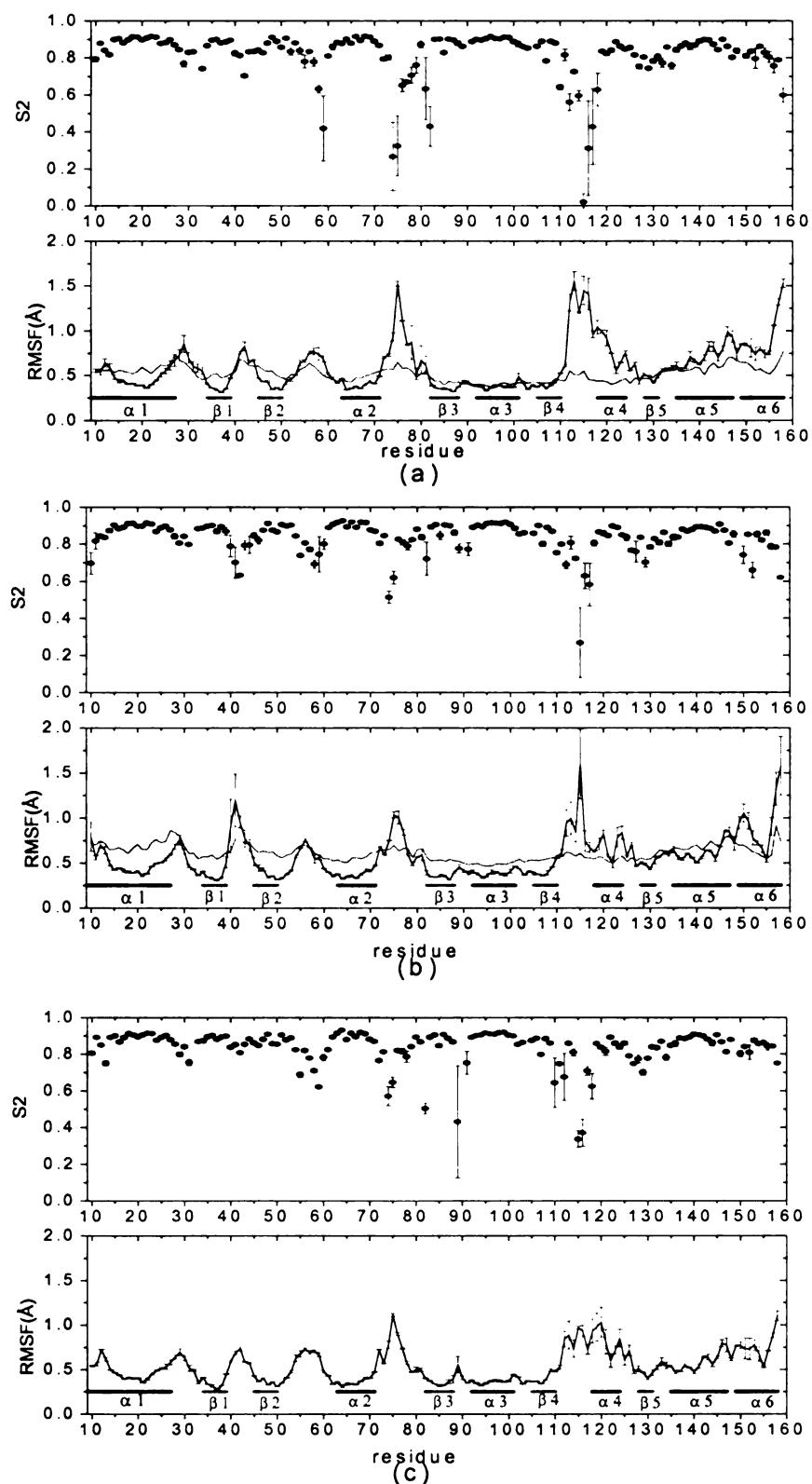


Figure 3-3 RMSF and order parameter of the (a), apo form. (b), inhibitor 2Py bound form. (c). product bound form. The x-ray B-factors of apo and 2Py bound form were converted and plotted in red.

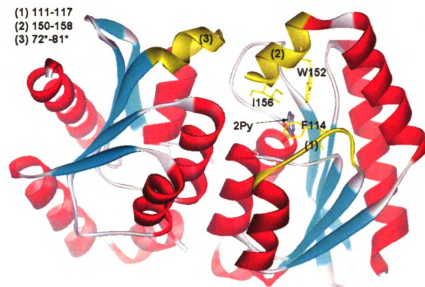


Figure 3-4 X-ray structure of yCD complexed with 2Py. Three regions 111-117, 150-158 and 72*-81* (from adjacent protomer) are highlighted in gold. Three residues F114, W152 and I156 which blocks the active site were also shown.

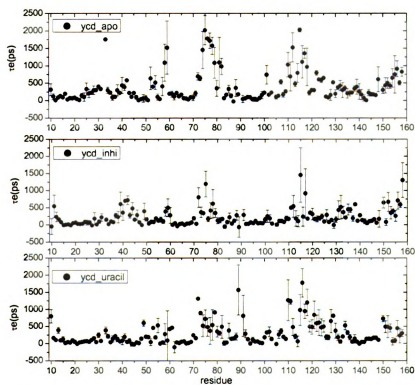


Figure 3-5 τ_e value calculated from MD simulation for the apo form, inhibitor 2Py complex and product complex.

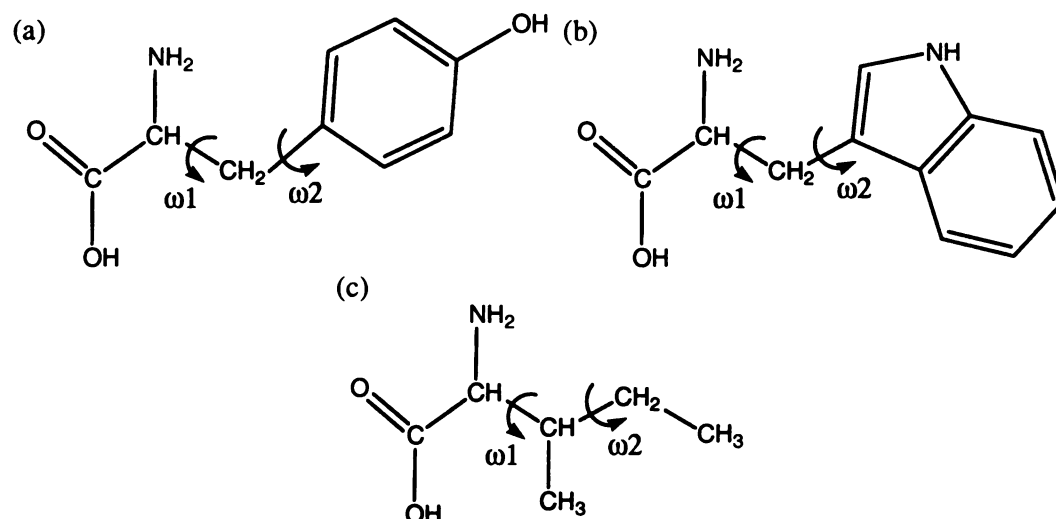


Figure 3-6 two sidechain dihedral angles defined in MD data analysis for (a). F114, (b). W152, (c). I156.

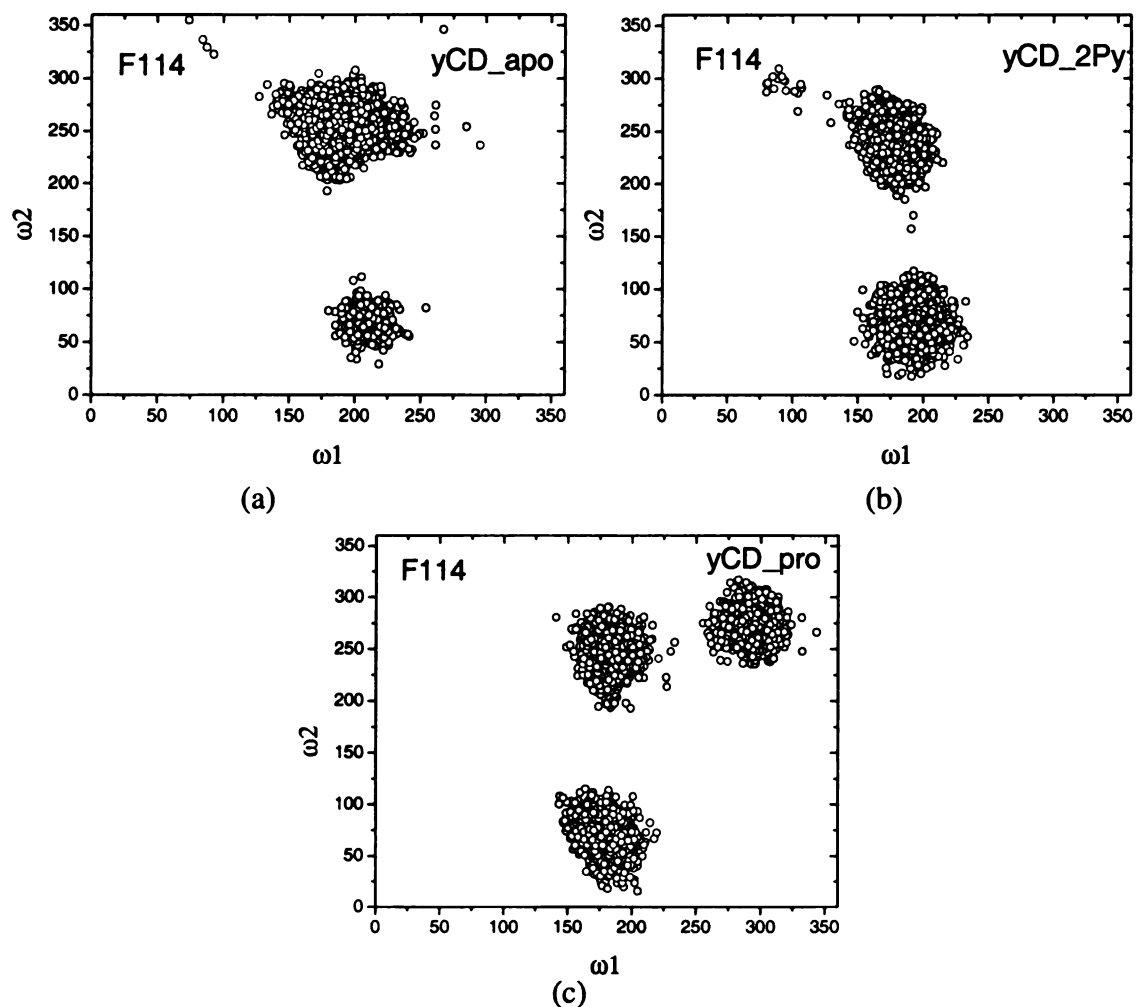


Figure 3-7 F114 sidechain dihedral angle ω_1 and ω_2 of the (a). apo form, (b). inhibitor 2Py bound form, (c). product uracil bound form in 10 ns MD simulation.

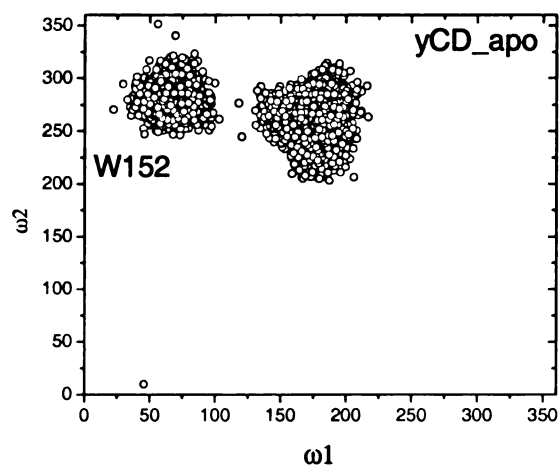


Figure 3-8 The sidechain dihedral angle ω_1 and ω_2 of W156 in apo form.

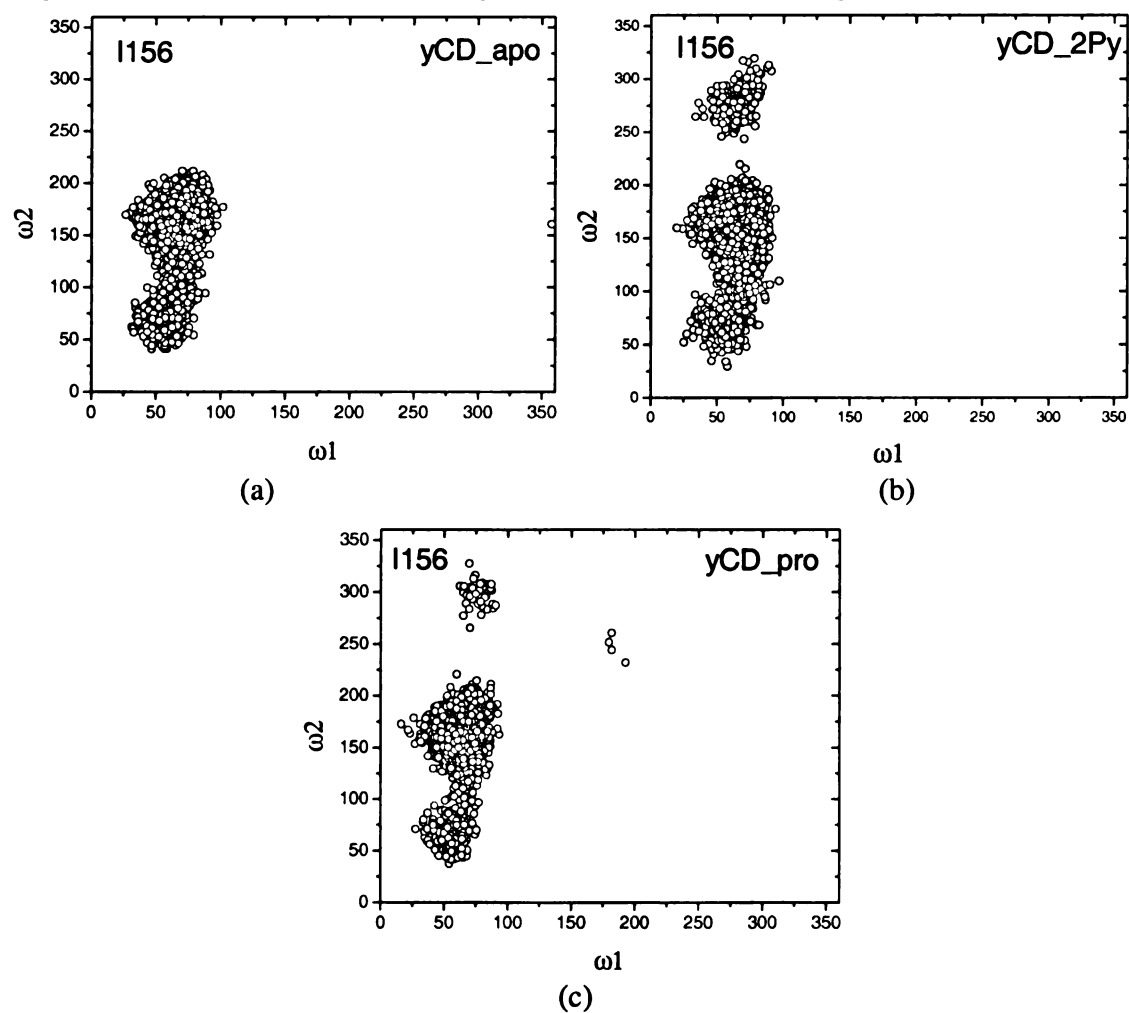


Figure 3-9 I156 sidechain dihedral angle ω_1 and ω_2 of the (a). apo form, (b). inhibitor 2Py bound form, (c). product uracil bound form in 10 ns MD simulation.

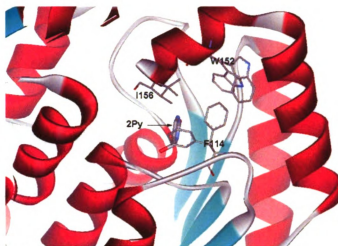


Figure 3-10 The sidechain conformations of F114, W152 and I156.

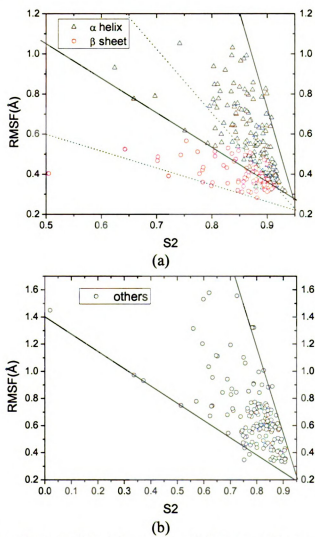


Figure 3-11 RMSF vs S2 plot of the (a) α helix and β sheet residues, (b) other region residues.

Chapter 4

The Dynamics Changes of YCD along Reaction Cycle Revealed by HD exchange

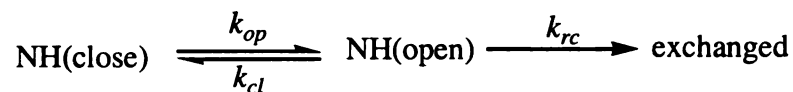
Introduction

Yeast cytosine deaminase (yCD) catalyzes the deamination of cytosine to uracil while it can also catalyze the deamination of the prodrug 5-fluorocytosine (5-FC) to the anticancer drug 5-fluorouracil (5-FU). One reaction mechanism was proposed (Sklenak et al. 2004) where cytosine (or 5-FC) is deaminated to uracil (or 5-FU) through a tetrahedral intermediate (Figure 4-1). High-resolution structures of yCD in the apo form (Ireton et al. 2003) and in complex with the inhibitor 2-pyrimidinone (2Py) (Ireton et al. 2003; Ko et al. 2003), a potent intermediate analog, were published. It is a surprise that the structure of apo yCD is essentially the same as that of the 2Py complex with the active site buried. A series of molecular dynamics (MD) simulations were performed along the catalytic cycle, which also show a buried active site (Yao et al. 2005). In order to bind the reactant or release the product, certain residues have to move. One possible ligand release path was found by using steered MD method (ms in preparation). The ligand releases along the path in between C-terminal helix (150-158) and the F114 loop (111-117). Certain residues in these two regions are crucial for ligand release including F114 and I156. Recent measurements have shown that product release is the rate-limiting step during the

activation of 5-FC (Yao et al. 2005), and 5-FU is in slow exchange with its bound form by the off-rate $\sim 13 \text{ s}^{-1}$. It is interesting that the binding of 5-FU is quite weak with a K_d only about 25 mM. It was proposed that the slow release was caused by high energy barrier during the bond cleavage between 5-FU O4 and catalytic Zn by using ONIOM computational method (Sklenak et al. 2004). But our experimental results suggest that the proposed mechanism occurs at a much slower rate than product release (unpublished data). An alternative mechanism of the Zn-O4 bond cleavage was later on demonstrated by combining MD and ONIOM, which shows that the barrier is rather small $\sim 3 \text{ kcal/mol}$. So it is more likely that the barrier comes from the rearrangement of the residues of the active site and along the product release path since the active site is completely buried. Understanding the product release or reactant binding process can certainly help us elucidate protein dynamics and its function, and improve this enzyme's catalytic efficiency, which is the essential goal of studying this enzyme.

Exchange with solvent hydrogens of backbone NH monitors the motion of protein at each of numerous NH sites along the protein backbone. A general two-process model (Loh et al. 1996; Qian and Chan 1999) was proposed to describe hydrogen exchange including: "local opening", a small-scale conformational fluctuation in the native state and "global unfolding", a large-scale conformational change from native state to unfolded state. The most slowly exchanging NH groups in native proteins appear to exchange through global unfolding, while the more rapidly exchanging NH groups exchange through local opening (Arrington and Robertson 2000). In this paper we focus on NHs where exchange appears to be dominated by local openings, and the protein β sheet core shows no hydrogen exchanges on the experimental time scale (see results below).

Proton exchange from protein amides to solvent is generally described by a two-state mechanism:



The apparent rate of exchange is

$$k_{ex} = (k_{op}k_{rc}) / (k_{cl} + k_{rc} + k_{op}) \quad (1)$$

Futhermore, k_{op} is generally much smaller than k_{cl} , because the protein is in the native state. Equation 1 can be simplified to

$$k_{ex} = (k_{op}k_{rc}) / (k_{cl} + k_{rc}) \quad (2)$$

Equation 2 can be further simplified by two limiting cases. The first one is called EX1 exchange where $k_{cl} \ll k_{rc}$, equation 1 can therefore be further reduced to

$$k_{ex} = k_{op} \quad (3)$$

which permits direct determination of the rate of opening. The second one is EX2 exchange where $k_{cl} \gg k_{rc}$, then equation 1 can be written as

$$k_{ex} = (k_{op}k_{rc}) / k_{cl} = k_{rc}K_{op} \quad (4)$$

where K_{op} is the equilibrium constant for the opening reaction. Therefore, an equilibrium constant can be measured from EX2. k_{rc} can be calculated by the method of Bai et al (Bai et al. 1993).

In this work we performed an HD exchange study on the apo enzyme, 5FPy (5-fluro-2-pyrimidinone, intermediate analog) complex and 5FU (5-fluro-uracil, product) complex at $PH \geq 7.0$ where exchange is mainly catalyzed by base (Arrington and Robertson 2000), the observed exchange rates are strongly pH dependent under EX2,

whereas under EX1 conditions the observed exchange rate is independent of pH (assuming k_{op} and k_{cl} are pH independent). Thus, by varying pH, EX1 and EX2 can be distinguished.

Materials and Methods

Construction, expression and purification of yCD and ^{15}N labeled yCD has been described (Yao et al. 2005). For ^2H , ^{15}N , ^{13}C protein, expression strain BL21(DE3)pLysS containing pET17b-yCD was grown in M9 medium (70% D_2O) with $^{15}\text{NH}_4\text{Cl}$ and ^{13}C -glucose as the sole nitrogen and carbon sources, respectively. The procedure for expression and purification is the same as that of the unlabeled protein.

NMR assignments of 5FPy bound form were made using the spectra of 1.8 mM ^2H , ^{13}C and ^{15}N triple labeled yCD or ^{15}N labeled yCD in 100 mM potassium phosphate, 100 μM NaN_3 , 20 μM DSS, pH 7.0, made with 93% H_2O /7% D_2O . Sequential assignments of 5FPy bound form were based on NOESY-HSQC spectrum and confirmed by the HNCACB experiment. Sequential assignments of apo and 5FU bound were based on NOESY-HSQC spectrum. Mixing time 70 ms were used for all three NOESY-HSQC spectra (^{15}N single labeled sample).

HD exchange samples were prepared by dissolving ^{15}N labeled yCD with or without ligand into pre-adjusted pH 100 mM potassium phosphate buffer. The concentration of 5FPy and 5FU is 80 mM and 20 mM respectively. Protein solutions were then lyophilized. After that, protein samples were added to D_2O . The final concentrations of yCD are about 1.5 mM. Sample pH was measured after experiments and the value reported in this paper was not corrected for isotope effect.

HD exchange rates were measured for individual NH groups by sensitivity enhanced ^1H - ^{15}N HSQC experiments performed on a 600 MHz Varian Inova spectrometer at 25 °C. Experimental dead time (time between the beginning of dissolving yCD in D_2O and the start of HSQC acquisition) is ~5 min. For the first 2 hours, HSQC spectra were acquired continuously. For the next 3 hours, the interval between two consecutive HSQC spectra was elongated to 30 min. Thereafter, the time interval is set larger to 1 hour and then 2 hours. The experimental time for apo, 5FU and 5FPy bound form are ~1 day, 3 days and 3 days respectively. Longer experimental time seemingly didn't increase the number of exchanged residues. The exchange experiments were performed for (1). apo form at pH 7.16 and 8.05. (2). 5FPy bound form at pH 7.10, 7.62 and 8.50. (3). 5FU bound form at pH 7.10 and 7.50. The spectrum widths were 9476 and 1900 Hz for ^1H and ^{15}N dimensions, respectively; 32 (t_1 , ^{15}N dimension) and 1946 (t_2 , ^1H dimension) complex data points were recorded for each spectrum. The number of transients was 4 for each FID with a 1.0 s delay between transients. It takes ~ 5 min for each spectrum.

The NMR data were processed with NMRPipe (Delaglio et al. 1995). The peak heights were measured by using NMRView (Johnson and Blevins 1994). A three parameter exponential equation were used for data fitting:

$$I = a \exp(-t/T_{ex}) + c \quad (6)$$

where I is the peak height, $T_{ex} = 1/k_{ex}$ is the apparent exchange time constant, a and c the two parameters. Origin was used for data fitting and the error of T_{ex} was estimated based on quality of fitting.

Two nanosecond molecular dynamics (MD) of apo yCD was performed with explicit solvent water molecules. The details of parameterization of the Zn bound complex and MD simulation are described elsewhere (Yao et al. 2005). The first 500 ps were discarded as the equilibration period and the later 1.5 ns were used for data analysis in this paper. The solvent accessibility analysis and hydrogen bond analysis were performed by using PTRAJ module in Amber 7.0 program (Pearlman et al. 1995). We assign hydrogen bonds when the distance of two heavy atoms (O or N) is less than 4.0 Å and the angle (heavy atom-hydrogen-heavy atom) is greater than 120°.

Results

Chemical shift perturbation

Chemical shift difference was calculated for each assigned NH in apo and 5FPy bound form by equation:

$$\delta_{NH} = |\delta_H| + |\delta_N| / 4.77$$

4.77 is the spectrum width ratio between nitrogen and proton dimension. 145 backbone NHs were assigned (all Non-proline residues except K9) for 5FPy bound form. K9 is the N-terminal residue in the protein. The large flexibility of the N-terminal might cause the disappearance of the K9 NH signal in backbone HSQC spectrum (Yao et al. 2005). 142 residues were assigned for the apo form. The unassigned residues were K9, N51, H62 and C91. N51 is from one loop. H62 and C91 are from the active site, both coordinated to Zn. The chemical shift differences due to the binding of 5FPy were plotted in Figure 4-2a. The average difference is 0.13 ± 0.16 . Most residues show little perturbation as shown. Significant differences are seen for G35, E64, L88, M93, Y101, K115 and I156 (with the

difference more than the 2 times standard deviation from average value) as shown in Figure 4-3a. Most of them are from regions around the active site.

139 residues were assigned for the 5FU bound form. The unaassigned residues were K9, V31, H62, S66, T67, and C91. The main reason for unassigned residues is probably the lack of NOE in NOESY-HSQC spectrum. H62, S66, T67 and C91 are from active site, which might be more flexible in the apo and 5FU bound form than in 5FPy bound form. The chemical shift changes between the apo and 5FU complex forms are shown in Figure 4-2b and 3b. The average change is 0.14 ± 0.29 similar to that of the 5FPy binding, but with larger fluctuations. The major changes are from regions around the active site (Figure 4-3b). The residues with significant chemical shift difference include G63, E64, T95 and K115.

HD exchange of the apo form

Non-exchange residues and water accessibility

The yCD monomer contains a central β sheet ($\beta 1 \sim \beta 5$, parallel to each other) with two α helices ($\alpha 1$ and $\alpha 5$) on one side and four α helices ($\alpha 2 \sim \alpha 4$ and $\alpha 6$) on the other side. In the apo form, the regions without exchange in the experimental period mainly include (Figure 4-4a): (1) residues 18~25 which is part of α helix 1 (11~27); (2) residues 33~38 from β strand 1 (34~39); (3) residues 46~49 from β strand 2 (45~50); (4) residues 82~89 from β strand 3 (82~88); (5) residues 94~97 which is part of α helix 3 (92~101); (6) residues 101~110 which covers part of α helix 3 and the whole β strand 4 (105~110); (7) residues 141~146 which is part of α helix 5 (135~147). There are totally 42 residues with no significant HD exchanges (Table 4-1) during experimental time. The strong hydrogen bond network among β strands makes the HD exchange hard to occur in these

regions. Helix 3 sits in the interface between two monomers which might prohibit the HD exchanges because of the hydrophobicity of the interface.

Figure 4-5 shows average number of water molecules within 3.5 Å from amide N in the 1.5 ns MD simulation of the apo form. It is clear that most non-exchange residues have little water accessibility. This might explain why some parts of α helices 1, 3, and 5 have no exchanges but not other parts or other helices. Interestingly, β 2 (45~50) and β 5 (128~131) both have water molecules around (with 0.42 water molecules for the former and 0.95 for the latter on average) but the former doesn't have HD exchange while the latter has a relatively fast exchange. Apparently only having water accessibility is not sufficient for HD exchanges. On the other hand, residue 51~54 (a loop) and 60~69 (α 2) have little water accessibility, but both regions have HD exchanges occurring. These two regions have to go through a relatively large conformation changes to exchange with water, or water has to penetrate into these two regions in a time scale longer than the ns MD simulation.

Fast exchange residues

There are 38 (Table 4-1) fast exchange residues that are spread among the whole protein (Figure 4-1a). Most of them come from loops or turn regions including 41~43, 57~59, 62~63, 73~78, 113~116, and 131~134. Some of them come from helical regions including 10~14 (α 1), 117~125 (α 4), 135~136 (α 5) and 150~158 (α 6). One comes from a β strand 129~131 (β 5). Figure 4-6a shows percentage of hydrogen bond formation between amide (as donor) and other residues (as acceptor) in the protein. The average percentage is 104.8 ± 49.7 , indicating that quite a lot of amides (93 out of 145) form hydrogen bonds with more than one acceptor and many of the amides are well protected.

However there are 31 residues with less than 80% hydrogen bond formation, among which 19 residues have fast exchanges. Apparently, the amides with weak hydrogen bond formation are easier to exchange with solvent water. But on the other hand, another 19 fast exchange residues have hydrogen bond formation larger than 80%. Though these residues seem to be protected well, the competition from solvent might make these residues exchange quickly.

Figure 4-6b shows the percentages of H-bond formation between backbone amide NH (as donor) and solvent water molecules (as acceptor) in the MD simulation. Only 60 residues form such a hydrogen bond with the average percentage 29.4 ± 38.1 , much smaller than hydrogen bond formed in between residues. For the base-catalyzed HD exchange, the H-bond between amide and water can assist the exchange by transferring amide proton to water (or OH^-). It is not a surprise that 29 out of 38 fast exchange residues have H-bond interaction with solvent (Figure 4-6b). But not all the amides hydrogen bonded to solvent have fast exchanges. On the other hand, except W10, the other 8 fast exchange residues are not solvent accessible including K13, N39, F54, G63, E64, I98, W152 and I156. Among these, G63 and E64 are from the active site; W152 and I156 are from the C-terminal helix. Conformational fluctuations might be responsible for the fast hydrogen exchanges of these residues.

Slow exchange residues

40 residues with a slow exchange rate could be measured with average lifetimes of 270 min (Table 4-1) and most of these residues are water accessible. However, residues 51~54 (close to a small α helix, 53~55) as well as 63~69 (part of α helix 2, 63~71) have no (or few) water accessibility during the simulation, but exchanges were observed in

both regions. The latter region is close to the active site with H62 coordinated to Zn, G63 and E64 directly h-bonded to 5FPy in the complex form. The exchange data show that G63, E64 have fast exchanges, L68 has no exchange. The other 4 residues have average exchange times of 386 min. The exchange in this region reflects that the active site is flexible in the apo form.

The yCD crystal structure shows that the active site is covered by three regions (Ireton et al. 2003), 53~61 (loop 1, between β 2 and α 2), 111~117 (F114 loop) and 150~158 (α 6) (Figure 4-7). The average exchange time is 36.9 min for region 53~61 (including Q55, K56, S58, T60. Residues F54, G57, A59 have fast exchanges while F53 has slow exchange and L61 is overlapped). The average exchange time is 256 min for 111~117 (including V112, N113, and F114. Residues K115, S116 and K117 have fast exchanges; N111 is overlapped). The average exchange time for W152 and E154 is 70 min (D155 has no exchange), while N150, D151, F153, I156, G157 and E158 have fast exchanges. The exchange data indicate these three regions are flexible. The flexibility around these regions might be important for ligand binding and release. Figure 4-8a shows the natural logarithm of protection factor verse residues, which gives us a direct view of the flexibilities of different regions. The blue color represents residues with no hydrogen exchange, mainly from the β sheet core. The red one represents residues with fast exchanges (exchange time less than 5 min), from various regions mainly loops, turns and part of some helices. All the colors in between represent residues with slow exchange (from ice blue to orange with the decrease of protection factor).

The binding of 5FPy

After the binding of the intermediate analog 5FPy, the HD exchange becomes much slower in various regions (Figure 4-4b). First, there are 70 residues which have no exchanges compared with 42 in the apo form. All the non-exchange regions in the apo form are maintained and/or extended, which include residue 17~25, 30~39, 46~55, 82~91, 93~110 (except 106) and 141~146. Besides these regions, residues 60~62, 65~71 and part of 150~158 have no exchanges. The average exchange time for all measurable residues is 662 min compared with 270 min in the apo form. Figure 4-8a shows the exchange time changes due to the binding of 5FPy, "▲" represents residues with exchange time scale from fast to slow or from slow to non-exchange or even from fast to non-exchange. HD exchange rates can be measured for 28 residues in both the apo and 5FPy bound form, with the average exchange time in 5FPy bound form 8.7 fold larger than that in the apo form. On the other hand this increase is not uniform, since 14 of these residues only have less than 2 times exchange time increase in 5FPy bound form, suggesting that some regions are rigidified significantly.

The largest measurable exchange difference comes from 72~80 with the increase of exchange time about 44 fold (Figure 4-9a, including G72 (50 fold), Y79 (41 fold) and K80 (42 fold); L74 and G76 have no exchange in 5FPy bound form compared to the fast exchange in apo form.). From the structure, we find that this region is in close contact with loop 1 and the C-terminal helix in the adjacent subunit. Another region with significant exchange time change is from residue 63~69, which forms one part of the active site. It is quite flexible in the apo form, but significantly rigidified by the binding of 5FPy (Figure 4-9a). The binding of 5FPy slows down the conformational fluctuation of the active site and stabilizes it.

The binding of 5FPy also changes the residue flexibility around the active site. First, Exchanges of 53~61 are 4.8 times slower than the apo form including Q55, K56 while S58, T60 changes from slow exchange (43.2 and 20.2 min respectively) to non-exchange; F54 and A59 changes from fast exchange to non-exchange and slow exchange respectively (870 min); G57 is in fast exchange the same as the apo form and L61 is overlapped. If we simply assume that the fast exchange has exchange time shorter than 5 min (the starting measurement time for the first HSQC) and the non-exchange has the exchange time longer than 4000 min (The last HSQC collected in 5FPy bound form), the exchange time changes are quite significant (~180 fold) in this region and also quite diverse for different residues. Second, the exchange time for the loop 111~117 is about 23 times longer than the apo form (including N113 and F114). Residues K115, S116 and K117 have fast exchanges same as the apo form while V112 has no exchange compared with slow exchange in apo form (671 min). Third, the C-terminal helix 150~158 is also rigidified due to the binding of 5FPy. F153, I156 and E158 have no exchange compared with the fast exchange of F153, I156, and E158 in the apo form; W152 has no exchange but has slow exchange in the apo form (38.4 min); D151 and G157 have slow exchanges (27 and 2041 min) compared with fast exchange in apo form; The average exchange time for E154 is about 12.5 times slower in 5FPy bound form. On average the exchange time increase is more than ~300 folds at least for this region. Just like region 53~61 and 111~117, the exchange time changes of the C-terminal helix are also quite significant and varied.

Therefore, the binding of 5FPy changes the HD exchange property of yCD significantly. It stabilizes various regions of the protein especially the active site and the

residues around. The binding of 5FPy increases the exchange time significantly of loop 1 (53~61), residue 63-69, residue 72~80, the F114 loop (111~117) and the C-terminal helix (150~158). Figure 4-8b shows the protection factor of 5FPy complex. We can see clearly the exchange time changes in these five regions. Because 5FPy is a mimic of a reaction intermediate, the rigidification of the active site and regions around may be favorable for the catalytic reaction, which is consistent with MD simulations (Yao et al. 2005) and QM calculations (Sklenak et al. 2004).

The binding of 5FU

The HD exchange property was also measured for the 5FU bound form. Unlike 5FPy, the binding of 5FU does not change dramatically the HD exchange rate of yCD as shown in Figure 4-4c and Figure 4-9b. There are 34 residues in fast exchange, 46 in slow exchange and 39 in no exchange quite similar to that of the apo protein (Table 4-1). The average exchange time is 383 min comparable to 270 min in apo form. The binding of 5FU slows down the HD exchange process for certain residues but accelerates for others. Residues H50, M52, L74, S89, I145, F153, W158 have exchange time scale change from either fast exchange to slow exchange region or from slow exchange region to no exchange region. And residues F54, E119, Y121, V129 have exchange time change in the reverse way (Figure 4-9b). There are 32 residues which HD exchange rates can be measured in both apo and 5FPy bound form, with the average exchange time in 5FU bound form 1.78 folds larger than that in apo form, compared with 8.7 folds increase in 5FPy bound form. Residue 72~80, which has a dramatic increase of exchange time (44 fold) in the binding of 5FPy, has a 2.5 times increase of exchange time (including G72 2.4 fold, T79 0.7 fold and K80 2.7 fold) while R73, G76 and K77 have the fast exchanges

the same as the apo form (L74 has an exchange time 12 min compared to the fast exchange in apo form).

Figure 4-8b shows colored residues picture based on the logarithm of protection factor. As we can see the overall pattern of 5FU complex form is quite similar to apo form, though differences exist in specific regions. The comparison between 5FU bound form and apo form in regions around the active site gives us information about the flexibility changes due to the binding of 5FU. The exchange time for F114 loop (111~117) is about 2.3 times longer than apo form (including V112, N113, and F114. Residues K115, S116 and K117 have fast exchanges.). The binding of 5FU slows down slightly the exchange in this region compared with 23 fold increase due to 5FPy binding. The exchange time for C-terminal helix (150~158) is similar to the apo form: the average time is 1.4 times longer than apo form for W152 and E154; F153 and E158 change from fast exchange region to slow exchange with exchange time 18.3 min and 3 min respectively; N150, D151, I156 and G157 have fast exchanges and D155 has no exchange the same as the apo form. Exchanges of 53~61 are 2.8 times slower than apo form which includes Q55, K56, S58, T60 while residues F54, G57, A59 have fast exchange times, the same as the apo form (F53 and L61 are overlapped). Therefore the residues around active site are slightly rigidified by the binding of 5FU.

On the other hand, exchange rates of residues 64~67 can't be measured in the 5FU bound form mainly due to the weak peak intensity or lack of assignment. The ^{15}N NOESY-HSQC experiment shows the NOEs between amide protons in this region and other protons are very weak suggesting that the active site is quite flexible in the 5FU bound form. It appears that the binding of 5FU changes the HD exchange property

insignificantly compared with 5FPy binding. The flexibility of Loop 1, F114 loop and the C-terminal helix which covers the active site is maintained (Figure 4-7c) while it is decreased dramatically in the 5FPy bound form. The results imply that these three regions might be important for reactant binding and product release. In the apo (product) complex form, the protein is ready to bind (release) ligand.

Effect of pH

Apo form

HD exchange experiments were performed in both pH ~7.0 and ~8.0 phosphate buffer for apo form. The exact pH values were measured after the experiments. The uncorrected readings for the pHs are 7.16 and 8.05 respectively. According to the Linderstrom-Lang model (Linderstrom-Lang 1955), under the EX1 limit, the exchange rate is the measure of the conformational opening rate, which is independent of pH whereas under the EX2 limit, the exchange rate is limited by the HD intrinsic chemical exchange of the open conformation, which is dependent on pH since it is catalyzed by acid or base. In system with pH higher than 3~4, the exchange is catalyzed by base. Therefore, in our system, ideally one unit increase of pH should increase the exchange rate 10 times under EX2, or maintain the values under EX1. Since the pH was increased by 0.89 in the experiments we performed, the corresponding increase of exchange rate should be 7.8 fold.

Figure 4-10 shows the exchange time changes at pH 8.05 for the apo yCD compared with pH 7.16. 35 residues can be measured under both pHs. On average, the exchange time decreases 4.8 fold at pH 8.05. And it is interesting that pH change has quite a different effect on different residues. G29, R48, H50, G72 and D133 change from the slow exchange region to fast exchange region while D155 and L68 change from non-

exchange to slow exchange. On the other hand, F54 and I98 change from fast exchanges to slow and non-exchange respectively. For 35 measurable residues, the exchange time is also spread. In some regions, the exchange time doesn't change much such as 65~70 with the decrease only about 1.5 fold (including I65: 0.91, S66: 0.97, T67: 0.89, E69: 1.8, N70: 2.9). Apparently, the HD exchange rate in this region is determined by conformational changes (EX1). The average exchange time for this region is 346 min (including I65, S66, T67, E69, N70), which is equal to conformation opening time. This is a fairly slow process. Besides this region, some other individual residues also show EX1 exchanges with exchange rate change less than 2 fold (Figure 4-6a). But in another region 137~140 (part of $\alpha 5$), the exchange time decreases ~7.6 fold (C137: 7.2 fold, K138: 9.2 fold, K139: 7.4 fold, I140: 6.6 fold). The HD exchange rate in this region is more determined by the pH value (EX2). Therefore, different regions (residues) may have different exchange mechanism and many of them have neither EX1 nor EX2 mechanism but only in between this two. On the other hand, k_{op} and k_{cl} may be influenced by the pH changes for some residues in the apo form. For example, the T60 and K80 exchange rate increases 10.1 fold and 11.3 fold respectively (Figure 4-10), larger than the theoretical increase of 7.8 fold under the EX2 extreme. And the exchanges of F54, I98 and C106 are slowed down at the high pH.

5FPy complex form

Exchange rate at three pHs were measured including pH 7.10, 7.62 and 8.50. Under the EX2 extreme, the pH change from 7.10 to 7.62 corresponds to 2.6 fold increase of exchange rate. The exchange rate of 37 residues can be measured under these two pHs (Table 4-1), with average exchange time decrease 2.7 ± 0.9 fold. It seems that the overall

protein exchanges under the EX2 mechanism. But this exchange time ratio spans from ~0.8 fold to 5 fold (Figure 4-11a) suggesting no uniform pattern can be determined based on the data. One more exchange rate at pH 8.50 was measured for the 5FPy bound form to give a larger pH difference that can give us a better illustration of EX1 and EX2 differences. And usually the more basic solution is the more possible EX2 exchange occurs because of the strong pH dependence of k_{rc} . 30 residues have measurable exchange rates under pH 7.62 and 8.50 with the average exchange time decrease 9.8 ± 5.6 fold slightly larger than the theoretical decrease of 7.6 fold under the EX2 mechanism (Figure 4-11b). The large fluctuation of the ratio reminds us of the discrepancy between the experimental results and the model. Among these 30 residues, 12 have the ratio larger than 10 fold, which indicates at least for these residues k_{op} and/or k_{cl} might be different under these two pHs. So just like the apo form, different residues show quite different exchange time effects due to pH changes. It is very hard to interpret the data simply as EX1 or EX2 mechanism and the assumption about the independence of k_{op} and/or k_{cl} over pH can hardly be justified in the 5FPy bound form.

5FU complex form

Two exchange rates of 5FU yCD complex were measured at pH 7.10 and 7.50 respectively (Figure 4-12). Compared with pH 7.10, the exchange time decreases by 1.9 ± 0.8 fold on average for 38 measurable residues under both pHs. The exchange time ratio spans from ~0.5 to ~3.9 (Figure 4-12). pH has quite different influences in exchanges over different residues.

Summary

The backbone amide NH chemical shift has been assigned for apo, 5FPy and 5FU bound form. The binding of 5FPy and 5FU perturbs mainly the active site and it appears that 5FU perturbs the active site more than 5FPy. HD exchange experiments were performed for apo, 5FPy intermediate analog and 5FU product complexes in different pHs. Several regions show different exchange properties in different forms. In the apo form, 4 central β sheets as well as part of three α helices have no exchanges. Most of the no exchange residues have no/few water accessibility while most of the fast exchange residues form h-bonds with solvent water. But 63~69 (from the active site) have exchanges though it is not water accessible. Apparently conformation changes are needed to make the exchange occur. After the binding of 5FPy, the exchanges are significantly slower in this region. The binding of 5FU makes this region rather flexible, amide of H62, S66, T67 can not be assigned due to the lack of NOE and the exchange rate of E64 and I65 can't be measured due to the weak signal intensity. The second interesting region is loop 1 (53~61), a loop covering the active site. The binding of 5FPy increases the exchange time by more than ~180 fold, while the binding of 5FU only increases the exchange time ~2.8 fold. The third interesting region is residue 72~80 which shows a dramatic increase of exchange time in the 5FPy bound form but not in the 5FU bound form. The fourth interesting region is loop F114 (111~117), another loop covering the active site. The binding of 5FPy increases the exchange time for N113 and F114 about 20 fold compared with the binding of 5FU, which increases that time about 2.3 fold. The fifth interesting region is the C-terminal helix $\alpha 6$ (150~158) which have relatively fast exchanges in the apo form. The binding of 5FU maintains the exchange rate in this region. But the binding of 5FPy makes exchange in this region more than ~300 times

slower. Loop 1, loop F114 and the C-terminal helix, covering the active site, might be important for ligand binding and product release. Compared with the apo form, on average the 5FPy complex increases the exchange time 1.8 folds; 5FPy complex increases the exchange time 8.7 fold for measurable residues.

The pH effect over the exchange time is quite different for individual residues. For example, in the apo form residue 65 ~ 70 displays the EX1 exchange mechanism while residue 137 ~ 140 displays the EX2 exchange mechanism. But the majority of residues show neither EX1 nor EX2 mechanisms in apo, 5FPy and 5FU bound forms.

Reference

- Arrington, C. B. and A. D. Robertson (2000). Kinetics and thermodynamics of conformational equilibria in native proteins by hydrogen exchange. Energetics of Biological Macromolecules, Pt C. **323**: 104-124.
- Arrington, C. B. and A. D. Robertson (2000). "Microsecond to minute dynamics revealed by EX1-type hydrogen exchange at nearly every backbone hydrogen bond in a native protein." Journal of Molecular Biology **296**(5): 1307-1317.
- Bai, Y. W., J. S. Milne, et al. (1993). "Primary Structure Effects on Peptide Group Hydrogen-Exchange." Proteins-Structure Function and Genetics **17**(1): 75-86.
- Delaglio, F., S. Grzesiek, et al. (1995). "Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes." Journal of Biomolecular Nmr **6**(3): 277-293.
- Ireton, G. C., M. E. Black, et al. (2003). "The 1.14 angstrom crystal structure of yeast cytosine deaminase: Evolution of nucleotide salvage enzymes and implications for genetic chemotherapy." Structure **11**(8): 961-972.
- Johnson, B. A. and R. A. Blevins (1994). "Nmr View - a Computer-Program for the Visualization and Analysis of Nmr Data." Journal of Biomolecular Nmr **4**(5): 603-614.
- Linderstrom-Lang, K. (1955). "Deuterium exchange between peptides and water." Spec. Publ. Chem. Soc. **2**: 1-20.
- Loh, S. N., C. A. Rohl, et al. (1996). "A general two-process model describes the hydrogen exchange behavior of RNase A in unfolding conditions." Proceedings of the National Academy of Sciences of the United States of America **93**(5): 1982-1987.
- Pearlman, D. A., D. A. Case, et al. (1995). "Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules." Computer Physics Communications **91**(1-3): 1-41.
- Qian, H. and S. I. Chan (1999). "Hydrogen exchange kinetics of proteins in denaturants: A generalized two-process model." Journal of Molecular Biology **286**(2): 607-616.
- Sklenak, S., L. S. Yao, et al. (2004). "Catalytic mechanism of yeast cytosine deaminase: An ONIOM computational study." Journal of the American Chemical Society **126**(45): 14879-14889.
- Yao, L. S., Y. Li, et al. (2005). "Product release is rate-limiting in the activation of the prodrug 5-fluorocytosine by yeast cytosine deaminase." Biochemistry **44**(15): 5940-5947.

Yao, L. S., S. Sklenak, et al. (2005). "A molecular dynamics exploration of the catalytic mechanism of yeast cytosine deaminase." Journal of Physical Chemistry B **109**(15): 7500-7510.

Appendices

Table 4-1 The statistics of HD exchange times for the apo, 5FPy and 5FU complex in various pHs.

	number of residues				Average time (min)
PH 7.0 ^(a)	fast	slow	no	undetermined	
apo	38	40	42	25	270
5FU	34	46	39	26	383
5FPy	19	43	70	13	662
PH 8.0 ^(b)					
Apo	42	39	41	23	159
5FU	39	38	39	29	306
5FPy	24	43	63	15	459
PH 9.0 ^(c)					
5FPy	28	36	61	20	140

(a). The measured pHs are 7.16, 7.10 and 7.10 for apo, 5FPy and 5FU bound form.

(b). The measured pHs are 8.05, 7.62, and 7.50 for apo, 5FPy and 5FU bound form.

(c). The measured pH is 8.50 for 5FPy bound form.

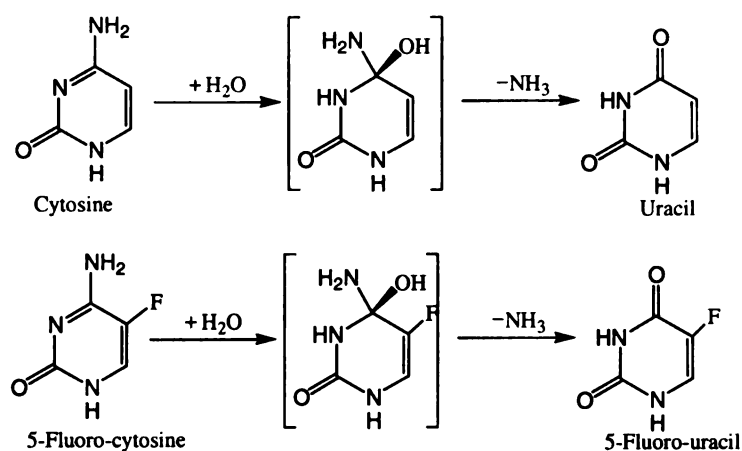


Figure 4-1 The deamination of cytosine/5FC to uracil/5FU through a tetrahedral intermediate.

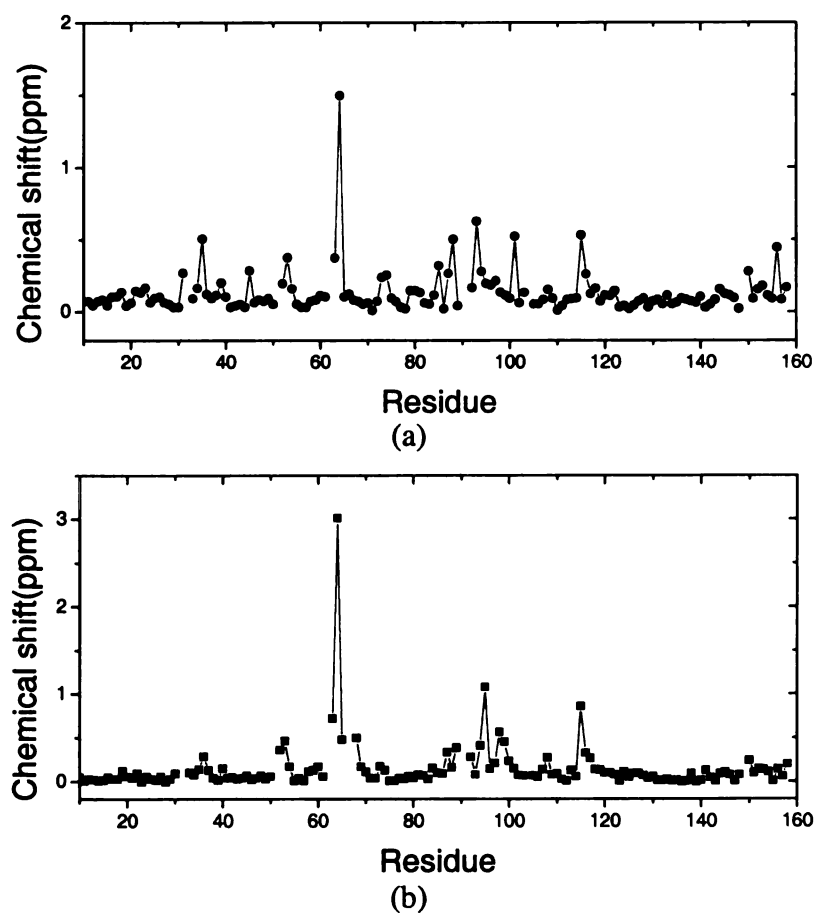
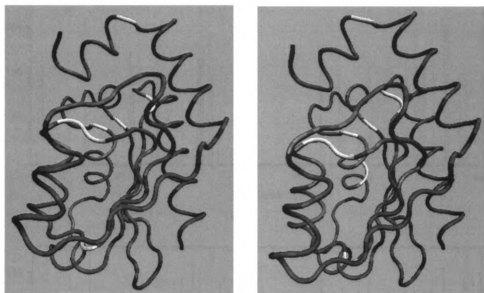
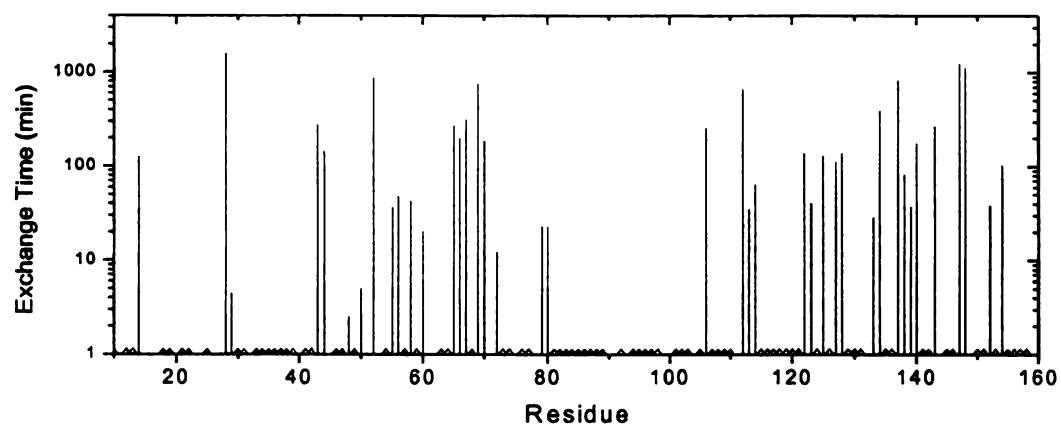


Figure 4-2 The chemical shift difference (a). in the apo and 5FPy bound form. (b). in the apo and 5FU bound form.

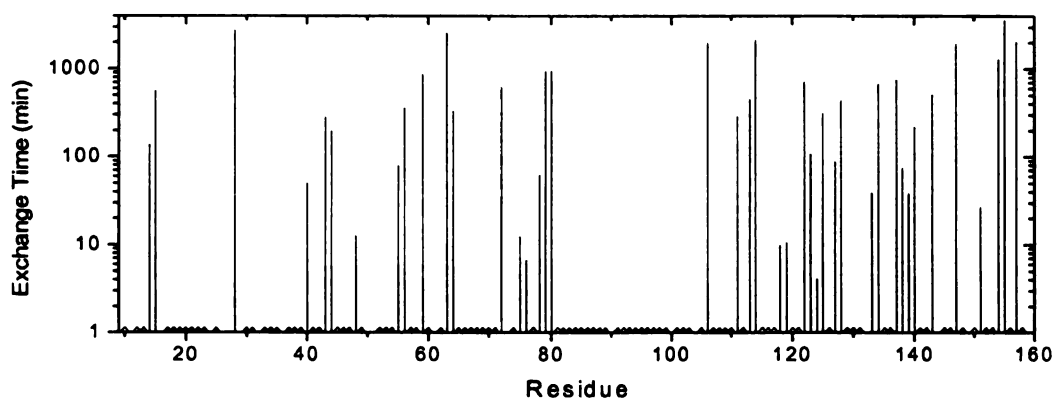


(a) (b)

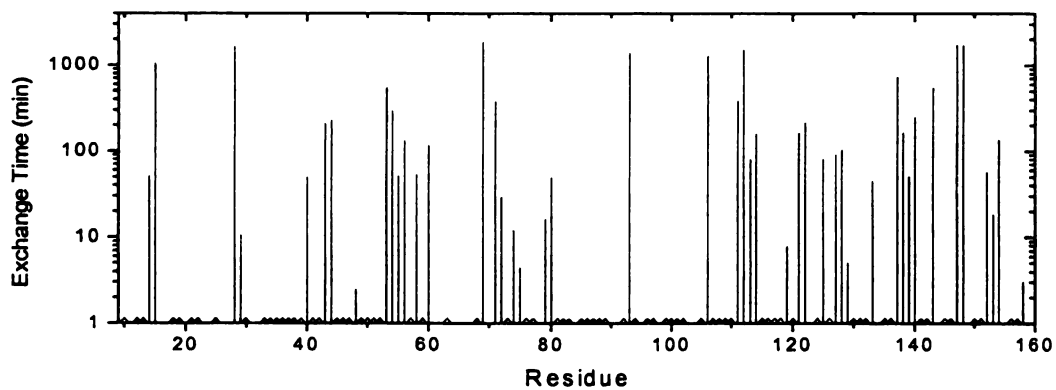
Figure 4-3 The colored tube plot of backbone CAs based on chemical shift difference (a). in apo and 5FPy bound form. (b). in apo and 5FU bound form. The difference increases with the color changing from blue to red. The residues with no assignment were painted white.



(a)



(b)



(c)

Figure 4-4 Exchange time for (a) apo, (b) 5FPy complex, (c) 5FU complex. \blacktriangle represents no exchanges (exchange time longer than ~ 2000 minutes (apo) or ~ 4000 minutes (5FPy and 5FU complex)), Δ represents fast exchanges (exchange time shorter than ~ 5 minutes).

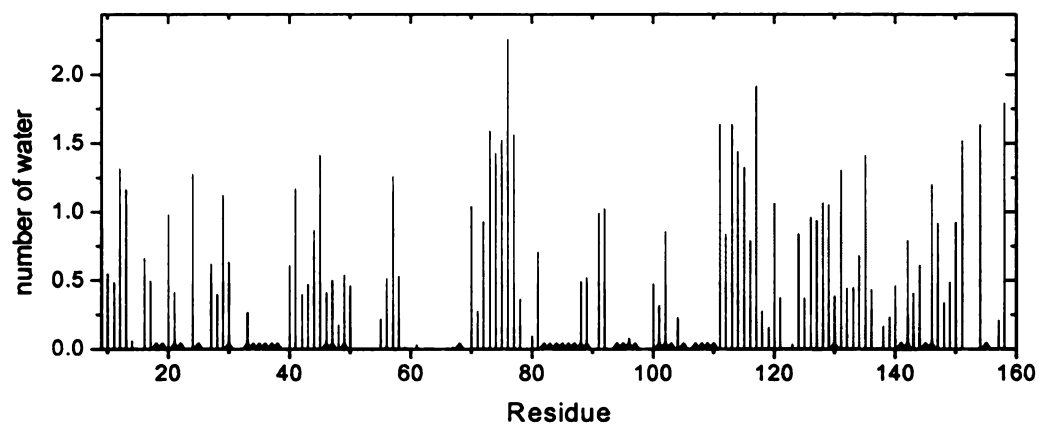
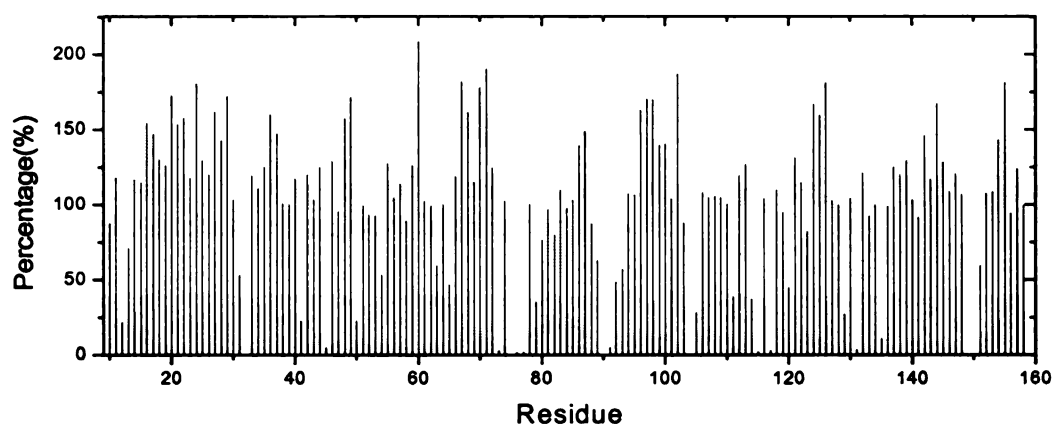
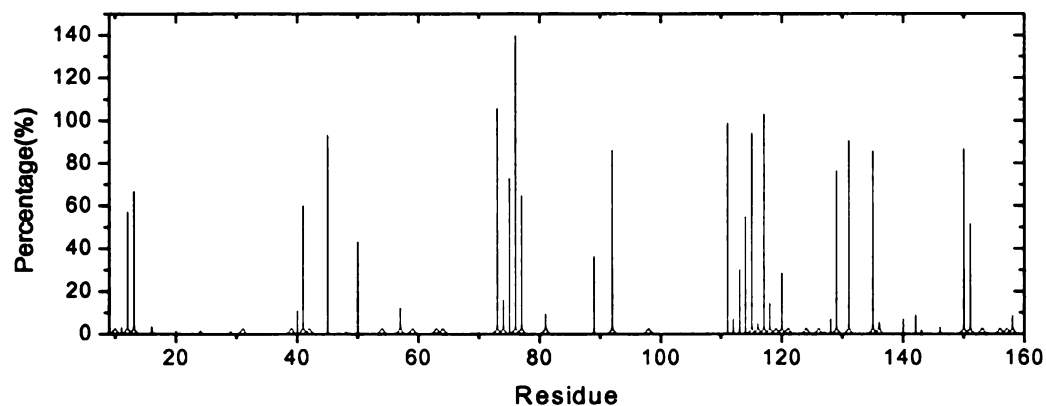


Figure 4-5 Number of water molecules around NH within 3.5Å from N atom in MD simulation. ▲ represents no exchanges residues (exchange time longer than ~2000 minutes).



(a)



(b)

Figure 4-6 (a). Percentages of H-bond formation between NH and other residues. (b). Percentages of H-bond formation between NH and solvent water during 1.5 ns MD simulation.

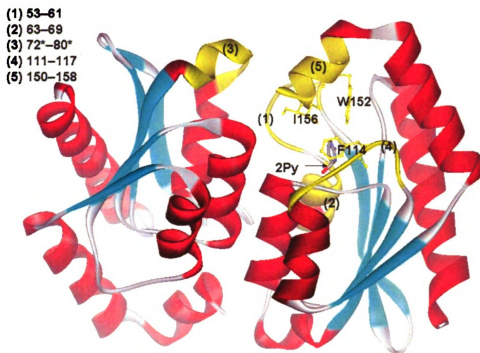


Figure 4-7 X-ray structure of yCD complexed with 2Py. Regions with significant increase of HD exchange time were highlighted in gold.

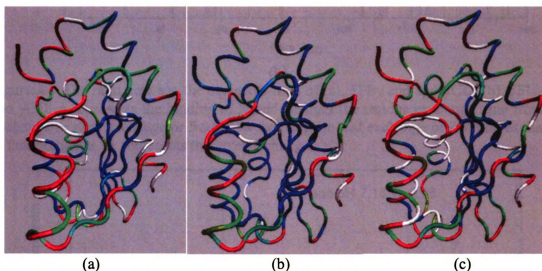
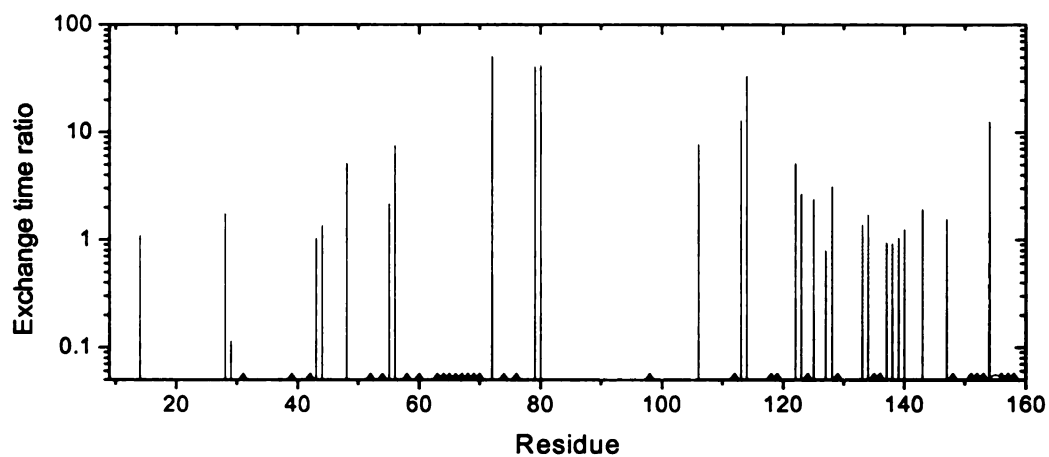
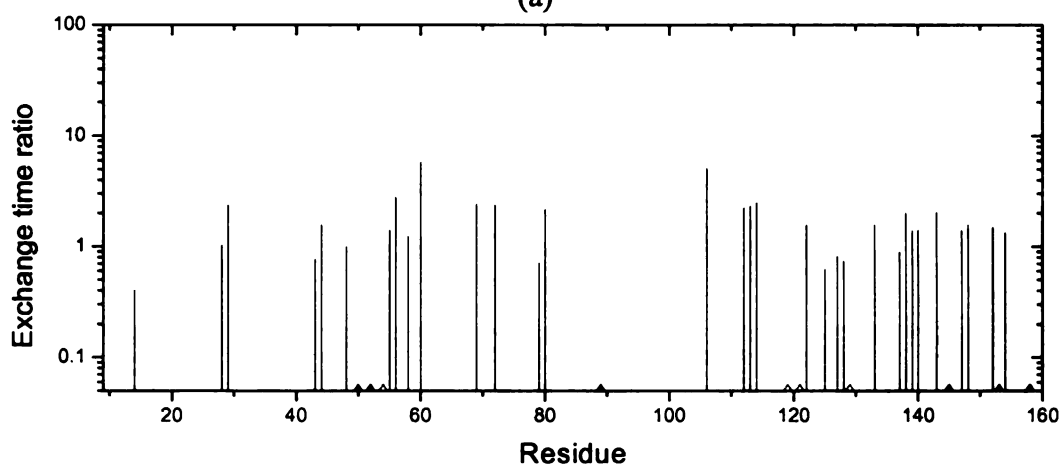


Figure 4-8 Summary of HD exchange data of yCD in (a). apo, (b). 5FPy complex and (c) 5FU complex forms at pH 7. Residues are colored according to the logarithm of protection factor: blue, residues without exchanges during experimental time ($\log P \geq 7.0$); red, residues with fast exchanges ($\log P \leq 2.0$); orange to iceblue, residues with slow exchanges ($7.0 > \log P > 2.0$). The crystal structure of apo form subunit one was used in the picture.



(a)



(b)

Figure 4-9 The ratio between the exchange time of (a). 5FPy and apo yCD. (b). 5FU and apo yCD. Δ represents the binding of ligand changes the residue exchange time from no exchange to slow exchange or from slow exchange to fast exchange or from no exchange to fast exchange. \blacktriangle represents the reverse change.

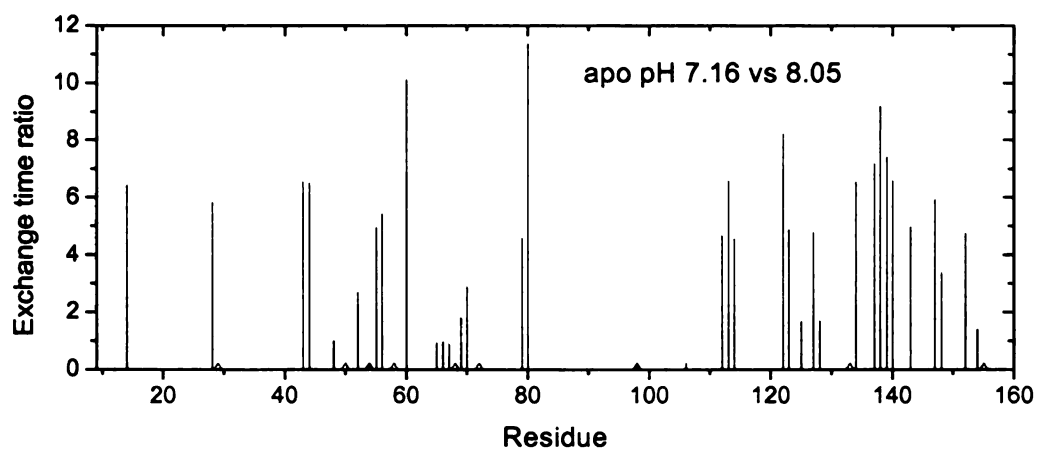


Figure 4-10 The exchange time ratio in apo form at pH 7.16 and 8.05.

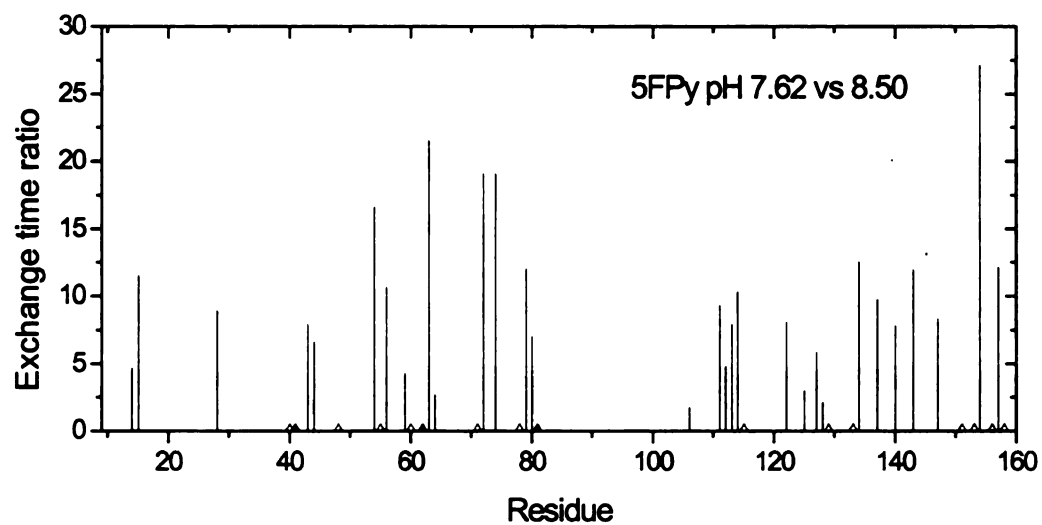
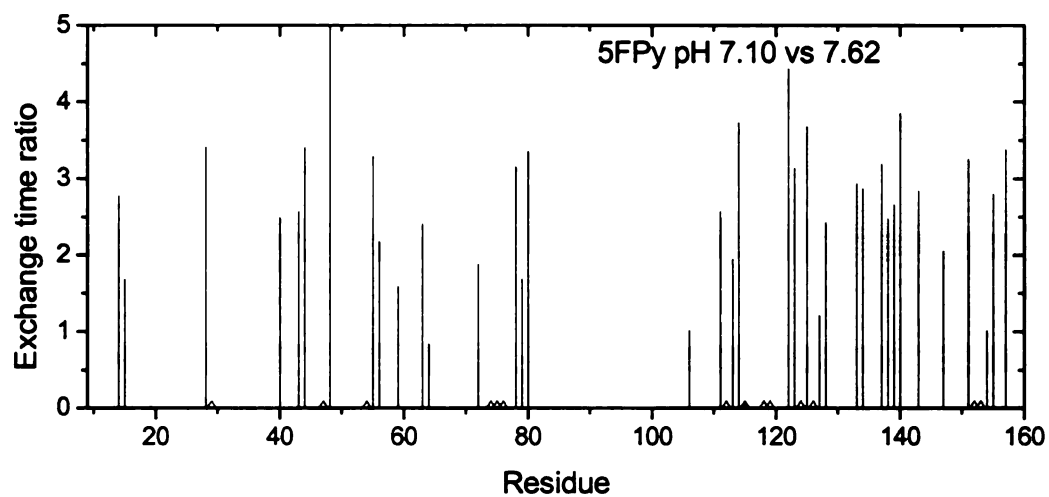


Figure 4-11 The exchange time ratio in 5FPy bound form at (a). pH 7.10 and 7.62, (b). pH 7.62 and 8.50.

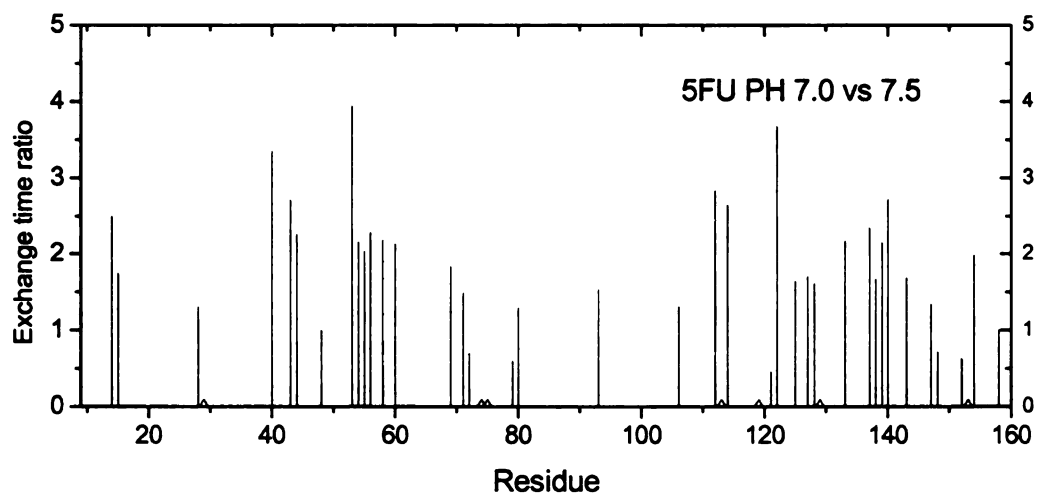


Figure 4-12 The exchange time ratio in 5FU bound form at pH 7.10 and 7.50.

Chapter 5

A Molecular Dynamics Exploration of the Catalytic Mechanism of Yeast Cytosine Deaminase

Introduction

Yeast cytosine deaminase (yCD) catalyzes the deamination of cytosine to uracil by nucleophilic attack of a Zn-bound hydroxide on the substrate, forming a tetrahedral intermediate that decomposes through the elimination of ammonia. Cytosine deaminase is present in bacteria and fungi, as part of the pyrimidine salvage pathway, but is absent in mammals.¹ yCD can also catalyze the deamination of the prodrug 5-fluorocytosine (5-FC) to the anticancer drug 5-fluorouracil (5-FU). Therefore, a potential system for gene-directed enzyme prodrug therapy combines yCD and the prodrug 5-FC, since the product, 5-FU, inhibits DNA synthesis and is thus a potent chemotherapeutic agent.^{2,3}

High resolution structures of yCD in the apo form⁴ and in complex with the inhibitor 2-pyrimidinone, a mimic of a reaction intermediate^{4,5}, are available. yCD is a homodimer, with each 158 residue subunit consisting of a central β -sheet flanked by two α -helices on one side and three α -helices on the other (Figure 5-1a). A single active site is present in each subunit. The active sites are separated by 14 Å, both are adjacent to the dimer interface, and no cooperativity has been reported for the enzyme.⁴ Each active center contains a single catalytic zinc ion that is tetrahedrally coordinated by a histidine (His62), two cysteines (Cys91 and Cys94), and a water molecule in the substrate-free

enzyme or the inhibitor in the complex (Figure 5-1b). The bound inhibitor is completely buried by a lid composed of Phe114 from the loop between $\beta 4$ and $\alpha 4$, and Trp152 and Ile156, both from the C-terminal helix. Surprisingly, the structure of apo yCD is essentially the same as that of the intermediate analog complex with a root-mean-squared deviation (RMSD) between the two structures of 0.23 Å for the backbone atoms.

The active site architectures of yCD and cytidine deaminase (CDA) from *E. coli* share a striking similarity.^{6,7} Superposition of their active sites reveals a very similar interaction network involving the ligated water molecule, the Zn ion ligated by two cysteine residues and one histidine, the Zn-bound ligand and a glutamic acid important to the catalytic mechanism.⁵ Based on the similarity of these active site structures and early studies of the mechanism of CDA catalysis, an analogous reaction mechanism was proposed for yCD.⁵ Our recent quantum chemical study⁸ using the ONIOM(B3LYP:PM3) method has revealed a complete path for the deamination reaction catalyzed by yCD. The cytosine deamination proceeds via a sequential mechanism involving the protonation of N3, the nucleophilic attack of C4 by the Zn-coordinated hydroxide, and the cleavage of the C4-N4 bond. Uracil is liberated from the zinc by an oxygen exchange mechanism that involves the formation of a *gem*-diol intermediate from the Zn bound uracil and a water molecule, C4-O^{Zn} bond cleavage, and regeneration of the Zn-coordinated water.

In this article, we explore aspects of the mechanism of the cytosine deamination reaction by running a series of MD trajectories (~2 ns) for the fully solvated enzyme in its free form, reactant complex form, product complex form, and four possible intermediate complex forms, with active site structures summarized in Figure 5-2. In particular, we

study the side-chain motion of Glu64 along the proposed⁸ reaction path and the role of several other residues in both ligand binding and catalysis. For each simulation, quantum chemical calculations are first carried out in order to obtain suitable charge parameters for the Zn ion, ligated residues, catalytic water, substrate and important neighboring residues. The protonation states of the Zn-ligated residues are critical for an accurate simulation, as is an account of the substantial charge transfer to the Zn (formally 2+) ion. The availability of the yCD crystal structures with the intermediate analog complex^{4,5} aided us in validating the charge parameters of the Zn bound complex by comparing the results of an MD simulation with these crystal structures.

The flexibility and average position of the loop containing Phe114, and the N- and C-terminal helices are monitored in the MD simulations of both the free and intermediate analog complex forms and compared with the corresponding crystal structures. That permits suggestions for possible ways the reaction products can exit the active site, of interest in view of the very similar crystal structures found for the apo and inhibitor-bound structures. The product complex simulation also suggests a binding position for ammonia and the exchanged water prior to release from the active site. In these ways, further insights into the catalytic and binding events taking place in the active site of yCD can be obtained.

Methods

Determination of the protonation state of the Zn bound complex

The protonation states of the Zn ligands, cysteine and histidine, are dependent on their surroundings.⁹⁻¹¹ Quantum chemical studies on model compounds show that the Zn – ligand distances are sensitive to their protonation states and, by comparison with

crystallographic data, the latter can be reliably assigned.^{9,10} We carried out B3LYP/6-31G** optimizations⁸ in the gas phase for the Zn complex starting from the crystal structure, modeling His62 by imidazole, Cys91 and Cys94 by -SCH₃, Glu64 by CH₃CH₂COO⁻, and included the Zn ion, a water and -OOCCH₂CH₂C(=O)CH₃, the last of which is a model for Asp155 that is hydrogen bonded to His62 (2.7 Å, a second shell interaction with the Zn) and, in the intermediate analog complex, also to the intermediate analog (Figure 5-1b). These calculations showed that to match the coordination geometry of the zinc ion in the crystal structure, the sulfur atoms of Cys91 and Cys94 must be deprotonated and His62 must be singly protonated.

Parameterization of the Zn bound complex

In metal-ligated simulations a choice between bonded versus non-bonded force fields must be made.¹² For the purposes of this work, where the crystal structures show strong four-coordinate ligation to the Zn ion, a bonded (covalent) force field is indicated. Thus, explicit bonds between the Zn cation and its ligands were used.¹³ The force constants for bonds and bond angles listed in Table 5-1 serve to preserve the experimentally observed tetrahedral Zn. They are consistent with force constants used for a Zn complex in a recent simulation by Suarez and Merz.¹⁴ The equilibrium bond distances and angles are taken from the apo yCD crystal structure.⁴ All the torsional parameters associated with the Zn-ligand interactions were set to zero as in Hoops et al.¹³

A single point QM calculation (B3LYP/6-31+G*) of the Zn complex was carried out based on the crystal structure of the yCD free form. Zn, its three ligated residues, His62, Cys91, Cys94, as well as Asp155 (all atoms), and one water molecule (coordinated with the Zn) are included in the calculation. Hydrogen atoms were added by

using the Insight II program. One hydrogen atom was added to each C-terminus and N-terminus of the four amino acids to saturate the heavy atoms. Atom-centered partial charges were derived by using the AMBER antechamber program (RESP methodology).¹⁵ Though HF/6-31G* is the method used for parameterization of the General AMBER Force Field (GAFF)¹⁶, in the case of the Zn-ligand bonds that have an intermediate character between a covalent and an ionic bond,¹⁷ B3LYP is a better method to calculate the RESP charges. For our test job of cytosine, the RESP charges derived from HF/6-31G* and B3LYP/6-31+G* produced a root-sum-of-square-deviation (RSSD) of only 0.071e for all the charges.

In the MD simulation, only the RESP charges of the substrate, intermediates and products, as well as those of the Zn-bound water, Zn and the *side-chains* of His62, Cys91, and Cys94 are re-evaluated by our quantum chemical calculations relative to their values in the AMBER database. The charges listed in Table 5-2 are slightly modified to preserve the charge neutrality of the Zn complex. Charges of Asp155, the backbone charges of His62, Cys91, and Cys94 are the same as those from the AMBER database. Since no new bond, bond angle or dihedral angle parameterization was performed for this system in our calculations, this procedure was used to minimally perturb the parameters from the AMBER database.

MD simulation of the free form of yCD

Starting coordinates for the protein atoms were taken from the free form yCD 1.43 Å resolution crystal structure.⁴ There is a second Zn in the active site of the crystal structure, which was deleted for the simulations. All the crystal water molecules were removed except the one bound to the catalytic Zn. The protonation states of the ionizable

residues were set to their normal values at pH 7. The protein was surrounded by a periodic box of 12.5 Å (~17,000) TIP3P water molecules. Na⁺ ions were placed by the Leap program¹⁵ to neutralize the -6 charge of the model system. The parm94 version of the all-atom AMBER force field¹⁸ was used for all the simulations.

MD simulations were carried out using the SANDER module in AMBER 7.0.¹⁶ The SHAKE algorithm¹⁹ was used to constrain all bond lengths involving hydrogen atoms permitting a 2 fs time step. A nonbonded pair list cutoff of 8.0 Å was used and nonbonded pair list was updated every 25 steps. The Particle-Mesh-Ewald method was used²⁰ to evaluate the contributions of the long-range electrostatic interactions. The pressure (1 atm) and the temperature (300 K) of the system were controlled during the MD simulations by Berendsen's method.²¹ The typical simulation time was 2 ns, with a 500 ps equilibration period. Coordinates are saved every 2 ps. All of the MD results were analyzed with the PTRAJ module of AMBER 7.0. In these analyses, we assign hydrogen bonds when the distance of two heavy atoms (O or N) is less than 3.2 Å and the angle (heavy atom-hydrogen-heavy atom) is greater than 120°.

MD simulation of the yCD intermediate analog complex

The starting structure for the yCD intermediate analog complex simulation was taken from the crystal structure.⁴ Subunit one has 156 residues (the first two are missing); subunit two has 161 residues because three extra residues are attached to the N-terminus. One more residue, Glu64, is included in the charge parameterization because it forms a strong hydrogen bond with the ligand in the crystal structure, and the inclusion of this residue may influence the RESP charges of the Zn-bound complex. The Glu64 charges used in the MD are the AMBER ones, since they are very close to those from the

quantum chemical calculation. The MD simulation protocols for the yCD intermediate analog were identical to those for the unbound model except the modeled system was maintained at constant volume and temperature, as was the case for the other simulations described below.

MD simulation of the yCD water/cytosine complex (1 in Figure 5-2)

Cytosine was docked to the yCD free enzyme crystal structure active site in the same position found for the intermediate analog. The Zn-bound water molecule was included in the simulation. All other crystal water molecules and the second Zn were removed, as in the free form simulation. The force field parameters of the yCD protein are the same as in the free yCD simulation. The RESP charges of cytosine are from a HF/6-31G* calculation.

MD simulation of the yCD hydroxide/Glu64H cytosine complex (2 in Figure 5-2)

The 500 ps instantaneous structure of the yCD water/cytosine system was adopted as the starting structure of the yCD hydroxide/cytosine system. The deprotonated Glu64 was mutated into a protonated glutamic acid, while the Zn complex water was mutated into a hydroxide group in this system. The new RESP charges of the Zn complex (Zn, His62, Cys91, Cys94, and the hydroxide ligated to Zn) were calculated based on a free form ONIOM(B3LYP:PM3) minimized structure with the proton restrained to Glu64.⁸

MD simulation of the yCD hydroxide/Glu64 cytosineH complex (3 in Figure 5-2)

The 1.0 ns snapshot structure of the yCD hydroxide/Glu64H cytosine system was used as the starting structure of yCD hydroxide/Glu64 cytosineH system. The proton of Glu64 was transferred to the N3 of cytosine in this system. The RESP charges of the Zn complex (Zn, His62, Cys91, Cys94, hydroxide) are the same as the yCD

hydroxide/Glu64H cytosine complex simulation, and the RESP charges of protonated cytosine were calculated based on HF/6-31G* QM calculation.

MD simulation of the yCD intermediate complexes (4 and 5 in Figure 5-2)

The starting structure for the yCD intermediate complex simulation was taken from the crystal structure of the yCD intermediate analog complex. The intermediate was docked to the same position where the intermediate analog stays. Since the intermediate could be either protonated, intermediate I, (with deprotonated Glu64, 4 in Figure 5-2) or deprotonated, intermediate II, (with protonated Glu64, 5 in Figure 5-2), two situations were implemented in the two different active sites. The RESP charges (see Table 5-2) of these two Zn complexes were calculated based on the yCD intermediate analog crystal structure.

MD simulation of the yCD product complexes (6 and 7 in Figure 5-2)

The starting structure for the yCD product complex simulation was taken from the yCD intermediate analog complex crystal structure. Intermediate II was converted to uracil/ammonia. Since our ONIOM(B3LYP:PM3)⁸ calculations showed that the Zn complex structure in this system was similar to that in the crystal structure of yCD intermediate analog complex, uracil was bound to the Zn atom for the simulation with a bond to the oxygen. This geometrical arrangement is similar to the cytidine deaminase-uridine complex crystal structure⁷ The new RESP charges of the Zn complex were calculated. In order to study the release of ammonia and its possible exchange with water in the active site, ammonia was substituted by water in one active site while it was kept in the other.

Potential of mean force (PMF) calculation of Glu64 COOH rotation in the yCD hydroxide/Glu64H cytosine complex

The simulations that we carry out show that the Glu64 (GluH64) carboxylate (carboxylic acid) moiety fluctuates around three orientations, which can be described by the dihedral angle φ defined by the CB-CD-CG-OE2 atoms of Glu64. The free energy of rotation was obtained from the potential of mean force (PMF) $W(\varphi)$ that is related to the probability $P(\varphi)d\varphi$ of finding the dihedral between φ and $\varphi+d\varphi$ according to $W(\varphi) = -k_B T \ln P(\varphi)$. The PMF was obtained by a free energy umbrella sampling method^{22,23} in which a series of harmonic restraint terms (windows) are added to the system's Hamiltonian in order to force the system to sample the desired regions of φ space, in this manner overcoming the poor sampling of high-energy conformations that would occur in a unbiased MD trajectory. The data from the windows were combined using the weighted histogram analysis method (WHAM)²⁴ to produce the PMF over the desired range of φ . The force constant was set to 20 kcal/(mol-rad²), which provides a rms restraint potential window width of approximately 10°, and the window size was set to 3.5° to ensure good overlap of data in neighboring windows. The PMF between state 1 and 2 was determined by running the simulations forward and backward with 5 ps of equilibration and 5 ps of data collection for each window (total 40 windows). The parameters we used were validated by testing the accuracy of the method on the dihedral PMF for ethane solvated in water. The results (not shown) indicate that this set of parameters can reproduce the PMF of the ethane dihedral rotation quite well. By using this set of parameters, our forward and backward simulations of the Glu64 CB-CG-CD-

OE2 dihedral rotation gave consistent results (see Results and Discussion). The PMF between state 2 and 3 was determined by starting the simulation from state 2.

Chemical alchemy simulation of NH₃ and H₂O exchange

Our ONIOM(B3LYP:PM3)⁸ calculations indicate that uracil is released by an oxygen exchange mechanism that uses a water molecule at the same location as the product ammonia in the active site. As a consequence, before the release of uracil, ammonia should exchange with the water molecule. Therefore, we evaluated the difference in binding free energy between water and ammonia by mutating water to ammonia (and ammonia to water) in both the protein active site and water solution. The RESP charges of ammonia were calculated based on a HF/6-31G* calculation. After ammonia was solvated in water (box side 12.5 Å) and equilibrated, it was mutated to a water molecule. The nitrogen and two ammonia hydrogen atoms were mutated to an oxygen atom and two water hydrogens while the third hydrogen was mutated to a dummy atom that has zero charge and zero van der Waals radius but maintains its internal interaction parameters. Twelve windows were used with 30 ps per window (15 ps for equilibration and 15 ps for data collection). SHAKE was not used considering that mutation of hydrogen atoms is involved, necessitating a time step reduction to 1 fs. Gaussian quadrature formulas were used to pick values for the ranges of these windows. The same procedure was used to mutate water (with a dummy atom attached) to ammonia. Thermodynamic integration was used to do the free energy calculation. The contribution to the free energy from the dummy atom was evaluated analytically.²⁵ A standard state of 1 M (1661 Å⁻³) was used for the free dummy atom. The results for the forward and backward mutation show excellent consistency (see Results and Discussion).

In the yCD product simulation, the water and ammonia were put into different active sites, and they were mutated to each other, respectively, by the procedure outlined above. The free energy difference in water and protein is the binding free energy. Assuming a small effect from the different active sites, the difference in binding free energies between water and ammonia is then available.

Results and Discussion

yCD free and intermediate analog complex compared with the crystal structures

yCD is a homodimer with 158 residues per subunit. Each subunit has one active site that is covered by Phe114 located in a loop consisting of residues 109-116, and by Trp152 and Ile156 contained in the C-terminal helix, residues 149-158. The data analysis is based on individual subunits in order to get better statistics, because the crystal structure shows that the configurations of the two subunits are the same, except for the N-terminus (~10 residues). We superimpose the MD snapshots on the crystal structure without using the N-terminal 14 residues.

The time evolution of the CA atom RMSD of the instantaneous structures (not including the first 14 residues) from the initial crystal structure for the free yCD and the intermediate analog complex indicate that the two systems are in equilibrium with respect to the RMS deviations during the analyzed trajectory (Figure 5-3). The mass weighted RMSD of all atoms relative to the crystal structure is, on average, 1.29 Å for subunit one and 1.60 Å for subunit two, in the free form simulation (Table 5-3), indicating that the structural changes in the protein were not large during the course of the simulation. The different RMSDs do indicate that some part of the protein in subunit two moved further away from the crystal structure than in subunit one. By superimposing the average structure of subunit one and subunit two (data not shown) we see that the largest

differences are in the N-terminal, C-terminal and Phe114 loop regions. The average RMSD data also suggest distinctions among these three regions (Table 5-3). The time evolution of the RMSD of the C-terminal helix shows that two sub-states were captured in the free yCD MD simulation (Figure 5-4a). Subunit one maintains the crystal structure of the C-terminal helix, while subunit two shifted the whole helix slightly away, and this movement lets the Trp152 side-chain occupy the active site. This movement is not surprising since, in the crystal structure, one additional Zn atom, which is bridged by a water molecule to the active site Zn, coordinates with three more water molecules.⁴ This Zn complex fills the active site; in the MD simulation the extra Zn with its coordinated waters was removed, providing space for some rearrangement around the active site.

Interestingly, the apo and intermediate complex forms have essentially the same crystal structure⁴, with a closed active site. Ligands cannot move in or out without moving the covering residues Phe114, Trp152 and Ile156. The time evolution of the Phe114 loop RMSD in the free form simulation highlights the difference in this region between subunit one and subunit two (Figure 5-4b). The Phe114 loop region is flexible, though a 2 ns simulation may not be long enough to describe a motion that could permit substrate entrance. To gain insight into the fluctuations of the yCD enzyme, the root mean square fluctuation (RMSF) was calculated by comparing the instantaneous protein structure with the average one for the free form (Figure 5-5a). Though the difference in RMSD is large between subunit one and subunit two compared with the crystal structure, the RMSF for CA atoms are almost the same, and consistent with the crystal B-factors. The residue RMSF values are quite small (most of them are around 0.5 Å), which

indicates that the yCD protein is quite rigid as a whole in the free form on the MD time scale.

As in the free form enzyme, the N-terminal residues are far more flexible than the rest of the protein in the intermediate analog complex (Table 5-3). The total RMSD (not including N-terminal 14 residues) is 1.46 Å for subunit one, and 1.24 Å for subunit two, quite similar to the free form values. The small RMSF values for the CA atoms indicate the rigidity of the intermediate analog complex on the MD time scale as well (Figure 5-5b). The differing RMSF's between subunits one and two (Figure 5-5b) in the intermediate analog complex are evidence that two different configurations are being accessed, at least on the 2 ns timescale. The results indicate that the Phe114 loop and the C-terminal helix are flexible, especially in subunit one, relative to the crystal structure.

Structure of the active site

A schematic representation and a typical snapshot of the active site of the apo and intermediate analog complex are displayed in Figure 5-6. In Table 5-4, the most significant H-bonds between the ligand (or, Zn-bound water in the free form) and the enzyme residues are characterized in terms of the distances between two heavy atoms and their percentages of occurrence. In the free form simulation, the Zn-bound water forms hydrogen bonds with OE1 and OE2 of Glu64, which are not present in the crystal structure, because the second Zn in the crystal structure blocks the formation of these hydrogen bonds. In the MD simulation, this second Zn is not present. There is ~70% hydrogen bond occurrence for OE2 of Glu64 and ~30% for OE1 in subunit two, which indicates that the carboxyl group of Glu64 can rotate in the free form simulation (Figure 5-7a). In the complex, the ligand forms hydrogen bonds with Asn51, Gly63, Glu64,

Cys91 and Asp155. The simulation results are in excellent agreement with the hydrogen bond network found in the crystal structure (Table 5-4), which supports the Zn complex parameters and MD protocols developed for this simulation.

Orientational changes of cytosine along the reaction path

Snapshots of the active site of the yCD water/cytosine reactant complex show, surprisingly, that cytosine rotates around the axis perpendicular to its pyrimidine ring to move its NH₂ group closer to the Glu64 carboxyl group (1 in Figure 5-2), relative to its starting position that is the same as the intermediate analog complex orientation in the crystal structure. This rotation helps the NH₂ group of cytosine form a hydrogen bond network with Glu64 (Table 5-5). The upper part of the pyrimidine ring moves away from the Zn-bound water, due to the van der Waals repulsion between them, while the bottom part is restrained by three hydrogen bonds between cytosine and Gly63, Asn51 and Asp155. The distance between the C4 of cytosine and O of the Zn-bound water is $3.26 \pm 0.17 \text{ \AA}$ in subunit one and $3.19 \pm 0.14 \text{ \AA}$ in subunit two. Cytosine orients somewhat differently in the two subunits (Table 5-5). The NH₂ group of cytosine in subunit one points directly at the carboxyl group of Glu64, and forms a hydrogen bond network using the two carboxylate oxygens and two amino hydrogens. But, in subunit two, NH₂ can also form a hydrogen bond with O of Ser89 that is oriented toward the upper right side of cytosine.

After protonation of Glu64 by proton transfer from the Zn bound water (yCD hydroxide/Glu64H cytosine model, 2 in Figure 5-2), cytosine rotates back to the orientation of the intermediate analog found in the crystal structure⁴ in the MD simulation. In this system, the protonated Glu64 forms one hydrogen bond with the

amino group of cytosine instead of the hydrogen bond network as in the previous carboxylate system (Table 5-6). The starting structure for this simulation was taken from one snapshot of the yCD water/cytosine simulation, with a slightly different geometry in subunits one and two. Correspondingly, these produce slightly different active site interactions. Notably, the distance between the C4 of cytosine and O of the Zn-bound hydroxide is 3.02 ± 0.13 Å in subunit one, but 3.33 ± 0.20 Å in subunit two. By superimposing the average structure of subunit one and two, we found that the cytosine in subunit two was pushed down by ~ 0.5 Å relative to subunit one. This is consistent with the formation of a hydrogen bond between the carbonyl O of Ser89 and the amino group of cytosine in subunit one, but not in subunit two (Table 5-6). The orientation of protonated Glu64 is also more favorable for proton transfer to the N3 of cytosine in subunit one than in subunit two (see discussion below).

After Glu64 transfers a proton to the N3 of cytosine (yCD hydroxide/Glu64 cytosineH), the distance between the C4 of cytosine and the O of Zn-bound hydroxide becomes 2.84 Å in subunit one and 2.80 Å in subunit two; these distances are about 0.4 Å shorter than in the initial yCD cytosine complex. Thus, the C4 atom of cytosine becomes well positioned for nucleophilic attack by the hydroxide. Therefore, subsequent to proton transfer from the Zn-bound water to the N3 atom, the reorientation of cytosine makes nucleophilic attack of the Zn-bound hydroxide group on C4 easier.

The proton shuttle with Glu64

Our quantum chemical calculations indicate that Glu64 in yCD acts as a proton shuttle with the Zn-ligated water.⁸ In the free form crystal structure, there is a second Zn that pushes Glu64 away from this water molecule, which makes proton transfer from it to

Glu64 improbable. In the free form MD simulation, where this second Zn atom was removed, the OE1 or OE2 of Glu64 forms a strong hydrogen bond with the Zn-bound water, with distance ~ 2.60 Å in both subunit one and subunit two (Table 5-4). Interestingly, the side-chain carboxyl group of Glu64 can rotate in subunit two despite the strong hydrogen bond (Figure 5-7a). Our calculations⁸ indicate that, in the apo enzyme, the proton stays covalently bound to the water, versus transferring to protonate Glu64.

After cytosine binds to the protein (yCD water/cytosine model, 1 in Figure 5-2), Glu64 still forms a strong hydrogen bond with the Zn-bound water (with distance OE1...O 2.59 Å, or OE2...O 2.60 Å) in subunit one, and OE2...O 2.60 Å in subunit two (Table 5-5). The difference between them comes from the rotation of the Glu64 carboxyl group in subunit one but not in subunit two (Figure 5-7b). As discussed above, after the proton transfers from Zn bound water to the carboxyl O of Glu64, the simulation (yCD hydroxide/Glu64H cytosine model, 2 in Figure 5-2) shows a slightly different active site conformation in subunit one and subunit two. The apparent difference is attributable to the dihedral angle φ formed by the Glu64H CB-CG-CD-OE2 atoms with $\varphi \sim 0^\circ$ (all four atoms in the same plane with the terminal atoms *cis* for subunit one, which we shall refer to as state 1 and $\varphi \sim 120^\circ$ for subunit two (state 2). In state 1, the C-OE1-OE2-H plane is close to parallel with the cytosine pyrimidine ring plane and OE2-H is hydrogen bonded to the Zn-bound hydroxide group as a donor. In state 2, the angle between these two planes is about 60° with the OE2-H hydrogen bonded to the same hydroxide group. In both cases, this hydrogen bond is quite strong, with distance OE2...O 2.60 Å in subunit one and 2.62 Å in subunit two. However, the ONIOM calculation⁸ indicates that these two protonated Glu64s are not stable states unless the OE2-H points to N3 of cytosine

and forms a hydrogen bond with it (**2'** in Figure 5-2), instead of pointing to the O atom of the Zn-bound hydroxide. This could be done simply by rotating the dihedral angle CB-CG-CD-OE2 of Glu64 to $\varphi \sim -50^\circ$. This geometry defines state 3, where OE2-H points to N3 and the C-OE1-OE2-H plane is again parallel with the cytosine pyrimidine ring.

From the MD perspective, where chemical species are fixed during a simulation, state 1 and state 2 can be stable in the sense of corresponding to minima on a free energy surface. Therefore, it is of interest to determine the potential of mean force (PMF) $W(\varphi)$ along the Glu64H carboxyl dihedral coordinate, φ , formed by the atoms CB-CG-CD-OE2. The umbrella sampling method that we use to obtain the PMF is described in the Methods Section. Figure 5-8 displays the PMF and shows that state 3 is only 2 kcal/mol higher than state 1. Because state 3 is not a minimum of $W(\varphi)$, configurations where the Glu64 carboxyl OH point toward the cytosine N3 atom in the simulation trajectory are transient. State 2 has the lowest free energy, but the barrier to rotate from state 2 to state 1 is about 6 kcal/mol, and that makes the dihedral angle rotation somewhat difficult. The rotation frequency estimated from transition state theory is around 0.5 ns^{-1} and this frequency is consistent with the one or no transition behavior observed in Figure 5-7. Thus, it is more likely that a proton transfers from the Zn bound water to form Glu64 carboxyl OH, followed by a dihedral angle CB-CG-CD-OE2 rotation of $\sim 50^\circ$ to form a stable state with the Glu64 OE2H pointing to the N3 of cytosine. That the MD force field does not produce a stable state for state 3 may be a consequence of the deficiency of the force field, or may reflect the short time scale of the simulation that is used to construct the PMF. After the proton transfers from Glu64 to the N3 of cytosine, (Figure 5-1b) a 2 ns simulation of the yCD hydroxide/Glu64 cytosineH complex (**3** in Figure 5-2), as

described in the Methods Section, was carried out. A hydrogen bond analysis of the trajectory shows that the Glu64 carboxyl group forms a stable hydrogen bond with the amino group of cytosine in both subunit one and subunit two (Table 5-7). Importantly, the geometry enforced by this new hydrogen bond places C4 of cytosine in a favorable position for nucleophilic attack by the Zn-bound hydroxide (The O-C4 distance is ~ 2.8 Å in both active sites.).

Following the nucleophilic attack of the Zn-bound hydroxide group on C4 of cytosine, the proton of the hydroxide group (now C4OH) could either stay with the tetrahedral intermediate I (4 in Figure 5-2) or transfer to the Glu64 intermediate II (5 in Figure 5-2). The intermediate I (intermediate II) state was modeled in the active site of subunit one (two) (see Methods Section) and a new MD simulation (yCD intermediate model) was performed. In the intermediate I state (Table 5-8), both OE1 and OE2 of Glu64 can form hydrogen bonds with the N3 and NH₂ of the intermediate I, which indicates that the carboxyl group of Glu64 can once again rotate during the simulation (Figure 5-7c). The Glu64 carboxylate forms a strong hydrogen bond with C4OH (distance 2.65 Å), suggesting that it is easier for the C4OH proton to transfer to the carboxyl group of Glu64 than to some other residue. After this proton transfer, to form the intermediate II state, the Glu64 OE2H forms hydrogen bonds with both N4 and O4, with a slight preference for N4 (occurrence 100%, distance 2.72 Å) than O4 (occurrence 84%, distance 2.83 Å), as shown in Table 5-8. Therefore, in a manner similar to the first proton transfer, after Glu64 receives the proton, the Glu64 CB-CG-CD-OE2 rotates to point OE2H to N4.

In summary, along the reaction path, the rotation of the Glu64 dihedral CB-CG-CD-OE2 is critical for enabling the various proton transfers required to form intermediate II. The free energy simulations show that such rotations can happen on a ns timescale, and suggest that Glu64 acts as a proton shuttle multiple times.

The hydrogen bond network

In addition to the critical interactions between Glu64, the Zn-coordinated water or hydroxide, and the substrate, intermediates or products, the enzyme maintains a network of hydrogen bonds for substrate binding and catalysis. The hydrogen bonds involving the side-chain amide of Asn51, the NH's of Gly63 and Cys91, and the carboxylate of Asp155 are maintained throughout the reaction path (Figure 5-2 and Tables 5-(5-9)). These hydrogen bonds help to stabilize all the complexes in the reaction path and therefore are important for both substrate binding and catalysis. In particular, the hydrogen bond between the NH group of Cys91 and the Zn-bound water (or hydroxide) may acidify the latter group and facilitate the proton transfers to Glu64. The hydrogen bond between the carbonyl O of Ser89 and the amino group of cytosine evolves during the reaction. The hydrogen bond is strongest both in terms of geometry and frequency of occurrence in the yCD hydroxide/Glu64 cytosineH complex (3 in Figure 5-2) and helps the positioning of cytosine for the nucleophilic attack by the Zn-bound hydroxide. As the reaction progresses further, the hydrogen bond weakens. The occurrence frequency of the hydrogen bond between the carbonyl O of Ser89 and ammonia is only 15%. The weakening of the hydrogen bond helps the release of the newly formed ammonia.

Binding mode and free energy difference of ammonia and uracil

After ammonia and uracil are formed, our ONIOM⁸ calculations indicate that uracil has a similar binding mode as the intermediate analog in the crystal structure. The product complex MD simulations (6 and 7 in Figure 5-2) show that uracil forms a hydrogen bond network that is similar to the other ligand complexes (Table 5-9). Ammonia stays on the top of uracil with distance between N and C4 of uracil about 3.0 Å, and forms a hydrogen bond with the carboxyl group of Glu64. When ammonia is exchanged with water, water also forms a similar hydrogen bond with Glu64.

As mentioned above, our recent ONIOM calculation indicates that uracil is released with the help of a water molecule inside the active site.⁸ To explore the possibility of ammonia-water exchange, the difference in binding free energy between water and ammonia was evaluated (see Table 5-10). The forward and backward mutations between water and ammonia in water solution show consistent results, with ~2.8 kcal/mol from water to ammonia and ~-2.7 kcal/mol from ammonia to water. The mutation between water and ammonia in the protein active site was performed with the same scheme. The forward mutation (water to ammonia), which was carried out in the active site of subunit two, gave ~6.4 kcal/mol while the backward mutation which was performed in the active site of subunit one gave ~-4.3 kcal/mol. This indicates that the free energy of chemical mutation is quite sensitive to the environment. The removal of the dummy atom (see Methods) costs the same free energy as that in water solution, which indicates that this contribution is not sensitive to the environment.

The (average) binding free energy difference between water and ammonia therefore is ~-2.6 kcal/mol, which shows that water is a better ligand than ammonia. And, considering that the reaction takes place in water, the water concentration is much higher

than that of the product, ammonia, which makes water movement into the active site easier. Thus, we may conclude that after the formation of ammonia, it is highly probable that ammonia exchanges with water in the active site.

Concluding remarks

The MD simulations of free form yCD, and in complex with its reactant (cytosine), product (uracil), several reaction intermediates, and an intermediate analog performed in this work provide insights into the reaction mechanism and structural requirements of the enzyme. Quantum chemical calculations to obtain appropriate charges for the simulations around the enzyme active site were carried out, since the presence of the Zn ion and its ligands leads to large electronic structure effects that are reflected in the MD force field parameterization. For the free enzyme and intermediate analog complex, the agreements between the simulation results and the X-ray structures are excellent, showing the importance of proper force field parameterization when metals are ligated with residues whose protonation states are dependent on the metal and the nature and number of ligands.

The simulations of the free enzyme and its intermediate analog complex show that the protein N-terminus is quite flexible, while all other parts of the enzyme are quite rigid on the MD time scale. The crystallographic data for the free and intermediate analog complex forms are very similar, raising the issue of how substrate enters and product leaves the active site. We do find that the Phe114 loop region and the C-terminal helix sample different configurations in the two subunits, even though the RMSFs for the CA atoms are almost the same and only around 1 Å. This may correspond to a more-or-less rigid body motion of these regions that could permit entrance of a substrate. The 2 ns

trajectories may not be long enough to provide enough sampling time to reveal the larger motions that may be required for substrate entry.

Rotation of the carboxyl group of Glu64 is critical to position it relative to the cytosine Zn(HOH/OH) complex and its intermediates to facilitate the various proton transfer steps along the reaction pathway. The MD potential of mean force calculation for Glu64 carboxylate rotation suggests that such motion is feasible on a ns time scale. The MD simulations of the reactant and a series of intermediates also reveal that cytosine adjusts its orientation and position to assist in the nucleophilic attack by the Zn(OH). With reference to Figure 5-2, the steps along the reaction pathway can be summarized as follows. First, cytosine and yCD forms an initial complex (1 in Figure 5-2) with cytosine having an orientation rather different from that of the intermediate analog found in the crystal structures. Second, the binding of cytosine facilitates proton transfer from Zn(H₂O) to form Glu64H and Zn(OH) (2 in Figure 5-2). The carboxylic acid group of Glu64H then rotates so that its OH group forms a hydrogen bond with N3 of cytosine (2' in Figure 5-2). Third, Glu64H then transfer its proton to N3 of cytosine to form the yCD hydroxide/Glu64 cytosineH complex (3 in Figure 5-2). The hydrogen bond between the carbonyl oxygen of Ser89 and the amino group of cytosine is strongest at this stage and helps position C4 of cytosine for the subsequent nucleophilic attack by Zn(OH). Fourth, after nucleophilic attack to produce intermediate I (4 in Figure 5-2), the MD shows that Glu64 hydrogen bonds to N3 (and NH₂) and can again rotate to be re-protonated by C4OH to form intermediate II (5 in Figure 5-2). Fifth, the MD trajectory of intermediate II leads to Glu64H hydrogen bonding to N4 and C4O, which implies that Glu64H has

rotated again to point to N4, facilitating decomposition of the tetrahedral complex to the product complex (6 in Figure 5-2).

In the MD simulation of the product complex, ammonia forms a stable hydrogen bonds with Glu64. On the ns timescale, it does not leave the active site. In view of our previous study⁸ that proposed a mechanism for the release of uracil that relies on a water molecule replacing ammonia in its active site, a water molecule was exchanged for ammonia (7 in Figure 5-2) and, in the subsequent MD simulation, this water forms a similar hydrogen bond with Glu64. Our evaluation of the binding free energy difference between water and ammonia shows that the difference is sufficiently small that, when coupled with the much greater water than ammonia concentration in the active site vicinity, it is likely for ammonia to readily exchange with water.

References

- (1) Nishiyama, T.; Kawamura, Y.; Kawamoto, K.; Matsumura, H.; Yamamoto, N.; Ito, T.; Ohyama, A.; Katsuragi, T.; Sakai, T. *Cancer Res* 1985, *45*, 1753.
- (2) Dipiro, J. T.; Talbert, R. L.; Yee, G. C.; Matzke, G. R.; Wells, B. G. *Pharmacology: A Pathophysiologic Approach*, 3rd ed.; Appleton and Lange: Stamford, CT, 1997.
- (3) Morris, S. M. *Mutat. Res.* 1993, *297*, 39.
- (4) Ireton, G. C.; Black, M. E.; Stoddard, B. L. *Structure* 2003, *11*, 961.
- (5) Ko, T. P.; Lin, J. J.; Hu, C. Y.; Hsu, Y. H.; Wang, A. H. J.; Liaw, S. H. *J. Biol. Chem.* 2003, *278*, 19111.
- (6) Xiang, S. B.; Short, S. A.; Wolfenden, R.; Carter, C. W. *Biochemistry* 1995, *34*, 4516.
- (7) Xiang, S. B.; Short, S. A.; Wolfenden, R.; Carter, C. W. *Biochemistry* 1997, *36*, 4768.
- (8) Sklenak, S.; Yao, L.; Cukier, R. I.; Yan, H. *J. Am. Chem. Soc.* 2004, *126*, 14879.
- (9) Dudev, T.; Lim, C. *J. Phys. Chem. B* 2001, *105*, 10709.
- (10) Dudev, T.; Lim, C. *J. Am. Chem. Soc.* 2002, *124*, 6759.
- (11) Simonson, T.; Calimet, N. *Proteins* 2002, *49*, 37.
- (12) Banci, L. *Curr Opin Chem Biol* 2003, *7*, 143.
- (13) Hoops, S. C.; Anderson, K. W.; Merz, K. M. *J. Am. Chem. Soc.* 1991, *113*, 8262.
- (14) Suarez, D.; Merz, K. M. *J. Am. Chem. Soc.* 2001, *123*, 3759.
- (15) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput Phys Commun* 1995, *91*, 1.
- (16) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; III, T. E. C.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsue, V.; Gohlke, H.; Radmer, R.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, C.; Weiner, P.; Kollman, P. A. AMBER7; University of California: San Francisco, 2002.
- (17) Diaz, N.; Suarez, D.; Merz, K. M. *J. Am. Chem. Soc.* 2000, *122*, 4197.

- (18) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* 1996, *118*, 2309.
- (19) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* 1977, *23*, 327.
- (20) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, G. L. *J. Chem. Phys.* 1995, *103*, 8577.
- (21) Berendsen, H. H. C.; Postma, J. P. M.; Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* 1984, *81*, 3684.
- (22) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Letts.* 1974, *28*, 578.
- (23) Frenkel, D.; Smit, B. *Understanding Molecular Simulation : from algorithms to applications*; Academic Press: San Diego, 1996.
- (24) Souaille, M.; Roux, B. *Comput Phys Commun* 2001, *135*, 40.
- (25) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* 2003, *107*, 9535.
- (26) Kraulis, P. J. *J. Appl. Crystallogr.* 1991, *24*, 946.
- (27) Bacon, D. J.; Anderson, W. F. *J. Mol. Graphics* 1988, *6*, 219.
- (28) Merritt, E. A.; Bacon, D. J. *Methods Enzymol.* 1997, *277*, 505.

Appendices

Table 5-1 Force constants for the Zn complex.

K_r (kcal/mol-Å ²)	Zn-S	Zn-O	Zn-N	
	300	300	300	
K_θ (kcal/mol-rad ²)	S-Zn-S	S-Zn-O	S-Zn-N	O-Zn-N
	75	75	75	75
	Zn-O-C	Zn-S-C	Zn-O-H	Zn-N-C
	35	35	35	35

Table 5-2 Charges (*e* units) of Zn complex used in the yCD simulations

		1	2	3	4	5	6
His62	CB	-0.168	-0.420	-0.708	0.047	0.129	-0.278
	HB2	0.082	0.105	0.266	0.063	0.045	0.150
	HB3	0.082	0.105	0.266	0.063	0.045	0.150
	CG	0.258	0.241	0.284	0.005	-0.051	0.074
	ND1	-0.283	-0.111	-0.303	-0.168	-0.159	-0.143
	CE1	-0.061	-0.195	0.021	-0.100	-0.082	-0.019
	HE1	0.120	0.175	0.159	0.168	0.169	0.133
	NE2	0.071	0.002	-0.336	-0.086	-0.119	-0.097
	HE2	0.233	0.276	0.416	0.326	0.335	0.281
	CD2	-0.427	-0.401	-0.168	-0.323	-0.291	-0.342
	HD2	0.313	0.268	0.195	0.218	0.220	0.254
Cys91/	CB	0.221	-0.006	0.215	0.08	0.059	0.027
Cys94	HB2	-0.035	0.016	-0.056	0.019	0.026	-0.002
	HB3	-0.035	0.016	-0.056	0.019	0.026	-0.002
	SG	-0.644	-0.568	-0.729	-0.703	-0.723	-0.673
Zn	Zn	0.762	0.609	1.027	0.876	0.887	0.668

(1) Data used in free form and water/cytosine active site simulations.

(2) Data used in intermediate analog complex simulation.

(3) Data used in hydroxide/Glu64H cytosine complex and hydroxide/Glu64 cytosineH complex simulations.

(4) Data used in intermediate I complex simulation.

(5) Data used in intermediate II complex simulation

(6) Data used in product uracil/ammonia (and uracil/water) complex

Table 5-3 RMSD for free and intermediate analog complex forms

RMSD (Å)				
	free form		intermediate analog complex	
	Subunit 1	Subunit 2	Subunit 1	Subunit 2
CA	0.72 (0.06)	1.01 (0.07)	0.90 (0.14)	0.77 (0.05)
Total ^(a)	1.29 (0.06)	1.60 (0.07)	1.46 (0.11)	1.24 (0.05)
N-terminal	2.28 (0.24)	3.43 (0.78)	2.65 (0.32)	3.66 (0.96)
C-terminal	0.98 (0.11)	1.50 (0.11)	1.40 (0.37)	1.32 (0.23)
Loop	1.41 (0.19)	1.93 (0.22)	1.97 (0.32)	1.20 (0.09)

^(a) Not including residues 1-14.

Table 5-4 Hydrogen bonds in yCD free and intermediate analog complex forms

Subunit 1	Occurrence (%)	Distance (Å)	Crystal structure
Intermediate complex			
Asn51–NH...O2	99.47	2.82 (0.10)	3.14
Asp155–OD...H–N1	100.00	2.70 (0.08)	2.69
Gly63–NH...O2	92.68	2.94 (0.12)	2.89
Glu64–OE1...H–O4	37.42	3.06 (0.11)	3.34
Glu64–OE2...H–O4	99.73	2.50 (0.07)	2.49
Cys91–NH...O4	22.90	3.10 (0.07)	3.02
Glu64–OE1...N3	99.87	2.79 (0.09)	2.80
Free form			
Glu64–OE2...H–WAT	100.00	2.62 (0.10)	
Subunit 2			
Intermediate complex			
Asn51–NH...O2	98.80	2.85 (0.11)	2.90
Asp155–OD...N1	100.00	2.69 (0.07)	2.69
Gly63–NH...O2	89.21	2.97 (0.12)	2.87
Glu64–OE1...H–O4	39.41	2.95 (0.24)	3.37
Glu64–OE2...H–O4	95.61	2.53 (0.11)	2.52
Cys91–NH...O4	21.04	3.12 (0.06)	3.03
Glu64–OE1...N3	91.48	2.80 (0.09)	2.78
Free form			
Glu64–OE2...H–WAT	73.60	2.57 (0.10)	
Glu64–OE1...H–WAT	28.93	2.60 (0.16)	

Table 5-5 Hydrogen bonds in yCD water/cytosine simulation

	Occurrence (%)	Distance (Å)	Occurrence (%)	Distance (Å)
	Subunit 1		Subunit 2	
Asn51–NH...O2	99.73	2.82 (0.10)	99.60	2.84 (0.11)
Asp155–OD...H–N1	99.07	2.84 (0.11)	100.00	2.79 (0.10)
Gly63–NH...O2	75.87	2.99 (0.12)	84.80	2.97 (0.12)
Glu64–OE1...H1–N4	8.27	2.94 (0.12)	88.00	2.88 (0.13)
Glu64–OE1...H2–N4	32.93	2.87 (0.14)		
Glu64–OE2...H1–N4	23.20	2.86 (0.13)	9.33	3.04 (0.10)
Glu64–OE2...H2–N4	50.67	2.84 (0.13)	31.07	2.88 (0.13)
Glu64–OE1...H–WAT	70.80	2.59 (0.09)		
Glu64–OE2...H–WAT	29.33	2.60 (0.09)	100.00	2.60 (0.10)
Cys91–NH...O–WAT	98.93	2.92 (0.10)	98.40	2.91 (0.09)
Ser89–O...H2–N4			49.20	2.91 (0.13)

Table 5-6 Hydrogen bonds in yCD hydroxide/Glu64H cytosine simulation

	Occurrence (%)	Distance (Å)	Occurrence (%)	Distance (Å)
	Subunit 1		Subunit 2	
Asn51–NH...O2	92.67	2.91 (0.12)	98.67	2.84 (0.11)
Asp155–OD...H–N1	99.33	2.78 (0.10)	36.27	2.79 (0.11)
Gly63–NH...O2	89.33	2.96 (0.12)	63.47	3.01 (0.11)
Glu64–OE1...H1–N4	79.73	2.94 (0.13)	79.33	2.90 (0.13)
Glu64–OE2H...OH	97.73	2.60 (0.10)	99.87	2.62 (0.10)
Cys91NH...OH	99.07	2.91 (0.10)	90.40	2.99 (0.11)
Ser89–O...H2–N4	79.33	2.96 (0.12)		

Table 5-7 Hydrogen bonds in yCD hydroxide/Glu64 cytosineH simulation

	Occurrence (%)	Distance (Å)	Occurrence (%)	Distance (Å)
	Subunit 1		Subunit 2	
Asn51–NH...O2	63.73	3.03 (0.11)	74.53	2.99 (0.11)
Asp155–OD...H–N1	100.00	2.70 (0.08)	100.00	2.71 (0.08)
Gly63–NH...O2	54.27	3.04 (0.10)	82.00	2.99 (0.11)
Glu64–OE1...H1–N4	100.00	2.73 (0.08)	99.87	2.72 (0.07)
Glu64–OE1...H–N3	44.67	3.07 (0.10)	38.40	3.07 (0.10)
Glu64–OE2...H–N3	99.87	2.80 (0.09)	100.00	2.79 (0.09)
Cys91NH...OH	100.00	2.82 (0.08)	100.00	2.83 (0.08)
Ser89–O...H2–N4	80.00	2.87 (0.13)	91.60	2.84 (0.12)

Table 5-8 Hydrogen bonds in yCD intermediate I/II simulations

	Occurrence (%)	Distance (Å)	Occurrence (%)	Distance (Å)
	Intermediate I		Intermediate II	
Asn51–NH...O2	99.07	2.86 (0.11)	100.00	2.84 (0.10)
Asp155–OD...H–N1	100.00	2.70 (0.08)	100.00	2.76 (0.09)
Gly63–NH...O2	88.00	2.97 (0.12)	89.07	2.96 (0.12)
Glu64–OE1...H1–N4	22.93	3.02 (0.10)		
Glu64–OE2...H1–N4	49.07	3.02 (0.11)		
Glu64–OE1...H–O4	34.80	2.65 (0.19)		
Glu64–OE2–H...O4			83.60	2.83 (0.13)
Glu64–OE2–H...N4			100.00	2.72 (0.11)
Glu64–OE1...H–N3	30.80	2.86 (0.10)	100.00	2.83 (0.10)
Glu64–OE2...H–N3	69.47	2.84 (0.10)		
Cys91NH...O4	96.13	2.97 (0.10)	97.47	2.96 (0.11)
Ser89–O...H2–N4	40.27	3.03 (0.11)	36.67	3.06 (0.10)

Table 5-9 Hydrogen bonds in yCD uracil/ammonia and uracil/water simulations.

	Occurrence (%)	Distance (Å)	Occurrence (%)	Distance (Å)
	uracil/ammonia		uracil/water	
Asn51–NH...O2	97.87	2.87 (0.12)	96.93	2.88 (0.12)
Asp155–OD...H–N1	100	2.69 (0.08)	100.00	2.70 (0.07)
Gly63–NH...O2	79.2	2.99 (0.11)	87.60	2.98 (0.12)
Glu64–OE2...H–N3	99.87	2.83 (0.09)	99.07	2.86 (0.11)
Cys91NH...O4	60.27	3.07 (0.09)	45.2	3.08 (0.08)
Glu64–OE1...H1–NH3 /(H1–WAT)	94.67	2.84 (0.11)	100.00	2.65 (0.11)
Ser89–O...H2–NH3 /(H2–WAT)	14.53	3.05 (0.10)	23.47	2.81 (0.17)

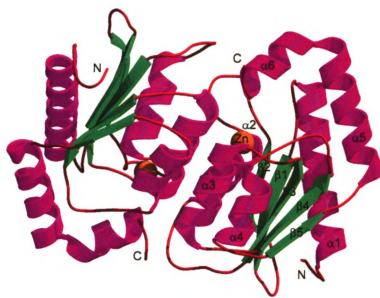
Table 5-10 Chemical mutation free energies between water and ammonia.

direction of mutation	In water		In protein	
	Alchemy	Restraint ^(a)	Alchemy	Restraint ^(a)
NH ₃ to H ₂ O (kcal/mol)	–2.74	–8.36	–4.32 ^(b)	–8.36
H ₂ O to NH ₃ (kcal/mol)	2.82	8.36	6.41 ^(c)	8.36

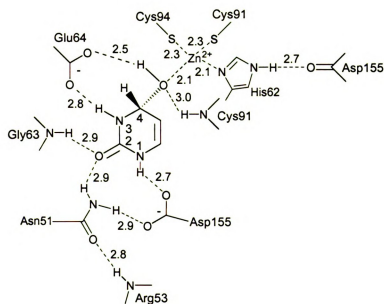
^(a) A 1M standard state was used as a reference state for the gas phase of the dummy atom.

^(b) Result from subunit 1.

^(c) Result from subunit 2.



(a)



(b)

Figure 5-1. (a) Ribbon representation of the crystal structure of yCD drawn according to the coordinate of the 1.14 Å crystal structure.⁴ The figure was prepared with Molscript²⁶ and Raster3D.^{27,28} (b) Schematic drawing of the coordination sphere of the catalytic zinc ion and polar interactions between yCD and a reaction intermediate analog as revealed by X-ray crystallography.^{4,5} The distances (Å) between the zinc ion and its ligands and between the heavy atoms involved in hydrogen bonds are obtained from one subunit (A) of the 1.14 Å crystal structure of yCD with the reaction intermediate complex.

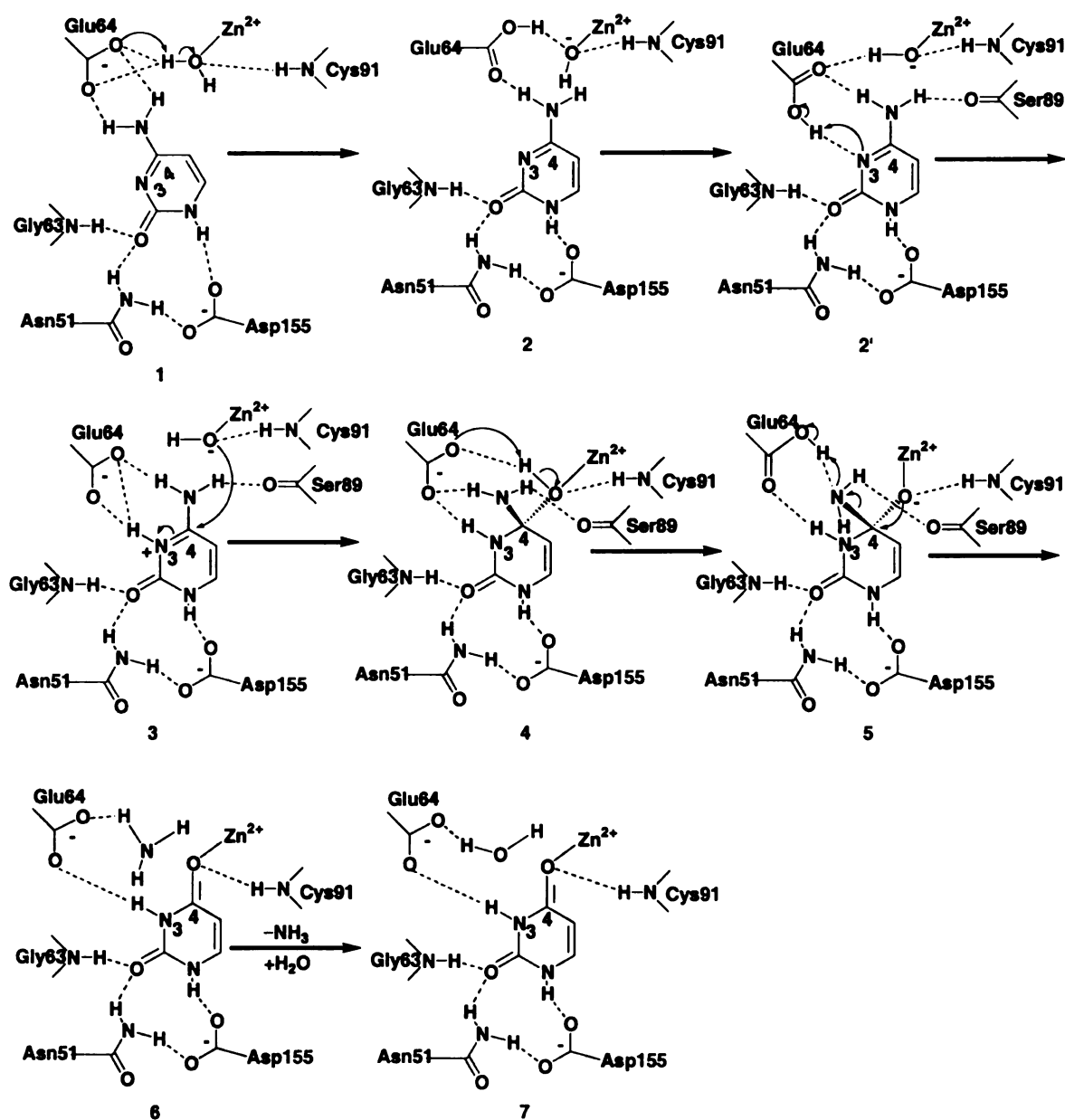


Figure 5-2 Schematic representation of the path of the yCD-catalyzed reaction as revealed by our quantum chemical calculations⁸ and molecular dynamic simulations (this work): 1. water/cytosine active site. 2. hydroxide/Glu64H cytosine complex. 3. hydroxide/Glu64 cytosineH complex. 4. intermediate I complex. 5. intermediate II complex. 6. uracil/ammonia complex. 7. uracil/water complex.

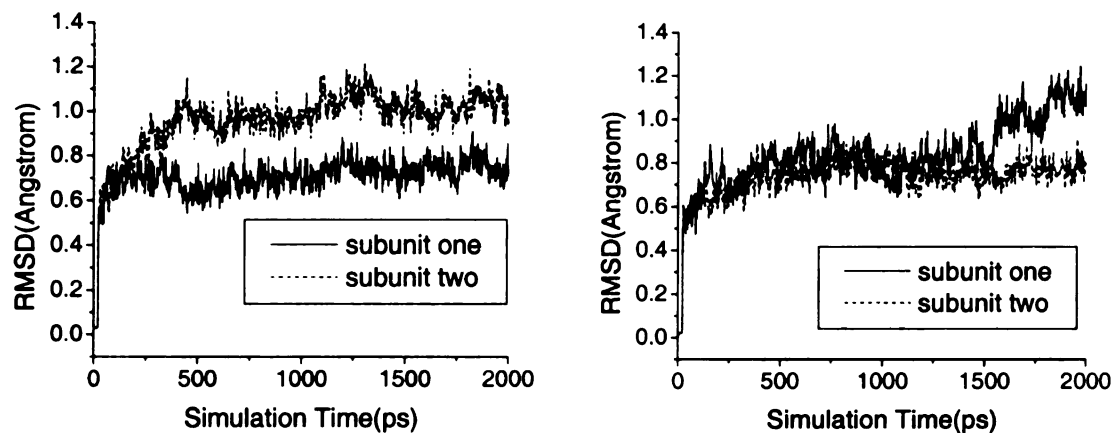


Figure 5-3 RMSDs of CA (residues 15-158) for subunit one and subunit two compared with the corresponding crystal structures. The first 500 ps were treated as an equilibration stage, while the last 1500 ps were used to do the data analysis. Left panel: free form, right panel: intermediate analog complex.

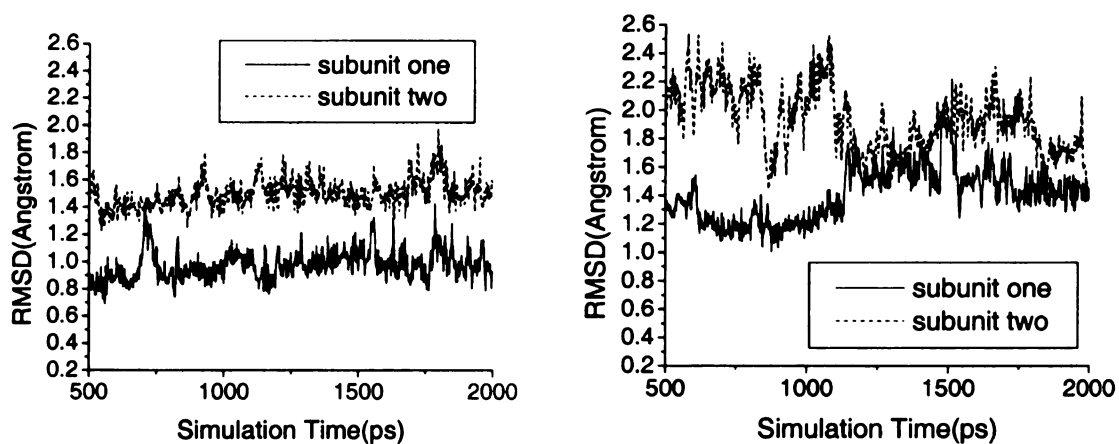


Figure 5-4 (a) RMSDs from the crystal structure of the C-terminal helix (left panel) and (b) the Phe114 loop (right panel) in the free yCD simulation, after equilibration.

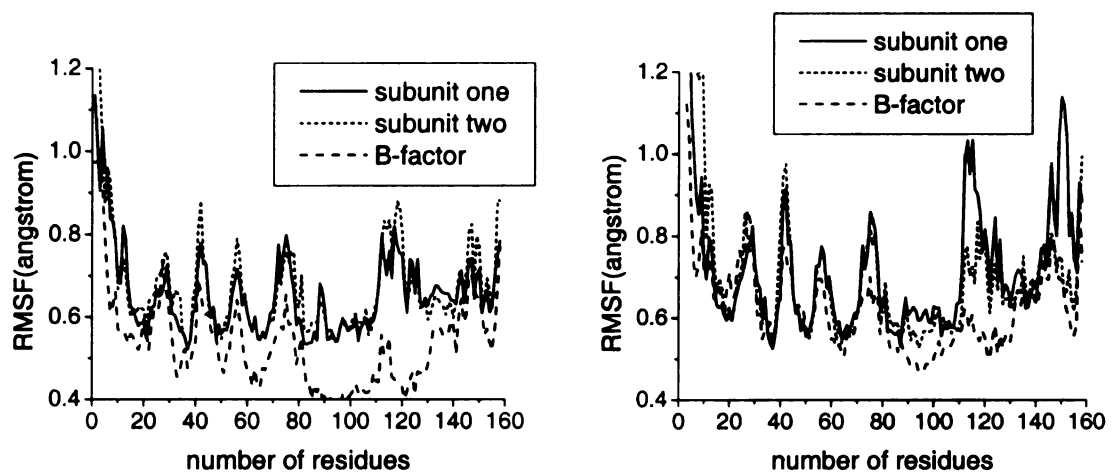


Figure 5-5 RMSFs of CAs in the yCD simulations compared with the crystal structure B-factors. (a) Left panel: free yCD simulation. (b) Right panel: yCD intermediate analog complex.

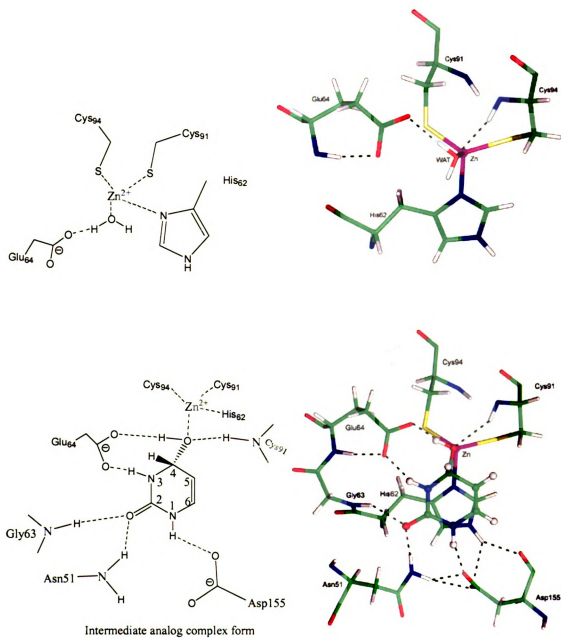


Figure 5-6 Schematic representation and MD snapshots of the yCD active site. Top panel: free form. Bottom panel: intermediate analog complex form.

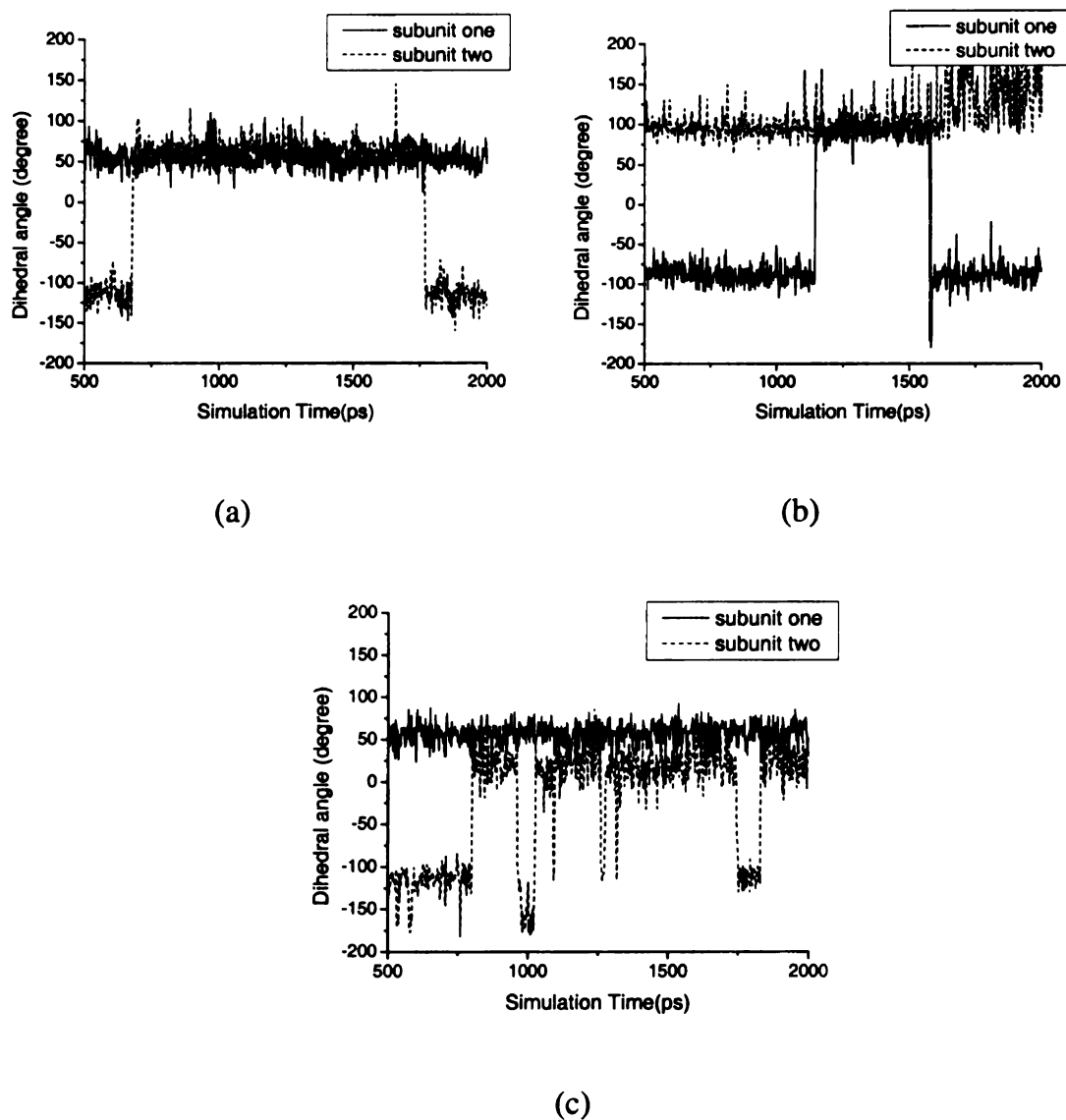


Figure 5-7 Changes in the CB-CG-CD-OE1 dihedral angle of Glu64. (a). Free yCD MD simulation. (b). yCD water/cytosine model (see 1 in Figure 2). (c) yCD intermediate I complex (see 4 in Figure 2).

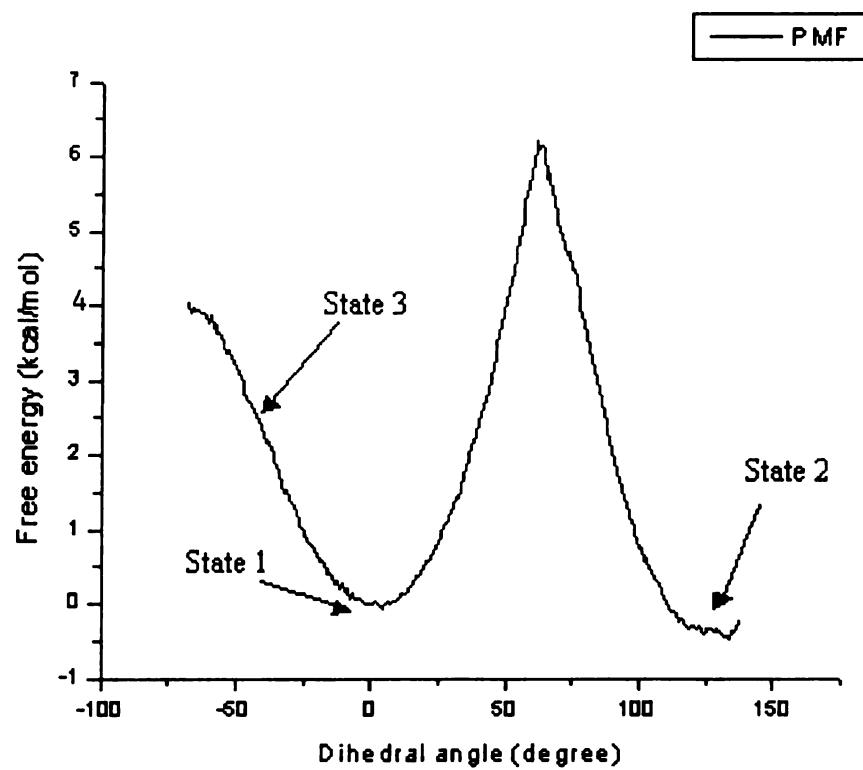


Figure 5-8 Potential of mean force for rotation around the CB-CG-CD-OE2 dihedral angle of Glu64H.

Chapter 6

Product Release Study of Yeast Cytosine Deaminase by a Combination Method of ONIOM and Molecular Dynamics Simulation

Introduction

Yeast cytosine deaminase (yCD) catalyzes the deamination of cytosine to uracil as well as 5-fluorocytosine (5-FC) to the anticancer drug 5-fluorouracil (5-FU). It has been observed that product release is the rate-limiting step during the activation of 5-FC (Yao et al. 2005). One reason for this slow product release might be the Zn-O4 bond formed between the protein and product 5-FU. Sklenak (Sklenak et al. 2004) showed that the direct Zn-O4 bond cleavage is nearly impossible based on two-layer ONIOM (B3LYP/PM3) quantum mechanics calculations (Maseras and Morokuma 1995; Svensson et al. 1996; Vreven et al. 2001; Vreven et al. 2003). Instead, an oxygen exchange mechanism was proposed to assist this bond breaking, which gives the highest barrier ~9 kcal/mol for the activation energy of the whole reaction cycle. However, our recent experiments (unpublished data) show that, though this oxygen exchange occurs, it is too slow and unlikely to be the mechanism for the Zn-O₄ cleavage.

One potential drawback of Sklenak's ONIOM study is that only ~25 residues around active site were used in calculations due to the high cost of QM methods. Since only a fraction of the protein was selected, atoms in the outer-layer have to be fixed, otherwise they will move away unrealistically, while atoms in the inner layer (including the ligand, Zn, Zn bound Cys91, Cys94 and His62, and catalytic Glu64) are optimized. If the rearrangement of the active site is rather small, this protocol might work well, which appears to be true in steps before Zn-O4 bond cleavage based on all atom molecular dynamics (MD) simulations (Yao et al. 2005). But, when the rearrangement is large, some atoms in the outer layer have to move, otherwise the ONIOM method used will fail. In this case, dynamics of the whole protein or at least the residues around the active site have to be incorporated into the calculation. QM/MM simulation methods have to be used to fully describe this process. But the challenge is that in order to model complexes containing Zn ions, reliable QM methods have to be used (Dudev and Lim 2001; Sklenak et al. 2004) which is very costly if used in simulation. For example, one single point calculation of the ONIOM yCD system takes ~80 min in a two-processor Linux workstation with 2.1 GHz CPU speed. With the MD time step 2 fs, a 1 ns QM/MD simulation will cost years. Therefore, an approximation has to be made to carry out the study with reasonable time cost. In this chapter, one protocol was proposed to calculate the potential of mean force for Zn-O4 bond cleavage by combining conventional MD simulation and ONIOM method.

Method

A two-step method combining MD and ONIOM

Assuming a protein molecule solvated with explicit water molecules in a box, one attempts to do a QM/MM simulation with the treatment of the ligand and catalytically important residues with a high level QM method (core layer), surrounding residues in the active site with a low level QM method (medium layer) and the rest of the protein including water with molecular mechanics (outer layer). With the Born-Oppenheimer approximation and assuming the electronic degrees of freedom have been integrated out, the Hamiltonian of the whole system is:

$$H_{tot} = H_h^q + H_l^q + H_{hl}^q + H_r^m + H_{hr}^{qm} + H_{lr}^{qm} \quad (1)$$

where superscript q means the term is treated with quantum mechanics and m means the term is treated classical mechanics and h , l and r denote high (quantum), low (quantum) and rest (molecular mechanics), respectively. The interaction between high/low level QM part and the rest of the system is described by QM/MM coupling term where the QM atoms electron clouds can see the MM atoms (classic particles with partial charges). For a system with N particles, the canonical ensemble distribution function is:

$$\rho(p^{3N}, q^{3N}) = \frac{\exp(-\beta H_{tot}(p^{3N}, q^{3N}))}{\int \exp(-\beta H_{tot}(p^{3N}, q^{3N})) dp^{3N} dq^{3N}}$$

where p^{3N} and q^{3N} are vectors defining momentum and displacement of the whole system, β is $1/kT$ in which k is Boltzmann constant, T is temperature in Kelvin. If we are interested in a chemical process in the core layer, we need to define one specific reaction coordinates r (e.g. distance between two atoms in a bond cleavage). One coordinate transformation needs to be introduced from Cartesian coordinates to a set of generalized coordinates that contains r :

$$q^{3N} \rightarrow (\mu^{3N-1}, r)$$

In addition, a corresponding transformation is made on the conjugate momentum according to:

$$p^{3N} \rightarrow (p_{\mu}^{3N-1}, p_r)$$

so that

$$dp^{3N} q^{3N} = dp_{\mu}^{3N-1} d\mu^{3N-1} dp_r dr$$

Thus the potential of mean force is:

$$\begin{aligned} PMF(R) &= -\frac{1}{\beta} \ln \int \exp(-\beta H_{tot}(p_{\mu}^{3N-1}, \mu^{3N-1}, p_r, r)) \delta(r-R) dp_{\mu}^{3N-1} d\mu^{3N-1} dp_r dr \\ &= -\frac{1}{\beta} \ln \int \exp(-\beta H_{tot}(p^{3N}, q^{3N})) \delta(r-R) dp^{3N} dq^{3N} \end{aligned}$$

The mean force is:

$$-\frac{\partial PMF(R)}{\partial R} = -\left\langle \frac{\partial H_{tot}}{\partial r} \right\rangle_{r=R} \quad (2)$$

in which $\langle \dots \rangle_{r=R}$ is a conditional ensemble average with the condition $r=R$. Since r is only a function of coordinates of certain atoms in the high-level QM part, equation (2) can be further simplified by

$$\frac{\partial PMF(R)}{\partial R} = \left\langle \frac{\partial (H_h^q + H_{hl}^q + H_{hr}^{qm})}{\partial r} \right\rangle_{r=R} \approx \left\langle \frac{\partial (H_h^q + H_{hl}^q)}{\partial r} \right\rangle_{r=R} \quad (3)$$

provided the interaction between core layer and outer layer is very small. The outer layer influences the mean force through the ensemble average. To estimate this average, certain number of MD snapshots need to be generated with the constraint $r=R$. Then, the mean force can be evaluated by equation 3. By integrating equation (3), PMF will be:

$$PMF(r_1) - PMF(r_0) = \int_{r_0}^{r_1} \frac{\partial PMF(r)}{\partial r} dr = \int_{r_0}^{r_1} \left\langle \frac{\partial H_h^q}{\partial r'} \right\rangle_{r'=r} dr \quad (4)$$

But, as discussed above, this QM/MM simulation is impractical with the current computer resources due to the expensive QM calculation. Certain approximations have to be introduced to evaluate the mean force more efficiently. We propose a two-step method to approximate this QM/MM simulation. 1. One MM simulation where all the atoms treated classically is performed with the restraint $r=R$. 2. A certain number of snapshots are chosen for a two-layer ONIOM optimization with the high and low level QM method, the same used in the "ideal" QM/MM simulation. In ONIOM optimization, the low level layer is fixed and the high level layer is fully optimized with the constraint $r=R$. Then the force between two atoms can be calculated and averaged over the snapshots by using equation (3). Repeating step 1 and 2 with different constraint distances, PMF can be calculated by using equation (4). Since the chemical bond cleavage process cannot be fully described by the MM method, the core layer has to be optimized using high level QM method before the force calculation which acts as a refinement of the core structure.

The main errors introduced by this approximation method come from the core and medium layer in the simulation. For the medium layer, the use of MM will certainly introduce an error. But since there is no chemical reaction occurring in the medium layer, MM should be a reasonable approximation for a low level QM method. If MM of the medium layer responds properly to the constraint changes in the core layer and generates corresponding conformational changes, the ONIOM optimization in step 2 will produce an accurate core region conformation which will give a good evaluation of the force. The second error comes from the MM treatment of the core layer in the simulation. The

inadequate treatment of the core in the simulation may not be able to generate good conformations for force calculation. Thus ONIOM optimization has to be performed in step 2. The third error is higher order error compared to the first two. The erroneous core conformations generated in simulation might cause the erroneous response of the medium layer which in turn will influence the second step ONIOM optimization. To guarantee this error is small, two conditions have to be fulfilled. 1. MM is a good method to describe the initial state, which is likely true with the current force field because the initial state is a well defined state. 2. During the reaction process, the modification of the core is small and then the quantum effect changes to the medium layer are insignificant. In a conclusion, the PMF quality depends primarily on the quality of MM treatment of the core system and the complexity of the reaction process in this two-step protocol.

MD simulation and ONIOM calculation of the YCD uracil complex

A 2-ns MD simulation was previously performed for the yCD uracil complex at constant temperature (300 K) and constant volume (Yao et al. 2005). The active site Zn and uracil O4 are covalently linked by a harmonic potential with the force constant 300 kcal/mol-Å². In order to see what happens after the Zn-O4 bond breaks, one 1ns MD simulation was carried out by starting from one snapshot with the harmonic potential removed. It appears that the Zn-O4 distance increased to ~3.6 Å. One MD snapshot was selected with the Zn-O4 distance 3.86 and trimmed for a two-layer ONIOM (B3LYP/PM3) in Gaussian 03 (Frisch) optimization with the same setup in the previous (Sklenak et al. 2004) study. The outer layer is fixed and treated by PM3, including Ile33, Asn51, Thr60, Leu61, Gly63, Ile65, Leu88, Ser89, Pro90, Asp92, Met93, Thr95, Phe114, Trp152, Phe153, Glu154, Asp155, and Ile156. The inner layer is fully optimized by using

B3LYP(Becke 1993), including uracil, Glu64, Zn and Zn bound residues His62, Cys91, Cys94. The ONIOM optimized Zn-O4 distance is similar to that from MD. Then this distance was scanned from 3.8 Å to 2.0 Å to see whether it can reproduce the relative energy profile from the crystal structure calculation.

Another 1ns MD simulation was performed with the harmonic potential between Zn-O4 maintained but the equilibrium distance changed 0.1 Å every 50 ps from 2.0 Å to 3.8 Å starting from one MD snapshot of the first simulation. The coordinates were saved every 2 ps, the first 20 ps was used as the equilibration period. For each Zn-O distance, 5 snapshots were chosen evenly and optimized by the ONIOM method as described above; then the force between Zn and O4 was calculated based on a Hessian matrix and averaged over 5 snapshots. The two step process is demonstrated in Figure 6-1. The error of the force average is rather small based on the standard deviation (see Results) except at a distance of 2.3 Å where 10 snapshots were used in the ONIOM calculation.

Results and discussion

Scan with the fixed outer layer

After the Glu64 transfers two protons from Zn bound water molecule to cytosine, NH₃ is formed and released, but O4 of uracil is still covalently bound to Zn atom (Sklenak et al. 2004). It appeared that the direct Zn-O4 bond cleavage is extremely difficult in our previous calculation (Sklenak et al. 2004). But one potential problem of the calculation is that the outer layer of the protein was fixed according to the crystal structure. If some of these residues need to rearrange themselves during Zn-O4 bond break, fixing the outer layer will make the barrier artificially high due to the steric clashes. The ONIOM optimizations were carried out to scan the Zn-O4 distance from 2.0

Å to 3.8 Å, based on the Zn-O4 bound product complex structure generated from yCD inhibitor crystal structure (Ireton et al. 2003; Ko et al. 2003; Sklenak et al. 2004). One MD simulation was performed with the bond restraint between Zn and O4 removed, which effectively gives a relaxed system with Zn-O4 bond broken. Then the same ONIOM optimizations were performed based on the MD snapshot. By comparing these two energy profiles, we can see whether the constraint of the outer layer will introduce any significant artifact. As shown in Figure 6-2, the ONIOM energy for the crystal structure shows a monotonically increase from 0 kcal/mol to 15 kcal/mol with Zn-O4 distance changing from 2.0 Å to 3.8 Å, while the ONIOM energy for the MD snapshots decreases along the increasing distances. It suggests that the active site has rearranged itself in MD simulation when Zn-O4 bond is cleaved. Further investigation shows that OE1 of Glu64 moves close to the Zn atom after the cleavage of Zn-O4 bond, due to the electrostatic attraction (Figure 6-3). The partial charge of Zn is +0.67 e and the charge of OE1 of Glu64 is -0.82 e. The average distance between Zn and OE1 is 1.99 ± 0.09 Å while the distance between Zn and O4 is 3.59 ± 0.27 Å in the MD simulation. So OE1 acts as the fourth atom coordinated with Zn after the loss of the Zn-O4 interaction. The water molecule forming the hydrogen bond with the Ser89 backbone carbonyl and Glu64 carboxyl in the starting point was pushed away during the MD simulation (Figure 6-3). This process was not observed during the ONIOM distance scanning of the crystal structure, because the backbone of Glu64 was treated as the outer layer and fixed, which prohibits O4 from moving close to Zn and the water molecule is contained by the outer layer residues. On the other hand, during the Zn-O4 distance scanning of the MD snapshots, the outer layer was also fixed but relaxed for the cleavage of the Zn-O4 bond.

That is probably why the system seems to be more stable with the Zn-O4 bond cleaved in the ONIOM calculations of the MD snapshot. Therefore, the ONIOM calculations based on the crystal structure and MD snapshot are two extremes: the crystal structure is the representation of the covalently bound form Zn-O4 complex, the MD snapshot is the representation of the covalently unbound form of the Zn-O4 complex. One has to scan the distance between Zn-O4 and rearrange the outer layer residues at the same time to generate the energy profile correctly.

Combining ONIOM with MD

As discussed above, the ONIOM method with the static outer layer can't demonstrate the Zn-O4 bond cleavage process. MD has to be combined with ONIOM to take care of the changes. First MD was used to generate snapshots with different Zn-O4 distance restrained, and then ONIOM was used to optimize these snapshots and calculate the forces. It appears that the average of forces at individual Zn-O4 distance converges quickly as shown in Figure 6-4, so only 5 snapshots are used for each distance except 2.3 Å where 10 snapshots are used. The average force decreases dramatically from 9.2 kcal/mol-Å to -12.4 kcal/mol-Å as the distance increases from 2.0 Å to 2.2 Å. Then the force increases to -2.0 kcal/mol-Å at 2.4 Å, after that it increase slowly to ~2 kcal/mol-Å at 2.7 Å and fluctuates around till 3.5 Å when it drops to ~0 kcal/mol-Å. The force curve passes zero three times, the first one in between 2.0 Å and 2.1 Å indicating the first minimum in potential of mean force curve, the second one in between 2.6 Å and 2.7 Å a maximum in the potential curve, the third one in between 3.4 Å and 3.5 Å another minimum in potential. The potential of mean force was calculated by using the discrete summation of the average force as shown in Figure 6-5. It is a typical chemical reaction

process with one transition state which has the Zn-O4 distance approximately 2.6-2.7 Å. The barrier for the reaction is ~2.9 kcal/mol. The corresponding three states are shown in Figure 6-6. The transition state has the Zn penta-coordinated with His62, Cys91, Cys94, Glu64 and uracil because OE1 of Glu64 moves close and coordinates to Zn. In fact the bond between Zn and OE1 is formed at the beginning of Zn-O4 bond cleavage as shown in Figure 6-7. The distance between Zn-OE1 decreases to 2.1 Å when Zn-O4 distance increases to 2.4 Å. In the transition state Zn has a weak interaction with O4 of uracil since the distance is already ~2.6-2.7 Å. After the reaction, the backbone amide of Cys91 forms hydrogen bond with OE1 similar to what occurs in MD simulation (Figure 6-3). Several hydrogen bond interactions between uracil and active site residues were maintained during the reaction, including Glu64 OE2 with N2H, Gly63 NH with O2, Asn55 NH2 with O2, Asp155 OD1 with N1H, consistent with MD simulation (Figure 6-3).

Conclusion

In this study we revisit the Zn-O4 bond cleavage mechanism during uracil release. The mechanism proposed previously is an oxygen exchange mechanism based on a two-layer quantum mechanics ONIOM method (Sklenak et al. 2004), which was observed in experiments (unpublished data). But the rate for the exchange is very slow, and therefore unlikely to be the Zn-O4 cleavage mechanism. It was seen in the previous ONIOM calculation that the direct Zn-O4 bond breaking needs to overcome a huge energy barrier. In this study, we demonstrate that this barrier mainly comes from the constraint of the outer layer that prohibits the proper response of active site residues, especially the E64 carboxyl group. The increase of Zn-O4 distance based on the crystal structure results in a

monotonic increase of energy. But, the increase of Zn-O4 distance based on the MD simulation shows a steady decrease in energy. The reason is that ONIOM optimization from the crystal structure is a good approximation of the Zn-O4 covalently linked state while the ONIOM from MD snapshot is a proper approximation of the Zn-O4 bond broken state. The constraint of the outer layer can't describe the intermediates in between correctly, which makes the ONIOM calculation always favor the starting point. One strategy to solve this problem is to use QM/MM simulation method. But the presence of Zn in the system requires a high level QM method, which makes the simulation impractical. Instead, we proposed a two-step process to mimic this target simulation by combining MM simulation with a two-layer ONIOM QM calculation. We demonstrated that this process can describe the response of the outer layer properly. Zn-O4 bond cleavage occurs through a penta-coordinated Zn complex. And the barrier for this process is rather low, only about 2.9 kcal/mol.

Reference

- Becke, A. D. (1993). "Density-Functional Thermochemistry .3. The Role of Exact Exchange." Journal of Chemical Physics **98**(7): 5648-5652.
- Dudev, T. and C. Lim (2001). "Modeling Zn²⁺-cysteinate complexes in proteins." Journal of Physical Chemistry B **105**(43): 10709-10714.
- Frisch, M. J. "Gaussian 03."
- Ireton, G. C., M. E. Black, et al. (2003). "The 1.14 angstrom crystal structure of yeast cytosine deaminase: Evolution of nucleotide salvage enzymes and implications for genetic chemotherapy." Structure **11**(8): 961-972.
- Ko, T. P., J. J. Lin, et al. (2003). "Crystal structure of yeast cytosine deaminase - Insights into enzyme mechanism and evolution." Journal of Biological Chemistry **278**(21): 19111-19117.
- Maseras, F. and K. Morokuma (1995). "Imomm - a New Integrated Ab-Initio Plus Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition-States." Journal of Computational Chemistry **16**(9): 1170-1179.
- Sklenak, S., L. S. Yao, et al. (2004). "Catalytic mechanism of yeast cytosine deaminase: An ONIOM computational study." Journal of the American Chemical Society **126**(45): 14879-14889.
- Svensson, M., S. Humbel, et al. (1996). "ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition." Journal of Physical Chemistry **100**(50): 19357-19363.
- Vreven, T., B. Mennucci, et al. (2001). "The ONIOM-PCM method: Combining the hybrid molecular orbital method and the polarizable continuum model for solvation. Application to the geometry and properties of a merocyanine in solution." Journal of Chemical Physics **115**(1): 62-72.
- Vreven, T., K. Morokuma, et al. (2003). "Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints." Journal of Computational Chemistry **24**(6): 760-769.
- Yao, L. S., Y. Li, et al. (2005). "Product release is rate-limiting in the activation of the prodrug 5-fluorocytosine by yeast cytosine deaminase." Biochemistry **44**(15): 5940-5947.
- Yao, L. S., S. Sklenak, et al. (2005). "A molecular dynamics exploration of the catalytic mechanism of yeast cytosine deaminase." Journal of Physical Chemistry B **109**(15): 7500-7510.

Appendices

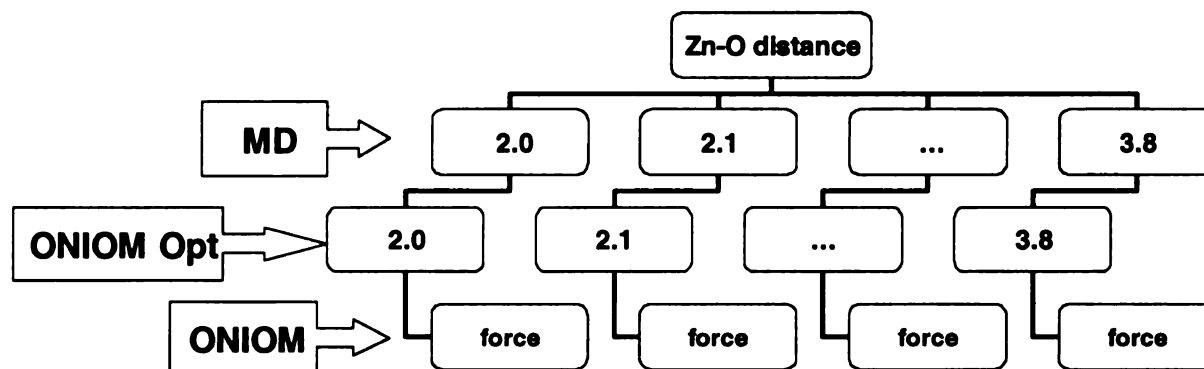


Figure 6-1 The flow chart of MD ONIOM combination used in Zn-O bond cleavage process.

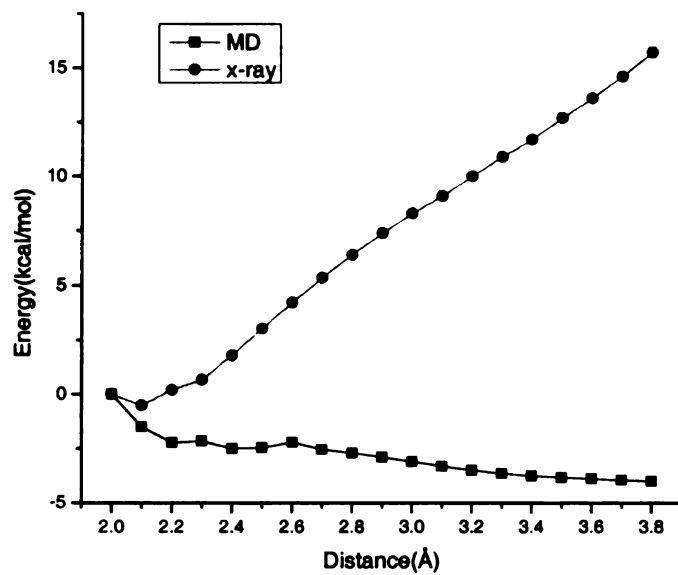


Figure 6-2 The ONIOM energy changes along the Zn-O4 distance scan.

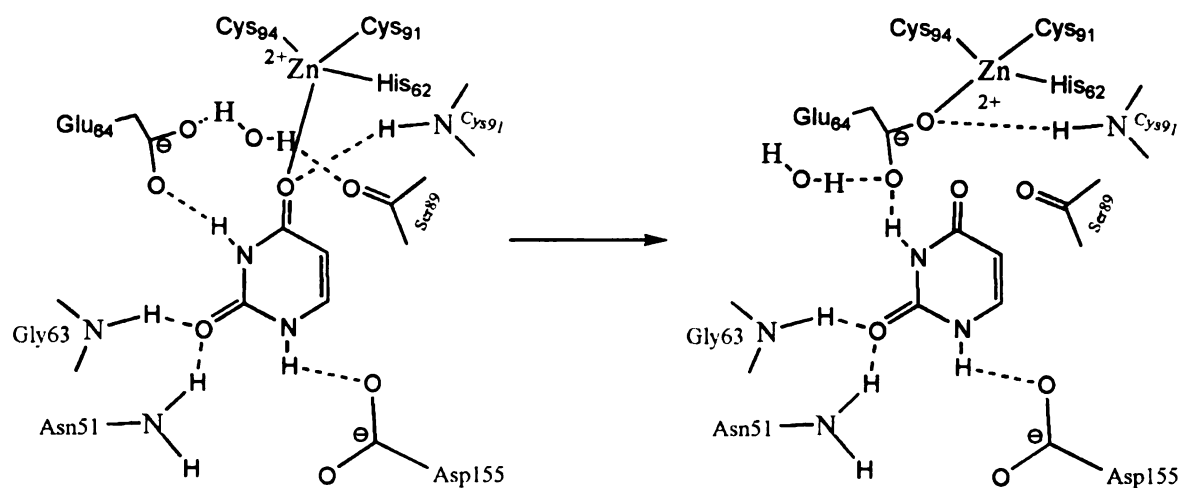


Figure 6-3 The rearrangement of the active site during MD simulation after the Zn-O4 bond cleavage.

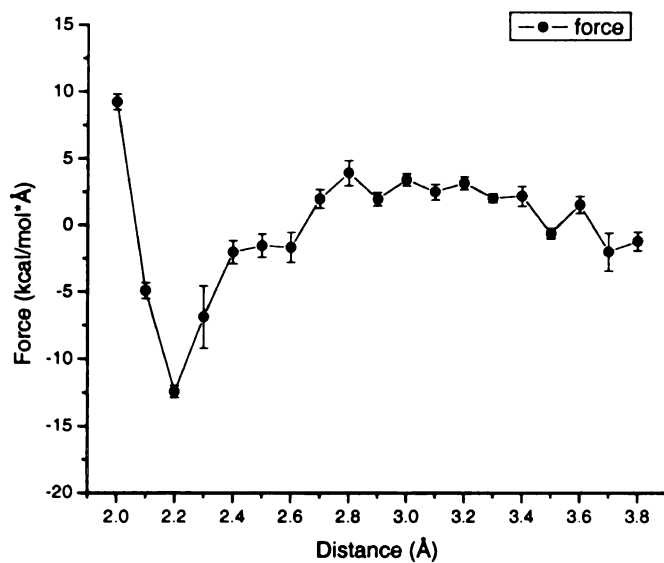


Figure 6-4 The average forces between Zn and O4. The bars in the plot represent the errors.

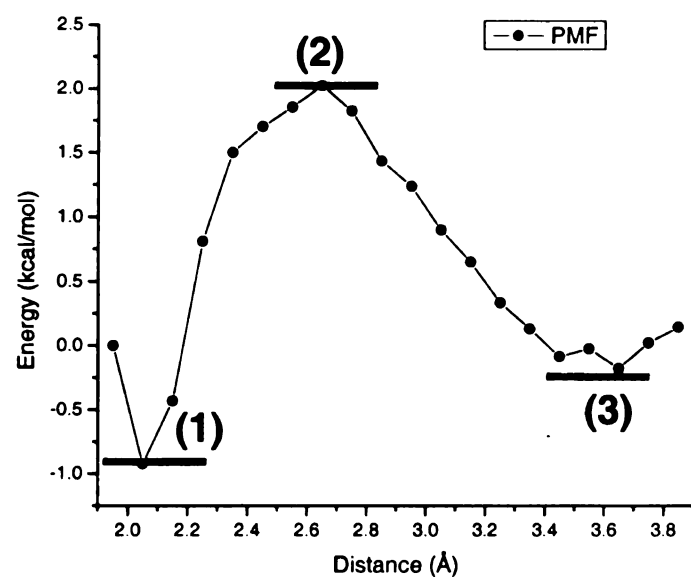


Figure 6-5 The potential of mean force along the Zn-O4 bond cleavage.

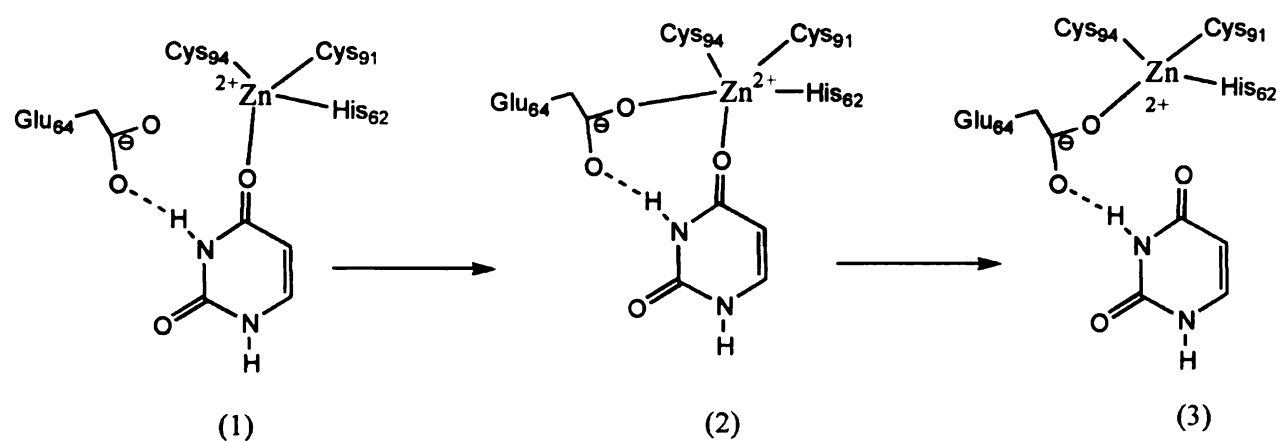


Figure 6-6 The three states defined based on PMF during Zn-O4 bond cleavage.

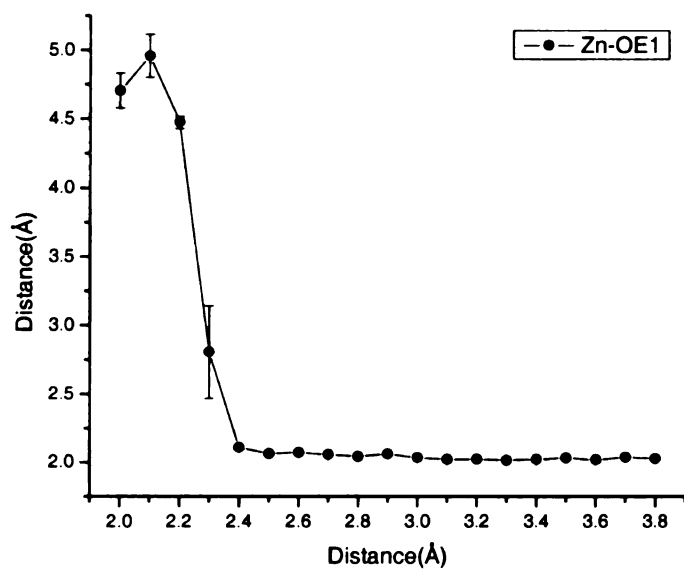


Figure 6-7 The distance change between Zn-OE1 during the cleavage of Zn-O4.

Chapter 7

A Molecular Dynamics Study of The Ligand Release Path in Yeast Cytosine Deaminase

Introduction

Yeast cytosine deaminase (yCD) catalyzes the deamination of cytosine to uracil. Cytosine deaminase is present in bacteria and fungi, as part of the pyrimidine salvage pathway, but is absent in mammals (Nishiyama et al. 1985). yCD can also catalyze the deamination of the prodrug 5-fluorocytosine (5-FC) to the anticancer drug 5-fluorouracil (5-FU). Thus, a potential system for gene-directed enzyme prodrug therapy combines yCD and the prodrug 5-FC, since the product, 5-FU, inhibits DNA synthesis and is thus a potent chemotherapeutic agent (Morris 1993).

X-ray structures of yCD in the apo form (Ireton et al. 2003) and in complex with the potent inhibitor 2-pyrimidinone (Ireton et al. 2003; Ko et al. 2003), are available. yCD is a homodimer with one active site in each protomer. The yCD protomer contains a center β sheet ($\beta 1 \sim \beta 5$, parallel to each other) with two α helices ($\alpha 1$ and $\alpha 5$) on one side and four α helices ($\alpha 2$, $\alpha 3$, $\alpha 4$ and $\alpha 6$) on the other side. Each active center contains a single catalytic zinc ion that is tetrahedrally coordinated with a histidine (His62), two cysteines (Cys91 and Cys94), and a water molecule in the substrate-free enzyme or the inhibitor in the complex. Surprisingly, the structure of apo yCD is essentially the same as that of the transition state analog complex with the active site completely buried.

Previous MD simulations for the apo form, reactant, various intermediate, and product complexes (Yao et al. 2005), all showed a buried active site during the ~2.0 ns simulations. The overall protein is quite rigid in those simulations, consistent with NMR backbone NOE data and order parameters, which describe motion in the ps-ns time scale (chapter 3). The question that we address in this work is how does the reactant bind and the product release in the catalytic cycle, in view of the buried active site?

Quench-flow experiments performed to study the reaction mechanism indicate that product release is the rate-limiting step during the activation of 5-FC (Yao et al. 2005). A slow product release process was also demonstrated by a ^{19}F NMR study, with the product release rate determined to be $\sim 13.1 \text{ s}^{-1}$. But, it is interesting that the K_d of 5-FU is only about $\sim 25 \text{ mM}$, suggesting that it is not a good binding ligand. Sklenak et al. (Sklenak et al. 2004) proposed a two-step process for uracil as well as 5-FU release. The first step is bond cleavage between the O4 atom of 5-FU (Figure 7-1) and the catalytic Zn atom through an oxygen exchange mechanism in which 5-FU exchanges its O₄ with a water molecule. In the second step, non-covalently bound 5-FU moves out of the active site. The first step gives a rather high activation barrier (Sklenak et al. 2004), which was thought to be the cause of the slow release of 5FU. However, our recent experiments (unpublished data) show that, though this oxygen exchange occurs, it is too slow and unlikely to be the mechanism for the Zn-O₄ cleavage. A new mechanism was proposed in which the Zn-O₄ bond was broken through a 5-coordinated Zn transition state, and the barrier for the Zn-O₄ bond cleavage found to be quite low, only $\sim 3 \text{ kcal/mol}$ (chapter 6). Therefore, it may be that rather than rate limitation by a chemical step, the barrier for product release comes from the loss of non-covalent interactions (such as hydrogen

bonds) with the protein and the steric hindrance due to the closed nature of the active site. If product release is rate limiting, knowledge of how product moves out or reactant moves into the active site would certainly help us understand this process, and lead to insights that would permit protein engineering to make this process faster, and consequentially improve the efficiency of the enzyme.

In this work, we explore several possible ligand release paths by pushing cytosine out of the active site, with the use of a restraint method. The methodology is quite similar to steered molecular dynamics (SMD), which has been used in studying protein ligand binding and unbinding process in various systems (Isralewitz et al. 2001; Shen et al. 2003; Xu et al. 2003). But, before pushing the ligand out of the active site, potential paths must be chosen. Typical ns simulations at 300 K can hardly give us valid information about the ligand release path and the involved residues. So, the strategy we adopted is to perform one regular simulation for the apo enzyme at 300 K and another at 320 K. By comparing these two trajectories, we might be able to find out the “soft spot” of the protein. As is well known, the higher temperature gives larger thermal fluctuations, and therefore a better chance for the protein to overcome certain barriers to reach some less-populated conformations; in this case, the opening of the active site. The simulation results show that the flexibility of certain residues does increase significantly more than others. In particular, the F114 loop (111-117) and C-terminal helix (150-158) that cover the active site do fluctuate more when comparing the 300 K and 320 K simulations. The motions of these two regions at 320 K open the active site, and permit water molecules to diffuse into and out of the active site through two paths. Based on the identification of these paths, cytosine was pushed out of the active site along them, and the residues that

respond to this release monitored. In this manner, motions of yCD regions required to release the ligand could be suggested.

Methods

MD simulation of the YCD apo form

The parameterization of the force field for Zn complexes was described in detail in a previous article (Yao et al. 2005). The protonation states of the ionizable residues were set to their normal ionization states at PH 7. The protein was surrounded by a periodic box of approximate dimensions $90 \times 88 \times 70$, leading to solvation by ~17,000 TIP3P water molecules. Six Na^+ ions were placed by the Leap program (Pearlman, Case et al. 1995) to neutralize the -6 charges of the model system. The parm94 version of the all-atom AMBER force field was used to model this system.

MD simulations were carried out using SANDER in AMBER 7.0 (Pearlman, Case et al. 1995) at constant pressure and temperature. The time step was 2 fs and the SHAKE algorithm was used to constrain all bonds involving hydrogen atoms. A nonbond pair list cutoff of 8.0 Å was used and nonbond pair list was updated every 25 steps. The pressure and the temperature (300 K and 320 K) of the system were controlled during the MD simulations by Berendsen's method. To include the contributions of long-range electrostatic interactions, the Particle-Mesh-Ewald method was used. A 2 ns simulation was performed for the 300 K simulation, with 500 ps of equilibration time. The 320 K simulation protocol is slightly different from 300 K's. After a 1 ns simulation, 3 ns of simulation was continued. Meanwhile, four other 1 ns independent simulations were started by reassigning the atom velocities to the last snapshot of the 1st ns simulation based on the Maxwell-Boltzmann distribution at 320 K. So effectively, an 8 ns simulation

was done at 320 K. The first 500 ps of the regular simulation and the first 200 ps of the velocity-reassigned simulations were treated as equilibration periods, and therefore not included in the data analysis.

The coordinates were saved every 2 ps, and were analyzed with the Moil-view program and the PTRAJ module of AMBER 7.0. Principle component analysis (PCA) was performed by using the gcovar and ganeig modules in gromacs to filter out the fast motions (Lindahl et al. 2001).

MD simulation of the YCD cytosine complex

A 2-ns MD simulation was previously performed for the yCD cytosine complex at constant temperature (300 K) and constant volume (Yao et al. 2005). The simulation shows the active sites in both protomers are buried and the overall protein is quite rigid, similar to the 300 K 2 ns simulation of the apo form.

In order to explore how cytosine gets out of the active site, a restraint was introduced to push the cytosine out through two paths based on the results from a 320 K apo form MD simulation. The first path is in between the F114 loop (111-117) and C-terminal helix (150-158) and the second path is in between loop 1 (53-61) and the C-terminal helix. A harmonic potential was added between the center of mass (COM) of heavy atoms of the cytosine ring and the COM of backbone heavy atoms of His50 (Ser89) in path 1 (2). The reasons to choose His50 are: (1) it is from the core region (β 2) and quite rigid and (2) it is along the path 1 direction of ligand release. Similar reasoning was used in choosing Ser89 in path 2. The equilibrium distance was increased gradually to give an effective force to push cytosine out. The starting structure was taken from one snapshot of the cytosine complex MD simulation.

Cytosines were pushed out 13 Å along path 1 in protomer 2 and path 2 in protomer 1 in a total of 15.6 ns simulation time. 130 windows were used with 0.1 Å per window and a force constant of 20 kcal/(mol-Å²). The force constant corresponds to a thermal fluctuation of distance ~0.25 Å. This pushing rate gives us a continuous sampling of the cytosine release path. Test simulations showed that this force constant is large enough to push cytosine out efficiently, similar to the pushing simulations performed on another system we studied (DHNA ref), and consistent with what other researchers used in a ligand unbinding study (Xu et al. 2003) (Gerini et al. 2003; Shen et al. 2003). The restraint force was calculated and averaged over each window. The first 20 ps of simulation were treated as the equilibration period and therefore not included in the force calculation. But, in the RMSD calculation between the instantaneous MD trajectory and the crystal structure, all the snapshots were used since no apparent difference can be seen between the equilibration and data collection period in terms of structure.

The pushing velocity is another important parameter in the simulation. Considering that the experimental rate of product release is 13 s⁻¹, which is impossible to simulate with current computer resources, we use a pushing speed as slow as is practical to mimic the unbinding process. Our test simulations with a faster pushing velocity (0.1 Å per 14 ps), a typical speed used in protein ligand unbinding in the literature, starting from different initial structures, showed no sign of distortion of the protein overall structure. Thus, from the structure point of view, the velocity we adopted is a reasonable compromise between the computational cost and the genuine property of the system. Moreover, the aim of this study is to address the potential ligand release paths and protein dynamics rather than evaluate the potential of mean force for ligand release.

Since the two ligands were pushed out simultaneously along two different paths, it might cause some synchronized artificial effect. In order to make sure this is not the case, a simulation of pushing a single ligand out along path 1 was carried out. No effect of dynamics changes in the active site of the adjacent protomer were found, probably because the two active sites are ~ 18 Å apart, and the pushing simulation results in the ligand further away from the adjacent active site.

Results and Discussion

Apo form simulation at 300 K and 320 K

yCD is a homodimer with 158 residues in each subunit. The data analysis was based on each subunit to improve the statistics, because the crystal structure shows that the two subunits are the same, except for several N-terminal residues. Because previous MD simulations showed that the N-terminus is far more flexible than the rest of the protein (Yao et al. 2005) and is also far away from the protein active site, the first 14 residues were not included in the fitting for the calculation of the root mean square deviation (RMSD) from the crystal structure or the root mean square fluctuation (RMSF) from the average structure.

The time evolutions of the CA-based RMSD of MD snapshots from the crystal structure stabilize after 500 ps, indicating that the two systems are in equilibrium during the analyzed trajectory (data not shown). Figure 7-2 shows RMSF plot at 300 and 320 K for two different subunits as well as the RMSF calculated from the crystal B-factors. As can be seen, they are consistent with the crystal B-factors. The overall protein is quite rigid with average RMSF ~ 0.41 Å and ~ 0.44 Å for protomer 1 and 2 respectively at 300 K. The increase of temperature from 300 K to 320 K causes a dramatic increase of

protein flexibility (Figure 7-2). The average RMSFs are ~ 0.62 Å for protomer 1 and ~ 0.67 Å for protomer 2 at 320 K as approximately 50% larger than those at 300 K. This flexibility increase mainly comes from the relatively flexible regions at 300 K, among which the C-terminal helix shows the largest RMSF increase in both protomers (Figure 7-2). Since the C-terminal helix covers the active site, as shown in the crystal structure (Ireton et al. 2003), the dramatic increase of RMSF in this region might be important to open up the active site and let the reactant bind.

In order to explore more details about the protein fluctuation changes and potential motion mode changes due to the increase of temperature, principle component analysis (PCA) was used to analyze the MD trajectories. PCA attempts to decompose the overall atom fluctuations into a small set of modes that encompass a large percentage of the overall RMSF², and a much larger set of modes that contribute a relatively small amount. In this way, we can focus on the small set of modes that represent slow, large-scale motions, because large portions of the flexibility increase usually comes from these modes. By comparing these significant motion modes at the two temperatures, we can obtain instructive information about large-scale motions that might be responsible for ligand binding or product release.

As shown in Table 7-1, the total variance of CA atoms (residue 15-158) is ~ 25 Å² and ~ 29 Å² for protomer 1 and 2 respectively at 300 K. The 320 K simulation gives much bigger variances with ~ 65 Å² for protomer 1 and ~ 76 Å² for protomer 2. It is a surprise that the total fluctuations are approximately doubled even though the temperature only increases $\sim 7\%$. Assuming harmonic vibrational motion for the CA atoms, the increase of total variance should scale with the temperature increase. The large deviation from this

expectation might come from two main reasons: 1. The increase of temperature gives the MD a better chance to overcome certain barriers and therefore explore phase space more effectively. 2. The longer simulation time at 320 K (8 ns) can also cover more phase space than the 2 ns simulation at 300 K. Nevertheless, it is a surprise that a small increase of temperature can trigger such a dramatic fluctuation change. New, slow motion modes might be captured in the 320 K simulation.

To see whether this is true, PCA was used to filter out the fast motions and keep the slow ones which we are primarily interested in. The first 5 modes contribute 24% and 29% at 300 K compared with 48% and 47% at 320 K to the total motion for protomer 1 and protomer 2 respectively. The fluctuation difference in the first 5 modes between these two temperatures is $\sim 25 \text{ \AA}^2$ ($\sim 27 \text{ \AA}^2$) for protomer 1 (2), contributing 63% (73%) of the total fluctuation difference. Thus the major motional differences between 300 K and 320 K were captured in the first 5 modes.

The RMSF plots of individual residues at these two temperatures can provide us information about which regions have large flexibility changes. Figure 7-3 shows the CA RMSF of the first 5 modes; the main difference between 300 K and 320 K is from the flexible regions. Though the trajectories of these two protomers show some differences of the flexibility increase in certain regions, such as the 120~125 (part of helix 4) and 144~149 (part of helix 5) probably due to the incomplete sampling of conformation space, both protomers show consistently large RMSF differences from the C-terminal helix (150-158) and F114 loop (111-117). In order to get a better view of the flexibility changes, a schematic graph was generated by superimposing the 50 MD CA snapshots projected onto the first 5 modes for individual protomers at these two simulation

temperatures (Figure 7-3a). Clearly, the C-terminal helix and F-114 loop show much larger fluctuations at 320 K than those at 300 K. Furthermore, it reveals that the motion of these two regions can expose the active site to the solvent, whereby water molecules can move freely in and out of the active site, unlike the 300 K simulation in which the active site is completely buried (Figure 7-4a). We identified a total of 27 water molecules that diffuse into the active site and 11 water molecules that diffuse out through the path in between the C-terminal helix and F114 loop in both protomers (Figure 7-4b). The 300 K MD simulation shows that the side-chains of F114, W152 and I156 cover the active site, while F114 and I156 block this path. The increase of flexibility of the C-terminal helix and F114 loop at 320 K moves these two side-chains slightly away from their original positions; as a consequence the active site is opened up. Three residues C91, F114 and I156 were used to define qualitatively the entrance of this water path (Figure 7-4b). Labeling the C91 CA-F114 CZ distance as d_1 , C91 CA-I156 CD distance as d_2 , and F114 CZ-I156 CD distance as d_3 , parametrically plotting these distances for 300 K and 320 K simulations in 3D reveals at 300 K the distances are restrained in the low left corner while at 320 K they are spread from low left to up right (Figure 7-4c). The distances for d_1 , d_2 and d_3 are 5.25 ± 0.72 Å, 8.04 ± 1.04 Å and 4.65 ± 0.64 Å at 300 K compared with 7.73 ± 1.90 Å, 10.20 ± 1.28 Å and 6.12 ± 1.20 Å at 320 K in protomer 1. All the distances and their fluctuations are increased at 320 K, indicating the mouth of the active site is opened up and the breathing motion is larger at this temperature. The 320 K data, when projected onto the 3 2D planes formed from these distances, shows that the two state behavior in Figure 7-4c mainly arises from the C91 CA-F114 CZ distance

coordinate. Thus, there are barriers that are overcome during the simulation at the higher temperature.

It is interesting that one water molecule moves out of the active site in protomer 2 through the path in between the C-terminal helix and the loop between $\alpha 2$ and $\beta 2$ (51-61, we name it loop 1 (Figure 7-4b)). As shown in Figure 7-3, loop 1 is also relatively more flexible at 320 K. Thus, two water diffusion paths were identified. But it seems that the path in between the F114 loop and C-terminal helix (path 1) is more favorable than the path in between the C-terminal helix and loop 1 (path 2), particularly considering either the natural reactant cytosine or the product uracil is much larger than a water molecule. Further investigations of these two paths were launched by pushing the reactant cytosine out of the active site along these two paths as described below.

To summarize the apo simulation results, the 320 K MD simulation increases the flexibility of the protein, and most of the increase comes from the large-scale (slow) motion modes that have relatively large barriers to overcome. The use of high temperature and longer simulation time gives us a greater chance to overcome these barriers and give a better description of the protein dynamics. Two pathways were identified for water diffusion into the active site, which are the potential paths for ligand binding, due to the increased motion of loop 1, the F114 loop and the C-terminal helix.

Our NMR experiments using HD exchange of backbone amides (chapter 4) confirm the motions of these three regions in the apo form, but on a much slower time scale, primarily in seconds, which is comparable to the product release rate. T_1 , T_2 relaxation, NOE and T_2 relaxation dispersion experiments show no apparent sign of large motions of these regions in the ps-ns and μ s-ms time scales (unpublished data). So, it appears that

the opening of the active site is quite slow, and probably limits the product release. By pushing cytosine out of the active site, we can discover details of the motions of individual residues that are important for product release or ligand binding.

Pushing cytosine out of the active site

In order to discriminate whether the two water paths identified in the regular simulation are sensible paths for product release, the reactant cytosine was pushed 13 Å away from the active site through these two paths in a 15.6 ns MD simulation. The structure and flexibility changes for the overall protein and individual regions during the pushing simulation were investigated and the forces used for the cytosine pushing were calculated to better understand these two paths and their influence on the overall protein and its various regions. Several residues were identified and studied which might be the potential ones that limit the slow product release process.

Path 1 In path 1, one harmonic restraint was added in between the COM of the cytosine heavy atoms and the COM of the backbone heavy atoms of His 50. This residue is from the $\beta 2$ region, which is quite rigid, and aligned along the path. So, this residue was selected as the anchor to implement the restraint force.

Figure 7-5a shows the RMSD plot of all the CAs versus time. Over the first 7 ns, the RMSD is stable at ~ 0.9 Å and then increases slightly to ~ 1.1 Å. The whole protein is quite rigid and the perturbation of overall structure is rather small during the ligand release. Figure 7-5b shows the average pushing forces in each window. In the beginning, the force increases to ~ 7 kcal/mol-Å at the 3rd ns, then fluctuates around and drops back to ~ 0 kcal/mol-Å at about the 7th ns. After that, the force fluctuates and goes up to ~ 8 kcal/mol-Å then rapidly falls to ~ 0 kcal/mol-Å at the 12th ns. Hereafter cytosine is out of

the active site and on the surface of protein, so the force fluctuates around 0 kcal/mol-Å. From the RMSD plot and the force plot (Figure 7-5a and b), it appears that there are two periods for the product release, the first period is from the beginning to the ~7th ns and the second period from the 7th ns to the end. In the first period the overall structure changes are quite small. The main barrier comes from the direct hydrogen bonding interactions lost between cytosine and the residues in the active site, which can be confirmed by a hydrogen bond analysis. Figure 7-5c shows the lifetime of the hydrogen bonds of the pushing simulation. There are 5 hydrogen bonds between cytosine and the protein, including N51, G63, E64 and D155, when cytosine is well bound in the active site. All these hydrogen bonds are lost during the first period of the pushing simulation. The hydrogen bond with G63 was lost first at ~3 ns primarily because G63 sits at the bottom of the active site, and the hydrogen bond donor, the backbone amide, is quite rigid. Then E64 and N51 lose their three hydrogen bond interactions with cytosine at ~5 ns. Lastly, the hydrogen bond between D155 and cytosine is broken at ~7 ns. Apparently, the loss of five hydrogen bonds demands the external force to do the work and compensate the energy. During the second period, since all the hydrogen bonds with cytosine have broken, the force needed to push the ligand out probably mainly comes from steric hindrance and conformational rearrangements of the residues along the path.

To understand how the cytosine-pushing procedure affects the dynamics of individual regions or residues, CA RMSFs were plotted in Figure 7-6. Figure 7-6a shows the CA RMSFs of the regular MD simulation, which are quite similar to the regular simulation of the apo enzyme at 300 K, with a rigid backbone. During the pushing of cytosine, the overall protein becomes more flexible as shown in Figures 6b and 6c, and

the RMSF increases when the first and second stages of the pushing are compared with the regular simulation. The total fluctuations of the CA atoms are 35.8 Å², 50.6 Å² and 59.3 Å² for the regular, stage 1 and stage 2 simulations, respectively. Since the active site is buried during the regular simulation, the protein has to move certain regions or residues to open up the active site to let cytosine out, which causes the fluctuation increases. Figure 7-6b and 6c show quite dramatic C-terminal helix and F114 loop RMSF increases in the pushing simulation, consistent with the 320 K apo form simulation results (Figure 7-2). The increases in the RMSFs of these two regions suggest that: 1. The motions of these two parts are important for ligand release. 2. The motion of protein required for ligand release is local, and the structural perturbation is small. It appears that the motion of the C-terminal helix is larger in stage 2 than that in stage 1. The MD trajectory shows that at the end of stage 2 cytosine moves out of the active site completely. The tip of the C-terminus moves close to the active site, partially to occupy the empty space, and consequentially block the active site, which effectively imparts to the C-terminal helix a relatively large motion. The unbinding process of cytosine can be better illustrated by the schematic graph of the CA fluctuations. Just like the data analysis carried out in the apo form simulations, PCA was used to filter out the high frequency motions and only the first 5 modes were used in the schematic plot that is composed of the superposition of 50 MD snapshots. As shown in Figure 7-7a, the whole protein is rather rigid, including the F114 loop and C-terminal helix in the regular 2 ns MD simulation of the cytosine complex. Figure 7-7b shows the increase of the fluctuation in these two regions in the first stage of pushing, while Figure 7-7c describes a similar increase, but some of the C-terminal helix snapshots are closer to the active site in the

second stage. A movie of the trajectory reveals that, in stage 2, the C-terminal helix and F114 loop undergo a concerted flapping motion, which effectively opens up the active site and then closes it after the release of cytosine (see discussion below). Besides the changes of these two regions, a small helix ~75-80 also undergoes fluctuation perturbations during the pushing simulation (Figure 7-6b and c). Further inspection shows that this region is in close contact with the C-terminal helix of the adjacent protomer, and suggests that this fluctuation increase might be caused by the motion of that C-terminal helix where cytosine was pushed out along path 2 simultaneously.

The MD trajectory shows that the cytosine release path is slightly different from the water diffusion path in the apo form simulation. Unlike water diffusing through the triangle mouth defined by C91, F114 and I156, the cytosine release path drifts slightly away and is in between F114 and I156 (Figure 7-8). The mass weighted RMSDs of F114 and I156 are plotted in Figure 7-9. The RMSD of F114 changes dramatically during the pushing simulation. Over the first 6 ns, it fluctuates around ~1.5-2.0 Å, then increases to ~6.0-7.0 Å and stabilizes at ~9-13 ns, after that it drops to ~2.5 Å at ~14 ns. Compared to F114, the motion of I156 is smaller (Figure 7-9), with a similar pattern to the whole C-terminal helix, and at the end it also goes back to ~1 Å, indicating that this residue returns to its original conformation. Apparently, F114 undergoes a large flapping motion during stage 2, which opens up the active site and assists the release of cytosine, while the C-terminal helix moves slightly away to make the entrance larger, and then both come back to cover the active site after the release of cytosine. It appears that the C-terminal helix moves as a whole block, while F114 moves in addition to the loop movement because

there is no secondary structure restraint in the loop region. As a result, the motion scale of F114 is much larger than I156.

Path 2 Cytosine in protomer 1 was pushed out along path 2 that is defined as in between the C-terminal helix (150-158) and loop 1 (53-61) (Figure 7-8). According to the 320 K simulation of the apo enzyme, it seems that this path is less popular for water to diffuse through than path 1. Figure 7-10a shows the RMSD of all CA atoms superposition with the crystal structure, which increases from ~ 0.8 Å to ~ 1.4 Å. This change is larger than that of path 1, which increases from 0.9 Å to 1.1 Å, suggesting that the overall protein has to move more when cytosine releases through this path. This could make the release harder. The total CA fluctuation is 92.5 Å², much larger than that in the regular simulation (25 Å² (29 Å²) for protomer 1 (2)), with the difference (for protomer 2) presented in Figure 7-10c. It appears that many regions in the protein have larger fluctuations, especially those flexible regions identified in the regular 320 K apo form simulation (Figure 7-2). Therefore, it seems that the protein has to rearrange many regions in a concerted way when cytosine releases through this path, unlike in path 1 where only the F114 loop and C-terminal helix regions need to move away, and the perturbation is rather local and small. An analysis of the MD trajectory shows that along this path cytosine first pushes aside V31, N51, D155 and breaks the hydrogen bonds with the later two residues, and then moves away from D151 (salt bridged with R148). After that, it rearranges T54's side-chain to gain enough space to exit. It appears rather hard for a ligand to move through this path, more likely for water but not for cytosine. The force used to push cytosine out is presented in Figure 7-10b. The average force is ~ 5.6 kcal/mol-Å, about two times larger than that in path 1 (~ 3.2 kcal/mol-Å), implying that

more work has to be done in path 2. The root mean square force fluctuation is also larger in path 2 (4.5 kcal/mol-Å) than path 1 (2.8 kcal/mol-Å), so the energy surface seems rougher along path 2.

Thus, from energy and structural points of view, path 2 is not favorable compared with path 1 in the current simulation, although path 2 cannot be excluded on the basis of the simulations alone. In order to release the ligand along path 2, many regions have to move, including the C-terminal helix and F114 loop, which are the only parts required to move significantly when cytosine is released along path 1. In fact, the F114 loop has the largest fluctuation changes in the path 2 simulation, because cytosine pushes the C-terminal helix close to this loop to move it away.

It appears that in both pathways, the C-terminal helix is very important, acting like a lid covering the active site. During the pushing of cytosine along both paths, the C-terminal helix moves as a whole block. The X-ray structure shows that this helix is negatively charged (−4, including D151, E154, D155, E158), and there are four salt bridges with residues from other regions, including D151-R148, E154-R53, E154-R73* (from the adjacent protomer) and D155-R53, which restrain this helix. During the pushing simulation, all these salt bridge are well maintained. It is reasonable to predict that cytosine pushing would be easier if some of these salt bridges were weakened or eliminated. On the other hand, mutation of F114 and/or I156 to smaller residues, without changing the dynamic property of the C-terminal helix, may also improve the product release rate. Experiments are underway with mutants of some of these residues to explore these suggestions.

Conclusions

In this study, we used Molecular Dynamics to explore possible ligand release paths from yCD. The 300 K apo enzyme simulation shows a closed active site, consistent with the crystal structure. Overall the protein is quite rigid, a feature that has been confirmed by NMR backbone relaxation experimental data (unpublished results) {ref}. An 8 ns 320 K simulation was performed, which shows a strikingly large increase of flexibility in some regions of the protein. The increases of flexibility in the C-terminal helix and F114 loop regions, which open up the closed active site, are most significant. Water molecules diffuse into the active site through two paths. One path is in between the F114 loop and C-terminal helix, and the other is in between loop 1 and the C-terminal helix. The former path appears to be more popular, with the mouth of the entrance defined by a triangular area formed by the residues C91, F114 and I156. This area and its fluctuations are much larger at 320 K than at 300 K (Figure 7-4c), which allows water molecules to frequently move in and out of active site.

Cytosine was pushed out of the active site along these two paths by 13 Å in a 15.6 ns simulation. The results show that in path 1 the required motion of the protein is quite local, only involving the C-terminal helix and F114 loop. Two residues F114 and I156 are identified that have to be moved away in order to let cytosine out. While, in path 2, the protein has to rearrange itself quite globally, and the changes are also much larger compared to the path 1 simulation. The average force along the path and its fluctuation are larger in path 2 than in path 1, suggesting that the barrier is higher and the energy surface along the release is rougher in path 2. Several residues have to be moved away including V31, N51, and D155 first, and then D151 and finally T54. This path appears

rather difficult, and seems to be a path much less frequently used even for water molecules and not meant for cytosine because of its much larger size in comparison with water. However, path 2 cannot be excluded on the basis of the simulations alone.

Nevertheless, in both paths the C-terminal helix is critical for ligand release. Note that the C-terminal helix is well restrained by four salt bridges and several hydrogen bonds with residues in other regions. By weakening these salt bridge interactions, we could, in principle, increase the flexibility of the C-terminal helix, and that could make cytosine release faster. Or, on the other hand, making F114 or I156 (residues along the path) smaller could also assist the release of the ligand. As discussed above, the active site opening limits the ligand escape process. From the experimental point of view, by modifying the protein as described above, we should be able to improve the product release rate and therefore enhance the protein's activity. Experiments are under the way to test these hypotheses, and with more data available, we will be able to correlate our experimental results with our simulation data.

Reference

- Gerini, M. F., D. Roccatano, et al. (2003). "Molecular dynamics simulations of lignin peroxidase in solution." Biophysical Journal **84**(6): 3883-3893.
- Ireton, G. C., M. E. Black, et al. (2003). "The 1.14 angstrom crystal structure of yeast cytosine deaminase: Evolution of nucleotide salvage enzymes and implications for genetic chemotherapy." Structure **11**(8): 961-972.
- Isralewitz, B., M. Gao, et al. (2001). "Steered molecular dynamics and mechanical functions of proteins." Current Opinion in Structural Biology **11**(2): 224-230.
- Ko, T. P., J. J. Lin, et al. (2003). "Crystal structure of yeast cytosine deaminase - Insights into enzyme mechanism and evolution." Journal of Biological Chemistry **278**(21): 19111-19117.
- Lindahl, E., B. Hess, et al. (2001). "GROMACS 3.0: a package for molecular simulation and trajectory analysis." Journal of Molecular Modeling **7**(8): 306-317.
- Morris, S. M. (1993). "The Genetic Toxicology of 5-Fluoropyrimidines and 5-Chlorouracil." Mutation Research **297**(1): 39-51.
- Nishiyama, T., Y. Kawamura, et al. (1985). "Antineoplastic Effects in Rats of 5-Fluorocytosine in Combination with Cytosine Deaminase Capsules." Cancer Research **45**(4): 1753-1761.
- Shen, L. L., J. H. Shen, et al. (2003). "Steered molecular dynamics simulation on the binding of NNRTI to HIV-1 RT." Biophysical Journal **84**(6): 3547-3563.
- Sklenak, S., L. S. Yao, et al. (2004). "Catalytic mechanism of yeast cytosine deaminase: An ONIOM computational study." Journal of the American Chemical Society **126**(45): 14879-14889.
- Xu, Y. C., J. H. Shen, et al. (2003). "How does Huperzine A enter and leave the binding gorge of acetylcholinesterase? Steered molecular dynamics simulations." Journal of the American Chemical Society **125**(37): 11340-11349.
- Yao, L. S., Y. Li, et al. (2005). "Product release is rate-limiting in the activation of the prodrug 5-fluorocytosine by yeast cytosine deaminase." Biochemistry **44**(15): 5940-5947.
- Yao, L. S., S. Sklenak, et al. (2005). "A molecular dynamics exploration of the catalytic mechanism of yeast cytosine deaminase." Journal of Physical Chemistry B **109**(15): 7500-7510.

Appendices

Table 7-1 Total CA fluctuations and contributions from first 5 modes at 300 K and 320 K.

	300 K		320 K	
	prot 1	prot 2	prot 1	prot 2
CA fluctuation (\AA^2)	25.4	29.4	65.1	76.1
First 5 modes	24%	29%	48%	47%

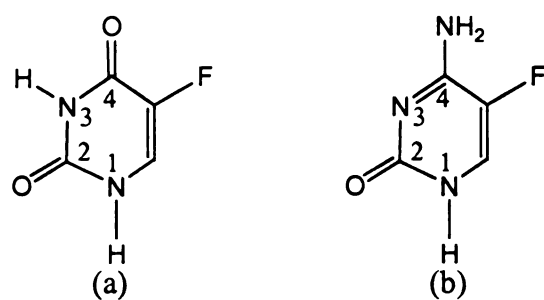
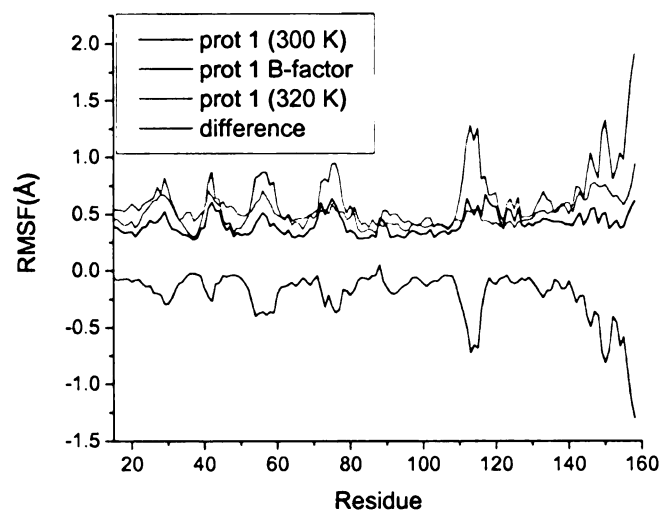
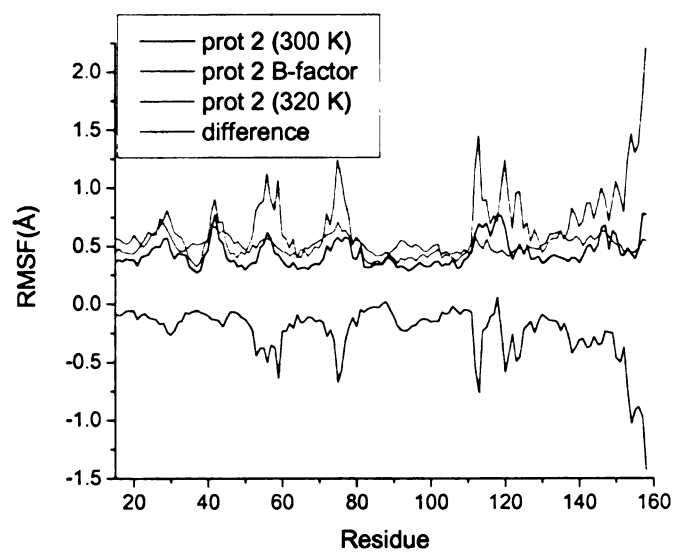


Figure 7-1 a. 5-fluoro-uracil (5FU) b. 5-fluoro-cytosine (5FC)

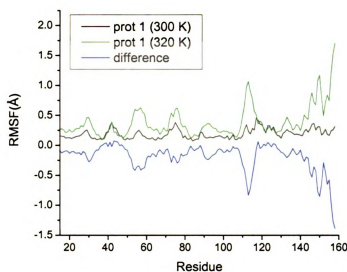


(a)

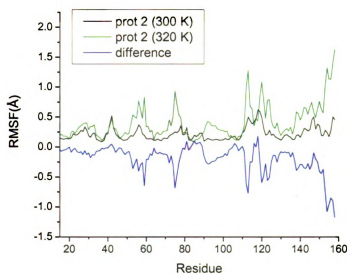


(b)

Figure 7-2 CA RMSFs of: (a). protomer 1. (b). protomer 2 at 300 K (black) and 320 K (green). The crystal B-factors were converted to RMSFs (red). The difference (blue) is the RMSF difference of the 300 K and 320 K data.



(a)



(b)

Figure 7-3 CA RMSF plots of the first 5 PCA modes at 300 K and 320 K and their difference: (a) protomer 1. (b) protomer 2.

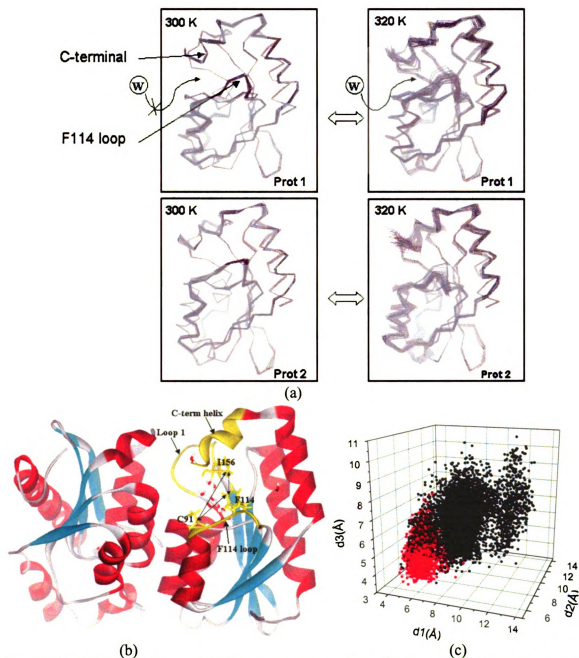
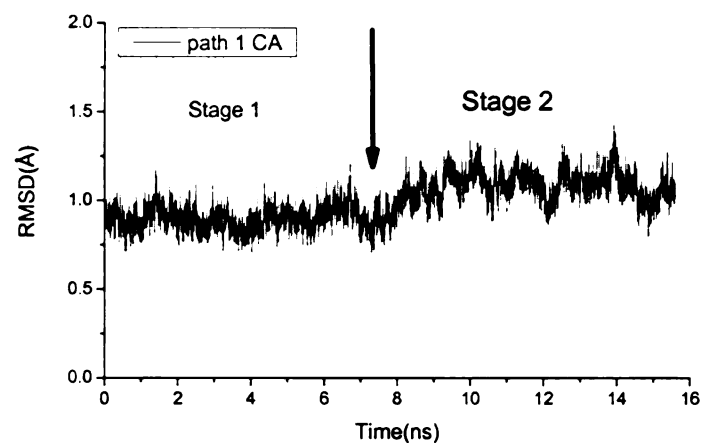
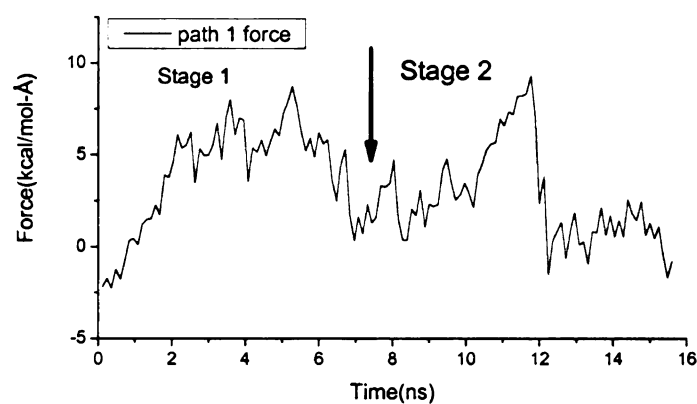


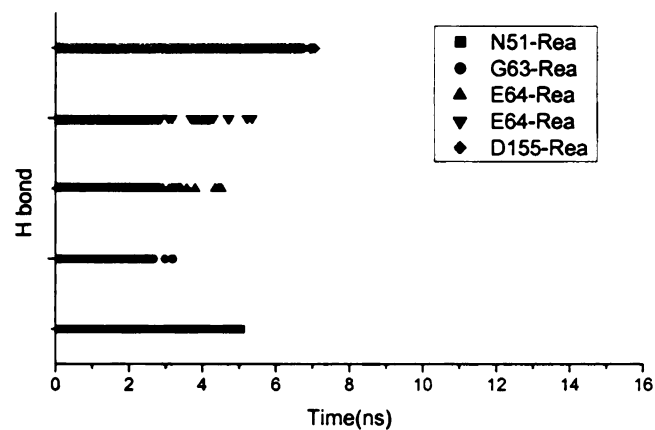
Figure 7-4 (a) Schematic graph of 50 trajectory snapshots of CA atoms projected out of the first 5 PCA motion modes at 300 K and 320 K. The F114 loop and C-terminal helix were labeled to give a better view. Water molecules can diffuse into the active site at 320 K through the path in between F114 loop and C-terminal helix, but not at 300 K. (b) One snapshot of apo MD simulation at 320 K. Water molecules diffuse into the active site through the triangle mouth defined by C91, F114 and I156. (c) 3D plots of distances $d1$ (C91 CA-F114 CZ), $d2$ (C91 CA-I156 CD) and $d3$ (F114 CZ-I156) at 300 K (red) and 320 K (black).



(a)

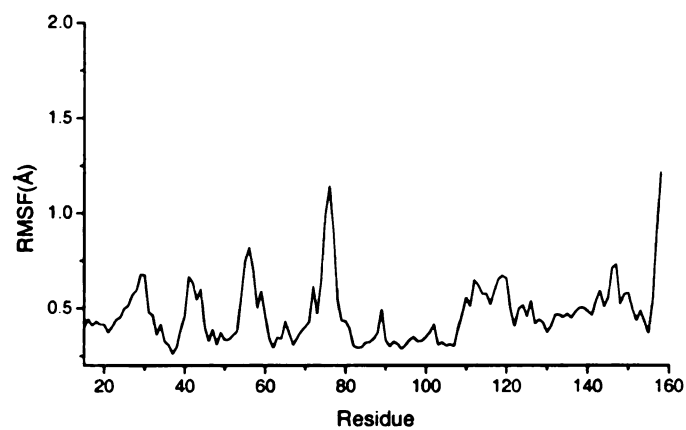


(b)

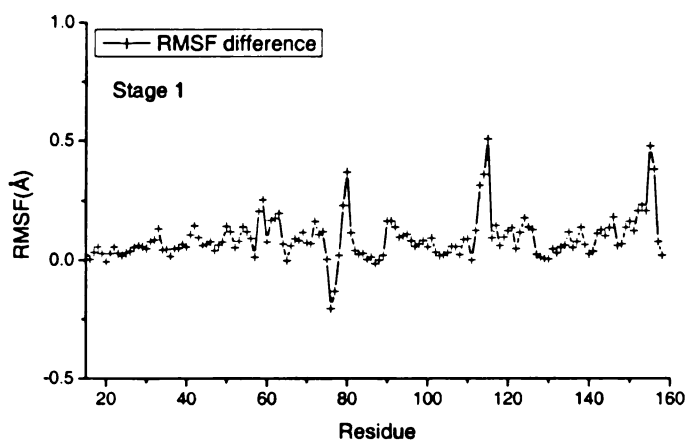


(c)

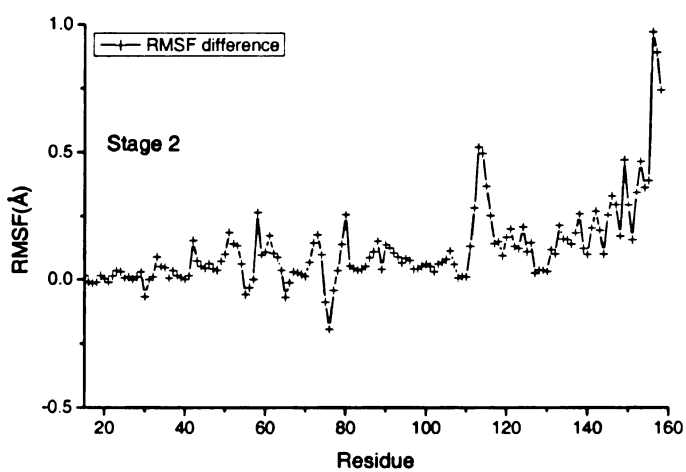
Figure 7-5 (a). Total CA RMSD versus time obtained by comparing the MD snapshots with the crystal structure. (b). The average restraint force between cytosine and the protein in each window. (c). Hydrogen bond lifetimes between cytosine and the protein during the cytosine pushing along path 1.



(a)



(b)



(c)

Figure 7-6 (a). CA RMSFs in 2 ns regular MD simulation. (b). RMSF difference by comparing the first stage of cytosine pushing with the regular simulation along path 1. (c). RMSF difference by comparing the second stage with the regular simulation.

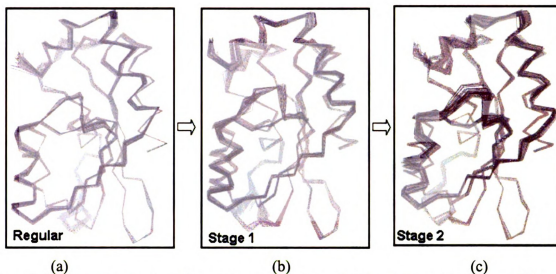


Figure 7-7 Schematic graph of 50 trajectory snapshots of CA atoms projected out of the first 5 PCA motion modes: (a). Regular simulation. (b). Pushing simulation stage 1. (c). Pushing simulation stage 2.

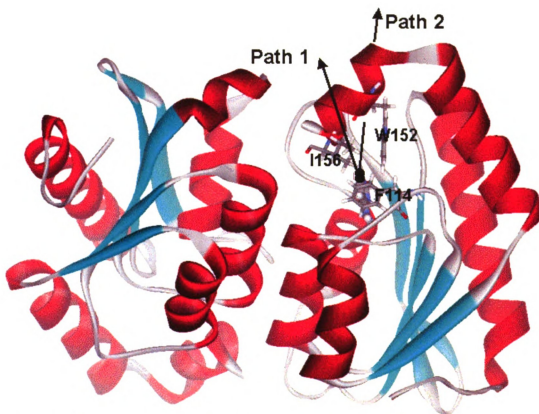


Figure 7-8 The cytosine release paths from the pushing simulation.

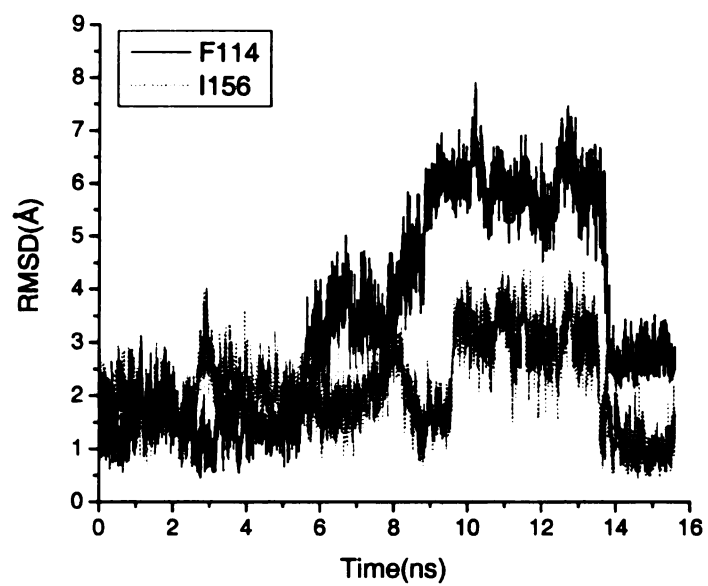
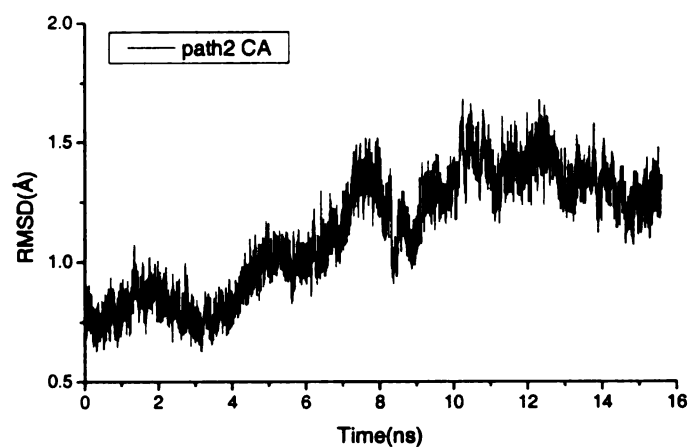
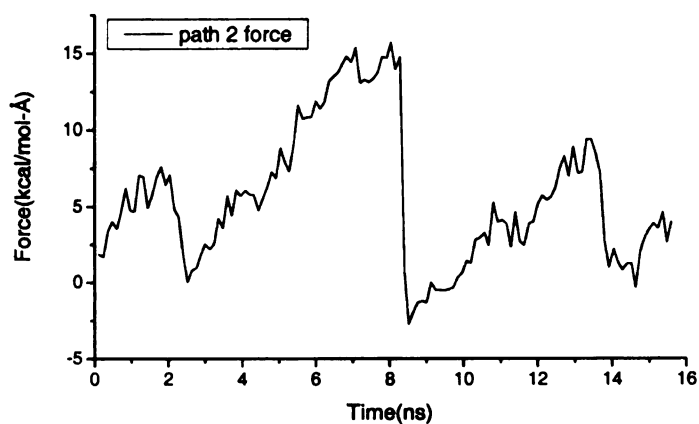


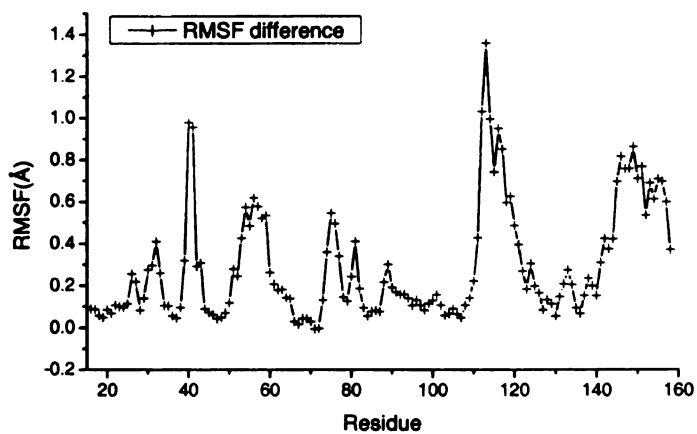
Figure 7-9 Mass weighted RMSD plot of F114 and I156 along the pushing path of cytosine (path 1).



(a)



(b)



(c)

Figure 7-10 (a). RMSD of all CA atoms compared with the crystal structure. (b). The **average** force calculated for the cytosine pushing along path 2. (c). The CA fluctuation **differences** between the regular simulation and the cytosine pushing.

Chapter 8

Reaction Mechanism of Guanine Deaminase: An ONIOM and Molecular Dynamics Simulation Study

Introduction

Guanine deaminase (GD), a Zn metalloenzyme, catalyzes the hydrolytic deamination of guanine into xanthine (Figure 1) and therefore plays an important role in nucleotide metabolism. The crystal structure of GD from *Bacillus subtilis* (bGD) complexed with imidazole was solved by Liaw and coworkers and his colleagues at 1.17 Å resolution (Liaw et al. 2004). bGD forms a homodimer with 156 residues in each monomer. The overall structure of each monomer consists of a central five-stranded β -sheet sandwiched by six helices (Liaw et al. 2004). The homodimer GD contains two active sites, including one Zn atom in each of the active sites. The Zn atom is coordinated with His53, Cys83, Cys86 and one water molecule. It is interesting that the active site is buried by a swapped C-terminal tail from the other subunit. It was proposed that this tail not only seals the active site but also is used to recognize specific substrate (Liaw et al. 2004). The bGD-catalyzed reaction is believed to proceed through a tetrahedral transition state, while Glu55 (Figure 8-2) acts as the proton shuttle. But the detailed mechanism is still unknown.

In this work, we performed a series of quantum calculations and molecular dynamics simulations to study the reaction mechanism. It is very challenging to carry out an accurate and reliable quantum mechanical study for an enzymatic catalytic reaction

due to the size of the system. Here we adopted a two-layered ONIOM (Humbel et al. 1996; Svensson et al. 1996; Dapprich et al. 1999; Vreven and Morokuma 2000) method implemented in Gaussian 03 (Frisch) to do the quantum calculation. The ONIOM method is a hybrid computational method allowing different levels of theory to be applied to different parts of a molecular system. In the two-layered ONIOM method, the molecular system was divided into an inner layer (the core region, usually at high level theory) and an outer layer (surrounding environment, usually at low level theory). The ONIOM method allows one to perform a high-level calculation on just a small but critical part of the molecular system while incorporating the effects of the environment at a low level theory, which makes large molecular system study affordable with good accuracy. Our ONIOM study of the reaction mechanism catalyzed by yeast Cytosine Deaminase (yCD) (Sklenak et al. 2004) correlates quite well with experimental results (Yao et al. 2005) (Yan's lab data).

But, in ONIOM calculations, one drawback is that only the static structure is used due to the high computer time cost. On the other hand, molecular dynamics simulation based on a classic force field can efficiently provide important dynamic property of the protein, which can be used to complement the ONIOM calculation (Yao et al. 2005). In addition, MD simulation can be performed for the whole protein solvated with explicit water, while ONIOM calculation is typically used for only for a selected part of the system (generally the active site and the immediate surrounding residues). MD simulation can be used to evaluate the environmental effect of the rest of the system and provide important information about the size of the selected part that should be used in ONIOM.

Methods

ONIOM calculation

The computational models studied in this work were based on the 1.17 Å imidazole bound GD crystal structure (Liaw et al. 2004). The active site model was generated from subunit one of the crystal structure. In order to get guanine bound in the active site, the imidazole ring of guanine was superimposed to the bound ligand imidazole. The missing hydrogen atoms were built using InsightII (InsightII). The protonation state of guanine was determined based on pKa of its individual atoms. In neutral pH, N1 and N9 are protonated while N7 is deprotonated in aqueous based on pKa of these three atoms (N1: 12.3-12.4, N9: 9.2-9.6, N7: 3.2-3.3) (Jang et al. 2003). This is consistent with the hydrogen bond analysis of the active site (see results part). N1 forms a hydrogen bond with carboxyl group of Glu55 as a donor, N9 forms a hydrogen bond with carboxyl group of Asp114 as a donor, and N7 forms one hydrogen bond with phenol group of Tyr156 as an acceptor. O6 forms two hydrogen bonds with backbone amide of Ala54 and sidechain amide of Asn42 as an acceptor, so O6 should be in ketone form instead of enol form (Figure 8-2).

The enzyme was modeled as 26 residues within 6 Å of the bound guanine, including Gly24, Pro25, Phe26, Gly27, Ala28, Glu41, Asn42, Asn43, Val44, Ala52, His53, Ala54, Glu55, Val56, Thr78, Cys80, Glu81, Pro82, Cys83, Cys86, Ala107, Phe112, Asp113, Asp114, Trp*92 and Tyr*156 (* represents residues from adjacent subunit, same below.). MD simulation shows that these residues contribute most of the interaction energies between ligand and surroundings (see results). The systems were divided into two layers. The inner layer, which was treated at high level of theory, was composed of the substrate, intermediate, or the products, Zn, the imidazole ring (model

for His53), $\text{CH}_3\text{CH}_2\text{COO}^-$ (model for Glu55), SCH_3 (model for Cys83 and Cys86), CH_3COO^- (model for Asp114), and the water coordinated to Zn. The rest of system was treated at a lower level. Quantum chemical studies on model compounds show that the Zn-ligand distances are sensitive to their protonation states (Dudev and Lim 2001; Dudev and Lim 2002). It appears that the Zn complex in bCD is identical to that in yCD active site, with very similar distances (Ireton et al. 2003). The four distances in bCD (yCD) are Zn-O 2.03 (2.07) Å, Zn-ND 2.06 (1.99) Å, Zn-SG 2.34 (2.30) Å and Zn-SG 2.28 (2.28) Å. Therefore, the protonation states of Zn bound ligands are set the same as yCD complex with SG of Cys83, Cys86 and ND of His53 deprotonated. The imidazole bound form active site structure was reproduced by ONIOM method showing good agreement with the crystal structure (Imidazole was treated at high level.) that confirms the protonation state assignment. B3LYP (Lee et al. 1988; Becke 1993) with the 6-31G* basis set was used as the high level, while Am1 was used as low level (Dewar et al. 1985). Energy barriers were estimated by scanning the reaction coordinates. Therefore the barriers are the upper bounds of real reaction barriers.

The geometry of the inner layer was optimized for all the species while the atoms of the outer layer were fixed at their crystal structure positions (Sklenak et al. 2004). MD simulation shows the protein is quite rigid; especially the active site in both the reactant guanine and the product xanthine bound form, which suggests freezing the outer layer is a reasonable approximation (see Results).

MD simulation of apo and guanine bound form

In the MD simulation, the RESP charges of Zn, Zn bound water, the side-chains of His53, Cys83, and Cys86 as well as the bond, angle, dihedral force constants of Zn

complex are taken from yCD apo form simulation considering the same architecture of Zn complex in the two crystal structures. Atom-centered partial charges of guanine were derived by using the AMBER antechamber program (RESP methodology)(Pearlman et al. 1995) based on HF/6-31G* quantum calculation.

Starting coordinates for the protein atoms were taken from the crystal structure. Imidazole was removed from the active site of subunit one, while it was mutated to guanine in the active site of subunit two. So, subunit one is the apo form while subunit two is the guanine complex. All the crystal water molecules were maintained. The protonation states of the ionizable residues were set to their normal values at pH 7. The protein was surrounded by a periodic box of 12.5 Å with ~18,000 TIP3P water molecules. Na⁺ ions were placed by the Leap program (Pearlman et al. 1995) to neutralize the -12 charge of model system. The parm94 version of the all-atom AMBER force field (Cornell et al. 1996) was used for all the simulations.

MD simulations were carried out using the SANDER module in AMBER 7.0 (Pearlman et al. 1995). The SHAKE algorithm was used to constrain the bond lengths of all bonds involving hydrogen atoms permitting a 2 fs time step (Ryckaert et al. 1977). A nonbonded pair list cutoff of 8.0 Å was used and nonbonded pair list was updated every 25 steps. The volume and the temperature (300 K) of the system were controlled during the MD simulations by Berendsen's method (Berendsen et al. 1984). The Particle-Mesh-Ewald method was used to include the contributions of long-range electrostatic interactions (Essmann et al. 1995).

The simulation time was 2 ns with a 500ps equilibration period. Coordinates were saved every 2 ps. All of the MD results were analyzed using PTRAJ module of AMBER

7.0. In these analyses, hydrogen bonds were assigned when the distance of two heavy atoms (O or N) is less than 3.5 Å and the angle (heavy atom -hydrogen-heavy atom) is greater than 120°.

MD simulation of xanthine bound form

In order to understand whether water assists the Zn-O2 bond cleavage during xanthine release, MD simulation of xanthine bound form was performed. One xanthine was docked to each crystal structure active site based on the ONIOM minimized active site structure. In subunit one, the distance between Zn and O2 is restrained at 3.10 Å with force constant 100 kcal/(mol-Å²) mimicking the transition state of Zn-O2 bond cleavage. In subunit two, the distance was restrained at 3.80 Å with the same force constant mimicking the xanthine bound form with the Zn-O2 bond broken. Both distances are from the corresponding ONIOM calculations.

The same force constants for Zn complex were used for the apo and guanine form. RESP charges were fitted by using the same program as done for guanine, but based on B3LYP 6-31+G* calculations of ONIOM optimized active site structures which were trimmed to only include His53, Glu55, Cys83, Cys86, Asp114, Zn and xanthine (Yao et al. 2005). Only the charges of the side chains of His53, Cys83 and Cys86 as well as charges of Zn and xanthine were modified. Product ammonia was substituted by one water molecule in each active site considering the release of ammonia should be easier than xanthine. 2-ns MD was performed by using the same protocol as described above, while the data were analyzed by using the PTRAJ program.

Results

Reproduction of the crystal structure active site

In order to validate the ONIOM method used in this work, the crystal structure of the active site was reproduced. The calculated distances and angles for Zn complex are summarized in Table 8-1 comparing with those in the crystal structure. They are in good agreement indicating that the ONIOM method can reproduce the active site Zn architecture; therefore it is reasonable to assume that it can also give us fairly good energies for various species along the reaction path. To elucidate the reaction mechanism catalyzed by bGD, a series of ONIOM calculations were performed for various species as shown below.

Substrate Binding

A schematic representation of the active site of the substrate binding form is given in Figure 8-2. The bound substrate is stabilized by a hydrogen bond network and a π -stacking interaction with the imidazole ring of His53. The hydrogen bonds between substrate and surrounding residues are listed in Table 8-2 with distances between two heavy atoms. The amino group of guanine forms hydrogen bonds with the carboxyl group of Glu55 and the carbonyl group of Glu81. These two hydrogen bonds can assist the positioning of C2 for nucleophilic attack by the Zn-bound water in the following steps. The carboxyl group of Glu55 also forms one hydrogen bond with N1-H and another one with Zn bound water molecule, while the latter also forms one hydrogen bond with N3 of guanine. In the ONIOM optimized structure, guanine forms an additional four hydrogen bonds with Gln42, Ala54, Asp114 and Tyr156* (Figure 8-2, Table 8-2).

To form an active enzyme-substrate complex, a proton must transfer from the Zn bound water to N3 of guanine. The hydrogen bond between OH of water and N3 seems to favor the direct proton transfer. But the angle between the OH vector and purine plane

of guanine is $\sim 90^\circ$ which makes the direct proton transfer extremely difficult. Our ONIOM calculation shows that the barrier for direct transfer is ~ 42 kcal/mol. Thus, this transfer is unlikely to occur. Alternatively, this proton transfer may be assisted by the carboxyl group Asp114. But the distance between the O of water and the OD1 of Asp114 is quite far, ~ 4.2 Å when guanine binds, which makes the proton transfer from water to Asp114 quite difficult. However, the energy of the protonated Asp114 and Zn bound hydroxide is 0.7 kcal/mol higher than the deprotonated Asp144 and Zn bound water (complex 1 \rightarrow complex 2, Figure 8-3) when the substrate binds. So from a thermodynamics point of view, the Zn bound water can transfer its proton to OD1 of Asp114. This process will be discussed later.

The protonation of guanine

After the protonation, OD1 of Asp114 forms a new hydrogen bond with N3 of guanine (complex 2 in Figure 8-3) with heavy atom distance 2.81 Å while other hydrogen bonds between the ligand and enzyme are maintained. This new hydrogen bond stabilizes protonated Asp114. The distance between N2 and O of Zn bound hydroxide is shortened to 2.46 Å, compared with 2.78 Å before the proton transfer from the Zn bound water to Asp114. Apparently this transfer favors the hydroxide nucleophilic attack.

Then Asp114 transfers its proton to guanine (complex 3 in Figure 8-3) with a barrier of 4.5 kcal/mol when OD1-H distance is 1.25 Å. The barrier was calculated as following. The distance between OD1 and the proton was scanned by 0.1 Å per step, so the maximum energy could be located approximately. Then 0.05 Å was used for the scan around the maximum. Complex 3 is 0.7 kcal/mol more stable than complex 2. The proton

transfer from Asp114 to guanine makes the distance between C2 and O of Zn bound hydroxide even shorter (~ 2.29 Å) due to the positive charge of the ligand.

The formation of tetrahedral intermediates

After the proton transfer from the Zn bound water to N3 through Asp114, the hydroxide is well positioned for a nucleophilic attack. The shortened distance between C2 and Zn bound hydroxide due to the proton transfer makes the nucleophilic attack facile. The barrier for this process is 1.5 kcal/mol with the distance between C2 and O of Zn bound hydroxide 1.85 Å. The barrier was calculated by using the same method as described above. Such a low barrier suggests this process can occur fairly easily. The new species (complex 4 in Figure 8-3) is 0.1 kcal/mol more stable than complex 3. This process also shortens the distance between O of Zn bound hydroxide and OE2 of Glu55 from 2.88 Å to 2.63 Å and one hydrogen bond is formed between them. This distance shortening assists the next step of the reaction — the proton transfer from Zn bound hydroxide to the carboxyl of Glu55.

The proton transfer from Zn bound hydroxyl to Glu55 is composed of two sub-steps: (1). Rotation of the amine group of the tetrahedral intermediate to point N2-H1 to other directions so that OE2 of Glu55 can extract the proton from hydroxyl. (2). Proton transfer to OE2 of Glu55. The new state (complex 5 in Figure 8-3) has 3.3 kcal/mol higher energy than complex 4. Since the reaction coordinates for this process are rather complicated, the barrier was estimated based on the stepwise process, ~ 4.2 kcal/mol with OE2-H distance 1.20 Å from sub-step 2. Considering the energy difference between complex 5 and complex 4 is 3.3 kcal/mol, the real barrier should be between 3.3 kcal/mol and 4.2 kcal/mol.

After the proton transfer from the Zn bound hydroxyl to Glu55, the carboxyl group of Glu55 rotates $\sim 30^\circ$ to point OE2H to the amine group and then transfers this proton to the amine group. The barrier for this step is 2.6 kcal/mol with a distance between N2 and the proton of 2.05 Å. This new complex (complex 6 in Figure 8-3) is 4.1 kcal/mol more stable than complex 5.

The bond length between N2 and C2 was monitored along the reaction so far. Interestingly this bond was elongated quite a bit after the nucleophilic attack. This bond length is 1.37 Å (complex 1), 1.36 Å (complex 2), 1.35 Å (complex 3), 1.44 Å (complex 4), 1.46 Å (complex 5) and 1.56 Å (complex 6). The first dramatic change (from complex 3 to complex 4) may be caused by the hybridization change of C2 from sp^2 to sp^3 , which weakens the correlation between N2 π electrons and the purine ring π electrons, while the second dramatic change (from complex 5 to complex 6) may be caused by the fact that N2 orbital hybridization changes from sp^2 to sp^3 . After all, this 0.2 Å C2-N2 bond elongation should make the C2-N2 bond cleavage easier in the next step.

The formation of products

Since the amino group has been formed after the proton transfer from Glu55, the C2-N2 bond can be cleaved to form Zn bound xanthine and ammonia. The distance between C2 and N2 was scanned, the maximum was found at the distance 2.10 Å with the barrier 8.8 kcal/mol. The product complex (complex 7 in Figure 8-2) is 5.9 kcal/mol less stable than complex 6. The stationary position was found for ammonia at the distance 2.79 Å forming two hydrogen bonds with OE2 of Glu55 and carbonyl of Glu81. The two hydrogen bonds distances are 2.97 Å (between N of ammonia and OE2 of Glu55) and 3.17 Å (between N of ammonia and O of Glu81), indicating that ammonia is loosely

bound in the active site, which should favor complex **7** more than complex **6** from entropic point of view. Once ammonia is released, this reaction will not be reversible.

Therefore, from complex **1** to complex **7**, the overall reaction is endothermic with ~5.0 kcal/mol energy differences and the rate-limiting step is the C2-N2 bond cleavage with the barrier 8.8 kcal/mol (complex **6** to complex **7**).

Release of xanthine

Zn and O2 of xanthine are still covalently bonded in complex **7**; this bond has to be broken in order to release xanthine. The Zn-O bond was scanned and the maximum energy occurs at distance 3.10 Å with the barrier 8.4 kcal/mol slightly smaller than C2-N2 bond cleavage barrier in the previous step. Xanthine becomes stable when this distance increases to 3.79 Å with the energy 7.9 kcal/mol higher than the covalently bonded form (Figure 8-3). After the break of the Zn-O bond, xanthine is free to release, and then water comes in and coordinates with Zn to complete the catalytic cycle.

Or alternatively, water might access to Zn and assist the bond breaking which would make the cleavage of Zn-O bond and the binding of water a concerted step. In order to identify whether this is the case and the possible water path for this process, MD simulation of Zn bound xanthine was performed with Zn-O distance restrained at 3.10 Å in subunit 1 to mimic the transition state and 3.80 Å at subunit 2 to mimic the free xanthine complex. During 2-ns simulation, no water was observed to get close to Zn within 3.5 Å in either active site. Though three water molecules are present in both active sites, but none can get access to Zn because the path is blocked by xanthine (data not shown). Then the distance between Zn and O2 of xanthine was reduced to 2.0 Å in subunit 1 (mimic complex **8**), and enlarged to 4.8 Å in subunit 2, again no water was

observed close to Zn in either active site in 1-ns simulation time. Therefore, all these suggest that water binds to Zn more likely after the release of xanthine, or at least after the relocation of xanthine in the active site.

The protonation of Asp114

As discussed above, the first step for the mechanism is the proton transfer from Zn bound water to carboxyl group of Asp114. This happens either before or after guanine binds to the active site. In order to elucidate this process, the crystal structure was optimized by ONIOM without the bound imidazole, but keeping all the crystal water molecules and residues within 6 Å around the active site. Interestingly, one water molecule bridges the Zn bound water and carboxyl OD1 of Asp114 by forming one hydrogen bond with the former as an acceptor and one with the latter as a donor (Figure 8-5). The bridging water also forms a hydrogen bond with the carbonyl of Glu81. The heavy atom distances are 2.64, 2.60, and 3.23 Å for these three hydrogen bonds. This indicates that this bridging water should be rather stable. The Zn bound water is also hydrogen-bonded to Glu55 (Figure 8-5).

Interestingly, the MD simulation of the apo bGD also shows the bridging water molecule with similar interaction, but more dynamical. During the first 150 ps of the MD simulation, one (the first) water molecule bridges Zn bound water and Asp114, the same as seen in the ONIOM calculation, but during the next 175 ps, another (the second) water comes in and forms a two-water bridge. Then in the next 680 ps one-water bridge is resumed and this second water molecule moves away to form a hydrogen bond with OD2 of Asp114. This one-water bridge is broken during the next 370 ps and the first water moves close and hydrogen bonded to the carboxyl group of Glu55. Thereafter this

second water moves back to OD1 and recovers the one-water bridge, at the same time a new water molecule migrates from solvent to the active site. Overall one-water (two-water) bridge exists during 70% (10%) of the 1.5 ns simulation.

Since a similar water bridge was observed in both ONIOM and MD calculations, this water bridge should be stable and might assist the proton transfer from the Zn bound water to Asp114. The distance between the Asp114 carboxyl OD1 and the proton it is hydrogen bonded to was scanned in an ONIOM calculation; the maximum energy occurs at the distance 1.20 Å with the barrier 7.0 kcal/mol (Figure 8-5). In the end, the bridging water transfers its proton to OD1 and extracts one proton from the Zn bound water. The energy of the protonated Asp114 with deprotonated Zn bound water is 5.20 kcal/mol higher than the energy of deprotonated Asp114 with Zn bound water. So it is more likely that the proton stays with the Zn bound water instead of Asp114 in the apo enzyme. However, from the previous calculations, we know that after the binding of guanine, the energy difference between these two states is only 0.70 kcal/mol in favor of the Zn bound water. So, the binding of guanine shifts the equilibrium to the right side of Figure 8-5. It stabilizes the protonated Asp114 probably by two hydrogen bonds between N3 and OD1-H, N9-H and OD2. But the question comes up as to whether it is possible that the proton transfer from the Zn bound water to Asp114 through the water bridge occurs after the binding of guanine.

In order to answer this question, we first tried to insert a water molecule in between Asp114 and the Zn bound water after guanine binds in the ONIOM calculation. After the energy minimization, the bridging water forms one hydrogen bond with the Zn bound water as an acceptor and another hydrogen bond with N3 instead of OD1 of Asp114.

Different orientations of the inserted water were tried, but the same minimized structure was obtained. Thus, it is more likely that the proton transfers through the water directly to N3 instead of to Asp114 in this case. The proton of the Zn bound water was transferred to N3 of guanine by shortening the distance between N3 and the proton of the bridging water. The proton transfer from the Zn bound water to the bridging water occurs at the distance 1.20 Å with a barrier of 14.7 kcal/mol and the end state is 8.2 kcal/mol higher than the starting state. This large barrier makes the proton transfer through the water bridge unlikely to happen compared with the previous proposed process. Furthermore, the inserted water molecule makes the region quite crowded such that the distance between O of the bridging water and carbonyl O of Glu81 is only 2.31 Å, which makes the van der Waals interaction very unfavorable and the bridging water less stable.

Since during an ONIOM calculation, the outer layer residues are fixed, which makes the space tight around guanine, the motion of these residues, especially Glu81, might be able to accommodate one water molecule in between the Zn bound water and Asp114. But MD simulation shows that the water bridge doesn't exist when guanine binds to the active site. Though two water molecules are seen around the carboxyl group of Asp114, none of them forms hydrogen bonds with the Zn bound water because the presence of guanine makes the space too small for water molecules.

Therefore, all these results suggest that it is unlikely that the proton transfers to Asp114 (or to guanine) after guanine binds. Instead it is more likely that proton transfer to Asp114 occurs through the water bridge right before the positioning of guanine in the binding pocket as seen in ONIOM, which then pushes away the bridging water and

stabilizes the protonated Asp114 by forming two hydrogen bonds with it. After this, the subsequent reactions occur as described above.

Comparison of MD and ONIOM

The outer layer was frozen in the ONIOM calculation, but the active site structure and interactions might change when allowing its movement. One good way to investigate this is by comparing the active site from the MD simulation with the corresponding structure from ONIOM. The hydrogen bond network of the guanine bound enzyme complex (complex 1) was investigated in MD simulation, which appears to be the same as that from the ONIOM calculation (Table 8-2). For the Zn-xanthine product complexes (complex 8 and complex 9) MD simulations, the same hydrogen bond interactions were also observed as in the ONIOM calculations (Table 8-3,4). Generally the distances between heavy atoms in MD are quite consistent with quantum results, and most distances from ONIOM can be covered by the corresponding MD numbers within their fluctuations (Table 8-2, 3, 4). This indicates that fixing the outer layer in ONIOM doesn't alter the active site interactions significantly in the reactant and the product bound form. Another way to justify the constraint of the outer layer is by checking its rigidity in the MD simulations. The Root Mean Square Fluctuation (RMSF) calculations show that the all-atom RMSF of the 26 residues in the active site (which were included in ONIOM) are quite small; only 0.45 Å for complex 1, 0.41 Å for complex 8 and 0.44 Å for complex 9. Thus, it is reasonable to fix the outer layer in these ONIOM calculations.

The outer layer of ONIOM has to be large enough to cover most of the environmental effect, but as small as possible to minimize the computational cost. In order to identify whether 26 residues in ONIOM are sufficient, the interaction energies

between guanine and its surroundings are calculated from snapshots of 1.5-ns MD simulation. The average interaction energy is -103.9 ± 4.42 kcal/mol between guanine and those 26 residues compared with 3.54 ± 3.62 kcal/mol between guanine and the rest of the system (including solvent). So the effect of the rest of the system is rather small, only ~3%. Thus, it is sufficient to include those 26 residues in the ONIOM calculations.

Alternative mechanism

As described above, the first step in the reaction, proton transfer from Zn bound water to Asp114 is important for the following steps. But, this step cannot be studied directly by using ONIOM, because it seems to be quite dynamical. On the other hand, it is possible that in the apo form Asp114 is already protonated, but the proton comes from somewhere else instead of the Zn bound water. Considering that this residue is buried in the active site that is sealed by the C-terminal helix (Liaw et al. 2004), the pKa of Asp114 might be different from those of model peptides (~4.0). A similar mechanism to the one above can be proposed based on protonated Asp114, as shown in Figure 8-6. Only the thermodynamic energies are evaluated. First, Asp114 transfers its proton to guanine, which increases the energy by ~9.6 kcal/mol (from complex 1' to complex 2', Figure 8-6). Second, Glu55 extracts one proton from the Zn bound water and forms a hydrogen bond with the guanine amine group. At the same time, the Zn bound hydroxide attacks C2 to form a tetrahedral intermediate, which increases the energy by ~7.4 kcal/mol (from complex 2' to complex 3', Figure 8-6). Alternatively, first Glu55 extracts one proton from the Zn bound water, which increases the energy by ~13.6 kcal/mol (complex 1' to complex 2'', Figure 8-6); second, Asp114 transfers its proton to N3 and the Zn bound hydroxide attacks C2 which gives the same intermediate (complex 3', Figure 8-6) as

above. That increases the energy by 3.4 kcal/mol. Third, Glu55 transfers its proton to the guanine amine, and the energy goes up by another 1.2 kcal/mol (complex **4'**, Figure 8-6). Fourth, Glu55 extracts another proton from the Zn bound hydroxide and at the same time the C2-N2 bond breaks automatically, which drops the energy by ~14.0 kcal/mol (complex **5'**, Figure 8-6). Finally, we forced the proton to attach bind to ammonia using a constraint. But, once this constraint is removed, the proton jumps back to Glu55, suggesting that this proton prefers to stay with Glu55. So, the final products are xathine, ammonia, and protonated Glu55. The overall energy change for this mechanism is ~4.2 kcal/mol endothermic, which is comparable with the ~5.0 kcal/mol from the previous mechanism. But, all the intermediates have much higher energies. The complex **4'** is 18.2 kcal/mol less stable than complex **1'** (Figure 8-6) compared with its analog complex **6** that is 0.9 kcal/mol more stable than complex **1** (Figure 8-2) in the previous mechanism. This 18.2 kcal/mol difference would make this reaction far less efficient. Therefore, this new mechanism based on protonated Asp114 is less likely to occur than the mechanism based on deprotonated Asp114.

The mechanism proposed (Figure 8-3) with deprotonated Asp114 as the starting point is reminiscent of the reaction mechanism proposed for the deamination reaction catalyzed by yeast Cytosine Deaminase (Sklenak et al. 2004). In the yCD system, two protons are transferred by Glu64 (like Glu55 in bGD) from the Zn bound water to cytosine (like guanine in GD). Here, Asp114 transfers the first proton and Glu55 transfers the second proton to guanine. In fact, if N3 is protonated (complex **b** in Figure 8-7) instead of N1 (complex **a** in Figure 8-7), the same mechanism can be proposed for the bGD catalyzed reaction up to the release of xanthine. Here we want to address whether

this is a possible mechanism. It has been predicted that complex **a** is 5.2 kcal/mol more stable than complex **b** in aqueous systems (Jang et al. 2003). The relative stability of these two compounds in the enzyme was also evaluated. Just like in aqueous system, ONIOM calculation shows complex **a** is 5.8 kcal/mol more stable than complex **b** in GD active site. Considering complex **b** is a minor form in solution and that the enzyme also prefers to bind complex **a** instead of **b**, the probability for complex **b** binding should be much lower which makes the yCD-like mechanism unlikely occur.

Discussion

This ONIOM study provides a detailed mechanism of the guanine deamination reaction, with the energetics of all the stable species as well as the barriers along the reaction path. First, one proton is transferred from the Zn bound water to Asp114. This process might be quite complicated and rather dynamic. In the apo form proton transfer is possible through the water bridge but not favorable (the energy increases 5.2 kcal/mol), while in the guanine bound form it is less unfavorable (the energy only increases by 0.7 kcal/mol) but difficult since no water bridge is present. Water entry in the location that was observed in ONIOM and MD might happen just before the positioning of guanine. The positioning of guanine breaks the water-bridge and stabilizes protonated Asp114 by forming two hydrogen bonds with it. Second, Asp114 transfers its proton to N3. The first two steps shorten the distance between C2 and Zn bound hydroxide and favor the nucleophilic reaction. Third, Zn bound hydroxide attacks C2 to form a tetrahedral intermediate with a rather low barrier, which also moves Zn bound hydroxide closer to OE2. Fourth, Glu55 transfers one proton from Zn bound hydroxide to N2 by rotating its dihedral CB-CG-OE1-H $\sim 30^\circ$. Till now, the C2-N2 bond length has been elongated 0.2

Å, and that makes the C2-N2 bond cleavage more facile. Fifth, the C2-N2 bond is broken and ammonia forms, which gives the highest barrier ~8.8 kcal/mol so far. Sixth, ammonia leaves the active site and xanthine is freed by the cleavage of the Zn-O2 bond with the barrier ~8.4 kcal/mol. Therefore, along the reaction path the highest barrier comes from the C2-N2 bond cleavage, while the barrier from the cleavage of the Zn-O2 bond is slightly smaller. Then, xanthine leaves the active site and water moves close and binds to Zn to finish the whole enzymatic reaction cycle. Along this path, Glu55 and Asp114 play an important role by acting as proton shuttles.

One alternative mechanism was proposed that involves protonated Asp114. Even though the overall reaction energy difference between product and reactant is comparable to that with deprotonated Asp114, the energies of various intermediates along the reaction path are much higher. This makes this mechanism less likely happen in the reaction.

One tautomer of guanine with deprotonated N1 and protonated N3 was also discussed, which would give a similar mechanism as the proposed (Sklenak et al. 2004) yCD catalyzed reaction if it is present in the active site. However, this tautomer is 5.8 kcal/mol less stable than guanine in the active site, similar to what was found in an aqueous environment (5.2 kcal/mol less stable). The probability to bind this tautomer is much lower than guanine, which makes the yCD-like mechanism unlikely.

In the ONIOM calculation, only 26 residues of the enzyme around the active site were included. Our MD simulation of guanine bound GD shows that these 26 residues contribute 97% of the interaction energies between the ligand and its surroundings. It suggests that it is sufficient to include just this part of the enzyme in the ONIOM

calculation. But, since only part of the protein is used, the energy minimization of the outer layer without constraints might give an unrealistic expansion of the protein. So, the outer layer was fixed in the quantum calculations. Our MD simulations of the protein with explicit water molecules lead to the same hydrogen bond network in the active site as in the ONIOM calculations, for both the reactant and product complexes, with consistent distances between heavy atoms. Therefore, the interaction between ligand and its surroundings is well defined and quite stable. The constraint on the outer layer residues didn't introduce artifacts in the hydrogen bond analysis of the active site.

The flexibility of active site was also investigated in the MD simulations. The all-atom RMSFs of those 26 residues are only ~ 0.4 Å in both reactant and product complexes. The rigidity of the active site in MD explains why MD gives similar active site interaction compared with ONIOM. This also suggests that it is a reasonable approximation to freeze the outer layer in the ONIOM calculations.

Conclusion

One reaction path of the deamination of guanine to xanthine catalyzed by GD was proposed based on a two-layer ONIOM and MD study. The highest barrier comes from C2-N2 bond cleavage. The Zn-O2 bond can be broken without the assistance of water during the release of xanthine. Glu55 and Asp114 were identified to assist two proton transfers from Zn bound water to the ligand.

Reference

Becke, A. D. (1993). "Density-Functional Thermochemistry .3. The Role of Exact Exchange." Journal of Chemical Physics **98**(7): 5648-5652.

Berendsen, H. J. C., J. P. M. Postma, et al. (1984). "Molecular-Dynamics with Coupling to an External Bath." Journal of Chemical Physics **81**(8): 3684-3690.

Cornell, W. D., P. Cieplak, et al. (1996). "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995)." Journal of the American Chemical Society **118**(9): 2309-2309.

Dapprich, S., I. Komaromi, et al. (1999). "A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives." Journal of Molecular Structure-Theochem **462**: 1-21.

Dewar, M. J. S., E. G. Zoebisch, et al. (1985). "The Development and Use of Quantum-Mechanical Molecular-Models .76. Am1 - a New General-Purpose Quantum-Mechanical Molecular-Model." Journal of the American Chemical Society **107**(13): 3902-3909.

Dudev, T. and C. Lim (2001). "Modeling Zn²⁺-cysteinate complexes in proteins." Journal of Physical Chemistry B **105**(43): 10709-10714.

Dudev, T. and C. Lim (2002). "Factors governing the protonation state of cysteines in proteins: An ab initio/CDM study." Journal of the American Chemical Society **124**(23): 6759-6766.

Essmann, U., L. Perera, et al. (1995). "A Smooth Particle Mesh Ewald Method." Journal of Chemical Physics **103**(19): 8577-8593.

Frisch, M. J. "Gaussian 03."

Humbel, S., S. Sieber, et al. (1996). "The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for n-butane conformation and S(N)₂ reaction: RCl+Cl." Journal of Chemical Physics **105**(5): 1959-1967.

InsightII "InsightII."

Ireton, G. C., M. E. Black, et al. (2003). "The 1.14 Å crystal structure of yeast cytosine deaminase evolution of nucleotide salvage enzymes and implications for genetic chemotherapy." Structure **11**: 961-972.

Jang, Y. H., W. A. Goddard, et al. (2003). "pK(a) values of guanine in water: Density functional theory calculations combined with Poisson-Boltzmann continuum-solvation model." Journal of Physical Chemistry B **107**(1): 344-357.

Lee, C. T., W. T. Yang, et al. (1988). "Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density." Physical Review B **37**(2): 785-789.

Liaw, S. H., Y. J. Chang, et al. (2004). "Crystal structure of Bacillus subtilis guanine deaminase - The first domain-swapped structure in the cytidine deaminase superfamily." Journal of Biological Chemistry **279**(34): 35479-35485.

Pearlman, D. A., D. A. Case, et al. (1995). "Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules." Computer Physics Communications **91**(1-3): 1-41.

Ryckaert, J. P., G. Ciccotti, et al. (1977). "Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes." Journal of Computational Physics **23**(3): 327-341.

Sklenak, S., L. S. Yao, et al. (2004). "Catalytic mechanism of yeast cytosine deaminase: An ONIOM computational study." Journal of the American Chemical Society **126**(45): 14879-14889.

Svensson, M., S. Humbel, et al. (1996). "ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition." Journal of Physical Chemistry **100**(50): 19357-19363.

Vreven, T. and K. Morokuma (2000). "On the application of the IMOMO (integrated molecular orbital plus molecular orbital) method." Journal of Computational Chemistry **21**(16): 1419-1432.

Yao, L. S., Y. Li, et al. (2005). "Product release is rate-limiting in the activation of the prodrug 5-fluorocytosine by yeast cytosine deaminase." Biochemistry **44**(15): 5940-5947.

Yao, L. S., S. Sklenak, et al. (2005). "A molecular dynamics exploration of the catalytic mechanism of yeast cytosine deaminase." Journal of Physical Chemistry B **109**(15): 7500-7510.

Appendices

Table 8-1 Comparison of ONIOM optimized structure with crystal structure of the imidazole inhibitor complex

Internal Coordinate	ONIOM	X-ray Structure
Zn-O (wat)	1.96	2.03
Zn-ND (His53)	2.03	2.06
Zn-SG (Cys83)	2.41	2.34
Zn-SG (Cys86)	2.37	2.28
SG(Cys83)-Zn-SG(Cys86)	115.6°	130.7°
SG(Cys83)-Zn-ND(His53)	101.9°	105.0°
SG(Cys86)-Zn-ND(His53)	105.9°	112.7°
SG(Cys83)-Zn-O(wat)	110.5°	106.8°

Table 8-2 Hydrogen bonds in GD substrate bound form

	ONIOM	MD	
	Distance (Å)	Distance (Å)	Occurrence (%)
Gln42-NDH...O6	3.09	3.00(0.18)	97.3
Ala54-NH...O6	3.01	3.03(0.16)	99.7
Glu55-OE1...H-N1	2.84	2.88(0.16)	80.1*
Glu55-OE2...H1-N2	3.05	3.11(0.22)	86.5*
Glu81-O...H2-N2	3.46	3.16(0.19)	74.1
Zn-WAT-OH...N3	3.05	3.07(0.16)	98.3
Asp114-OD1...H-N9	2.73	2.96(0.16)	98.0 ^a
Tyr156-OH...N7	3.32	3.03(0.19)	92.3

* Due to the rotation of carboxyl group, OE2(OE1) also forms hydrogen bond with H-N1(H1-N2).

^a In MD OD2 also forms hydrogen bond with H-N9 with distance 2.96(0.18) and occurrence 96.8%.

Table 8-3 Hydrogen bonds in complex **8** bound form

	ONIOM	MD	
	Distance (Å)	Distance (Å)	Occurrence (%)
Gln42-NDH...O6	3.06	2.90(0.14)	97.8
Ala54-NH...O6	3.15	3.15(0.16)	95.2
Glu55-OE1...H-N1	2.73	2.79(0.12)	83.2
Asp114-OD1...H-N3	2.68	2.96(0.18)	78.8
Asp114-OD2...H-N9	2.69	2.78(0.10)	100
Tyr156-OH...N7	2.82	2.94(0.17)	97.8

Table 8-4 Hydrogen bonds in complex **9** bound form

	ONIOM	MD	
	Distance (Å)	Distance (Å)	Occurrence (%)
Gln42-NDH...O6	3.20	2.92(0.14)	99.7
Ala54-NH...O6	3.11	3.20(0.15)	91.9
Glu55-OE1...H-N1	2.77	2.79(0.09)	100
Asp114-OD1...H-N3	2.64	2.79(0.09)	100
Asp114-OD2...H-N9	2.73	2.81(0.11)	100
Tyr156-OH...N7	2.99	2.89(0.16)	99.2

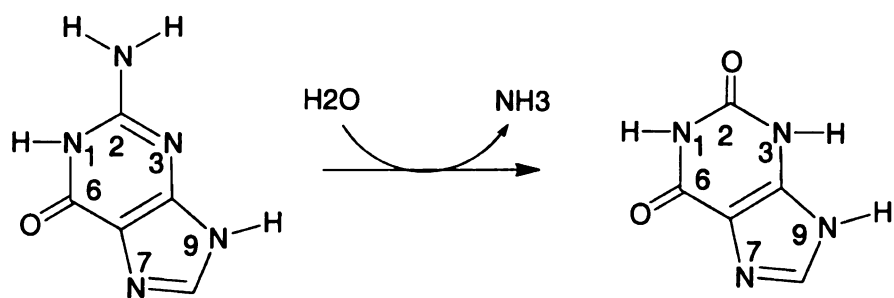


Figure 8-1 The deamination reaction catalyzed by GD.

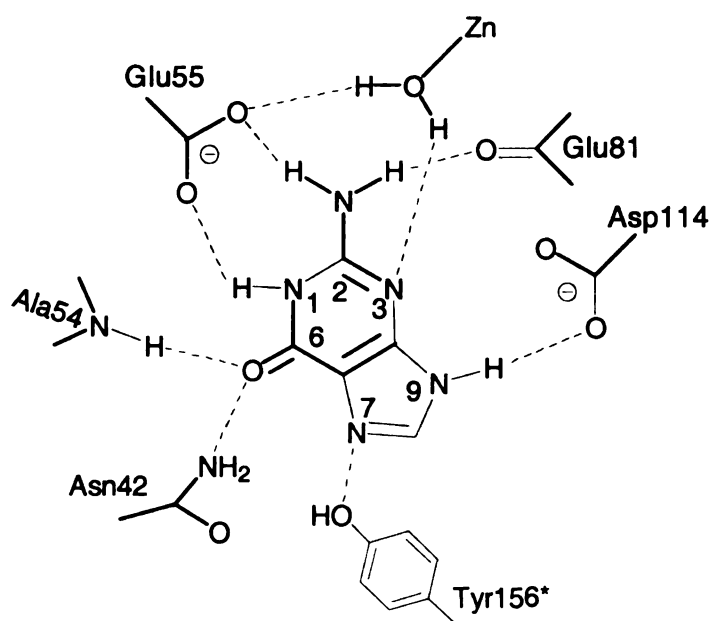


Figure 8-2 Active site interaction of the guanine bound complex from the ONIOM calculation.

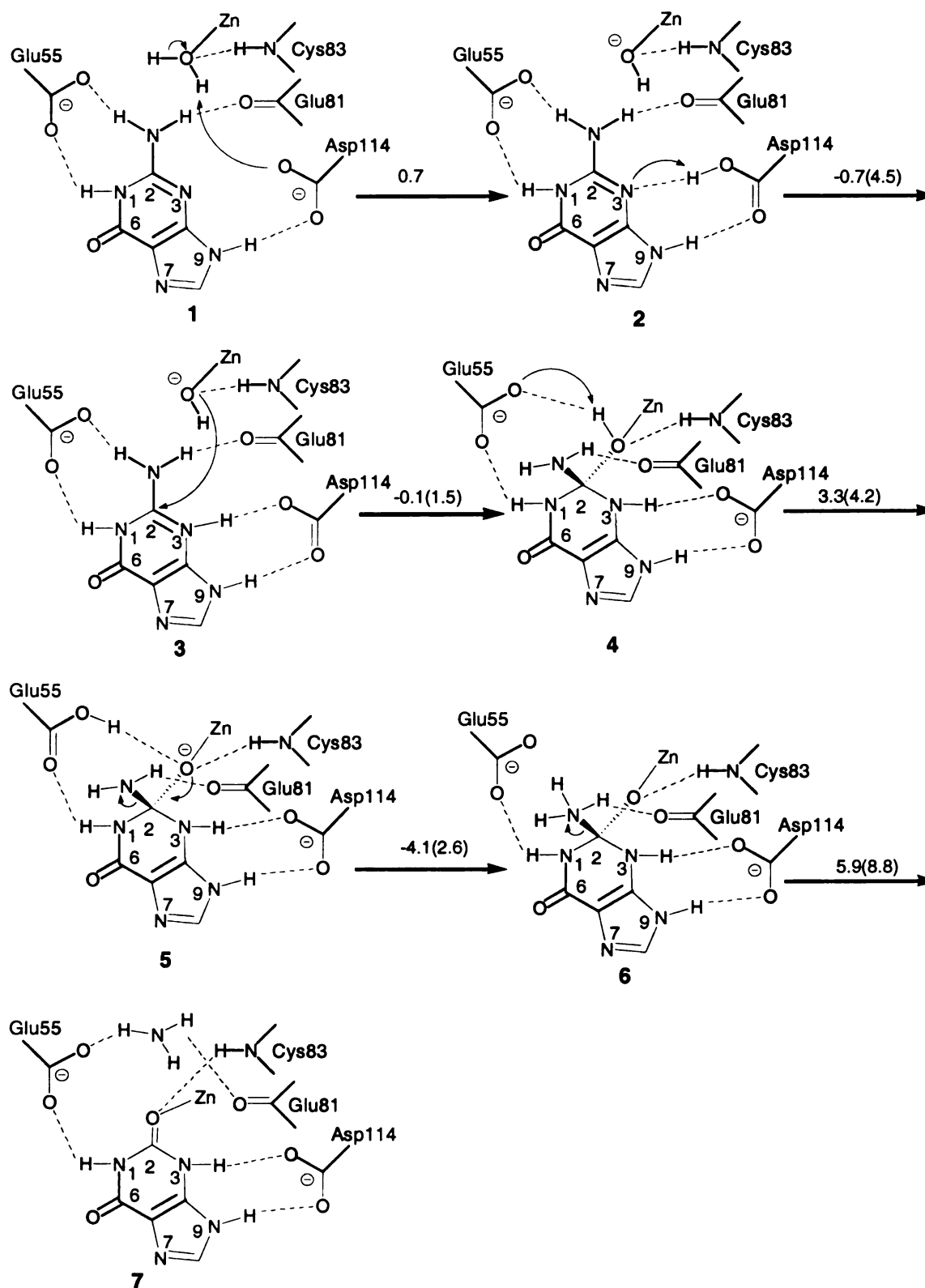


Figure 8-3 Reaction mechanism proposed for deamination of guanine to xanthine catalyzed by GD. All the species are labeled with numbers (see text).

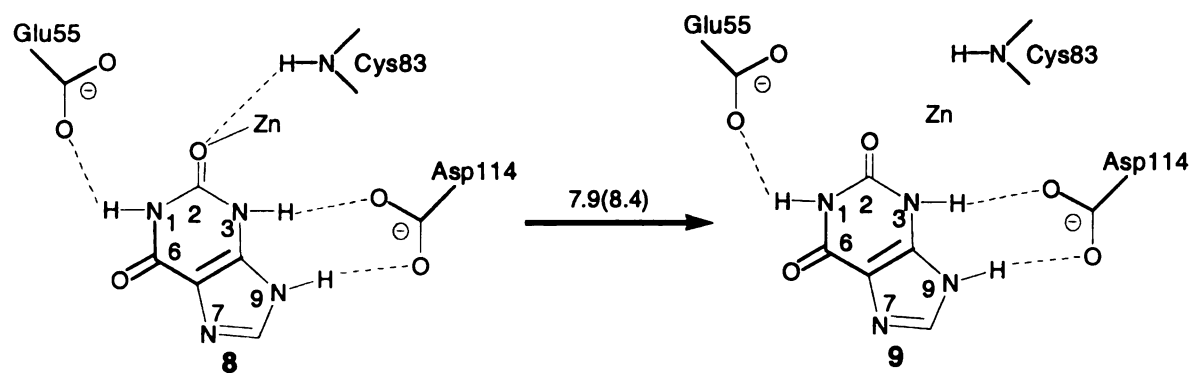


Figure 8-4 Zn-O bond cleavage.

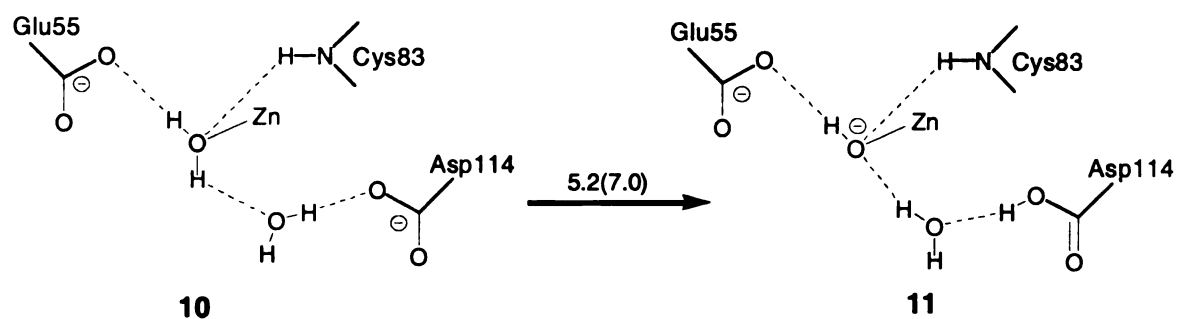


Figure 8-5 The proton transfer from Zn bound water to Asp114 through a water bridge.

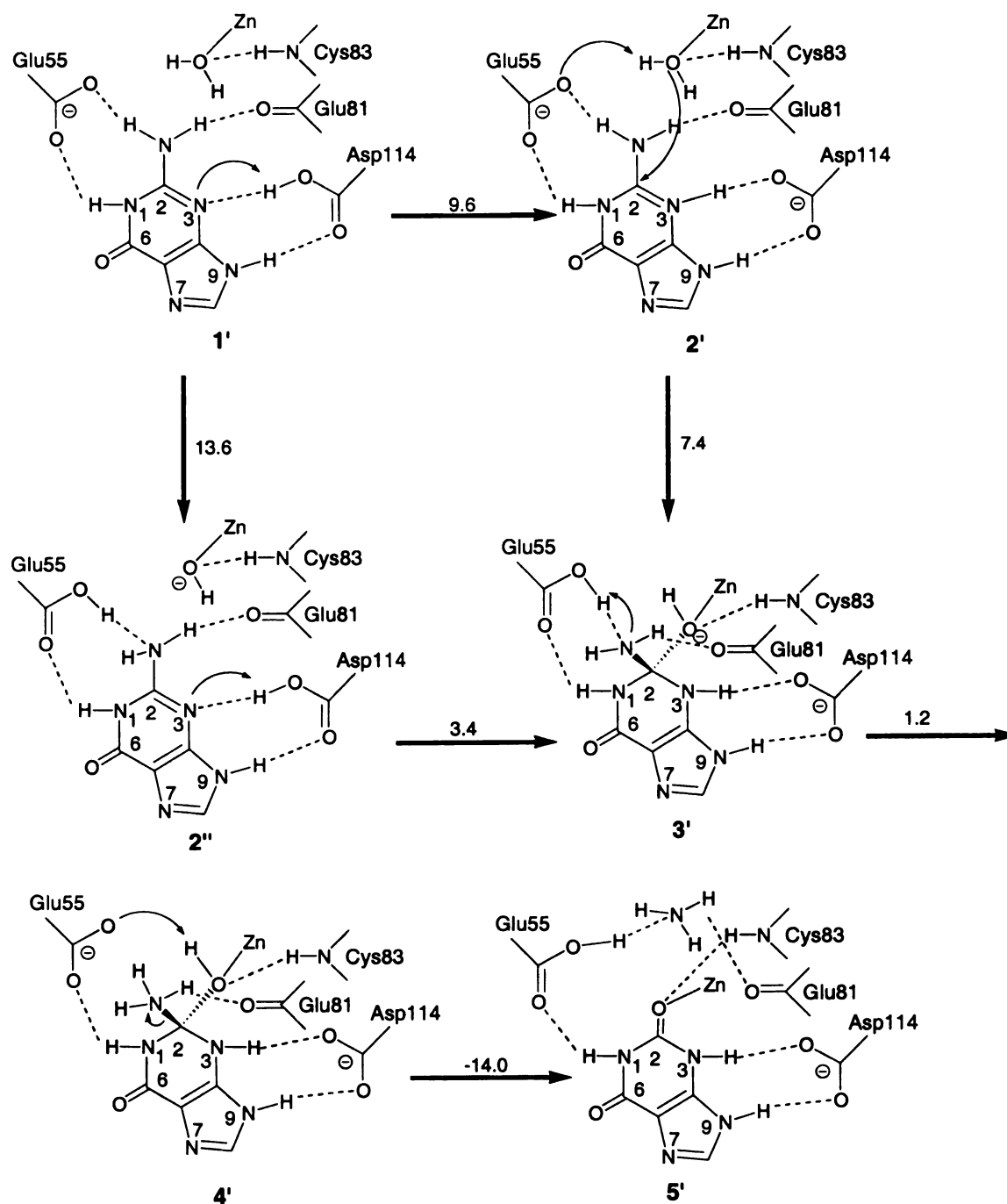


Figure 8-6 Alternative mechanism with protonated Asp114.

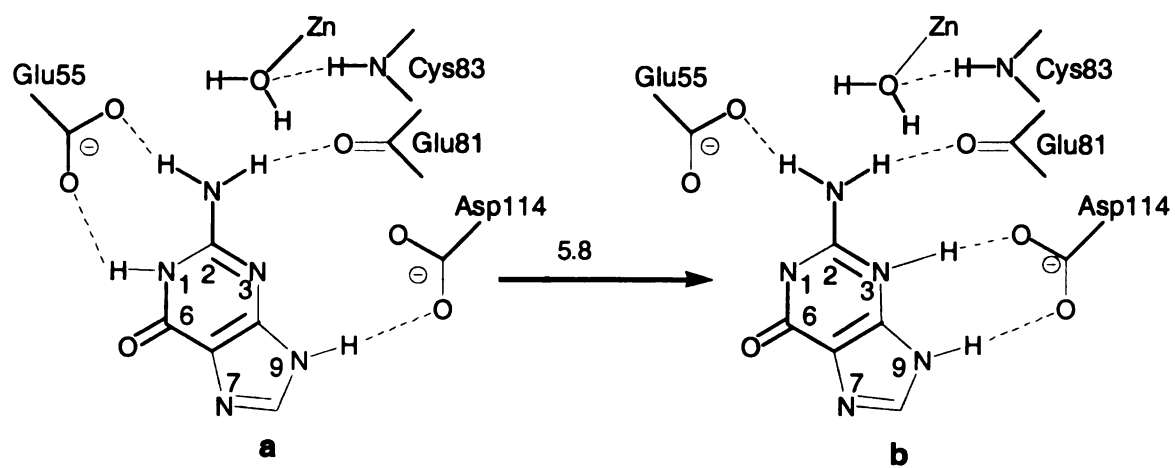


Figure 8-7 The tautomerization of guanine

Chapter 9

Mechanism of Dihydroneopterin Aldolase: A Molecular Dynamics Study of the Apo Enzyme and Its Product Complex

Introduction

Dihydroneopterin aldolase (DHNA) catalyzes the conversion of 7,8-dihydroneopterin into 6-hydroxymethyl-7,8-dihydropterin (HP) and glycolaldehyde that may occur by a retroaldol-based, carbon-carbon bond cleavage mechanism.¹ Plants and most microorganisms obtain folate cofactors by *de novo* synthesis, while vertebrates must depend on nutritional sources.^{2,3} Therefore, DHNA, along with other enzymes in the folate pathway, is a potential target for antimicrobial and anti-parasitic drugs.³ The particular enzyme under study, DHNA from *Staphylococcus aureus*, has recently been explored for this purpose.⁴

S. aureus DHNA crystallizes as a D₄-symmetric homo-octamer.^{5,6} A solution NMR study also leads to an octameric protein structure.⁷ Structures are available for the apo and HP-bound forms of DHNA.⁵ There are eight active sites that form at the interfaces between pairs of adjacent protomers. The bottom of the active site has a glutamate (Glu74) that serves as an “anchor” for binding and there are residues lining the active site that stabilize HP. Why DHNA and at least one other folate pathway enzyme are

octameric proteins is not clear. There is a suggestion that because DHNA can be a partner in a bi-functional enzyme, an active site formed from a noncovalent interface between protomers may aid in product/reactant transport between the active sites.⁸

In this article, DHNA and its product-bound complex are studied by molecular dynamics (MD) simulations and complemented by a series of simulations based on pushing HP out of its active site. The octameric structure with its eight active sites provides the advantage that for a given simulation length, assuming independence of the active sites (in apo and HP-bound forms), there is eight times the data that would be available from a monomeric protein simulation. Comparing statistical measures of protein residue fluctuations that are based either on the entire octamer or on the individual dimeric units provides a convenient way to study the active sites of DHNA. In particular, the issue of the flexibility of residues forming the active site and their rigidification by HP binding can be addressed. The binding of HP does indeed rigidify DHNA both in DHNA's flexible regions and in its more rigid regions. Another view of such changes in DHNA can be obtained by Principal Component Analysis (PCA)⁹, which decomposes the fluctuations of a protein into, typically, a relatively small set of modes that account for a great deal of the motion and a remainder, large set of small fluctuations that contribute little to the total fluctuation.¹⁰ As applied to apo and HP-bound DHNA, the PCA analysis shows that in both structures there are small sets of modes that do dominate the total protein fluctuation. Furthermore, HP binding substantially reduces the total fluctuation relative to the apo form, again indicating the rigidification of the protein induced by HP. The character of the PCA modes and their differences between apo and HP bound forms provides insight into the nature of the binding site.

As noted above, the active site is formed by the noncovalent interaction of protomer pairs and that provides impetus for exploring the exit pathway of bound HP. Such a study requires determination of an exit pathway and construction of the potential of mean force along the path. These are both challenging tasks since the true de-binding path most likely is a high dimensional one that is not determinable exogenously and, once chosen, constructing a potential of mean force to cover the bound to free form is computationally expensive.¹¹⁻¹⁴ Here, a compromise approach will be used where, by examination of the crystal structure bound form, a possible exit pathway is found and implemented by restraining the distance between one atom in DHNA and one in HP. Provided that the atoms are suitably chosen, and the distance between the atoms is increased at a sufficiently slow rate, the protein structure should not be distorted by the added force required to push out the ligand. By using a restraining potential between only a pair of atoms, HP should be free to translate and rotate to find an exit path from the bound state that should resemble the true path.

Instead of trying to determine directly the free energy along the exit pathway, only the energetics of the HP-DHNA interaction will be evaluated with averages over time windows that are chosen to obtain reliable energy averages. This energetics will serve to show which residues are involved in gradually releasing HP from the binding pocket. The internal energy of HP along the exit pathway is also evaluated since it could provide a substantial energetic contribution from the release of strain energy as HP exits from its binding site. An estimate of the free energy can be obtained by incorporating entropic contributions from the increasing freedom of HP as it is released from the binding site and the difference in the DHNA entropy of the apo and HP-bound forms. The summation

of these entropic and energetic terms then provides an estimate of the kinetic barrier to HP release.

Methods

Starting coordinates for the MD simulations of the apo and HP product complex forms of DHNA were taken from crystal structures.⁵ The protein atom coordinates of the two forms are essentially the same, with a ~ 0.3 Å root-mean-square deviation (RMSD) for all CA atoms. The N5 of HP was deprotonated in accordance with the spectroscopic evidence¹⁵ and all residues were assumed to be in their standard ionization states at pH 7. Atom-centered partial charges were derived by using the AMBER antechamber program (RESP methodology).¹⁶ The RESP charges for HP were obtained at the HF/6-31G* level. MD simulations for apo and HP complex forms were performed using the SANDER module in the AMBER 7.0 program, with the amber94 force field. All crystallographic waters were removed and the system solvated with $\sim 19,500$ TIP3P water molecules and 32 Na⁺ ions were added to neutralize the system. The MD simulations were run at constant volume for 5 ps to heat up the system from 100 K to 300 K, then at constant temperature and pressure for ~ 25 ps to get the proper density, and then at constant temperature and volume for 2 ns. The simulation box is approximately a cube of side 90 Å, and the octameric protein is cylindrical in shape with diameter and major axis approximately 70 Å. All velocities were assigned according to the Maxwell-Boltzmann distribution at 100 K; and the system was coupled to a temperature bath with a time constant of 0.5 ps.¹⁷ The Particle-Mesh-Ewald method¹⁸ was used to evaluate the contributions of the long-range electrostatic interactions. A nonbonded pair list cutoff of 8.0 Å was used and nonbonded pair list was updated every 25 steps. All bonds to

hydrogen atoms were constrained by using the SHAKE algorithm¹⁹ allowing a time step of 0.002 ps. The coordinates were saved every 2 ps, and the trajectories after 500 ps were analyzed with the Moil-view program and the PTRAJ module of AMBER 7.0. Hydrogen bonds were assigned when the distance of two heavy atoms (O or N) is less than 3.2 Å and the angle (heavy atom-hydrogen-heavy atom) is greater than 120 °.

The starting structure for pushing HP out of the active site was taken from a snapshot of the HP-DHNA complex simulation. A harmonic restraint was added between the backbone N of Arg118 and C10 of HP. The reasons to choose the N of Arg118 are: 1. Arg118 is in the C-terminal beta sheet, which is quite rigid (see discussion below). 2. The vector between Arg118 N and HP C10 points along the channel between protomer pairs that is quite open and water accessible. Different force constant values were tested in the simulation. Using too small a constant may not push out the ligand efficiently, while using too large a constant may take too long to equilibrate the system. The force constant finally used was 20 kcal/mol-Å². The equilibrium distances were changed by 0.2 Å every 50 ps. Since the force constant used corresponds to rms fluctuations around the equilibrium distance of ~0.25 Å, this pushing rate should give us continuous sampling along the exit path. HP was pushed out by 8 Å in 2 ns. It moves out of the active site and stays on the surface of the protein. A simulation was also run that starts with all sites occupied by HP but only pushes out one of them, otherwise using the same protocol and parameter values. The HP protein interaction energetics along the trajectory was found to be similar to that for pushing all out at the same time. The energetic fluctuations for pushing out one protomer were greater than when averaged over the eight protomers, as expected.

Root-mean-square fluctuations (RMSF) were analyzed in two ways that are based on different averaging procedures, stimulated by the feature that DHNA is a homo-octamer. Denote the vector location of atom r for the t th configuration in the i th protomer as $r_t^i = \{x_t^i, y_t^i, z_t^i\}$. The trajectory average over the N_C configurations ($t = 1, 2, \dots, N_C$) for the i th protomer is

$$\langle r_t^i \rangle_c = \frac{1}{N_C} \sum_{t=1}^{N_C} r_t^i, \quad (1)$$

and the further average over the N_P protomers ($i = 1, 2, \dots, N_P$) is

$$\langle \langle r_t^i \rangle_c \rangle_p = \frac{1}{N_P} \sum_{i=1}^{N_P} \langle r_t^i \rangle_c. \quad (2)$$

Also denote deviations from these two averages as $\delta_i r_t^i = r_t^i - \langle r_t^i \rangle_c$ and

$\delta r_t^i = r_t^i - \langle \langle r_t^i \rangle_c \rangle_p$, with δ_i distinguishing which protomer average is used in the

former definition. The first RMSF we define is

$$RMSF1 = \left\langle \left\langle (\delta_i r_t^i)^2 \right\rangle_c^{1/2} \right\rangle_p \equiv \langle RMSF_i \rangle_p. \quad (3)$$

RMSF1 is the average over protomers of the i th protomer $RMSF_i$ and will be referred to as the average protomer RMSF. The second RMSF we define is

$$RMSF2 = \left\langle \left\langle (\delta r_t^i)^2 \right\rangle_c^{1/2} \right\rangle_p. \quad (4)$$

RMSF2 is an RMSF obtained by treating the trajectory as one record over the entire $(N_p)*(N_c)$ data, and will be referred to as the octamer RMSF. RMSF2 does not distinguish atom r among the different protomers. To the extent that the averages in these deviations are different, there is no rigorous connection between the two RMSFs. If, however, the protomer averages are independent of protomer, $\left\langle r_t^i \right\rangle_c = \left\langle \left\langle r_t^i \right\rangle_c \right\rangle_p$

($i = 1, \dots, N_p$), then they can be related as follows. Using the protomer $RMSF_i$ definition in Eq. (3), the difference between RMSF2 and RMSF1 can be expressed as

$$\Delta RMSF = \left[RMSF_2^2 - RMSF_1^2 \right]^{1/2} = \left\langle RMSF_i^2 \right\rangle_p - \left\langle RMSF_i \right\rangle_p^2 > 0 \quad (5)$$

Thus, $\Delta RMSF$ is the RMS fluctuation over different protomers of the individual protomer RMSFs. This indicates that, to the extent that the average values of the vector positions are independent of protomer, as they will be approximately (when referenced to the same coordinate origin), the RMSF2s will be larger than the RMSF1s. This difference should be magnified for the atoms that are intrinsically more mobile, and this feature will be useful for distinguishing fluctuations of given atoms in the apo versus bound forms of DHNA. When the protomer averages are dependent on protomer, the inequality in Eq. (5) may be violated, and RMSF1 could be slightly larger than RMSF2.

The principal component analysis was carried out with the gcovar and ganeig modules of GROMACS 3.0.²⁰ The PCA method diagonalizes the covariance matrix $\sigma_{ij} = \left\langle \delta x_i \delta x_j \right\rangle$ of the CA atom fluctuations from their trajectory-averaged $\left\langle \dots \right\rangle$ values, where $\delta x_i = x_i - \left\langle x_i \right\rangle$ and x_i denotes the Cartesian components of the i th CA atom.¹⁰ For

this analysis, $\langle \dots \rangle = \langle \langle \dots \rangle_c \rangle_p$; the entire trajectory is used and corresponds to the octamer-based RMSF approach. The overall translational/rotational motion of all the snapshots is removed relative to a reference structure before matrix diagonalization. The sum of the eigenvalues of the covariance matrix corresponds to the total variance of the motion over the trajectory. Ordering the eigenvalues from large to small can, in favorable cases, provides a small set of modes that captures most of the protein's fluctuations. Another application of PCA is to the evaluation of entropy.²¹ Assuming that the covariance matrix σ_{ij} describes the protein dynamics permits the entropy change in DHNA for binding of HP to be estimated as:

$$\Delta S = \frac{k_B}{2} \ln \left[\frac{\det(\sigma_a)}{\det(\sigma_b)} \right], \quad (1)$$

where k_B is Boltzmann's constant, and σ_a and σ_b are the above covariance matrices in the bound and free forms respectively. After diagonalization of these two matrices, the entropy change can be obtained since the determinants in Eq. (1) are expressible as the products of their PCA eigenvalues.

The entropy of HP was estimated by using the quasih module of AMBER 7.0 to obtain the normal modes that can be used in standard formulas from statistical thermodynamics.

Results and Discussion

1. Apo DHNA and its HP complex

A. Flexibility analysis of DHNA and its HP complex based on the octamer and its protomers. DHNA is an octamer, with its eight active sites that bind HP formed by interfaces between pairs of adjacent protomers. There is no cooperation among the active

sites based on kinetic studies and, therefore, they can be treated as statistically independent units, indicating that more conformational space can be explored than in a usual MD simulation of a given duration. The RMSDs of the CA atoms of the instantaneous MD structures from the initial X-ray structure for the apo DHNA and its HP complex are ~ 1.1 - 1.5 Å between 0.5 and 2.0 ns (data not shown). For this comparison, the superposition of structures for the RMSD evaluation was based on the octamer providing an effective simulation time of 12 ns (8×1.5 ns of time after 0.5 ns for equilibration).

The RMS fluctuations (RMSF) of the CA atoms for each residue (Figure 9-1a and 1b) were first calculated by the average protomer RMSF method, as discussed in the Methods. This approach evaluates each protomer RMSF and averages them over the eight protomers to obtain the average protomer RMSF. Overall, the protein is quite rigid with an average RMSF of ~ 0.6 Å. The binding of HP apparently does not change the RMSF of the protein on the MD time scale. The RMSF variations along the amino acid sequence obtained from the MD simulations and experimental crystal B-factors are quite similar, with correlation coefficient 0.79 for the apo form and 0.77 for the HP complex.

The MD trajectory of the DHNA octamer was then partitioned into the coordinate sets for the individual protomers and combined to obtain a data record of length $N_P \times N_C$ with averages and fluctuations from averages based on all the data for a given CA. This octamer-based RMSF is the second approach discussed in the Methods and, as noted there, will tend to emphasize the fluctuations. These octamer RMSF values, compared with the previous average protomer RMSF, are plotted in Figure 9-2. The new RMSFs display a similar pattern to the previous ones, but with larger variations, especially for the

apo enzyme (Figure 9-2a), in some regions. Strikingly, all these regions are around the active sites, at the interface of two subunits (Figure 9-3), including residues 15-25, 45*-55*, 68-74 and 100-110. (We will henceforth denote residues in the adjacent subunit that together form the active site by a “*”, as in the notation 45*-55*). This difference reflects the limited 2 ns sampling space around the initial subunit structures. The “extra” flexibility around the active site might assist in ligand binding and release, as discussed below.

The increase in RMSF on octamer versus average protomer basis is unanimous in the apo enzyme simulation. In the HP complex simulation, the analogous increase is smaller and even negative (decreased relative to the average protomer-based method) in the rigid regions of the complex (Figure 9-2b), which indicates that HP binding rigidifies the enzyme not only in the flexible regions but also in the rigid regions. This feature can be seen clearly in Figure 9-4a, where the apo and HP-bound RMSF differences are compared for octamer and average protomer methods. Thus, while RMSF differences based on the average protomer RMSF would not support the hypothesis that the binding of HP rigidifies the protein, the differences based on the octamer RMSF show clearly that the binding of HP rigidifies DHNA, and does so not only in the flexible regions, but also in the rigid regions.

While most parts of DHNA are rigidified by binding HP, residues 49-51 seem more flexible after HP binding (Figure 9-4a, dotted line). Further investigation indicates that the ψ dihedral angle of Asp*50 and the ϕ dihedral angle of Thr*51 can change cooperatively to give a different conformation for Asp*50 and Thr*51. Figure 9-5 shows the time evolution of these two dihedral angles in the different subunits displayed

sequentially as one trajectory. In subunits 1, 2, 3, 4, 7, 8, ϕ and ψ remain at $\sim 0^\circ$ and -140° , respectively; but, in subunits 5 and 6 (especially 6), ψ changes from 0° to -40° , at the same time that ϕ changes from -140° to -70° . These simultaneous changes of dihedral angles may be caused by the new interactions between the Thr*51 side chain and its surroundings (see discussion below). After excluding subunit 6 from the trajectory, the new RMSF differences between apo enzyme and HP bound enzyme show no positive spike around residue 50 (Figure 9-4b).

B. Principal component analysis. The rigidification of DHNA upon HP binding can also be investigated by the PCA method, which attempts to decompose the total variance of the atom RMS fluctuations over the MD trajectory into a contribution from a small set of modes that have a relatively large contribution to this variance and a large set of modes with a small summed contribution.¹⁰ Because the above analysis does show that there are flexible regions of DHNA that become rigidified upon HP binding, a comparison by PCA of the two forms should be instructive. The trajectory of subunit 6 was excluded in the PCA analysis, because Asp*50 and Thr*51 in subunit 6 have backbone conformations quite different from those in the other subunits. In order to highlight the flexible regions in the PCA analysis, only CAs of rigid regions (with RMSF less than 0.7 \AA in the free form) were fitted to remove the overall translation/rotational motion. The new fitting gives total variance 96.6 \AA^2 for all CAs in the apo form and 67.2 \AA^2 in the bound form. The first four modes contribute 56% of the variance in the apo and 39% in the bound form. The remaining modes contribute $\sim 43 \text{ \AA}^2$ in the apo form and $\sim 41 \text{ \AA}^2$ in the bound form. Thus, as indicated in Figure 9-6, the first four modes incorporate most of the flexibility difference between the apo and bound forms. That a few modes

capture most of the difference in RMS fluctuations between the apo and bound forms reinforces the data obtained above that a limited number of residues are responsible for the changes in DHNA flexibility upon HP binding. The first mode has variance $\sim 25 \text{ \AA}^2$ in the free form but only $\sim 9 \text{ \AA}^2$ in the HP bound form.

The RMS fluctuations of each CA corresponding to the trajectory based on mode 1 for the apo and HP-bound forms are displayed in Figure 9-7. They show that the enhanced flexibility of the apo relative to bound form is concentrated in the three region, 45*-55*, 68-74 and 100-110, which were identified in Section A. These data, along with the data for modes 2 and 3, suggest that examination of three distances between three residues, one in each of these regions, should provide a refined perspective on the active site fluctuations. Labeling the Tyr*54-Ile105 distance as d_1 , the Tyr*54-Ala69 distance as d_2 , and the Ala69-Ile105 distance as d_3 , and parametrically plotting the data for all eight active sites in 3D reveals that the most interesting differences are in the d_2 and d_3 directions, as displayed in Figure 9-8. The fluctuations in distance along the d_1 direction are quite similar for the apo and HP-bound forms, and its scale is about half of the total d_2 and d_3 apo fluctuation scale. The apo form has a major and a minor state, while the HP-bound form has only one state, which is confined within the major apo state. The apo motion is correlated in the sense that on average an increase in d_2 is accompanied by an increase in d_3 , indicating a kind of breathing motion (both distances have a common residue) that is, approximately, an equally weighted combination of the two distances.

The estimated entropy loss of DHNA due to HP binding for CA atoms is about 4.9 cal/mol-deg based on the first four modes and about 13.1 cal/mol-deg including all the 357 PCA-determined modes, when evaluated from Eq. (1). It is quite interesting that even

though the first several modes contribute most of the flexibility changes in the binding, ~60% of the entropic loss comes from the small eigenvalue (high frequency) contribution. This is consistent with the fact that the binding of HP not only rigidifies the flexible regions but also the rigid regions.

C. Active site analysis. Since the most significant flexibility changes are localized in the active site region, it is useful to analyze the interactions between the active site residues and between the active site residues and bound HP. In the active site of apo DHNA (Figure 9-9a), the carboxylate of Glu22 forms a salt bridge with Lys100 and hydrogen bonds with the backbone amides of Ala18 and Leu19 and with the side chain amide of Gln27. The carboxylate of Glu74 forms hydrogen bonds with the hydroxyl of Thr*51 and the phenoxyl of Tyr*54 from the adjacent subunit and the backbone amides of Leu73 and Glu74 from the same subunit. The strong hydrogen bond networks formed by the carboxylates of Glu22 and Glu74 with their surroundings indicate that these two residues are restrained in the apo form. By comparing the hydrogen bond interactions in different active sites (Table 9-1), the hydrogen bond strengths are shown to be more diverse around Glu22 than Glu74, which may be caused by the difference in flexibility between these two regions. The different active sites between different subunits show quite consistent hydrogen bond interactions (Table 9-1), which suggests that the active sites are well defined in the apo form MD simulation. All the hydrogen bonds in Table 9-1 are similar to those found in the crystal structure, with the one exception that the Y54* to Glu74 hydrogen bond, which spans two subunits, is longer in the crystal structure.

The hydrogen bond interactions after binding HP are listed in Tables 2 and 3 and a typical snapshot displayed in Figure 9-9b. The results are in good agreement with those

from the crystal structure. Compared with the apo form, the strong hydrogen bond between the amide and carboxyl of Glu74 is maintained, as well as the hydrogen bond between the side chain hydroxyl of Thr*51 and the Glu74 carboxyl group. The Glu74 carboxyl forms two new strong and stable hydrogen bonds with N2H1 and N3H of HP. This suggests that Glu74, which sits at the bottom of the active site and is well restrained by the hydrogen bonds with surrounding residues, acts as an anchor to fix the orientation of HP. The side chains of His*53 and Tyr*54 sandwich the HP pterin ring forming a π - π stacking interaction in some active sites, while in others the His53 imidazole ring is angled away. In the crystal structure, the Tyr*54 phenol is stacked with the pterin ring, but the His*53 imidazole is not. HP also forms hydrogen bonds with the carbonyl of Val*52 and amide of Tyr*54 from the adjacent subunit.

The O4 atom of HP forms a stable hydrogen bond with a water molecule that is hydrogen bonded to the carbonyl of Asn71 and the side chain amine of Lys100 (Table 9-2, Figure 9-9b), which is similar to the bound water interaction found in the HP crystal structure. Since all crystallographic water molecules were removed before the start of the MD simulation, the appearance and consistent presence of water in this position in all the active sites suggests that it is a stable point for water, and that the simulation time was sufficient to organize proper active sites. All of the above-noted interactions that do not exist in the free form enzyme contribute to make the complex more stable and rigid.

It is sensible that HP binding rigidifies the region around residue 70 (Figure3) considering that HP forms two hydrogen bonds with Glu74, one with Leu73, and one hydrogen bond network with Asn71 through the trapped water. Since this region is quite close to residues around 45 from the adjacent subunit (Figure 9-3), these interactions may

also contribute to rigidifying the region around residue 45*. Further investigation indicates that the amides of Asn71 and Leu72 form a hydrogen bond network with the carboxyl of Asp*46 from the adjacent subunit (Table 9-4), which might be the reason that residues around 45* are also rigidified.

The strong hydrogen bonds between HP and Glu22 certainly stabilize the region around residue 22. Interestingly, the largest rigidity change due to HP binding comes from residue 25, which does not interact with HP directly. Ile25 shows great flexibility in the free form simulation with RMSF equal to 1.7 Å that might occur because Ile25 is at the end of helix 1 (residues 20-24) and close to the beginning of beta sheet 3 (residues 27-37). In the bound form, the Ile25 RMSF is only about 0.9 Å. Further inspection shows that the carbonyl of Ile25 forms one hydrogen bond with the amide of Ala21 and one weak hydrogen bond with the amide of Glu22, while Glu24 forms two hydrogen bonds with Ser20 and Ala21 (Table 9-5). Clearly, the rigidity increase around Glu22 also stabilizes Glu24 and Ile25.

D. Product binding energy decomposition. The nature of the binding pocket can be defined by examining which residues contribute the most to stabilizing HP. As shown in Figure 9-9b, the direct interactions with HP typically include hydrogen bonds with the carboxyl groups of Glu22 and Glu74, the carbonyl oxygen of Val*52, and the amide nitrogens of Tyr*54 and Leu73. In order to assess how the individual residues contribute to the binding of HP, the interaction energies between HP and residues within 5 Å of HP, averaged over the 2 ns trajectory and the eight active sites, are shown in Figure 9-10. These residues contribute about -80 kcal/mol of electrostatic energy (97% of the total

over all residues) and about -23 kcal/mol of van der Waals energy (95% of the total over all residues).

The highly conserved residues Glu22, Glu74 and Lys100 contribute about -68 kcal/mol of the electrostatic energy (82% of the total over all residues electrostatic energy) and clearly are essential to fix the orientation of HP within the active site. The electrostatic interaction (-35 kcal/mol) between Glu74 and HP is the largest among all individual residues due in part to the two strong hydrogen bonds between the carboxyl group and HP. Glu22 contributes -11 kcal/mol of electrostatic stabilization in a similar way. Though Lys100 does not form a direct hydrogen bond with HP, being water-mediated, the positively charged amine group of Lys100 does have a strong interaction with the partial negative charges of O4 and N5 of HP, which contributes the second largest electrostatic energy of -23 kcal/mol. His*53 and Tyr*54 contribute -9.3 kcal/mol (37%) van der Waals interaction due to the fact that HP pterin ring is sandwiched by the side chains of these two residues in four of the active sites. In the other four active sites, the pterin ring is tilted slightly and moves closer to Lys100 and Glu22. Tyr54 is maintained in a similar position but His53 moves away. This might be caused by the strong interaction between HP and Lys100 and Glu22. The solvent-HP interaction provides about -14 kcal/mol electrostatic and -1 kcal/mol van der Waals of stabilization energy. A major contributor (6-7 kcal/mol) to this energy is the hydrogen bond between the HP O3 atom and the water bound to Lys100.

2. Exit path of HP from DHNA

A. Pushing HP out of the active site: Fluctuation analysis. In order to investigate how the HP release process changes the structure and flexibility of DHNA, HP was

pushed out by 8 Å in all the active sites. The pushing force was imposed between the backbone N of Arg118 and C10 of HP. The trajectory average RMSD of 1.56 Å is quite stable and slightly larger than the 1.42 Å of the regular HP complex simulation, indicating that no significant changes of the overall backbone structure occurred during the push. The average RMSD of the N of Arg118 is 1.29 Å during this pushing simulation compared with 1.10 Å in the regular HP simulation. As shown in Table 9-6, the RMSF of this N atom in the eight different subunits increases slightly overall in the pushing MD simulation, with the average increase ~0.07 Å. Both RMSD and RMSF measures show that the backbone N of Arg118 does not move much during the pushing simulation, which confirms that the restraint between this N atom and C10 of HP and the time scale used led to a sensible product release trajectory.

Figure 9-11 shows the by residue RMSF comparison (based on superposition of individual subunits) between the pushing and regular MD simulations. The product release does not change the backbone protein structure and flexibility significantly, though slight changes occur around the flexible regions, residues 15-25, 45*-55*, 68-74 and 100-110, which may reflect the disturbance that the exit of HP creates in this regions. The total fluctuation of the CA atoms is 84.9 Å², larger than the 65.6 Å² of the regular HP complex simulation, but comparable with that of apo enzyme simulation's 86.5 Å². This is consistent with the previous simulation result that product binding rigidifies the protein; therefore, the release of HP should increase the enzyme flexibility.

In order to understand how protein structure and flexibility change along the HP release trajectory, the data were also analyzed based on individual time windows. The average protomer structure in each window (an average over all subunits over the

simulation time window of 50 ps) was compared with the average protomer structure (average over all subunits over the 1.5 ns simulation time) in the regular HP complex simulation (Figure 9-12). The first window corresponds to the first 50 ps of the regular simulation. The CA RMSD changes from 0.2 Å to 0.4 Å during the first 2 Å of pushing, and then fluctuates around 0.4 Å. The initial window difference of 0.2 Å is a consequence of the 50 ps versus 1.5 ns averaging time. The 0.2 Å subsequent RMSD increase again illustrates the minimal overall protein disturbance, considering that the CA RMSD between the average structure of the regular HP complex and that of the apo form simulation is only 0.44 Å.

The CA fluctuation in each window is plotted along the exit path of HP in Figure 9-13. Unlike the average structure plot (Figure 9-12), the fluctuation increases from $\sim 65 \text{ Å}^2$ to $\sim 92 \text{ Å}^2$ along the HP exit path from 1 to 6 Å, and then stabilizes after 6 Å. Interestingly, the average fluctuation between 6 Å and 8 Å is slightly larger than that of the apo form of $\sim 87 \text{ Å}^2$. This extra flexibility may be caused by the weak DHNA-HP interaction perturbing the energy surface of the protein and assisting it to access more conformation space. Note that the difference in fluctuation magnitude is probably not a sampling deficiency, though it is hard to get converged fluctuations in short time simulations, because the fluctuation in the first window of the HP push is 67.0 Å^2 , which is consistent with the 65.6 Å^2 in the regular HP complex enzyme simulation.

The window RMSD calculations were also carried out on the CA of the four fragments that correspond to the flexible regions: 15-25, 45*-55*, 68-74, 100-110, as displayed in Figure 9-14. The CA RMSD of residues 15–25 increases from 0.3 Å to 0.7 Å (Figure 9-14a) over the first 2 Å push, but then drops back to 0.3 Å around the distance

3.5 Å and thereafter fluctuates around 0.5 Å. This region may undergo a conformational change during the exit of HP associated with its strong interaction with Glu22. The CA RMSD of residues 45–55 (Figure 9-14b) increases from 0.2 Å to 0.5 Å, and then fluctuates before increasing to 0.8 Å around distance 7–8 Å. This residue range includes Thr*54 that is π -stacked with HP. The CA RMSD of residues 68–74 (Figure 9-14c) increases from 0.2 Å to 0.5 Å, and then goes back to 0.3 Å at the distance 6 Å, thereafter fluctuating around 0.4 Å. The RMSD of CAs of residues 100–110 (Figure 9-14d) goes up to 0.5 Å, and then fluctuates.

While the RMSD data provides information about the drift of protein structure during the HP exit, the fluctuation data indicates how flexibility changes in these regions. Residues 15–25 show the largest CA fluctuation changes, from ~ 10 to ~ 22 Å², (Figure 9-15a), while residues 45*–55*, 68–74 and 100–110 show slight increases of flexibility along the path (Figure 9-15b, c, d). By summing the fluctuations in these four regions, as shown in Figure 9-15e, the overall fluctuation increases from ~ 30 Å² to ~ 50 Å² which contributes $\sim 75\%$ of the increase in protein flexibility. Therefore, the major increase in flexibility comes from these more mobile regions around the active site.

B. Pushing HP out of the active site: Energetic analysis. The local nature of the HP-DHNA interactions suggests that an exit path should show a relatively fast transition between bound and weakly bound behavior. Figure 9-16a plots the electrostatic interactions between HP and Glu22, Glu74 and Lys100. The local nature of the electrostatic interactions between HP and the protein noted in Section 1D is consistent with the fairly well defined transition behavior in this plot. The magnitude of the electrostatic interaction decreases mainly due to the breaking of hydrogen bonds between

Glu22, Glu74 and HP and the increase of distance between Lys100 and HP. Figure 9-16a also shows that the hydrogen bonds between HP and Glu22 and Glu74 break when HP is 2-3 Å away from its original bound position. Interestingly, the interaction between Glu22 and HP changes from -10 to 6 kcal/mol during the first 4 Å of HP exit, indicating that the motion of HP pushes this residue away. Then, this interaction energy decreases to -4 kcal/mol, which suggests that Glu22 favorably interacts with HP afterward. Figure 9-16b displays the interaction energies of the flexible regions with HP. The interaction between HP and residues 15-25 shows a similar pattern to that between HP and Glu22, with a lower maximum (-2 kcal/mol) at a distance of 4 Å, and strong interaction of -10 kcal/mol after 6 Å. This feature suggests that the interaction between HP and residues 15-25 can assist in breaking the hydrogen bond between HP and Glu22 and also that this region undergoes a conformational change to accommodate the release of HP. The trend in energy for residues 15-25 as a function of distance is consistent with the RMSD trend summarized by Figure 9-14a. The interaction energy between HP and residues 45*-55* fluctuates around -20 kcal/mol until the distance increases to 6 Å. The strong stabilizing van der Waals interaction as well as electrostatics between this region and HP (Figure 9-10) was maintained in the first 4 Å of pushing, with the residues moving with HP. Then, the van der Waals energy increases and the electrostatic energy drops to maintain the overall energy until about 6 Å. The residue regions 68-74 and 100-110 show similar interaction energy changes as those of Glu74 and Lys100, respectively. Along the exit path, the interaction between HP and DHNA weakens from -108 to -33 kcal/mol while the HP solvent interaction strengthens from -16 to -61 kcal/mol (mainly electrostatic

energy), the latter compensating for the loss of interactions between HP and the enzyme (Figure 9-16b).

The total interaction energy between HP and its environment along the product release path is plotted in Figure 9-17, which would imply a barrier to HP release of ~32 kcal/mol. That is a rather large barrier to product release; however, it does not account for the internal energy of HP. If the internal energy of HP along the path is also included (averaging the HP internal energy over the 50 ps windows) the barrier drops by ~7 kcal/mol and lowers the ligand release barrier to ~25 kcal/mol. The decrease of internal energy of HP helps to compensate for the loss of interaction between HP and the enzyme relative to its solvation energy. Note that the reaction coordinate is defined by the N of Arg118 to C10 of HP distance, so that 0 Å corresponds to 17.5 Å. When HP is pushed out by restraining this distance, the Glu74 to HP distance, for example, does not necessarily immediately respond. Furthermore, various orientations are being sampled at each distance. That the energetic change does not occur immediately (between 0-2 Å) can be attributed to such effects. Most of the decrease in internal energy occurs between 2~4 Å and is a combination of the loss of interaction energy of HP with the protein and the release of strain energy of the bound product.

C. Pushing HP out of the active site: Free energy analysis. A binding constant is of course related to the free energy difference between bound and solvated states, and a kinetic constant for dissociation is related to the free energy barrier separating bound and un-bound states. Above, we focused on the energy of pushing HP out of the binding pocket. In principle, the exit pathway strategy pursued here could be used to obtain directly the free energy along the chosen coordinate. However, that would require an

averaging period for each of the 40 windows that cannot be practically reached in current MD simulations. An approximate free energy profile may be obtained by including the entropy associated with HP along the exit path and the corresponding protein entropy. The HP entropy may be estimated by a quasi-harmonic analysis since, with the exception of the ring NH_2 and the CH_2OH tail, it is a quite rigid structure. The line in Figure 9-17 labeled as $-\text{TS}$ is this vibrational contribution to the HP entropy. The ΔS of ~ 13 cal/mol-deg obtained in Section 1B for DHNA between bound and apo forms provides a less than 4 kcal/mol decrease in free energy for un-binding. Including these two entropic contributions leads to a barrier to HP release of ~ 15 kcal/mol around 4 Å.

Conclusions

The eight catalytic active sites of DHNA are formed by the noncovalent association of protomers. The presence of eight active sites permits a more efficient investigation by molecular dynamics of the binding site, and of how HP is released from DHNA. The reliability of the simulations was assessed based on RMSD from the crystal structures, the active site hydrogen bonding patterns, and the conformations of the active sites compared with those of the crystal structures. The RMSD and RMSF data confirm that the simulations maintain the basic DHNA conformation. The overall DHNA structure is quite rigid, and the binding of HP does not cause significant changes in the structure.

By analyzing the trajectory data with the two RMSF methods denoted as average protomer RMSF and octamer RMSF, the more flexible regions of DHNA were identified. In particular, the four regions spanning residues 15-25, 45*-55*, 68-74 and 100-110 are the more flexible regions of apo DHNA. These regions include residues that are part of the active site. The binding of HP rigidifies DHNA, not only in the above flexible regions

but also in the more rigid regions that do not directly participate in HP binding. All these conclusions relied on a comparison of the octamer and average protomer method changes in residue RMSFs for the apo and HP-bound forms; it would not have been evident from just the average protomer-based method. The network of hydrogen bonds formed by HP and the protein, and also induced in the protein by HP, are schematized in Figure 9-9b, and specified, for each dimer, in Tables 9-(2-5). These interactions between HP and active site residues might explain the increase in rigidity of the enzyme when HP is bound.

The PCA results of apo and HP-bound DHNA complements those of the RMSF analyses. The rigidification, on average, of DHNA upon HP binding is evident in the total PCA variance of 86.5 \AA^2 for the apo and 65.6 \AA^2 for the bound form. The data in Figure 9-6 show that the first four modes not only capture a significant fraction of the total fluctuation in both apo and bound form, but that most of the effect of rigidification upon HP binding is captured in those four modes, suggesting that a limited number of residues are responsible for the change in DHNA flexibility upon HP binding. That picture is reinforced by the projections of the atom motions onto the first few principal components, as shown for the first mode in Figure 9-7, and leads to the identification of residues in the three regions, 45*-55*, 68-74 and 100-110, which can provide a contrast between the fluctuations in the apo and HP bound forms, as displayed in Figure 9-8.

An energetic analysis of the residues that contribute to binding HP reveals that the binding is localized, as summarized by the data in Figure 9-10. It is remarkable that the HP binding energy is so local, even though it is dominated by the nominally long-range electrostatic interaction energy contribution. Residues within 5 \AA of HP contribute about

97% of the total electrostatic and 95% of the total van der Waals energy. Just three residues, Glu22, Glu74 and Lys100 contribute most (82%) of the electrostatic energy, which emphasizes the importance of these residues for binding HP and for creating a catalytically competent structure. Glu74, which has the largest interaction with HP (−35 kcal/mol), sits at the bottom of the active site and may act as an electrostatic attractor for HP. Both Glu22 and Lys100 are implicated in the catalytic mechanism of HP. Note that Lys100 is not directly hydrogen bonded to HP; but, there is one water molecule that bridges Lys100 and O4 of HP, which is also hydrogen bonded to Asn71. This water is found in the crystal structure. In the simulation, which is carried out in the absence of crystallographic waters, a water molecule from the solvent migrates to this position and persists in all the eight active sites (Table 9-2). This is about half the total HP water interaction energy, which indicates the importance of this water for binding HP. The hydrogen bond between Lys100 and the trapped water may also serve to decrease the pK_a of the water molecule as part of the catalytic mechanism of DHNA. The electrostatic interaction energy between Lys100 and HP of −23 kcal/mol may be important for binding the pterin ring, versus modifying the N5 pK_a , as noted in a Raman spectroscopic study that found N5 to be deprotonated even at pH 6.5.¹⁵

The flexibility of DHNA is 87 \AA^2 in the apo and 65 \AA^2 in the HP bound form, providing a measure of how HP binding increases the rigidity of DHNA, which complements the view obtained from specific interactions as detailed in this work. It would be possible, of course, for HP to rigidify DHNA locally but still lead to an overall increase in DHNA flexibility. The increase in rigidity that is found is qualitatively related to an entropic loss upon binding that needs to be compensated by a sufficient increase in

DHNA-HP interaction energy. It is interesting that during the course of the HP exit path simulation, DHNA increases its flexibility to 92 \AA^2 that is even larger than the free form value. This “extra” flexibility may be due to the surface-bound HP interacting with the protein and helping it sample more conformation space. Much (75%) of the flexibility increase comes from the four flexible regions around the active site that have been identified in the simulation. Among them, residues 15-25 contribute the most, about 40%, to the increase. Since the binding of HP is so local, it is sensible that the flexibility decrease upon binding should also be concentrated in this active site region.

The direction and speed of pushing that we chose for the HP exit path did not significantly disturb the overall DHNA structure, as summarized in Figures 11–12. As noted above, HP is bound strongly by Glu74 at the bottom of the binding site, and examination of HP relative to its surroundings along the exit pathway, shows that it moves out with the pterin ring first parallel with its bound position followed by a gradual upward rotation as it exits. That the active site is formed by the noncovalent association of two protomers indicates that it is mainly HP-residue hydrogen bonds that are being broken in the release process. The HP-DHNA interaction energies, as displayed in Figures 16 and 17, show a fairly abrupt transition region around the 2-4 \AA range, which, again, is consistent with the local binding energetics (Figure 9-10) mainly involving Glu22, Glu74, Lys100. It appears (Figure 9-16a) that Lys100 must be pushed aside as HP exits the active site. The local binding energetics of HP may be important to the release mechanism because it lets HP easily leave once those local interactions are broken.

The large decrease in HP-DHNA stabilization energy during the exit of HP needs to be compensated, if the barrier to HP release is not rendered too large. Partly, this comes

from the increase of the HP-solvent interaction, as shown in Figure 9-16b. At the same time, the internal energy of HP also decreases (Figure 9-17), which suggests that HP relaxes a conformationally based strain during the release. Thus, HP binding destabilizes itself, which may be viewed as a trade-off associated with positioning HP in the active site to permit catalytic activity.

There are, of course, entropic terms that contribute to the free energy barrier for HP release. The contribution associated with the HP internal fluctuations during the exit path as obtained from a quasi-harmonic analysis is shown in Figure 9-17, and clearly lowers the barrier to and contributes a stabilizing influence on HP release. In view of the enhanced freedom of HP when it is in the solvent, relative to its bound state, an entropic penalty of binding is expected. The change in protein entropy that is obtained from the PCA eigenvalues (based on the CA atoms) provides less than 4 kcal/mol of free energy protein stabilization in favor of the apo protein. The change in protein entropy found most likely is an upper bound because the initial state is all HP bound and the final state is all HP free. If fewer than eight ligands were released, the change in protein entropy should be reduced. The experimental data²² indicate that the rate constant for product release is $\sim 10 \text{ s}^{-1}$, which is consistent with the barrier to HP release obtained here.

Reference

- (1) Illarionova, V.; Eisenreich, W.; Fischer, M.; Haussmann, C.; Romisch, W.; Richter, G.; Bacher, A. *J. Biol. Chem.* 2002, 277, 28841.
- (2) Henderson, G. B.; Huennekens, F. M. *Methods Enzymol.* 1986, 122, 260.
- (3) Bermingham, A.; Derrick, J. P. *Bioessays* 2002, 24, 637.
- (4) Sanders, W. J.; Nienaber, V. L.; Lerner, C. G.; McCall, J. O.; Merrick, S. M.; Swanson, S. J.; Harlan, J. E.; Stoll, V. S.; Stamper, G. F.; Betz, S. F.; Condroski, K. R.; Meadows, R. P.; Severin, J. M.; Walter, K. A.; Magdalinos, P.; Jakob, C. G.; Wagner, R.; Beutel, B. A. *J. Med. Chem.* 2004, 47, 1709.
- (5) 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. *J. Mol. Biol.* 1999, 287, 211.
- (6) Bauer, S.; Schott, A. K.; Illarionova, V.; Bacher, A.; Huber, R.; Fischer, M. *J. Mol. Biol.* 2004, 339, 967.
- (7) Salzmann, M.; Pervushin, K.; Wider, G.; Senn, H.; Wuthrich, K. *J. Am. Chem. Soc.* 2000, 122, 7543.
- (8) Lopez, P.; Lacks, S. A. *J. Bacteriol.* 1993, 175, 2214.
- (9) Cox, T. F.; Cox, M. A. A. *Multidimensional scaling*, 2nd ed. ed.; Chapman & Hall: Boca Raton, 2001.
- (10) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Structure, Function, and Genetics* 1993, 17, 412.
- (11) Kollman, P. *Chem. Rev.* 1993, 93, 2395.
- (12) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* 1997, 72, 1047.
- (13) Boresch, S.; Archontis, G.; Karplus, M. *Proteins* 1994, 20, 25.
- (14) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* 2003, 107, 9535.
- (15) Deng, H.; Callender, R.; Dale, G. E. *J. Biol. Chem.* 2000, 275, 30139.
- (16) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* 1995, 91, 1.

- (17) Berendsen, H. H. C.; Postma, J. P. M.; Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* 1984, *81*, 3684.
- (18) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, G. L. *J. Chem. Phys.* 1995, *103*, 8577.
- (19) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* 1977, *23*, 327.
- (20) Lindahl, E.; Hess, B.; van der Spoel, D. *Journal of Molecular Modeling* 2001, *7*, 306.
- (21) Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D. *Macromolecules* 1984, *17*, 1370.
- (22) Yan, H. (unpublished results).

Appendices

Table 9-1 Active Site Interactions in Apo DHNA

Subunit	1	2	3	4	5	6	7	8
Glu74- OE1...	2.71 (0.13)	2.70 (0.12)	2.68 (0.13)	2.76 (0.16)	2.68 (0.13)	2.71 (0.13)	2.68 (0.12)	2.71 (0.15)
Thr*51- OGH	100%	100%	100%	100%	100%	100%	100%	100%
Glu74- OE2...	2.67 (0.12)	2.63 (0.10)	2.93 (0.25)	2.62 (0.10)		2.66 (0.12)	2.65 (0.12)	2.62 (0.10)
Tyr*54- OH	100%	87%	35%	100%		100%	100%	100%
Glu74- OE2...	3.23 (0.16)	3.21 (0.16)	3.24 (0.16)	3.28 (0.15)	3.27 (0.16)	3.24 (0.15)	3.17 (0.16)	3.25 (0.15)
Leu73- NH	72%	82%	61%	41%	37%	75%	76%	63%
Glu74- OE2...	2.78 (0.09)	2.76 (0.08)	2.76 (0.09)	2.78 (0.10)	2.77 (0.09)	2.78 (0.09)	2.77 (0.09)	2.76 (0.09)
Glu74- NH	100%	100%	100%	100%	100%	100%	100%	100%
Glu22- OE1...	2.77 (0.11)	2.80 (0.12)	2.74 (0.12)	2.90 (0.18)	2.93 (0.22)	2.75 (0.10)	2.74 (0.09)	2.86 (0.16)
Lys100- NZ ^a	100%	98%	100%	100%	77%	100%	100%	99%
Glu22- OE2...	3.12 (0.20)	3.05 (0.21)	2.96 (0.18)	2.88 (0.15)	3.06 (0.22)	3.12 (0.19)	2.96 (0.17)	3.01 (0.20)
Ala18- NH	51%	60%	81%	97%	100%	100%	91%	78 %
Glu22- OE2...	3.26 (0.16)		3.04 (0.18)	3.13 (0.19)	2.99 (0.16)	2.98 (0.16)	2.98 (0.15)	3.15 (0.18)
Leu19- NH	34%		91%	36%	93%	99%	99%	73%
Glu22- OE1...	2.94 (0.17)	2.88 (0.15)		2.90 (0.17)	2.82 (0.12)	2.81 (0.11)	2.80 (0.11)	2.90 (0.15)
Gln27- NEH	96%	97%		53%	99%	100%	100%	97%

^a Only distance criterion was used for the salt bridge analysis.

Table 9-2 Hydrogen Bonds between HP and Its Surroundings in HP complex ^(a)

Subunit	1	2	3	4	5	6	7	8
N1...	3.15			3.25	3.21	3.25	3.23	
Tyr ⁵⁴ -	(0.16)			(0.15)	(0.15)	(0.15)	(0.15)	
NH	94%			31%	84%	25%	26%	
N2H1...	2.80	2.80	2.80	2.75	2.81	2.78	2.80	2.80(0.1
Glu74-	(0.10)	(0.10)	(0.10)	(0.09)	(0.10)	(0.10)	(0.11)	0)
OE1	100%	100%	100%	100%	100%	100%	100%	100%
N2H2...	3.06	3.15	3.07	2.97	3.04	3.13	3.04	3.08
Val ⁵² -	(0.19)	(0.20)	(0.20)	(0.15)	(0.18)	(0.19)	(0.18)	(0.20)
CO	87%	47%	61%	98%	95%	59%	95%	52%
N3H...	2.79	2.80	2.80	2.99	2.91	2.81	2.83	2.79
Glu74-	(0.08)	(0.09)	(0.09)	(0.17)	(0.14)	(0.10)	(0.13)	(0.09)
OE2	100%	100%	100%	98%	100%	100%	99%	100%
O4...	2.96	2.99	2.98	3.17	2.97	3.03	3.03	2.90
Leu73-	(0.15)	(0.16)	(0.16)	(0.18)	(0.16)	(0.16)	(0.17)	(0.13)
NH	99%	99%	97%	81%	97%	98%	95%	100%
I' OH...	2.79	2.62	2.80	2.64	2.67	2.69	2.75	2.63
Glu22-	(0.23)	(0.12)	(0.26)	(0.11)	(0.17)	(0.22)	(0.23)	(0.10)
OE1	91%	100%	82%	95%	88%	67%	73%	100%
Glu22-	2.94	3.17	2.93	3.14	3.07	2.77	2.97	3.24
OE2...	(0.27)	(0.19)	(0.30)	(0.35)	(0.34)	(0.31)	(0.26)	(0.16)
I'OH	78%	36%	69%	18.4%	57%	45%	54%	54%
O4...	2.85	2.80	2.88	2.80	2.85	2.79	2.77	2.86
H ₂ O	(0.22)	(0.18)	(0.20)	(0.20)	(0.22)	(0.17)	(0.16)	(0.19)
	103%	100%	60%	106%	106%	94%	94%	91%
Asn71-	2.85	2.84	3.05	2.74	2.84	2.88	2.79	2.85
O...	(0.20)	(0.19)	(0.25)	(0.13)	(0.19)	(0.18)	(0.17)	(0.19)
H ₂ O	90%	99%	11%	100%	86%	77%	97%	96%
Lys100-	2.86	2.92	2.95	2.86	2.88	2.88	2.86	2.87
NZ...	(0.14)	(0.15)	(0.19)	(0.11)	(0.14)	(0.13)	(0.13)	(0.12)
H ₂ O	98%	98%	61%	100%	85%	100%	100%	100%
I'OH...	2.85	2.91	2.90	2.85	2.91	3.21	2.83	2.80
H ₂ O	(0.22)	(0.22)	(0.22)	(0.18)	(0.21)	(0.22)	(0.19)	(0.17)
	101%	68%	136%	55%	35%	8%	69%	113%

^(a) Some percentages are greater than 100% since more than one atom can hydrogen bond at a given time.

Table 9-3 Active Site Interactions in HP Complex

Subunit	1	2	3	4	5	6	7	8
Glu74- OE1...	2.64 (0.10)	2.65 (0.11)	2.68 (0.11)	2.98 (0.26)	2.83 (0.21)	2.64 (0.10)	2.68 (0.14)	2.65 (0.11)
Thr*51- OGH	100%	100%	100%	63%	95%	100%	100%	100%
Glu74- OE2...				2.90 (0.23)	3.11 (0.25)		2.96 (0.24)	
Tyr*54- OH				75%	74%		11%	
Glu74- OE2...	3.29 (0.13)	3.32 (0.12)	3.30 (0.13)	3.29 (0.14)	3.27 (0.14)	3.31 (0.13)	3.29 (0.16)	3.32 (0.13)
Leu73- NH	56%	43%	51%	55%	59%	46%	54%	40%
Glu74- OE2...	2.74 (0.08)	2.76 (0.08)	2.78 (0.09)	2.86 (0.11)	2.80 (0.10)	2.76 (0.08)	2.78 (0.09)	2.75 (0.08)
Glu74- NH	100%	100%	100%	100%	100%	100%	100%	100%
Glu22- OE1...	2.84 (0.18)	2.84 (0.15)	2.90 (0.22)	2.88 (0.16)	2.88 (0.20)	2.86 (0.18)	2.80 (0.12)	2.82 (0.13)
Lys100- NZ	67%	99%	22%	91%	100%	100%	100%	100%
Glu22- OE2...	2.94 (0.17)	2.98 (0.19)	2.92 (0.16)	2.97 (0.18)			2.96 (0.18)	2.90 (0.14)
Ala18- NH	32%	62%	91%	86%			83%	98%
Glu22- OE2...	3.09 (0.22)	3.12 (0.19)	3.06 (0.19)	3.16 (0.19)			3.03 (0.18)	3.01 (0.17)
Leu19- NH	21%	65%	88%	60%			88%	96%
Glu22- OE1...		2.91 (0.17)	3.02 (0.20)	2.87 (0.14)	2.84 (0.14)	2.84 (0.12)	2.86 (0.14)	2.89 (0.15)
Gln27- NEH		97%	39%	87%	95%	93%	99%	98%

Table 9-4 Hydrogen bonds between Asp*46, Asn71, and Leu72 in HP Complex

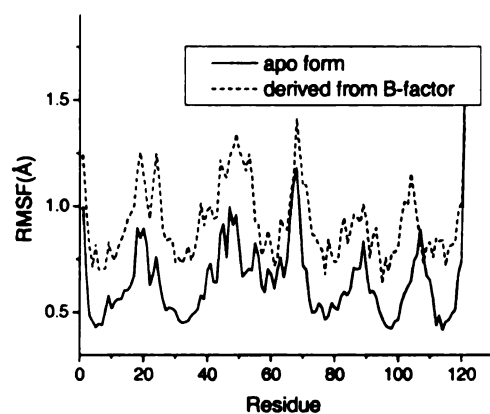
Subunit	1	2	3	4	5	6	7	8
Asp*46- OD1...	2.85 (0.14)	2.85 (0.14)	2.91 (0.16)	2.87 (0.15)	2.84 (0.14)	2.86 (0.14)	2.87 (0.15)	2.86 (0.14)
Asn71- ND2H1	96%	88%	43%	81%	99%	89%	96%	94%
Asp*46- OD1...	3.17 (0.18)	3.13 (0.18)	3.06 (0.18)	3.12 (0.18)	3.12 (0.18)	3.09 (0.18)	3.09 (0.18)	3.08 (0.18)
Asn71- NH	85%	91%	96%	92%	88%	93%	95%	95%
Asp*46- OD2...	2.92 (0.14)	2.98 (0.17)	2.94 (0.15)	2.91 (0.13)	2.94 (0.15)	2.97 (0.16)	2.93 (0.14)	2.98 (0.16)
Asn71- NH	100%	99%	100%	100%	100%	100%	99%	99%
Asp*46- OD2...	2.95 (0.15)	2.95 (0.19)	2.95 (0.15)	2.93 (0.14)	2.98 (0.16)	2.95 (0.14)	2.99 (0.16)	2.96 (0.15)
Leu72- NH	95%	99%	99%	100%	99%	99%	99%	99%

Table 9-5 Hydrogen Bonds with Glu24 and Ile25 in HP Complex

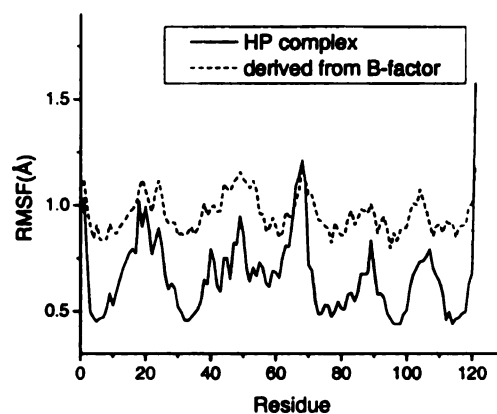
Subunit	1	2	3	4	5	6	7	8
Ser20-O...	3.05 (0.19)	3.06 (0.20)	3.00 (0.19)	3.10 (0.19)	3.00 (0.17)	3.17 (0.19)	3.06 (0.19)	3.19 (0.19)
Glu24-NH	88%	79%	86%	77%	96%	50%	84%	61%
Ala21-O...	3.17 (0.18)	3.11 (0.17)	3.20 (0.18)	3.13 (0.18)	3.22 (0.18)	3.12 (0.18)	3.11 (0.18)	3.04 (0.17)
Glu24-NH	65%	82%	60%	75%	56%	83%	85%	94%
Ala21-O...	3.17 (0.18)	3.02 (0.18)	3.00 (0.17)	3.07 (0.18)	3.24 (0.16)	3.18 (0.18)	3.07 (0.19)	3.08 (0.18)
Ile25-NH	57%	95%	83%	79%	46%	58%	84%	90%
Glu22-O...	3.22 (0.17)			3.25 (0.19)	3.32 (0.14)	3.30 (0.15)		
Ile25-NH	53%			11%	23%	15%		

Table 9-6 RMSF of Arg118 Backbone N Atom in the Regular and Pushing MD Simulations

Subunit	1	2	3	4	5	6	7	8
Regular MD (Å)	0.47	0.49	0.48	0.48	0.51	0.45	0.52	0.54
Pushing MD (Å)	0.50	0.61	0.51	0.64	0.56	0.61	0.56	0.50

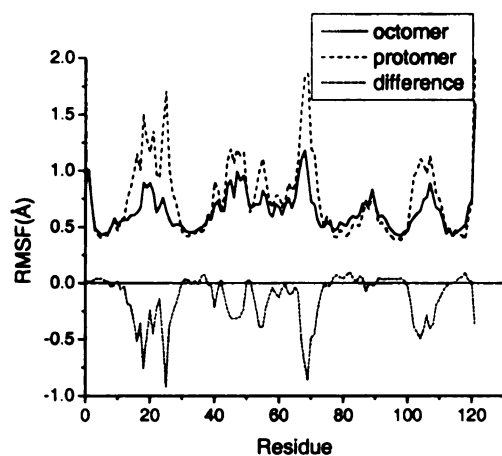


(a)

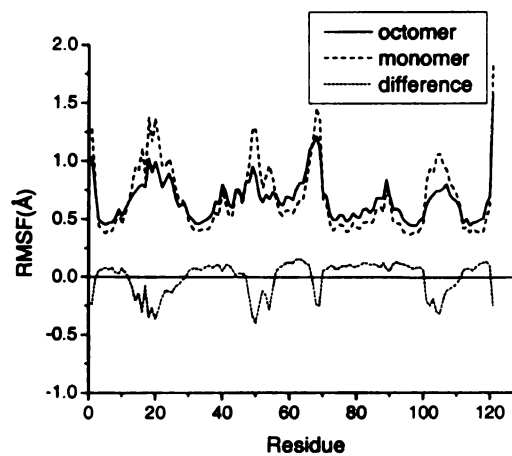


(b)

Figure 9-1 CA atom RMSDs from the simulations and corresponding crystal structure B-factors. (a) Apo form. (b) HP bound form.



(a)



(b)

Figure 9-2 Comparison of the RMSFs derived from the average protomer RMSF and the octamer RMSF (see text for the difference in these RMSFs). (a) Apo form. (b) HP bound form.

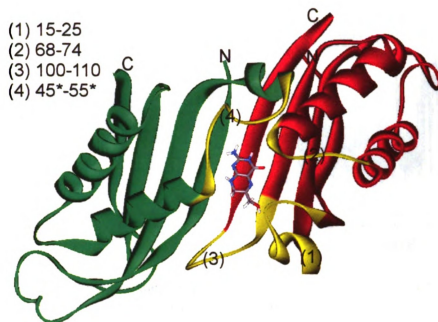


Figure 9-3 A representative snapshot of two adjacent subunits of apo DHNA from the MD simulation. HP is placed for the identification of the active site. One subunit is colored in green, and the other in red. The regions corresponding to large RMSF differences are highlighted in gold color. For clarity, only one set of the regions is highlighted.

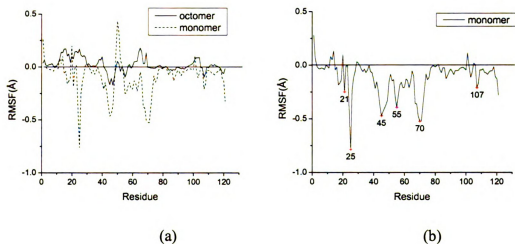


Figure 9-4 (a) RMSF differences between the apo DHNA and HP complex simulations. The differences based on the average protomer and octamer RMSF methods are displayed with solid and dashed lines, respectively. (b) RMSF differences between the apo DHNA and HP complex simulations excluding protomer 6, based on the octamer RMSF method.

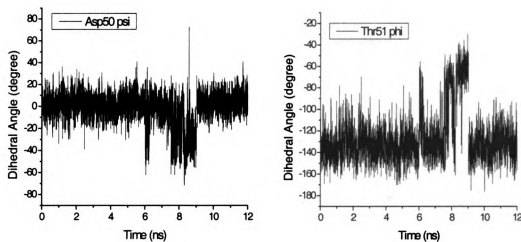


Figure 9-5 Time evolution of the ψ dihedral angle of Asp 50 and the ϕ of Thr51 in the eight subunits displayed sequentially as one trajectory.

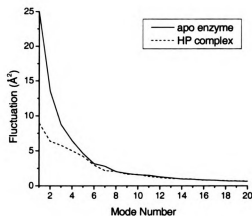


Figure 9-6 Principal Component Analysis eigenvalues of the first 20 modes in the apo and HP complex simulations. The total fluctuation over all the modes is 96.6 \AA^2 for the apo and 67.2 \AA^2 for the complex simulation.

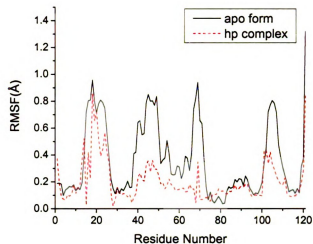


Figure 9-7 The CA atom RMSF projections onto the first principal component eigenvector for the apo and HP bound forms. The enhanced fluctuation regions of apo versus HP complex are evident and similar to those found as displayed in Figure 4.

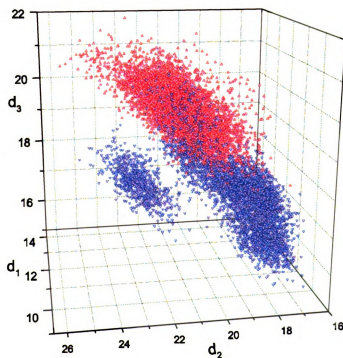


Figure 9-8 The CA distances: d_1 (Tyr*54 to Ile105), d_2 (Tyr*54 to Ala69) and d_3 (Ala69 to Ile105) parametric on the trajectory for the apo (blue) and HP bound (red) forms. There is a major and minor component for the apo form. The HP bound form distance fluctuations are smaller and fairly well confined to a part of the apo major state volume.

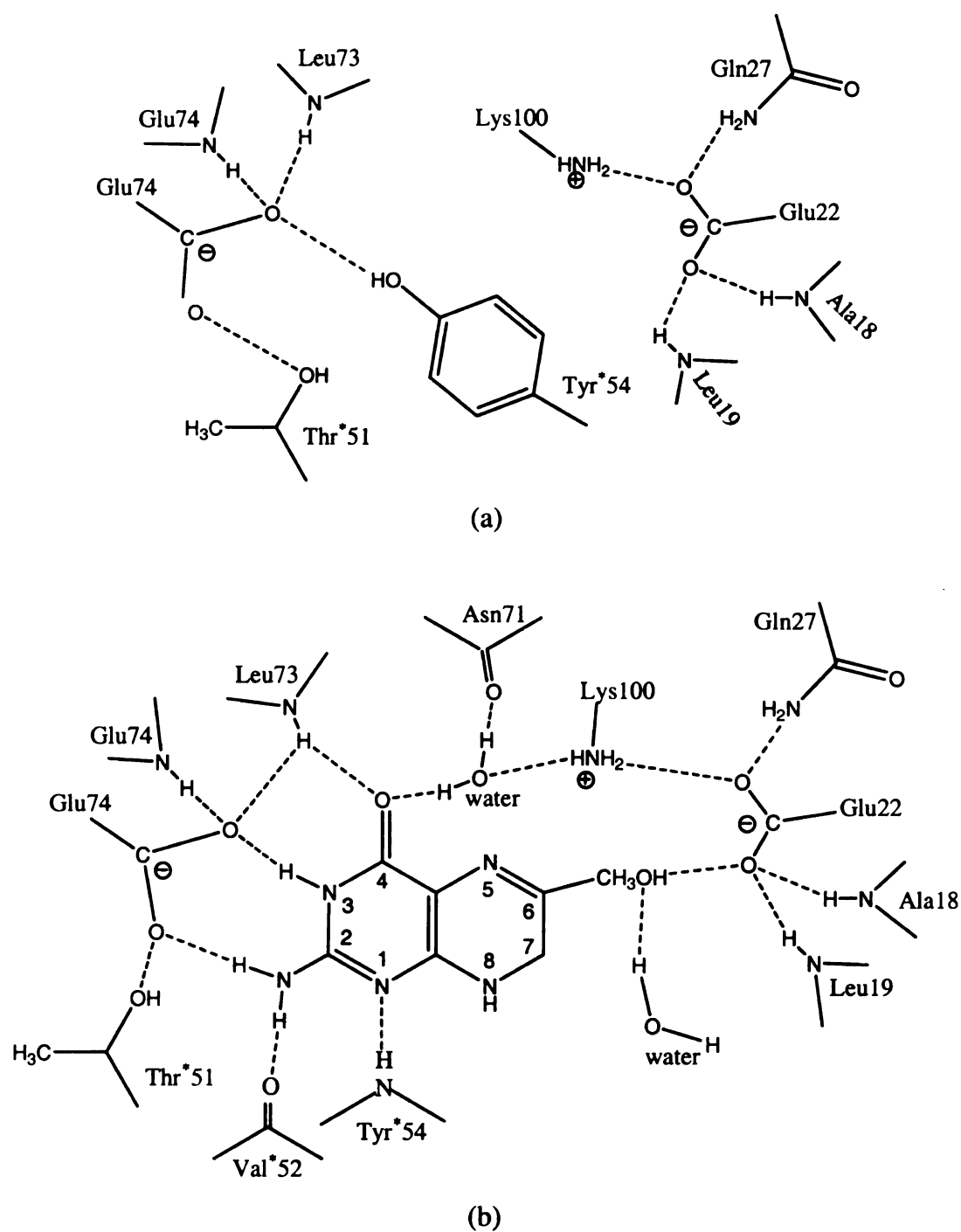


Figure 9-9 Active site interactions based on the MD simulations of apo enzyme (a) and HP complex (b). Residues in one protomer are distinguished from the other protomer by “*”s.

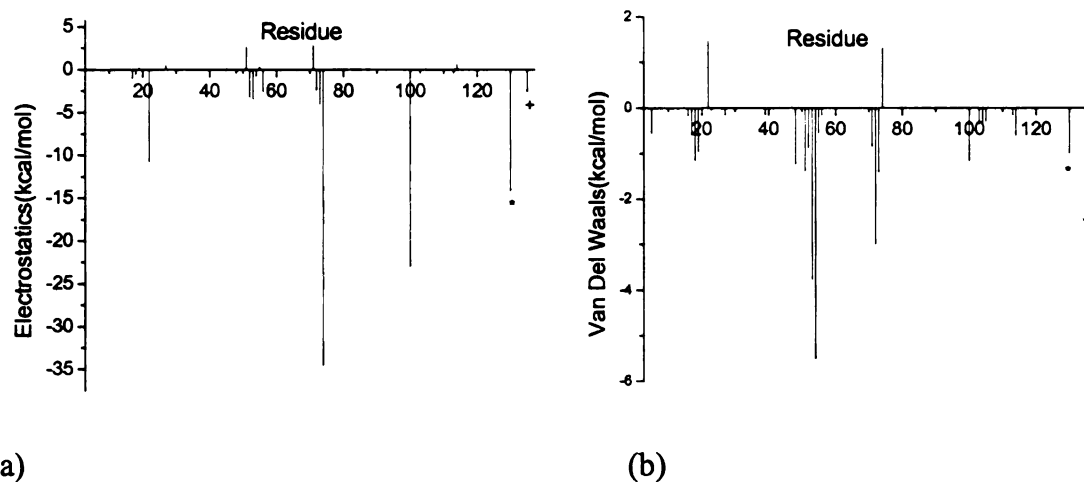


Figure 9-10 (a) Electrostatic and (b) van der Waals interaction energies between HP and all residues within 5 Å of HP. The energies are averages over the 2 ns simulation and the protomers.

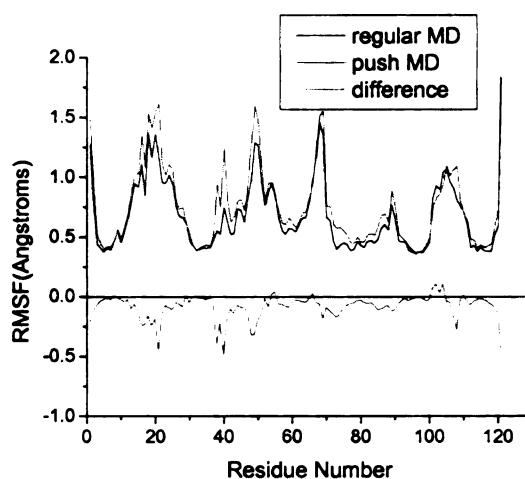


Figure 9-11 CA RMSF comparison between the push and regular simulations, indicating the small disturbance in the protein fluctuations from the exit of HP for the chosen force constant and pulling rate.

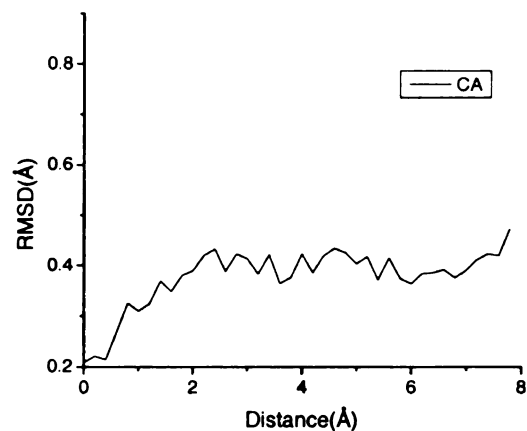


Figure 9-12 CA RMSD of the 50 ps averaged structure in each window relative to the average structure in the HP complex regular simulation.

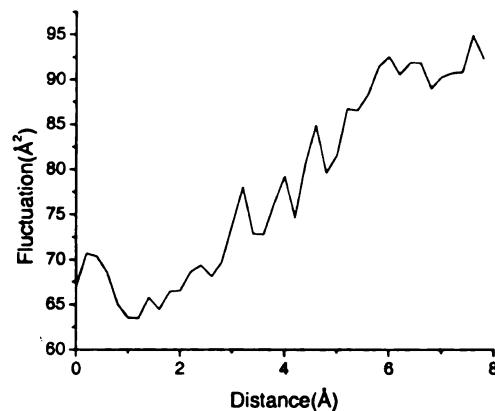
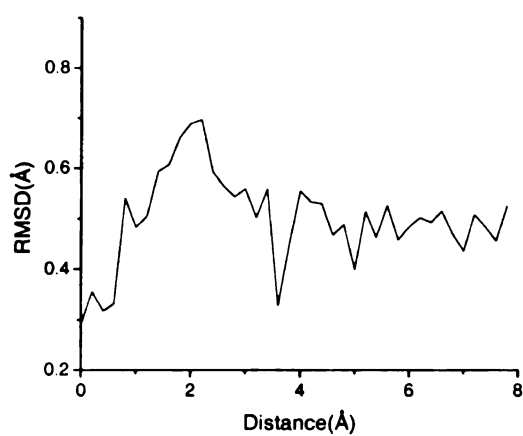
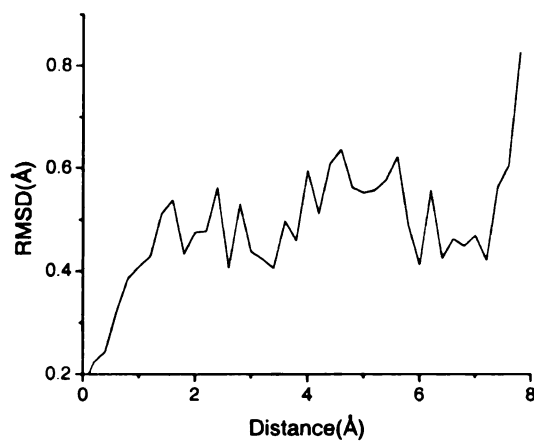


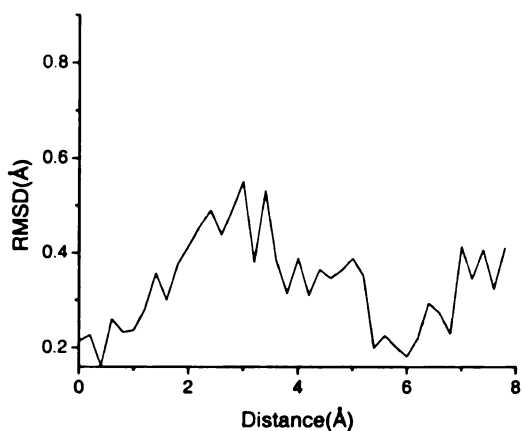
Figure 9-13 Total CA fluctuation of DHNA along the exit trajectory of HP showing a range of protein fluctuations that span the bound to free form results.



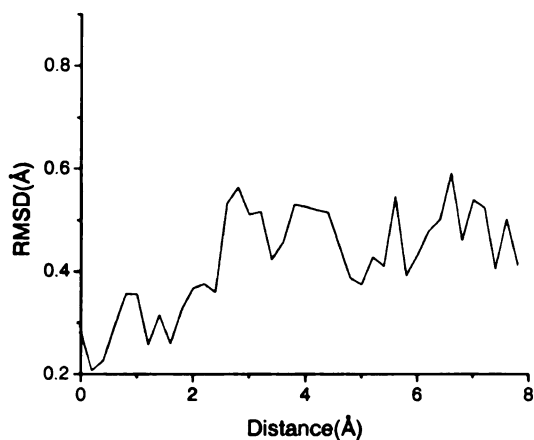
(a)



(b)



(c)



(d)

Figure 9-14 CA RMSD of the flexible regions of the 50 ps averaged structure in each window relative to the average fragments in the HP complex regular simulation. (a) Residues 15–25. (b) Residues 45*–55*. (c) Residues 68–74. (d) Residues 100–110.

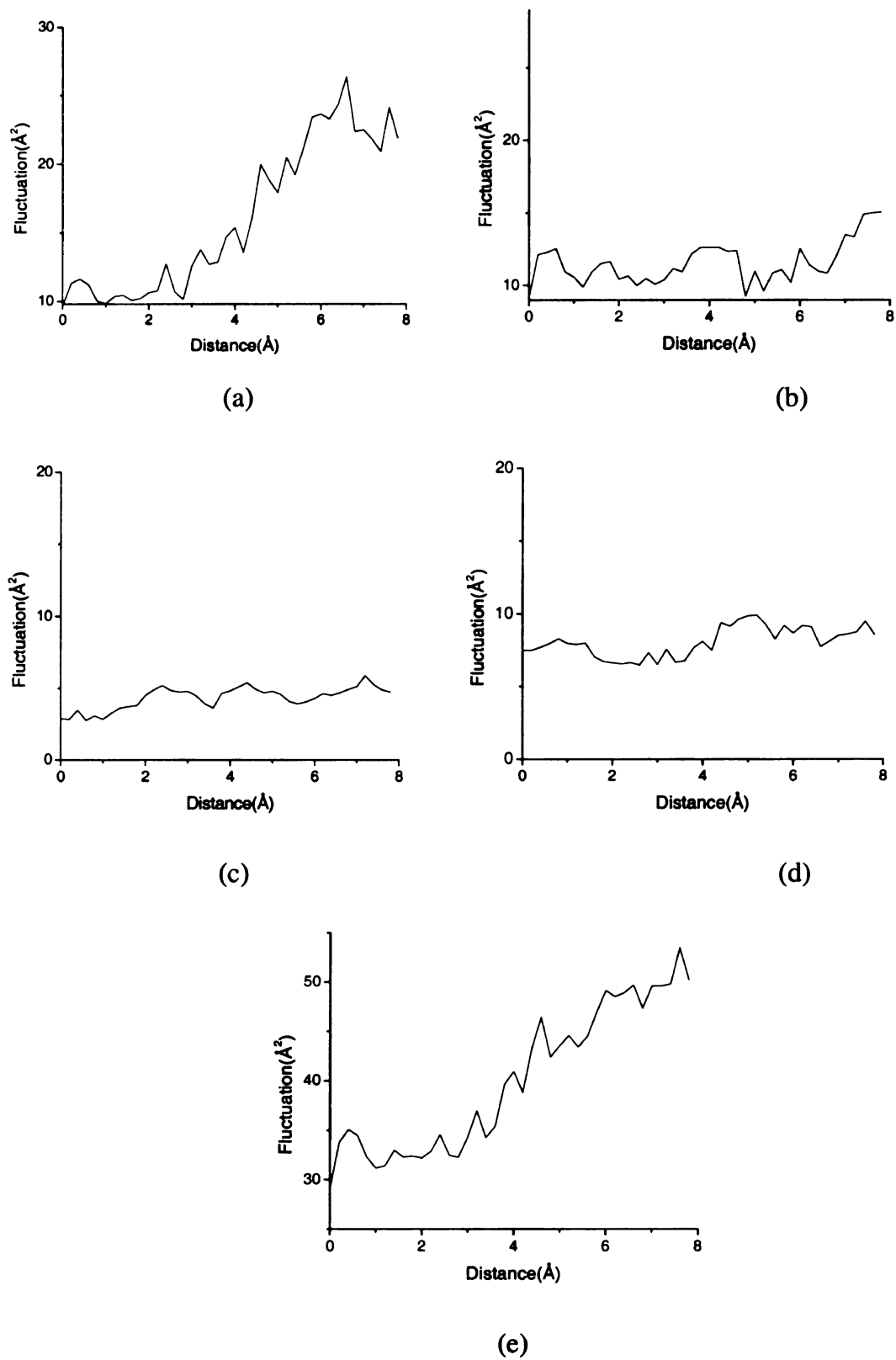


Figure 9-15 CA fluctuations of flexible regions along the exit path of HP. (a) Residues 15–25. (b) Residues 45*–55*. (c) Residues 68–74. (d) Residues 100–110. (e) sum of (a), (b), (c), and (d).

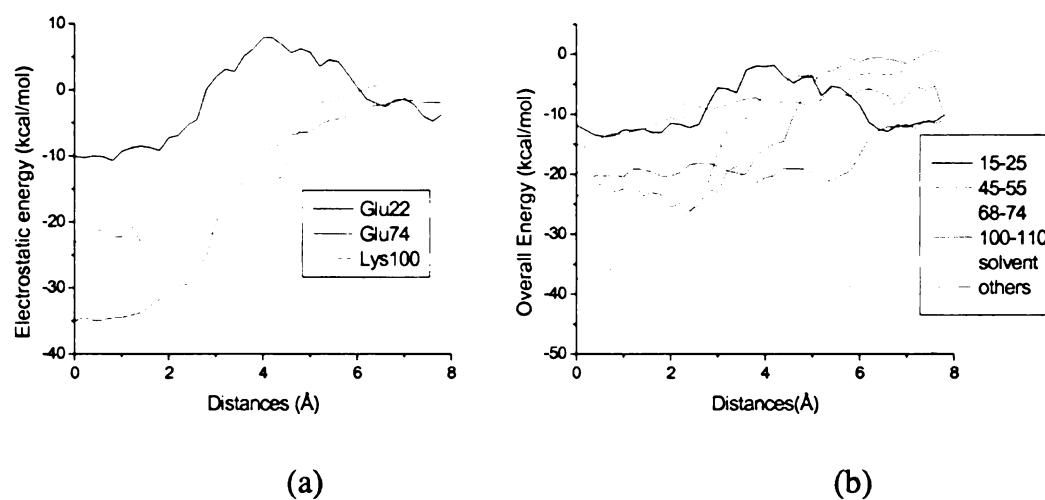


Figure 9-16 The interaction energies between HP and its surroundings over the exit pathway. (a) Electrostatic interactions between HP, and Glu22, Glu74 and Lys100. (The solvation energy of the starting point (-14.5 kcal/mol) was set to zero, to give a better view for this figure). (b) Overall interactions between HP and residues 15-25, 45*-55*, 68-74, 100-110, solvent and the other parts of the protein.

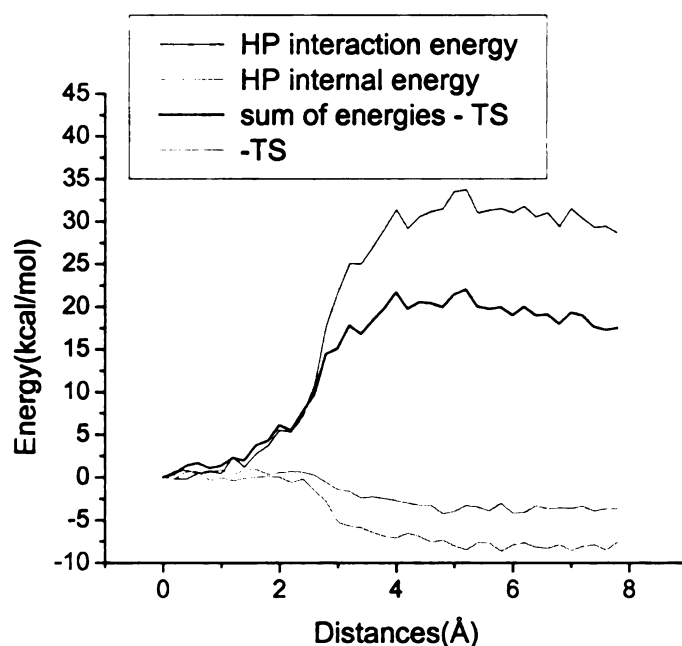


Figure 9-17 The intramolecular energetic and entropic contributions of HP, and its interaction energy with the environment and solvent, along the exit pathway.

Summary

The reaction mechanism catalyzed by and dynamics of yCD were studied with the use of experimental methods (NMR and quench-flow) and computational methods (MD and ONIOM). A complete reaction mechanism was proposed. The cytosine deamination proceeds via a sequential mechanism involving the protonation of N3, the nucleophilic attack of C4 by the Zn-coordinated hydroxide, and the cleavage of the C4-N4 bond. Two products, ammonia and uracil, are formed while the O4 of uracil is covalently linked to Zn. Thereafter, ammonia exchanges with bulk water, and then the Zn-O4 bond is cleaved through a 5-coordinated Zn complex transition state (the fifth coordinated atom is the carboxyl oxygen of Glu64). Along the reaction path, Glu64 is essential by acting as a proton shuttle to transfer protons from the Zn bound water to cytosine during the catalysis and substituting for uracil to coordinate with Zn during the uracil release.

The protein is quite rigid along the whole reaction path, and the active site stays completely buried, based on the MD simulation studies. The rigidity of the protein is confirmed by an NMR relaxation study (that describes motions on the ps-ns time scale) of the backbone amide groups in the apo and the transition state analog 5FPy complex. The quench-flow kinetics study shows that the product release is the rate limiting step in the activation of 5FC. The slow release process of the product was also studied by using NMR, which gives a release rate of 13 s^{-1} , though the dissociation constant K_d is only 22 mM. The opening of the active site apparently limits the release of the product. Two ligand releases paths were studied by using a steered MD method. Path 1 is in between the C-terminal helix and the F114 loop, which are required to move during the ligand release. Over this release path, the disturbance of the overall protein is rather small. Path

2 is in between the C-terminal helix and loop 1. When the ligand releases along this path, the rearrangement of the overall protein is much larger than that in the path 1 release. It appears that path 1 is more favorable than path 2, based on the MD simulation study.

The dynamics of yCD was also studied in apo, 5FPy and 5FU complexes, by using an NMR HD exchange method, which describes motion on the sec-min time scale. The results show that in the apo and the 5FU complex the C-terminal helix, F114 loop and loop 1 are more flexible than in the 5FPy complex. This suggests that yCD is ready to bind the substrate in the apo form and release the product in the product complex form, but is quite rigid during the chemical reaction process. It is striking that even though the apo form and the transition state analog complex have identical crystal structures, the dynamics of these two are quite different. The HD exchange study also confirms the product release path proposed on the basis of the MD simulation.

The catalytic mechanism of GD was studied by using the MD and ONIOM methods. The proposed mechanism is similar to but not the same as that of yCD. Two residues Glu55 and Asp114 act as the proton shuttle to transfer protons from a Zn bound water to guanine.

The dynamical properties of DHNA were studied in the apo and product, HP, bound form by using the MD method. The binding of HP rigidifies the protein's active site. The HP exit path way was studied by using steered MD. The chosen pathway leads to minimal structural disturbance of the protein. The analysis of the various components that contribute to the product release free energy provides an insight into the energy entropy compensation during this process.

Future work

The product release path from yeast Cytosine Deaminase has been studied by Molecular Dynamics. Future work for this system will focus on experimental mutagenesis studies of residues that the MD suggested as contributing to the product release process. The experimental studies would not only complement the computational results, but also potentially improve the enzyme activity, since product release is the rate-limiting step in the activation of 5FC.

In the Guanine deaminase ONIOM and MD study, two residues, Glu55 and Asp114, were proposed to be important for catalysis. Experimental mutagenesis studies of these two residues would be very interesting. They would possibly confirm the mechanism proposed in the ONIOM calculation.

A two-step method combining MD and ONIOM was introduced to study the Zn-O4 bond cleavage step in the uracil-yCD complex. While certainly approximate for obtaining a potential of mean force along a bond breaking reaction coordinate, it does permit the use of a high-level ONIOM method with an ensemble of structures obtained from MD simulations at a realistic computational cost. The method could be extended to study other reactions where a bond breaking/making reaction coordinate can be identified.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02845 4290