Chapter 4

Material and Method

4.1 Sample Preparation

4.1.1 Production of ^{15}N-Labeled TEP-I

E. coli thioesterase/protease I was isolated from the BL21 strain of *E. coli* containing a plasmid which carried the *testA/apeA* gene for producing thioesterase I, under the control of T7 polymerase[13; 42]. A string of six histidines was added to the C-terminus to facilitate protein purification by a Ni-NTA column [13]. The histidine tag was not cleaved from the C-terminus. The purified protein was checked by gel electrophoresis and was found to be better than 95% pure after elution from the Ni column. Typical protein yields were about 15mg/L for cells grown in M9 medium.

To obtain ^{15}N-labeled protein, the uniform ^{15}N-labeled protein samples were purified from *E. coli* cells grown in M9 medium, supplemented with 1g/L of $^{15}NH_4Cl$ and 0.5g/L of ^{15}N-labeled Isogro^{TM} (Isotec, OH). After cells had grown to an OD of 0.6, the ^{15}N-labeld amino acid was added, together with IPTG. The protein concentration was determined spectrophotometrically, using a molar extinction coefficient, $\varepsilon_{280}=34850M^{-1}cm^{-1}$. And the sample condition is: pH=6.02, concentration is around 0.8mM.

4.1.2 Michealis Complex and Tetrahedral complex formation

To produce the MC of TEP-I, we added DENP directly into TEP-I sample. The molar ratio of DENP and TEP-I is 10:1. The added sample was kept at room temperature for 5 minutes to ensure DENP binding to TEP-I. Because TEP-I with DENP is a slow formation of TC, this procedure can not make the sample going to TC state. Finally, using the centrifuge removes some un-binded DENP, and the force of centrifuge can not be higher than 1000G. This procedure is in order to ensure the sample in MC state during the following NMR relaxation measurements. All of relaxation measurements in MC state were at 290K. The sample can keep in MC state in 30hours. Before every relaxation measurement, we checked high field of a 1-D ^{1}H NMR experiment (1331) to confirm the sample is in MC state, or not.

After the relaxation measurements of MC of TEP-I, we added the same amount of DENP in the same sample to make the enzyme in TC state. Then, we put the sample at 310K in 30 minutes to increase the rate of reaction. Before relaxation measuring of TC state, we also checked the enzyme in TC state by 1-D ^{1}H NMR spectrum.

4.2 Data Acquisition and Process

All the NMR experiments were performed at 290K in Bruker AVANCE 600 AV spectrometer at magnetic static field of 14.1 Tesla with hydrogen Larmor frequency of 600.18MHz. And TXI ($^{1}H/^{13}C/^{15}N$) probe heads fitted self-shielded X, Y and Z-gradient coils were used. Standard pulse sequences were employed for measuring $^{15}N\text{-}T_1$, $^{15}N\text{-}T^2$, and steady state heteronuclear ($^{1}H\text{-}^{15}N$) NOE.[43; 44] Quadrature detection was achieved using States-TPPI method in F1 (nitrogen) dimension.52Water suppression was achieved using WATERGATE sequence.[45; 46] The spectral widths were set to 14 ppm in F2 (proton) dimension and 34 ppm in F1 (nitrogen) dimension. The proton carrier frequency was placed on the water peak during the pulse sequence, and was shifted to 8 ppm just before data acquisition. The nitrogen carrier frequency was set to 119 ppm. Time domain data were recorded with 2048 complex points in t2 dimension and 256 complex points in t1 dimension, with 16 (or 32 for NOE experiments) scans per t1 point. $^{15}N\text{-}T_1$ values were determined from spectra taken with ten different relaxation delays of 0.8, 51.8, 110.8, 176.8, 345.8, 456.8, 600.8, 803.8, 1149.8ms, whereas 8 different delays of 15.584, 31.168, 46.752, 62.336, 77.920, 93.504, 109.088 and 124.672ms were used during the CPMG period of the $^{15}N\text{-}T_2$ experiments. $^{1}H\text{-}^{15}N$ steady-state heteronuclear NOE values were measured by recording spectra in the absence and presence of a 1H saturation period of three seconds. This saturation was achieved by the application of 1H 120° pulses spaced at 20-ms intervals.[47] The recycle delay between individual scans was five seconds in the absence of 1H saturation, and two seconds in the other relaxation experiments. The steady-state heteronuclear NOE experiments were carried out in two times. All of the NMR spectra were processed using XWINNMR software (Bruker AG, Karlsruhe, Germany) on a UNIX-based Silicon Graphics Indigo 2 workstation.

The proton chemical shift values were referenced to DSS at 0 ppm.[48; 49] The ^{15}N chemical shift values were indirectly referenced using the following consensus Ξ ratio of 0.101329118 for $^{15}N/^{1}H$ [48]. The free induction decay in time domain was processed with squared-sine window function 60° shifted in t2 and t1 dimensions, phase-corrected, base-line corrected, and Fourier-transformed. The AURELIA(**AU**tomated **RE**sonance **LI**ne **A**ssignment) program was employed for peak picking and intensity measurements [50].

R_1 and R_2 were obtained by curve fitting to mono-exponential decay, $I(t)=I(0)exp(-R_{1,2}t)$, using Marqaurdt-based nonlinear least-squares algorithm in Prism. The uncertainties of R_1 and R_2 are errors in curve fitting. The reported $R_{1,2}$ values and their uncertainties were the mean values of two independent data sets. The reported XNOE ratios were the ensemble average of two pairs heteronuclear NOE experiments, and the errors were defined the standard deviation of two data sets.

4.3 Data Analysis

4.3.1 Rotational Diffusion Tensor

The program "Quadric Diffusion" version 1.12 from Arthur G. Palmer's Lab at Columbia University[51] was used to calculate the diffusion tensor. In this approach a local diffusion constant, $D_i=(6\tau_{ci})^{-1}$, is defined for the ith spin for i .1, . . ., N by fitting the Lipari-Szabo model-free spectral density function[24; 25] to the experimental R2, R1 and XNOE data.[52]

In this program, the diffusion tensor is determined by R_2/R_1 values and protein structure. The diffusion tensor is transformed to the principal frame and a linear least-square optimization technique is employed to determine the diffusion tensor and orientation with respect to the molecular frame. The diffusion tensor parameters ($\{D_{iso}\}$, $\{D_{\perp}, D_{//}, \theta, \phi\}$, and $\{D_{zz}, D_{yy}, D_{xx}, \theta, \phi, \psi\}$) for isotropic, axially symmetric and fully anisotropic diffusion models are fitted to the experimental data via a χ^2 minimization protocol[51; 52]. The χ^2 function is the sum, over all evaluated residues, of the squared difference between experimental and model-calculated D_i values, weighted by the squared experimental uncertainty in D_i. For model selection, the F-test, as defined by Lee et al.[51] was used to test the improvement in the statistical fit. Unlike the χ^2 test, which depends on the number of sites included in the calculation, the F-test takes into account the different number of degrees of freedom of the two models, and is used to test the improvement in the statistical fit. In practice, a value of $F > 4.5$ indicates that the improved fit is significant at a confidence level greater than 99 %.

4.3.2 Model-free Analysis

R_1, R_2 and NOE of each spin were used to calculated Model-free parameters by the software version 4.1 provided by A. G. Palmer's group and performed in IBM RS/6000 workstation. Model-free analysis has three steps:

(1.) The initial estimation of the overall correlation time or diffusion tensor,

(2.) Model selection,

(3.) The final optimization.

The initial estimation of overall motion was calculated from the program "Quadric Diffusion 1.12". For the diffusion model selection, the parameters were also estimated from "Quadric Diffusion 1.12": τ_m $(=(6D_{iso})^{-1})$ for isotropic model, τ_m, $D_{//}/D_{\perp}$, θ, ϕ for axially symmetric model.

The model selection in Model-free was fitted by relaxation data. There are five models in Model-free approach (Table 1). Γ_i is the sum squared-error residual for the *i*th spin. χ^2 in five models were also obtained to characterize the goodness of fit. The definition of χ^2 is:

$$\chi^2 = \sum_{i=1}^{N} \Gamma_i = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \frac{\left(R_{ij} - \hat{R}_{ij}\right)}{\sigma_{Rij}^2} \qquad (4.1)$$

in which the total number of spins is *N* and the total number of static magnetic fields for which data is available in *M*; R_{ij} is the *j*th relaxation parameter such as R_1, R_2, and NOE; $\hat{R}_{ij}$ is the corresponding fitted value; σ_{ij} is the experimental uncertainty in the *j*th relaxation parameter. In Monte Carlo simulation, 300 data sets including R_1, R_2, and NOE for every spin were synthesized by adding a random noise term to the best-fit relaxation parameters, $\hat{R}_{ij}$. The quality of the fit between the experimental data and theoretical model was assumed by comparing the optimal value of Γ_i derived from equation [4.1] with the α=0.1 critical value of Γ_i for each model.[31]

Then, simulated data generated using a particular model was fit using a second model containing one or more additional parameters. The two models were chosen to be nested: the parameters of the first model were a subset of the parameters of the second model. The simulation results were used to determine the cumulative probability distribution and α= 0.20 critical value of an *F*-statistic defined as

$$F_i = \frac{p_2\left(\Gamma_{1i} - \Gamma_{2i}\right)}{\left(p_1 - p_2\right)\Gamma_{2i}} \tag{4.2}$$

in which Γ_{1i} and Γ_{2i} are the sum of squared errors for models with p_1 and p_2 degrees of freedom ($p_1 > p_2$). The *F*-statistic measures the random statistical reduction in Γ_i obtained by incorporation of additional parameters when the simpler model correctly describes the data. Simulated data generated with model 1 (p_1 = 2) were analyzed using models 2 or 3 (p_2 = 1) to calculate the *F*-statistic comparing one and two parameter models. The Model selection is outlined in the flowchart presented in Figure 4.1. If Γ_i was much greater than the critical values for model 1, 2, and 3 (generally $\Gamma_I > 20$) and Γ_i was 0 for either model 4 or model 5.

After model selection, the optimization of the overall rotational diffusion parameters and internal parameters was performed. The optimization was to minimize χ^2 in equation [4.2]. The range of τ_e in grid search was between 0~200ps in steps of 5ps for model 2 and 0~5000ps in steps of 12.5ps for model 5, S_s^2 and S_f^2 from 0 to 1 in steps of 0.05, and R_{ex} from 0 to 15s^{-1} in steps of 0.3s^{-1}. In model 5, τ_e was required to be less than τ_m. In Monte Carlo simulations, 500 randomly distributed synthetic data sets were created.

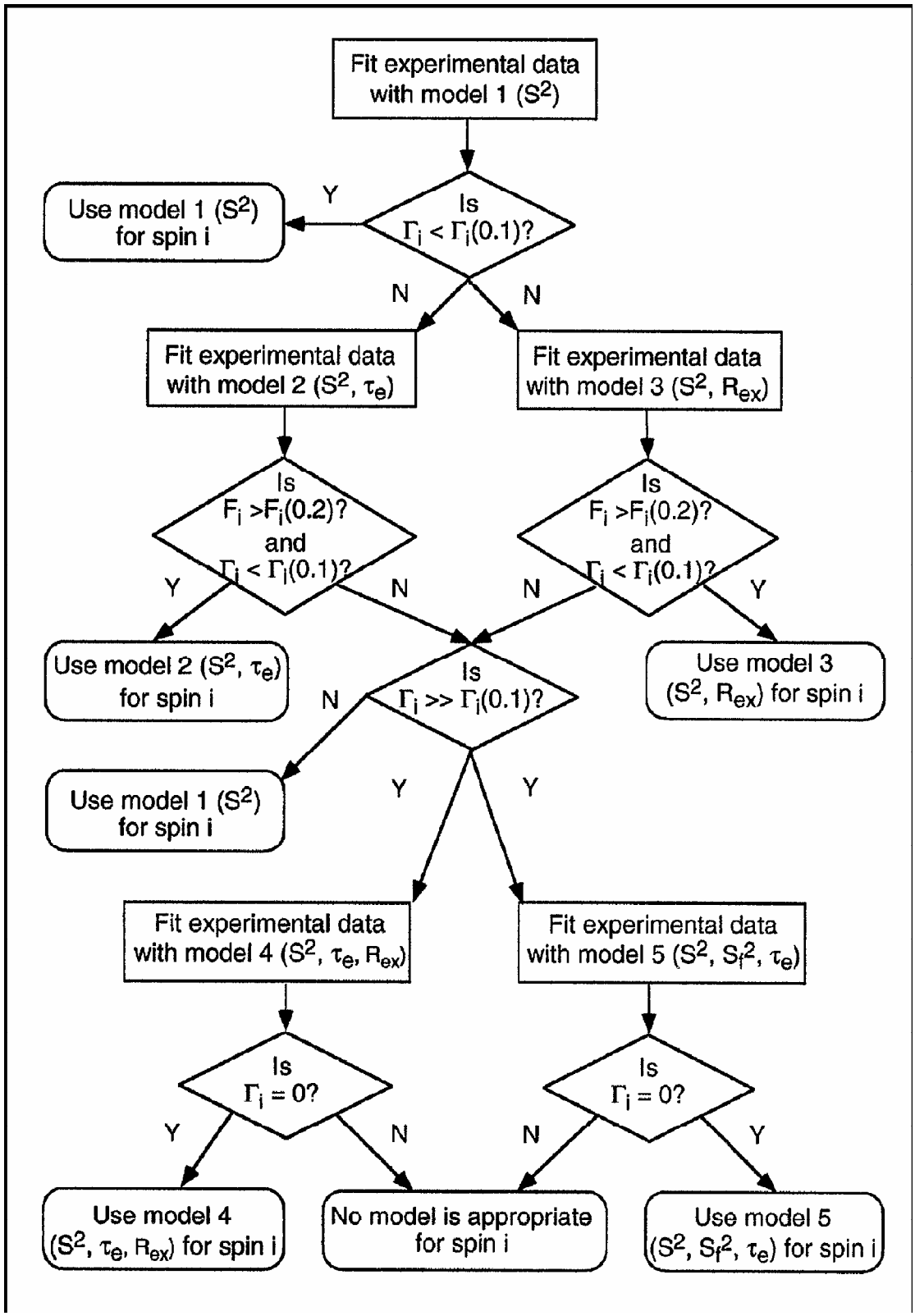

Figure 4.1

Flowchart of the model selection in Model-free analysis. [31]

4.3.3 Reduced Spectral Mapping

Equation (2.32) was applied to determine J(0), J(ω_N), and J($0.87\omega_H$). In all calculations, the following parameters were set as: $\Delta\sigma=\Delta\sigma_{//}-\Delta\sigma_{\perp}$=-170ppm for backbone amide nitrogen; r_{NH} = 1.02Å for the NH bond length.[53]

4.4 Analysis of Molecular Dynamics Simulation

4.4.1 Protocols of Molecular Dynamics Simulations

The explicit solvent molecular dynamics (MD) simulation of *apo* and TC of TEP-I was performed by using the program "sander" in AMBER8 with an all-atom force field PARM99, and Coulomb interactions were calculated using the particle-mesh Ewald method. The cutoff was set as 9Å. The atom initial coordinates of the *apo* and TC systems were taken from the X-ray crystal structure of TEP-I, PDB code are 1IVN and 1J00, respectively[16]. Hydrogen atoms were added using the sub-suite of AMBER, LEaP. To neutralize the system, we added two counter ions, Na^+. All bonds were restrained by the SHAKE algorithm. And the systems were solvated in TIP3P water box with 74.14×70.41×69.40(Å), the total number of water molecule are 8812 and 8398 in apo and TC states, respectively. In the system of *apo*-TEP-I, there are 2814 protein atoms and 28714 total atoms, and for the TC system, there are 2839 protein atoms and 27497 total atoms. The missing residue GLN32 in 1IVN and 1J00 was added by the "loopy" module of the Jackal suite. The environment of *apo* and TC is all set at 290K, which is the same as NMR setting.

In order to analyze the TEP-I in the MC state, initial MC structure was prepared by docking DENP to the apo TEP-1 structure using AutoDock. Although the DENP docked in the binding pocket, the hydrogen bonds between DENP and oxyanion holes of TEP-I did not form during the simulation. Thus, this MD simulation is so-called "preMC" state.

There is an unusual residue, SDP, in TC structure, which is the DENP formed covalent bond with Ser10 without phenyl group. The charge library of SDP was calculated by GAUSSIAN.

For running the energy minimization, the same strategy was applied for *apo* and TC for 40ps. The systems were heated slowly to 290K and equilibrated for additional 500ps. Then, we set isotropic pressure for running the MD simulation, the time step is 2fs and the duration time is 6ns.

4.4.2 Dynamic Parameters from MD Simulation

To understand the overall motion of molecular, we calculated two general dynamic parameters from our simulation, B-factor and order parameters. B-factor and order parameters are popular used to describe the overall motion from X-ray and NMR relaxation, respectively. [54]

X-ray crystallographic *B*-factors are sensitive to the mean square displacements of atoms. *B*-factors are available at nearly all heavy atom positions and consequently provide atomic-level detail on disorder in proteins.

$$B-\ factor = \frac{8\pi^2}{3}\left(\left\langle \vec{r}_i^{\,2}\right\rangle_t - \left\langle \vec{r}_i\right\rangle_t^{\,2}\right) \tag{3.2}$$

In which $\vec{r}_i$ is the *i* atom position, and the braket is the ensemble average by time. The right hand term of equation is proportional to the mean square displacement.

MD simulation provide a atomic level detail on the dynamical behavior of protein, and all quantities obtained from NMR laboratory frame spin relaxation measurements in principal and be calculated from and MD trajectory. Comparison between MD- and NMR- derived order parameters indicates that for relatively rigid proteins, quantitative agreement is obtained for ordered parts for protein.[55]

$$\mathrm{C_I}(\tau) = \left\langle \mathrm{P_2}(\hat{\mu}(t)\cdot\hat{\mu}(t+\tau))\right\rangle \quad \rightarrow \quad \mathrm{S}^2 = \lim_{\tau\to\infty} C_I(\tau)$$

Where $\hat{\mu}(t)$ is a unit vector, $P_2(\theta)$ is of the Legendre polynomial. $\theta = \hat{\mu}(t) \cdot \hat{\mu}(t+\tau)$

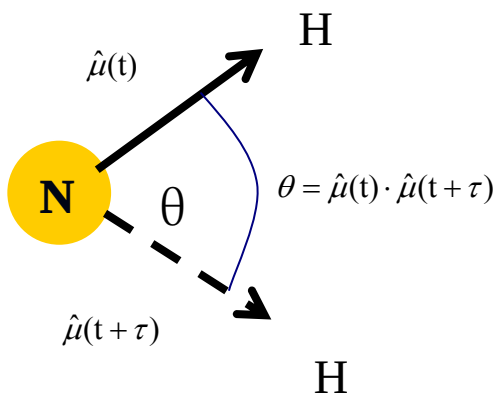

Figure 3.2

The simple scheme of NH vector.