Variation of Free-Energy Landscape of the p53 C-Terminal Domain Induced by Acetylation: Enhanced Conformational Sampling

Shinji Iida, Tadaaki Mashimo, Takashi Kurosawa, Hironobu Hojo, Hiroya Muta, Yuji Goto, Yoshifumi Fukunishi, Haruki Nakamura, and Junichi Higo

Introduction

An intrinsically disordered region (IDR) of a protein is highly flexible in physiological conditions unless interacting with its partner molecule. Structurally ordered proteins are well known to exert their biological functions through well-defined quaternary, tertiary, and secondary structures. In contrast, IDRs exert their functions actively using conformational flexibility. Signal transduction is a typical function of IDR, where a single IDR interacts with different partner molecules to regulate the signal transduction. This multipartner interaction property is called a hub property. IDRs are related to some diseases such as cancer, diabetes, and neurodegenerative disease. Therefore, IDRs are an important subject related to biophysics, and medical science.

Actually, IDRs can be characterized as (1) dynamic conformations (conformational ensemble), (2) post-translation modification (PTM), and (3) roles of IDR–partner interaction. (1) An IDR does not adopt a specific tertiary structure in the unbound state. Consequently, IDR is not characterized by a single tertiary structure but by a conformational ensemble. The relation between the conformational ensemble and its biological functions has not been understood sufficiently. Presumably, the conformational ensemble architecture affects the binding mechanism. (2) PTM is found frequently in IDR. Importantly, PTM such as methylation, phosphorylation, and acetylation alters the physicochemical properties of IDR and modulates the functionality. Consequently, it is interesting to investigate the variation of the conformational ensemble by PTM. (3) Some IDRs adopt specific tertiary structures when interacting with their partners. This phenomenon is known as coupled folding and binding. Two interaction mechanisms have been proposed for coupled folding and binding processes: induced folding and conformational selection. In induced folding, IDR
binds to the partner with conformations differing from that adopted in the complex (i.e., genuine conformation found in the native complex form). Subsequently, the IDR conformation varies until reaching the most stable complex structure. In conformational selection, the genuine bound conformation is involved in advance in the conformational ensemble of the unbound state. This bound form is used to bind to the partner. Recently, a computational report has described that coupled folding and binding takes place by a combination of the induced folding and the conformational selection.\cite{11} Another IDR–partner interaction picture is fuzzy-complex formation, by which IDR conformational disorder/flexibility emerges in both bound and unbound states. Enthalpic stability for the individual bound form is less important for discussing the complex formation. A fly casting mechanism provides an alternative viewpoint to the IDR–partner interaction scheme.\cite{12} An IDR has a greater interaction radius than that of a folded protein. It captures partners at a large distance like a fishing line.

Tumor suppressor protein p53 consists of four functional domains: the transactivation domain (TAD) [residues 1–43], the DNA binding domain (DBD) [100–300], the tetramerization domain (TET) [320–360], and the C-terminal negative regulatory domain (CTD) [363–393]. Both TAD and CTD are IDRs that interact with many molecules. These domains possess a hub property.\cite{13} To date, three TAD-partner complex structures have been solved, where TAD adopts helical structures on surfaces of the partners in all three cases.\cite{15–17} Four CTD-partner complex structures have been determined, where CTD adopts various structures: a helical structure to bind to S100B (PDB ID: 1DT7),\cite{18} a sheet to Sir2 (PDB ID: 1MA3),\cite{19} and fuzzy/coiled structures to CBP (PDB ID: 1JSP),\cite{20} and Cyclin A (PDB ID: 1H26).\cite{21} It is particularly interesting that CTD uses a common sequence [380–386] to bind to these four partners. In this sense, this short common binding region has the hub property. Furthermore, the states of histidine 380 (H380) and lysine 382 (K382) in CTD vary depending on the binding partner as follows: when binding to S100B, K382 is acetylated, although H380 can take either a positively charged or neutralized state depending on the pH condition.\cite{18,22} To Cyclin A, K382 is nonacetylated and H380 is neutralized.\cite{21} To Sir2, K382 is acetylated and H380 is positively charged.\cite{19} To CBP, K382 is acetylated and H380 is neutralized.\cite{20} The correspondence between the state of CTD and the partner is presented in Table 1.

Actually, CTD and its partner have been studied using molecular simulations: Allen et al. demonstrated trends of fluctuations around CTD binding sites of partners.\cite{23,24} Chen et al. assessed which population selection or the induced folding is plausible to bind to S100B using molecular dynamics (MD) with an implicit solvent model and some simplifications of the protein model.\cite{24} McDowell et al. showed heterogeneity of CTD and the binding mechanism of CTD to S100B.\cite{25} Staneva et al. investigated conformational preferences of the unbound CTD using an implicit-water Monte Carlo simulation.\cite{26} Although these studies provided beneficial knowledge for the conformational ensemble of CTD, they did not clarify the effects of acetylation on the CTD’s conformational ensemble. We consider that computation of a free-energy landscape is crucially important to elucidate the effects of acetylation on the conformational ensemble. To that end, a powerful conformational sampling method is necessary.

A multicanonical simulation has been introduced to enhance the sampling of complicated systems.\cite{27–29} Nakajima et al. introduced a multicanonical MD simulation (McMD) using Cartesian coordinates for dynamic variables. Adoption of Cartesian coordinates produces a multimolecular system that is tractable without special devices in a computer program. McMD generates various conformations under equilibrium conditions, which provides not only the most thermodynamically stable state but also intermediate states of the system. Importantly, a free-energy landscape at arbitrary temperature is computable from the resultant conformational ensemble. To increase the sampling efficiency of McMD, trivial trajectory parallelization McMD (TTP-McMD) was developed,\cite{30,31} and applied to systems consisting of an intrinsically disordered segment and its partner protein.\cite{11,32} Recently, a virtual system coupled McMD (V-McMD) was also developed to increase the McMD sampling efficiency.\cite{33} Terakawa et al. performed all-atom TTP-V-McMD of a p53 linker region (40 residues long) to design force-field parameters for a coarse-grained simulation model. The study reproduced an X-ray scattering profile using the force field.\cite{34}

For this study, we examined four CTD fragments in an unbound state (i.e., single-chain state) using TTP-V-McMD, where K382 was nonacetylated or acetylated and H380 was positively charged or neutralized. The system was treated with an all-atom model in an explicit solvent. The free-energy landscapes were computed from the sampled conformations. Results show that the free-energy landscape of the single-chain state varies by the K382 acetylation and the H380 neutralization. In fact, the IDR-partner binding is controlled not only by the finally formed complex structure but also by the conformational distribution of the unbound state. We suggest possible binding mechanisms of CTD to their partner molecules, S100B, Sir2, CBP, and Cyclin A with investigation of the single-chain free-energy landscapes.

### Theory

In a conventional (canonical) MD simulation of a biomolecular system, the force acting on an atom $i$ is computed as $f_i = -\nabla E_i(r)$, where $E_i(r)$ is a potential energy of the
From eq. (5), the energy distribution is calculated as

\[ P_c(E_R, T) = Z_c^{-1}n_R(E_R) \exp \left( \frac{-E_R}{R_{\text{gas}}T} \right), \]  

where \( R_{\text{gas}} \) represents the gas constant, \( n_R(E_R) \) denotes the density of states (DOS) of the system at the potential energy \( E_R \), \( T \) stands for temperature, and \( Z_c = \int n_R(E_R) \exp \left( -E_R/R_{\text{gas}}T \right) dE_R \) represents a partition function (i.e., a normalization factor) at \( T \).

McMD uses a modified potential energy \( E_{mc} \) instead of \( E_R \), which is defined formally as

\[ E_{mc}(E_R) = E_R + R_{\text{gas}} T \ln \left( \frac{P_c(E_R, T)}{P_c(E_R, T)} \right), \]  

\[ = R_{\text{gas}} T \ln \left[ n_R(E_R) \right]. \]  

The force acting on the atom \( i \) is represented as

\[ f_{mc}^i = -\frac{\partial E_{mc}}{\partial r_i}. \]  

An energy distribution \( P_{mc}(E_R) \) from the McMD simulation at \( T \) is given simply by replacing \( E_R \) by \( E_{mc} \) in eq. (1) as

\[ P_{mc}(E_R) = Z_{mc}^{-1}(T)n_R(E_R) \exp \left( -\frac{E_{mc}}{R_{\text{gas}}T} \right) \]  

\[ = Z_{mc}^{-1}(T) \frac{n_R(E_R)}{n_R(E_R)} = \text{const}, \]  

where \( Z_{mc} = \int n(R) \exp \left( -E_{mc}/R_{\text{gas}}T \right) dE_R \) is a partition function at \( T \). This equation guarantees that the energy distribution from a longtime McMD converges to a flat function. Then, the flatness of \( P_{mc}(E_R) \) is a measure to judge the convergence of sampling.

Virtual system coupled McMD (V-McMD) has been developed to increase the sampling efficiency of McMD, where a “virtual system” is introduced with setting its physical quantities arbitrarily. From here, we explain the framework of V-McMD: Assume that a virtual system exists in addition to the real system (i.e., the biomolecular system). We define an entire system as the sum of the real system and the virtual system. Here we assume that the virtual system does not interact explicitly with the real system. In other words, no cross-term exists in the potential energy of the entire system. Then, the total potential energy is given as \( E_{\text{total}} = E_R + E_v \), where \( E_v \) is the virtual system’s potential energy, a function of a virtual system’s coordinate \( v \). Accordingly, DOS of the entire system is represented as \( n_{\text{total}}(E_R, E_v) = n_R(E_R) n_v(E_v) \), where \( n_v \) is DOS of the virtual system. Using the arbitrary property of the virtual system, we simply set this DOS as \( n_v(E_v) = 1 \). Then, the multicanonical energy \( E_{mc} \) of the entire system is given as

\[ E_{mc}(E_R, E_v) = R_{\text{gas}} T \ln \left[ n_{\text{total}}(E_R, E_v) \right] \]  

\[ = R_{\text{gas}} T \ln [n_R(E_R) n_v(E_v)]. \]  

From eq. (5), the energy distribution is calculated as

\[ P_{vc}(E_R, E_v) = Z_{mc}^{-1}(T) n_{\text{total}}(E_R, E_v) \exp \left( -\frac{E_{mc}}{R_{\text{gas}}T} \right) \]  

\[ = Z_{mc}^{-1}(T) \frac{n_R(E_R)}{n_{\text{total}}(E_R, E_v)} = \text{const}, \]  

where \( Z_{mc} = \int n_{\text{total}}(E_R, E_v) \exp \left( -E_{mc}/R_{\text{gas}}T \right) dE_R dE_v \) is a partition function of the entire system at \( T \). As with eqs. (4) and (6) also ensures that the two-dimensional distribution \( P_{vc}(E_R, E_v) \) converges to an even function after a long V-McMD simulation.

Generally, it is more difficult to achieve convergence of the two-dimensional distribution than that of one-dimensional one. The virtual system can be set arbitrarily. Then, we discretize coordinate \( v \) to reduce the sampling space as \( v = v_i \), where \( i = 1, \ldots, n_v \) and \( n_v \) is the number of the virtual states allowed. We introduce the term “virtual state” to specify the state of the discretized virtual system. The \( i \)th virtual state corresponds to \( v_i \). Accordingly, \( E_v \) is discretized as \( E_{vi} \). Therefore, the time-evolution of the virtual system is done using the Metropolis Monte Carlo scheme.

Now, the system is specified by two quantities: \( E_R \) and \( v_i \). We set the action of the virtual system on the biomolecular system so that \( E_v \) is confined in a zone \( Z_v = [E_v^\text{min}, E_v^\text{max} \leq E_v^\text{min}(v_i)] \) when its virtual state is in \( v_i \). The method for the confinement is explained in an earlier report. We set the zones so that \( Z_v \) overlaps with \( Z_{v-1} \) and \( Z_{v+1} \), although \( Z_{v-1} \) and \( Z_{v+1} \) have no mutual overlap as presented in Figure 1. The actually used zones are given later. Assuming that the system is at the filled-circle position of Figure 1, for which energy \( E_R \) is involved in \( Z_t \) in the virtual state \( v_t \), then, the potential energy is also involved in \( Z_{v+1} \). Therefore, the state \( v_i \) can transition to \( v_{i+1} \) using the Monte Carlo method. The molecular configuration \( r \) is not changed (i.e., \( E_R \) is not altered) in this transition. The transition from \( v_i \) to \( v_{i-1} \)
might occur because $E_R$ is involved in $Z_i$ and $Z_{i-1}$ if the system is at the open-circle position of Figure 1. We set the transition probability between virtual states $v_i$ and $v_{i\pm 1}$ as shown below.

$$P_i(E_R, v_i \rightarrow v_{i\pm 1}) = \begin{cases} 1 & (E_R \in Z_{i\pm 1}) \\ 0 & (\text{otherwise}) \end{cases}$$

Equation (7) establishes that the transition occurs unconditionally. However, the probability can be set arbitrarily in general. We operate the virtual-state transition once every $N_{int}$ steps. Then, $r$ moves according to MD scheme from the $i$th to the $(i + N_{int})$th step without changing the virtual state $v_i$. Consequently, $v_i$ might transition to $v_{i-1}$ or $v_{i+1}$ at the $(i + N_{int})$th step without changing $r$. The actual value for the transition interval $N_{int}$ is given later. Because of eq. (7), the energy distribution for each virtual state converges to a constant.

$$P_{vmc}(E_R, v_i) = \begin{cases} \text{const} & (\text{in zone } Z_i) \\ 0 & (\text{outside of } Z_i) \end{cases}$$

### Materials and Methods

We constructed four fragment systems of p53 CTD, all of which consist of 17 amino-acid residues (residues 372–388 in UniProt\cite{35}) and amino acid sequences in single-letter codes as (i) Ace-KKGQTSRK--KKLMFKTE-NH2, (ii) Ace-KKGQTSRK--K-AKLMFKTE-NH2, (iii) Ace-KKGQTSRHKKLMFKTE-NH2, and (iv) Ace-KKGQTSRHKKLMFKTE-NH2, which are k, H\(^+\), Ace, and NH2, respectively, represent acetylated lysine, positively charged histidine, acetyl cap, and amine cap. As described in this article, we refer, respectively, to these fragments as (i) NonAc(H\(^+\)), (ii) Ac(H\(^+\)), (iii) NonAc, and (iv) Ac. These fragments include in common the binding regions (residues 380–386) to the four partners S100B, Sir2, CBP, and Cyclin A. We designate the regions as “common binding regions” in this study, which are shown as underlined in the sequences above. Remember that Table 1 presents correspondence between the fragment and its binding partner molecule.

We put each of the fragments in a periodic box (50 \(\times\) 50 \(\times\) 50 \(\AA\)\(^3\)) filled with water molecules and ions, where the number of ions is set so that the net charge of the whole system is zero and the ionic concentration is set to a physiological salt one. The conformations of the CTD fragments were the helical structure taken from the S100B-CTD complex. As explained later, these helical structures were randomized quickly in a preparative canonical MD simulation at a high temperature. Table 2 presents characteristics of the simulation systems.

Before V-McMD simulations, we performed a constant-pressure (NPT) simulation at 300 K and 1 atm to ascertain the periodic box size for each system. The resultant box sizes are presented in Table 2. Then, we conducted a long high-temperature (600 K) constant-volume (NVT) simulation (time step: 2.0 fs) to randomize the helical conformation of each system.

Generally, DOS ($\rho_{int}(E_R)$) is required to perform multicanonical sampling [see eq. (2)]. However, DOS is unknown a priori. Consequently, to begin with, we approximated DOS by performing conventional MD runs (i.e., canonical MD runs) at different temperatures covering a wide temperature range [280 K–600 K]. Then, as reported in the literature,\cite{33} the resultant canonical energy distributions at the various temperatures are integrated to approximate DOS, which is used to define $E_{vmc}$ for the first V-McMD simulations [see eq. (5)]. DOS was estimated in the range of 280 K to 600 K. Therefore, the subsequently performed V-McMD simulations aim to obtain a flat energy distribution [eq. (6)] in this range. The upper temperature limit (600 K) was set so that the fragment overcame various energy barriers. The Results and Discussion section shows that various conformations were sampled. The lower limit (280 K) was lower than a room temperature (300 K). Therefore, the obtained conformational ensemble involved conformations probable at 300 K.

The first V-McMD simulation was performed with using $E_{vmc}$ obtained above, where 32 runs were executed in parallel starting from the randomized conformations sampled from the high-temperature simulation. Therefore, we used the TTP procedure to perform V-McMD\cite{31,32} although we do not explicitly use the term “TTP-V-McMD” in this article. After the first V-McMD simulation, we updated $E_{vmc}$ according to the method presented in an earlier report,\cite{33} and performed the second V-McMD simulation with using the updated $E_{vmc}$, and so on. The initial conformation for one of the 32 runs in the 4th iteration is the last snapshot for the run in the $(i-1)$th iteration. We repeated this iteration procedure until the energy distribution converges to a function flat sufficient: $P_{vmc} \approx$ const. Then, the final iteration is the production run to collect snapshots for analyses. The numbers of iterations were 8 for Non-Ac(H\(^+\)), 9 for Ac(H\(^+\)), 8 for NonAc, and 8 for Ac, where the length of the production run was 320 ns for all the systems.

![Table 2. Simulation systems.](Image)
state zones \( Z \) for each of iterations. One might consider that the simulation length of 320 ns for the production run is too short to obtain a statistically significant ensemble. However, the enhanced sampling method has higher efficiency than conventional sampling does. Later, we discuss statistical properties of the resultant ensembles.

We used a computer program psygene-\( G^2 \)[36] for V-McMD with the SHAKE[37] method to fix the covalent-bond length related to hydrogen atoms, the zero dipole summation method[38–40] to calculate long-range electrostatic interactions, the velocity scaling method[41] to control temperature, TIP3P model[42] for water molecule, and an Amber-based hybrid force field for p53 CTD.[43] The Amber-based hybrid force field is defined as \( E_{\text{hybrid}}(\omega) = (1 - \omega)E_{94} + \omega E_{96} \), where \( E_{94} \) and \( E_{96} \), respectively, denote param 94 and param 96 AMBER force fields[44,45], and where \( \omega \) (0 \( \leq \omega \leq 1 \)) is a mixture weight for \( E_{94} \) and \( E_{96} \). Kamiya et al. confirmed that a proper range for \( \omega \) is 0.45 \( \leq \omega \leq 0.95 \).[43] Ikebe et al. showed that the larger the value of \( \omega \), the smaller the helical propensity in the resultant conformational ensemble at 300 K.[46] In our previous simulation of an IDP system,[31] we set \( \omega = 0.75 \) and obtained results comparable to experimentally obtained results. However, in our preparative simulation of the present systems with setting \( \omega = 0.75 \), the resultant ensemble exhibited a considerably high helical content. Then, we set \( \omega = 0.80 \) for the current study. The MD time step was set to 2.0 fs for all the simulations.

We constructed the force-field parameters of acetylated lysine (aK) as follows: The dihedral angle parameters are the same as those in lysine in the Amber force field. The atomic partial charges of aK were from a force field (Ref: http://pc164.materials.uoi.gr/dpapageo/amberparams.php).

Conformational ensemble and free-energy landscape

The V-McMD simulation of each system produces a conformational ensemble that consists of snapshots of various energies. The statistical weight assigned to a snapshot of energy \( E_k \) at temperature \( T \) is equivalent to the canonical energy distribution \( P_2(\epsilon, \lambda, T) \) [eq. (1)]. We denote this ensemble as \( Q_{\text{SYS}} \), where the notation “SYS” specifies the computed system as \( Q_{\text{SYS}} = \text{NonAc(H)}, \text{Ac(H)}, \text{NonAc}, \text{or Ac} \). Accordingly, the statistical weight of each system is denoted as \( \rho_{\text{SYS}}(\epsilon, T) \). The summation of the four ensembles is denoted as \( Q_{\text{sum}}: Q_{\text{sum}} = \sum_{\text{SYS}} Q_{\text{SYS}} \). In this study, we set \( T = 300 \) K to prepare \( Q_{\text{SYS}} \), although we do not mention explicitly that the statistical weight is set at 300 K.

To analyze the conformational ensemble, we generate a two-dimensional (2D) free-energy landscape using principal component analysis (PCA) as follows: First, we compute inter-\( C_\alpha \) atomic distances for each snapshot in \( Q_{\text{sum}} \), and define a vector \( \mathbf{q} = (q_1, q_2, \ldots, q_{N_{\text{pair}}}) \), where \( q_i \) is a distance for a C\( \alpha \) atomic pair and \( N_{\text{pair}} \) is the number of the pairs. Remember that the number of residues is 17 for all four systems. Consequently, \( N_{\text{pair}} \) is 136 (=17 \( \times \) (17-1)/2) for all systems. Then, we calculated a variance–covariance matrix \( A \) with elements \( (i,j) \) expressed as \( A_{ij} = \langle q_i q_j \rangle - \langle q_i \rangle \langle q_j \rangle \), where brackets are the ensemble average over conformations in \( Q_{\text{sum}} \). Diagonalizing this matrix, we obtained \( N_{\text{pair}} \) eigenvectors \( (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{N_{\text{pair}}}) \) and eigenvalues \( (\lambda_1, \lambda_2, \ldots, \lambda_{N_{\text{pair}}}) \), where \( \mathbf{v}_i \) and \( \lambda_i \) are paired satisfying an equation \( \mathbf{A} \mathbf{v}_i = \lambda_i \mathbf{v}_i \). The eigenvectors satisfy an orthogonal and normalized relation: \( \mathbf{v}_i \cdot \mathbf{v}_j = \delta_{ij} \). We presume that the eigenvalues are arranged in descending order as \( \lambda_1 > \lambda_2 > \ldots \).

We use \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \) to construct the 2D space (2D PCA space) by setting the coordinate axes to \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \), and to generate a conformational distribution by projecting conformations in \( Q_{\text{SYS}} \) to the 2D PCA space. The coordinate axis \( \mathbf{v}_i \) is designated as a principal component (PC) axis \( i \). We denote the \( k \)th conformation in \( Q_{\text{SYS}} \) as \( \mathbf{q}^{(k)} = (q_1^{(k)}, q_2^{(k)}, \ldots, q_{N_{\text{pair}}}^{(k)}) \). Then, the projection of \( \mathbf{q}^{(k)} \) to the axis \( \mathbf{v}_i \) is done by a scalar product: \( x_{PC}^{(k)} = \mathbf{q}^{(k)} \cdot \mathbf{v}_i \). The position of \( \mathbf{q}^{(k)} \) in the 2D PCA space is given by 2D coordinates \( (x_{PC1}^{(k)}, x_{PC2}^{(k)}) \). Repeating this procedure for all conformations in \( Q_{\text{SYS}} \), we obtain the distribution of conformations in the 2D PCA space: \( \rho_{\text{SYS}}(x_{PC1}, x_{PC2}, T) \), where conformation of energy \( E_R \) contributes to the distribution with the weight \( \rho_{\text{SYS}}(E_R, T) \). Finally, the potential of mean force (PMF) is computed as \( F_{\text{PMF}}(x_{PC1}, x_{PC2}, T) = -R_{\text{T2} \text{H}_2} \ln \rho_{\text{SYS}}(x_{PC1}, x_{PC2}, T) \) with spatial patterns that are called the free-energy landscape (FEL) in this study. From comparison of the spatial patterns of FEL among the four systems, we can discuss the difference of the architecture of FEL among the systems.

The ratio of contribution from the PC component \( i \) to the whole standard deviation is expressed as

\[
\Delta_{i} = \frac{\lambda_i}{\sum_{i=1}^{N_{\text{pair}}} \lambda_i}
\]

In PCA, the larger the eigenvalue assigned to the PC components \( i \) becomes, the greater the contribution: \( \Delta_{1} > \Delta_{2} > \ldots \). Therefore, the 2D PCA space constructed by \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \) is suitable to overview the conformational distribution. The contribution ratio by the PC components 1 and 2 is given simply as \( \Delta_{1} + \Delta_{2} \).

Synthesis of the nonacetyl and acetyl CTD fragments

CTD fragment with nonacetyl lysine (K382) was synthesized by the 9-fluorenylmethoxycarbonyl (Fmoc) method at a 0.05 mmol scale using a peptide synthesizer (Liberty Blue; CEM Corp., NC). After the completion of the peptide chain assembly, the obtained resin was treated with TFA containing 2.5% trisopropylsilane and 2.5% distilled water for 2 h. The crude peptide was concentrated by a nitrogen stream, precipitated by ether and purified by the reversed-phase HPLC to obtain nonacetyl peptide. ESI mass, found: 1038.7; calcd. for \([M + 2H]^2+\) 1038.8.

The synthesis of CTD fragment with acetyl-lysine (aK382) was also performed by the synthesizer, except that Lyt[11] was introduced using Fmoc-Lys(Aloc)-OH (Aloc: allyloxycarbonyl), activated by O-benzotriazol-1-yl-N,N',N''-tetramethyluronium hexafluorophosphate (HBTU)/N,N-diisopropylethylamine (DIEA). After the completion of the chain elongation, the solution of tetrakis(triphenylphosphine)palladium(0), phenylsilane, acetic anhydride in dichloromethane was added to the resin for 45
min to acetylate the side chain amino group of Lys selectively [11]. The deprotection and purification were performed in the same manner to obtain Ac peptide. ESI mass, found: 1059.7, calcd. for [M + 2H]2+: 1059.8.

Far-UV circular dichroism measurements

The nonacetyl and acetyl CTD fragments were dissolved in Milli-Q water at 200 μM. The peptide solutions were diluted to 40 μM with the desired concentration of trifluoroethanol (TFE) for circular dichroism (CD) measurements. To detect the helix propensity, we varied TFE concentrations of 0–50%. It is noteworthy that the TFE concentrations in this article are volume per volume percentages. The peptide solutions also contained 25 mM sodium phosphate (pH 7.0 without TFE) and 0.1 M sodium chloride. Because pH is 7.0, the histidine state is neutral. Consequently, the fragments treated in this CD measurement are also designated as NonAc and Ac.

We performed far-UV CD measurements of the peptide solutions. Far-UV CD spectra were recorded on a spectropolarimeter (JASCO J-820; Jasco Corp., Japan) at 25°C using a quartz cuvette with 1-mm path length. The spectra were expressed as the mean residue ellipticity and [θ] (deg cm² dmol⁻¹). All spectra were estimated iteratively 16 times. Furthermore, this procedure was repeated three times to compute the average and standard deviation of the spectrum.

Results and Discussion

The V-McMD production run generated conformational ensemble Q_SYS for the four CTD systems (SYS = NonAc(H⁺), Ac(H⁺), NonAc, and Ac) in the unbound state. Figure 2 demonstrates the flat distributions from the production runs. The flatness ensures that eq. (6) was satisfied (DOS was estimated accurately). Therefore, the sampling has been done with statistical significance. Below, we analyze effects of the acetylation of K382 and the charge neutralization of H380 on the free-energy landscape and secondary structure contents. Furthermore, we discuss possible binding mechanisms of the four CTD fragments to their partner molecules. Last, we again check the statistical significance of the resultant ensemble by demonstrating the convergence of FEL.

Free energy landscape of the full-length fragments

Figure 3 shows FELs at 300 K constructed in the 2D PCA space. The contribution ratios r₁ and r₂ from Q>SYS were, respectively, 41.4% and 18.8%. Then, the contributions from the PC 1 and 2 axes (r₁ + r₂) were 60.2%. In Figure 3, we refer to the clusters as G>SYS, where superscript SYS specifies the computed system and the subscript k is a label assigned to the clusters. Figure 4 demonstrates representative tertiary structures in each cluster. In all panels, G>SYS is assigned to the cluster of the global minimum PMF, which corresponds to a nearly complete helix (see structures in G>SYS in Fig. 4) located at the same position in the 2D PCA space. Figure 3 manifests that the clusters can transition mutually at 300 K: the free-energy barriers among the clusters are surmountable at 300 K, except for the cluster G>NonAc(H⁺). In other words, the CTD fragments are disordered.

All FELs involved not only the complete-helix cluster (G>SYS) but also partially helical ones. The tertiary structures are shown in Figure 4a, in G>NonAc(H⁺) and G>Ac(H⁺) in Figure 4b, in G>NonAc, G>NonAc, G>Ac, and G>Ac in Figure 4c, and in G>Ac, G>Ac, and

Figure 2. Flat energy distribution for the a) NonAc(H⁺), b) Ac(H⁺), c) NonAc, and d) Ac systems. Individual distributions (P(vmcm(E_R, vi), i = 1, ..., n_v) for the virtual states are integrated into the shown distribution P_mc(E_R) using the method presented in an earlier report. [33]
$G_{3}^{Ac}$ in Figure 4d. The $\beta$-hairpins are also found as $G_{3}^{NonAc(H^{+})}$ in Figure 4a, as $G_{3}^{Ac(H^{+})}$ in Figure 4b (this is a distorted hairpin-like structure), as $G_{3}^{NonAc}$, $G_{3}^{NonAc}$, and $G_{5}^{NonAc}$ in Figure 4c, and as $G_{5}^{Ac}$ in Figure 4d.

Although all the CTD fragments exhibited conformational diversity, the FEL shape is considerably different as shown in Figure 3. Remarkable differences are apparent for regions indicated by the solid-line and broken-line circles in the figure. Comparison of FELs between Figures 3a and 3b as well as between Figures 3c and 3d clarifies that the acetylation of K382 diminishes the probability in the region by a solid-line circle. We find that extended conformations are distributed in the solid-line circled region. It is likely that the acetylation facilitates hydrophobic-core formation by making the CTD fragment compact, which results in disappearance of the extended conformations. To verify this expectation, we calculated a radius of gyration of the fragments at 300 K only using hydrophobic atoms in the CTD fragment: $C_{b}, C_{c},$ and $C_{d}$ atoms of K372, K373, K381, K382, and K386; $C_{b}$ atom of H380; $C_{b}$ atom of L383; $C_{b}$ atom of M384; and $C_{b}$ atom of P385. The side-chain tip of the acetylated K382 in the $Ac(H^{+})$ and Ac systems was excluded from the computation of radius of gyration for the strict comparison because the side-chain tip does not exist in NonAc($H^{+}$) and NonAc. Table 3 presents the radius of gyration values ($R_g$), which demonstrates that the acetylation induces a compact hydrophobic core. We discuss this point further by viewing the tertiary structures of the CTD fragments below.

We computed the secondary-structure propensity of each residue in $Q_{SYS}$ using the DSSP program. Figure 5 depicts the secondary-structure contents along the sequence at 300 K. Comparison of Figures 5a and 5b reveals that K382 acetylation induces the helix propensity of the CTD fragment. This tendency is also apparent from comparison of Figures 5c and 5d. Furthermore, Table 4 presents the helix increment by the acetylation quantitatively.

Figure 6a presents a tertiary structure taken from the cluster $G_{3}^{Ac}$ ($H^{+}$), where the $aK382$ and L383 side-chains form a hydrophobic contact in the helix. Similarly, a conformation taken from $G_{5}^{Ac}$ shows that a hydrophobic contact is formed between $aK382$ and S378 (Fig. 6b). Figure 6c shows a hydrophobic contact between the side-chain stems of $aK382$ and K381 in the partially helical conformation taken from $G_{2}^{Ac}$ ($H^{+}$). Unless K382 is acetylated, a repulsive force acts between these two lysine residues. Furthermore, in this conformation, the oxygen atom in the $aK382$ side-chain and the nitrogen atom in the K381 side-chain interact electro-statistically. Figure 6d presents a conformation taken from $G_{2}^{Ac}$, where a hydrophobic core is formed by side-chain tips of $aK382$ and M384 and side-chain stem of R379. Consequently, the acetylation induces the hydrophobic core formation in helices. Then, the radius of gyration becomes small.
charged residues. Consequently, acetyl-lysine stabilizes helical structures more than nonacetyl lysine does.

As described above, comparison between Figures 3a and 3c as well as between Figures 3b and 3d clarified that the charge neutralization of H380 increases the probability in the broken-line circled region, where hairpin structures are distributed. Figure 5 and Table 4 also show that the charge neutralization enhances the hairpin formation. Figure 7a portrays a hairpin taken from cluster $G_{NonAc}^4$, where H380 and K381 form a hydrogen bond. Repulsive interaction acts between H380 and K381 if H380 is positively charged. Similarly, Figure 7b displays a distorted hairpin taken from $G_{NonAc}^4$, where the side-chain of H380 and main-chain of K382 form a hydrogen bond. If H380 is positively charged, then this distorted hairpin becomes unstable because a repulsive interaction between the positively charged H380 and K382 might break the hydrogen bond. Furthermore, a repulsive interaction between the positively charged H380 and R379 might also destabilize the $\beta$-hairpin structure. Consequently, the charge neutralization of H380 is necessary for stabilizing the hairpins in Figures 7a and 7b. Figure 7c displays a hairpin from $G_{Ac}^3$, where H380 and T377 form a hydrogen bond. These tertiary structures exemplify that the charge neutralized H380 serves the hydrogen bonds to stabilize turns in the hairpins.

Data shown in Figure 3 suggest visually that the Ac(H$^+$) system might have the narrowest structural varieties among the four systems. To elucidate this feature quantitatively, we computed the standard deviation $\sigma_{SYS}$ of the conformational distribution for Cz atoms for each system as

$$\sigma_{SYS} = \left[ \sum_i (q_i^2 - \langle q_i^2 \rangle) \right]^{1/2},$$

(10)

where the brackets are the ensemble average over conformations in each ensemble $Q_{SYS}$ weighted at 300 K. The resultant values are: $\sigma_{NonAc(H^+)} = 37.6$ Å, $\sigma_{Ac(H^+)} = 30.6$ Å, $\sigma_{NonAc} = 36.5$ Å, and $\sigma_{Ac} = 36.9$ Å. These values of structural fluctuations are consistent with the radii of gyration in Table 3. Consequently, the Ac(H$^+$) system has the smallest standard deviation. This reduction of the broadening results from the acetylation of K382. One might expect the Ac system to have a narrow

**Table 3.** Radius of gyration computed from selected hydrophobic atoms at 300 K.

|          | NonAc(H$^+$) | Ac(H$^+$) |
|----------|--------------|-----------|
| $R_g$ [Å] | 8.6 ± 1.2    | 8.2 ± 0.8 |
|          | NonAc       | Ac        |
| $R_g$ [Å] | 8.6 ± 1.0    | 8.1 ± 1.9 |
distribution because K382 is also acetylated in this system. However, as shown above, the neutralization of H380 induces hairpins, which prevents reduction of the distribution.

A temperature replica exchange MD simulation\cite{24,26} of a 14-residue p53 CTD fragment and a conventional MD simulation\cite{24,26} of a 15-residue p53 CTD fragment were performed to obtain a conformational ensemble in the unbound state. These two fragments are fully included in our 17-residue fragment, and H380 and K382 were positively charged, respectively, and K382 was nonacetylated. Consequently, those fragments are parts of the NonAc(H\textsuperscript{+}) fragment, and binds to the S100B molecule with adoption of a helical conformation. It is particularly interesting that, in these studies, the ensembles involved a helical fraction, which is consistent with our result for \(Q_{\text{NonAc}}\). In contrast, recent CD measurement of two 32-residue p53 CTD fragments, which involve our NonAc and Ac segments, showed that the conformations of the CTD fragments are randomized.\cite{48} Our computational results demonstrated that both \(Q_{\text{NonAc}}\) and \(Q_{\text{Ac}}\) contain a helical fraction and that the helical content of \(Q_{\text{Ac}}\) is larger than that of \(Q_{\text{NonAc}}\). This apparent inconsistency between the computational results and the CD experimental observation should be analyzed. We note, however, that the fragment length for the CD experiment is considerably longer than ours.

To link the computation and experiment, we conducted CD experiments of the NonAc and Ac fragments of 17 residues long. TFE enhances formation of secondary structures, especially of helix.\cite{49,50} To clarify the inherent helix propensity of the two fragments, the measurement was done at various TFE concentrations. CD spectra at zero TFE concentration have suggested that the overall structural feature of both fragments is characterized by a disorder state (data not shown), which is consistent to the preceding CD measurement.\cite{48} Therefore, the helical contents obtained from the simulations were larger than those from the CD experiments. In fact, although CD experiments are useful to discuss the secondary-structure properties of polypeptide qualitatively, the CD data might involve quantitative ambiguity in assessing the secondary-structure contents. Conversely, the simulation data might involve some errors. Therefore, we compare the simulation data with the CD data qualitatively. Figure 8 shows that the Ac fragment has a higher helix contents than the NonAc fragment at all examined TFE concentrations including the zero TFE concentration. This result was also supported by analysis of the CD spectra using "Bestsel" software\cite{51} (data not shown), which estimates the secondary-structure contents from CD spectra. Therefore, we conclude from both computation and

| Table 4. Secondary-structure contents at 300 K for entire (17-residue) CTD fragments. |
|-----------------|-----------------|-----------------|
|                 | Helix (%)       | Sheet (%)       | Other (%)       |
| NonAc(H\textsuperscript{+}) | 39.1            | 5.1             | 55.8            |
| Ac(H\textsuperscript{+}) | 51.1            | 1.6             | 47.3            |
| NonAc | 36.3            | 12.4            | 51.2            |
| Ac | 41.2            | 5.9             | 53.0            |

Figure 5. Secondary-structure contents along CTD sequence at 300 K: (a) NonAc(H\textsuperscript{+}), (b) Ac(H\textsuperscript{+}), (c) NonAc, and (d) Ac. Secondary structures are classified into three types: "helix" for \(\alpha\)-helix, \(\pi\)-helix, or 3-10 helix; "sheet" for \(\beta\)-bridge or extended \(\beta\) strand, and "other" for else. Red, blue, and black lines, respectively, present contents of helix, sheet, and other. [Color figure can be viewed at wileyonlinelibrary.com]
experimentation that the acetylation of K382 enhances the helix formation slightly.

As described earlier, NonAc(H\(^{+}\)) binds to S100B with adopting helical conformation and Ac(H\(^{+}\)) does not adopt a helical conformation to bind to a partner (Table 1). This experimentally obtained result might be inconsistent to the computational result that \(Q_{\text{NonAc}}(H^{+})\) contains a smaller helix content than \(Q_{\text{Ac}}(H^{+})\). We discuss possible binding mechanisms of the CTD fragments to their partner molecules in the next section.

### Free energy landscape of the common binding region

As described in the Introduction section, the p53 CTD has a hub property. Four CTD-partner complex structures were determined.

As described in the Materials and methods section, the residues 380–386 of CTD are the common binding region to all the four partners S100B, Sir2, CBP, and Cyclin A. We computed the 2D FEL for this common binding region (Fig. 9), where the variance-covariance matrix was computed only for the common binding region. This figure shows that various clusters are distributed in the 2D PCA space for all the ensembles. The contribution ratios are \(r_{c1} = 73.5\%\) and \(r_{c2} = 14.4\%\); then \(r_{c1} + r_{c2} = 87.9\%\).

We refer to the clusters as \(S_k^{\text{SYS}}\), where superscript SYS specifies the computed system and subscript \(k\) is a label assigned to clusters in Figure 9. Cluster \(S_1^{\text{SYS}}\) is assigned to the global minimum of PMF in all panels along with FEL for the full-length fragments. Figure 10 presents representative tertiary structures.
Figure 9. FEL, $F^{SYS}$, at 300 K for the common binding region (residues 380–386) constructed in 2D PCA space: a) $F^{NonAc(H^+)}$, b) $F^{Ac(H^+)}$, c) $F^{NonAc}$, and d) $F^{Ac}$. The lowest PMF site, $r_{min}$, is set as $F^{SYS}(r_{min})=0.0$ kcal/mol. Clusters $S_{1}^{SYS}$ are shown in FEL. Black filled circle, blue triangle, black filled square, and magenta colored diamond, respectively, denote positions of the bound forms $S_{1100B}$, $S_{11}$, $S_{CBP}$, and $S_{CBP}$. Table 1 presents the correspondence between the CTD-fragment type and the binding partner. PCA1 and PCA2 axes are computed from the entire ensemble (i.e., $Q_{un}$. Consequently, the PCA axes for all panels are common. [Color figure can be viewed at wileyonlinelibrary.com]
Figure 9 proposes possible mechanisms of CTD binding to their partner molecules. We infer that the main binding mechanism of NonAc(H\(^+\)) to S100B is the population selection because S\(_1\)\(^{\text{NonAc}(H^+)}\) is located at a fringe of the most stable cluster S\(_1\)\(^{\text{NonAc}(H^+)}\) (Fig. 9a). In other words, Q\(_{\text{NonAc}(H^+)}\) prepares the bound form in advance. Furthermore, because the free-energy barriers from the other clusters to S\(_1\)\(^{\text{NonAc}(H^+)}\) are low as described above, the bound form is recruited quickly when the bound form is exhausted to bind to S100B. The helical content of Q\(_{\text{NonAc}(H^+)}\) is smaller than that of Q\(_{\text{Ac}(H^+)}\). Figure 9a suggests that Q\(_{\text{NonAc}(H^+)}\) contains helical conformations sufficient for use for binding to S100B even if the helical content of Q\(_{\text{NonAc}(H^+)}\) is less than that of Q\(_{\text{Ac}(H^+)}\).

The population-selection mechanism might take place when the Ac fragment binds to CBP, where the conformations in the most stable cluster S\(_1\)\(^{\text{Ac}(H^+)}\) can transition readily to the conformation S\(_1\)\(^{\text{CBP}}\) (Fig. 9d). However, the cluster S\(_1\)\(^{\text{Ac}(H^+)}\) is not connected to the other clusters by low free-energy pathways. Consequently, the recruitment of conformations to S\(_1\)\(^{\text{Ac}(H^+)}\) from the other clusters might be slow. In other words, the rate constant for the Ac fragment binding to CBP might be smaller than that for the NonAc(H\(^+\)) binding to S100B if the population-selection mechanism occurs.

For binding of the Ac(H\(^+\)) fragment to Sir2, the conformation S\(_{\text{Sir2}}\) is located at a high free-energy site in Figure 9b.
Therefore, we presume that the fragment binds to Sir2 with a different conformation than $s_{Sir2}$, by which an encounter complex is formed. Then, the intermolecular interactions bring the fragment to the genuine complex structure. Therefore, we presume that the binding mechanism of this fragment belongs to the induced folding mechanism.

The NonAc fragment can bind to either S100B or Cyclin A (Table 1). The binding mechanism to S100B might be the population selection for the same reason for the NonAc($H^+$) fragment binding to S100B. The structure $s_{Cyclin}$ is located at a site with free energy of about 3 kcal/mol in Figure 9c. Therefore, a small fraction of the ensemble $Q_{NonAc}$ is a conformation close to $s_{Cyclin}$. Consequently, the binding mechanism of this fragment to Cyclin A might belong to the population selection. However, the main fraction of $Q_{NonAc}$ is far from $s_{Cyclin}$. Therefore, different conformations might be used to bind to Cyclin A. Consequently, the induced folding mechanism is also possible.

It is likely that the diversity of FEL modulated by the state variation of lysine and/or histidine induces the hub property of CTD. One can reasonably infer that different FELs have different interaction mechanisms to other molecules. A single protein segment can have multiple binding partners. This property is called a hub. It is noteworthy that the flexibility of the CTD segment is fundamentally important for the diversity of FEL. If CTD is a structurally well-defined portion of the protein, then FEL has no great diversity. Therefore, CTD might bind only to a single partner.

The binding mechanism proposed here is based only on the FEL of the unbound state, which means that no IDR-partner interactions are considered. To ascertain whether the proposed mechanism is correct or not, we should perform simulations of systems where CTDs and their partner molecules coexist, as in earlier studies. However, the current study is useful to investigate the variation of FEL in the presence or absence of the partner.

Finally, we confirmed the convergence of the sampled data. As reported in the convergence of sampling–section in Supporting Information, the convergence is good for all the systems.

Conclusions

To investigate a highly flexible biomolecular system, computational approaches are fundamentally important because experimental detection of large fluctuations at an atomistic resolution is still difficult. Because the high flexibility is an inherent property of IDR, investigation of the conformational ensemble is necessary to elucidate the nature of IDR. Therefore, a powerful conformational sampling method is required. We performed the enhanced conformational sampling method, V-McMD, to obtain the conformational ensembles of four p53 CTD fragments in the unbound state at the atomic resolution in an explicit solvent. Then, we constructed free-energy landscapes from the obtained conformational ensembles.

The shape of the free-energy landscape varied depending on the K382 acetylation and/or the H380 neutralization in CTD. It is particularly interesting that acetylation enhanced the helix propensity. This computational result was confirmed using CD experiments. We also demonstrated that acetylation induces the hydrophobic-core formation. The H380 neutralization has enhanced the hairpin formation of CTD. The helix content obtained from V-McMD tends to be larger than that from the CD experiment. This fact suggests that the force field is imperfect and that there have not been accurate force fields yet. Results from the CD experiment were explained by the McMD simulation with atomistic details. Therefore, we believe that our results are useful to discuss the variation of CTD’s conformational ensemble. Furthermore, the current results might assist in the generation of a general model for understanding the switching mechanism conducted by PTMs.

Each of the four CTD fragments has particular binding partner(s). We proposed possible binding mechanisms from the free-energy landscape of the unbound state. To judge whether the proposed mechanisms are correct or not, sampling of systems consisting of CTD and their partners is necessary for the next stage of research. However, as discussed in the Introduction, the binding mechanism of IDR is determined not only by the finally formed complex structure but also by the conformational distribution in the single-chain state. As discussed in Results, the spreading of the free-energy landscape and the free-energy barriers might affect the IDR-partner binding mechanism. Therefore, results of the current study of the unbound state are expected to be useful to investigate the variation of the free-energy landscape in the presence of the partner molecules. The results will provide useful knowledge to ascertain the hub property and coupled folding and binding of CTD.

Acknowledgments

I thank Dr. Narutoshi Kamiya whose instruction enables us to construct the force-field parameters for the acetylated lysine. We thank Prof. Jozsef Kardos (Eötvös Loránd University, Hungary) for helpful instruction for the computer program “Bestsel”.

Keywords: free energy landscape - p53 C-terminal - intrinsically disordered - post-translation modification - monocanonical

How to cite this article: S. Iida, T. Mashimo, T. Kurosawa, H. H. Hojo, H. Muta, Y. Goto, Y. Fukunishi, H. Nakamura, J. Higo. J. Comput. Chem. 2016, 37, 2687–2700. DOI: 10.1002/jcc.24494

Additional Supporting Information may be found in the online version of this article.

[1] V. N. Uversky, A. K. Dunker, Biochim. Biophys. Acta 2010, 1804, 1231.
[2] V. N. Uversky, C. J. Oldfield, A. K. Dunker, Annu. Rev. Biophys. 2008, 37, 215.
[3] A. Patil, K. Kinoshita, H. Nakamura, Int. J. Mol. Sci. 2010, 11, 1930.
[4] S. J. Metallo, Curr. Opin. Chem. Biol. 2010, 14, 481.
[5] F. Theillet, A. Binol, T. Frembgen-kesner, K. Hingorani, M. Sarkar, C. Kyne, C. Li, P. B. Crowley, L. Gierasch, G. J. Pielak, A. H. Elcock, A. Gershenson, P. Selenko, M. G. Hilliger, J. Mol. Biol. 2008, 377, 589.
[6] R. B. Berlow, H. J. Dyson, P. E. Wright, FEBS Lett. 2015, 589, 2433.
[7] K. Sugase, H. J. Dyson, P. E. Wright, Nature 2007, 447, 1021.
[8] J. Monod, J. Wyman, J. P. Changeux, J. Mol. Biol. 1965, 12, 88.
[9] K. Okazaki, S. Takada, Proc. Natl. Acad. Sci. USA 2008, 105, 11182.
