

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

Shinji Iida,^[a] Tadaaki Mashimo,^[b,c] Takashi Kurosawa,^[b,d] Hironobu Hojo,^[a] Hiroya Muta,^[a] Yuji Goto,^[a] Yoshifumi Fukunishi,^[b,e] Haruki Nakamura,^[a] and Junichi Higo^{*[a]}

The C-terminal domain (CTD) of tumor suppressor protein p53 is an intrinsically disordered region that binds to various partner proteins, where lysine of CTD is acetylated/nonacetylated and histidine neutralized/non-neutralized. Because of the flexibility of the unbound CTD, a free-energy landscape (FEL) is a useful quantity for determining its statistical properties. We conducted enhanced conformational sampling of CTD in the unbound state via virtual system coupled multicanonical molecular dynamics, in which the lysine was acetylated or nonacetylated and histidine was charged or neutralized. The fragments were expressed by an all-atom model and were immersed in an explicit solvent. The acetylation and charge-

neutralization varied FEL greatly, which might be convenient to exert a hub property. The acetylation slightly enhanced alpha-helix structures that are more compact than sheet/loop conformations. The charge-neutralization produced hairpins. Additionally, circular dichroism experiments confirmed the computational results. We propose possible binding mechanisms of CTD to partners by investigating FEL. © 2016 The Authors. Journal of Computational Chemistry Published by Wiley Periodicals, Inc.

DOI: 10.1002/jcc.24494

Introduction

An intrinsically disordered region (IDR) of a protein is highly flexible in physiological conditions unless interacting with its partner molecule.^[1] 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,^[2] where a single IDR interacts with different partner molecules to regulate the signal transduction. This multipartner interaction property is called a *hub* property.^[3] IDRs are related to some diseases such as cancer, diabetes, and neurodegenerative disease.^[4] Therefore, IDRs are an important subject related to biology, 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.^[5] The relation between the conformational ensemble and its biological functions has not been understood sufficiently.^[6] 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.^[2] 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.^[7] Two interaction mechanisms have been proposed for coupled folding and binding processes: induced folding^[8,9] and conformational selection.^[9,10] In induced folding, IDR

[a] S. Iida, H. Hojo, H. Muta, Y. Goto, H. Nakamura, J. Higo
Institute for Protein Research Osaka University, 3-2 Yamadaoka, Suita,
Osaka, 565-0871, Japan
E-mail: higo@protein.osaka-u.ac.jp

[b] T. Mashimo, T. Kurosawa, Y. Fukunishi
Technology Research Association for Next Generation Natural Products
Chemistry, 2-3-26 Aomi, Koto-Ku, Tokyo 135-0064, Japan

[c] T. Mashimo
IMS BIO Co, Ltd, Owl Tower 6F, 4-21-1, Higashi-ikebukuro, Toshima-ku,
Tokyo 170-0013, Japan

[d] T. Kurosawa
Hitachi Solutions East Japan, 21-1 Ekimaehoncho, Kawasaki-ku,
Kanagawa 210-0007, Japan

[e] Y. Fukunishi
Molecular Profiling Research Center for Drug Discovery (molprof), National
Institute of Advanced Industrial Science and Technology (AIST), 2-3-36,
Aomi, Koto-ku, Tokyo 135-0064, Japan

Contract grant sponsor: Japan Society for the Promotion of Science KAKENHI (to I. S. and H. N.), for Scientific Research on Innovative Areas; Contract grant number: 24118008; Contract grant sponsor: Ministry of Education, Culture, Sports, Science and Technology (MEXT) Japan (to J. H.); Contract grant number: 21113006; Contract grant sponsor: the Development of core technologies for innovative drug development based on IT from Japan Agency for Medical Research and Development, AMED [Scientific Research on Innovative Areas (to J. H.)]; Contract grant sponsor: TSUBAME2.5 supercomputer in the Tokyo Institute of Technology [HPCI System Research Project (Project IDs: hp150015)]

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2016 The Authors. Journal of Computational Chemistry Published by Wiley Periodicals, Inc.

Table 1. Binding partners of CTD fragments.

Partner	CTD fragment			
	NonAc(H ⁺)	Ac(H ⁺)	NonAc	Ac
S100B ^[a]	P ^[e]	N	P	N
Cyclin A ^[b]	N ^[f]	N	P	N
Sir2 ^[c]	N	P	N	N
CBP ^[d]	N	N	N	P

[a] Complex structures were referred from another report.^[18] Binding affinity of CTD to S100B was determined from an *in vitro* experiment.^[22] [b] Complex structures were referred from another report.^[21] [c] Complex structures were referred from another report.^[19] [d] Complex structures were referred from another report.^[20] [e] Mark "P" means that CTD can bind to the partner. [f] Mark "N" means that CTD cannot bind to the partner.

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.^[11] Another IDR-partner interaction picture is fuzzy-complex formation, by which IDR binds to its partner adopting multiple conformations.^[12] Therefore, in this mechanism, 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.^[13] 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.^[14] To date, three TAD-partner complex structures have been solved, where TAD adopts helical structures on surfaces of the partners in all three cases.^[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),^[18] a sheet to Sir2 (PDB ID: 1MA3),^[19] and fuzzy/coiled structures to CBP (PDB ID: 1JSP),^[20] and Cyclin A (PDB ID: 1H26).^[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.^[18,22] To Cyclin A, K382 is non-acetylated and H380 is neutralized.^[21] To Sir2, K382 is acetylated and H380 is positively charged.^[19] To CBP, K382 is acetylated

and H380 is neutralized.^[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.^[23] 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.^[24] McDowell et al. showed heterogeneity of CTD and the binding mechanism of CTD to S100B.^[25] Staneva et al. investigated conformational preferences of the unbound CTD using an implicit-water Monte Carlo simulation.^[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.^[27–29] Nakajima et al. introduced a multicanonical MD simulation (McMD) using Cartesian coordinates for dynamic variables. Adoption of Cartesian coordinates produced 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,^[30,31] and applied to systems consisting of an intrinsically disordered segment and its partner protein.^[11,32] Recently, a virtual system coupled McMD (V-McMD) was also developed to increase the McMD sampling efficiency.^[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.^[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 $\mathbf{f}_i = -\text{grad}_i E_R(\mathbf{r})$, where $E_R(\mathbf{r})$ is a potential energy of the

system as a function of the coordinates of the constituent atoms $\mathbf{r}=[\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N_{\text{atom}}}]$, and where N_{atom} is the number of atoms (polypeptide atoms plus solvent atoms) and $\mathbf{r}_i=[x_i, y_i, z_i]$ is the position vector of atom i , that is, the x , y , and z coordinates. An energy distribution obtained from the canonical MD converges to the following canonical distribution as

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

where R_{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 $Z_c = \int n_R(E_R) \exp[-E_R/R_{\text{gas}} T] dE_R$ 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

$$E_{\text{mc}}(E_R) = E_R + R_{\text{gas}} T \ln[P_c(E_R, T)] \\ = R_{\text{gas}} T \ln[n_R(E_R)]. \quad (2)$$

The force acting on the atom i is represented as

$$\mathbf{f}_i^{\text{mc}} = -\text{grad}_i E_{\text{mc}}. \quad (3)$$

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

$$P_{\text{mc}}(E_R) = Z_{\text{mc}}^{-1} n_R(E_R) \exp\left[-\frac{E_{\text{mc}}}{R_{\text{gas}} T}\right] \\ = Z_{\text{mc}}^{-1} n_R(E_R) \frac{n_R(E_R)}{n_R(E_R)} \\ = \text{const}, \quad (4)$$

where $Z_{\text{mc}} = \int n(E_R) \exp[-E_{\text{mc}}/R_{\text{gas}} T] dE_R$ 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 $P_{\text{mc}}(E_R)$ 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.^[33] 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 $E_{\text{tot}} = E_R + E_v$, 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 $n_{\text{tot}}(E_R, E_v) = n_R(E_R) n_v(E_v)$, 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_{vmc} of the entire system is given as

$$E_{\text{vmc}} = R_{\text{gas}} T \ln[n_{\text{tot}}(E_R, E_v)] \\ = R_{\text{gas}} T \ln[n_R(E_R)]. \quad (5)$$

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

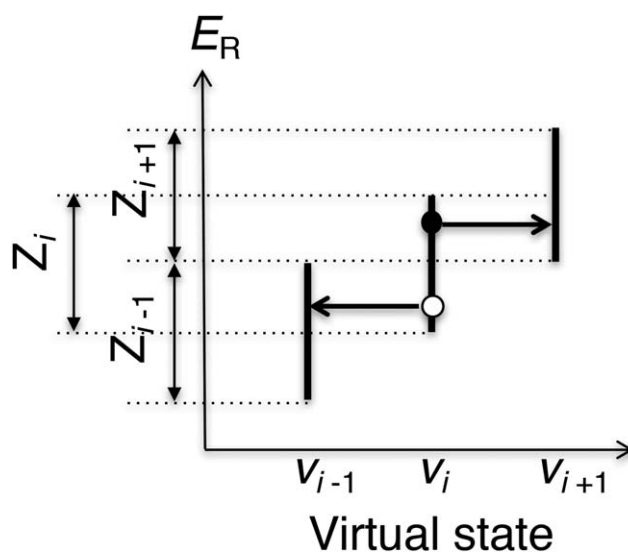


Figure 1. Transitions among adjacent virtual states. Presuming that the system is at the filled-circle position, where potential energy is E_R and the virtual state belongs to v_i , then, the virtual state might jump to v_{i+1} without changing atomic positions (i.e., without changing E_R). Its transition probability is given in eq. (7). However, when the system is at the open-circle position, the virtual state might transition to v_{i-1} .

$$P_{\text{vmc}}(E_R, E_v) = Z_{\text{vmc}}^{-1} n_{\text{tot}}(E_R, E_v) \exp\left[-\frac{E_{\text{vmc}}}{R_{\text{gas}} T}\right] \\ = Z_{\text{vmc}}^{-1} n_{\text{tot}}(E_R, E_v) \frac{n_R(E_R)}{n_R(E_R)} \\ = \text{const}, \quad (6)$$

where $Z_{\text{vmc}} = \int n_{\text{tot}}(E_R, E_v) \exp[-E_{\text{vmc}}/R_{\text{gas}} T] dE_R dE_v$ is a partition function of the entire system at T . As with eqs. (4) and (6) also ensures that the two-dimensional distribution $P_{\text{vmc}}(E_R, E_v)$ 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 $i = 1, \dots, n_{\text{vs}}$ and n_{vs} 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_{v_i} . 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_R is confined in a zone $Z_i = [E_R^{\text{min}}(v_i) < E_R < E_R^{\text{max}}(v_i)]$ when its virtual state is in v_i . The method for the confinement is explained in an earlier report^[33]. We set the zones so that Z_i overlaps with Z_{i-1} and Z_{i+1} , although Z_{i-1} and Z_{i+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_i in the virtual state v_i , then, this potential energy is also involved in Z_{i+1} . Therefore, the state v_i can transition to v_{i+1} using the Monte Carlo method. The molecular configuration \mathbf{r} is not changed (i.e., E_R is not altered) in this transition. The transition from v_i to v_{i-1}

Table 2. Simulation systems.

	System			
	NonAc(H ⁺)	Ac(H ⁺)	NonAc	Ac
Total atoms	12,196	12,196	12,197	12,197
Water molecules	3954	3953	3955	3954
Ions	Na: 11, Cl: 17	Na: 11, Cl: 16	Na: 11, Cl: 16	Na: 11, Cl: 15
Box size ^[a]	(49.3) ³	(49.2) ³	(49.1) ³	(49.2) ³

[a] Periodic box size is given in Å³ unit. Shown values are those obtained after NPT simulation at 300 K and 1 atm.

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.

$$P_t(E_R, v_i \rightarrow v_{i\pm 1}) = \begin{cases} 1 & (E_R \in Z_{i\pm 1}) \\ 0 & (\text{otherwise}) \end{cases} \quad (7)$$

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, \mathbf{r} moves according to MD scheme from the i th to the $(i+N_{\text{int}})$ th steps without changing the virtual state v_i . Consequently, v_i might transition to v_{i-1} or v_{i+1} at the $(i+N_{\text{int}})$ th step without changing \mathbf{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.

$$P_{\text{vmc}}(E_R, v_i) = \begin{cases} \text{const} & (\text{in zone } Z_i) \\ 0 & (\text{outside of } Z_i) \end{cases} \quad (8)$$

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^[35]) and amino acid sequences in single-letter codes as (i) Ace-KKGQSTSR-H⁺-KKL^{MF}KTE-NH₂, (ii) Ace-KKGQSTSR-H⁺-KaK-L^{MF}KTE-NH₂, (iii) Ace-KKGQSTSRHKKL^{MF}KTE-NH₂, and (iv) Ace-KKGQSTSRHK-aK-L^{MF}KTE-NH₂, where aK, 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 ($50 \times 50 \times 50$ Å³) 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 ($n_R(E_R)$) 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,^[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} ($=E_{\text{mc}}$) 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,^[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_{mc} according to the method presented in an earlier report,^[33] and performed the second V-McMD simulation with using the updated E_{mc} , and so on. The initial conformation for one of the 32 runs in the i th 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: $P_{\text{vmc}} \approx \text{const}$. Then, the final iteration is the production run to collect snapshots for analyses. The numbers of iterations were 8 for NonAc(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 S1 of Supporting Information presents the simulation lengths, inter-virtual state transition interval N_{int} , and virtual-

state zones Z_i 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^[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 ω ($0 \leq \omega \leq 1$) is a mixture weight for E_{94} and E_{96} . Kamiya et al. confirmed that a proper range for ω is $0.45 \leq \omega \leq 0.95$.^[43] Ikebe et al. showed that the larger the value of ω , the smaller the helical propensity in the resultant conformational ensemble at 300 K.^[46] In our previous simulation of an IDP system,^[11] 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 of lysine in the Amber force field. The atomic partial charges of aK were from a force field (Ref: <http://pc164.materials.uoi.gr/dpapego/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_R at temperature T is equivalent to the canonical energy distribution $P_c(E_R, T)$ [eq. (1)]. We denote this ensemble as Q_{SYS} , where the notation “SYS” specifies the computed system as $\text{SYS} = \text{NonAc}(\text{H}^+)$, $\text{Ac}(\text{H}^+)$, NonAc , or Ac . Accordingly, the statistical weight of each system is denoted as $P_c^{\text{SYS}}(E_R, T)$. The summation of the four ensembles is denoted as Q_{sum} : $Q_{\text{sum}} = \sum_{\text{SYS}} Q_{\text{SYS}}$. In this study, we set $T=300$ K to prepare Q_{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 α atomic distances for each snapshot in Q_{sum} , and define a vector as $\mathbf{q} = [q_1, q_2, \dots, q_{N_{\text{pair}}}]$, where q_i is a distance for a C α atomic pair and N_{pair} is the number of the pairs. Remember that the number of residues is 17 for all four systems. Consequently, N_{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_{sum} . Diagonalizing

this matrix, we obtained N_{pair} eigenvectors ($\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{N_{\text{pair}}}$) and eigenvalues ($\lambda_1, \lambda_2, \dots, \lambda_{N_{\text{pair}}}$), where \mathbf{v}_i and λ_i are paired satisfying an equation $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$. 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 $\lambda_1 > \lambda_2 > \dots$.

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_{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_{SYS} as $\mathbf{q}^{(k)} (= [q_1^{(k)}, q_2^{(k)}, \dots, q_{N_{\text{pair}}}^{(k)}])$. Then, the projection of $\mathbf{q}^{(k)}$ to the axis \mathbf{v}_i is done by a scalar product: $x_{\text{PC}i}^{(k)} = \mathbf{q}^{(k)} \cdot \mathbf{v}_i$ ($i=1, 2$). The position of $\mathbf{q}^{(k)}$ in the 2D PCA space is given by 2D coordinates $[x_{\text{PC}1}^{(k)}, x_{\text{PC}2}^{(k)}]$. Repeating this procedure for all conformations in Q_{SYS} , we obtain the distribution of conformations in the 2D PCA space: $P^{\text{SYS}}(x_{\text{PC}1}, x_{\text{PC}2}, T)$, where conformation of energy E_R contributes to the distribution with the weight $P_c^{\text{SYS}}(E_R, T)$. Finally, the potential of mean force (PMF) is computed as $F^{\text{SYS}}(x_{\text{PC}1}, x_{\text{PC}2}, T) \equiv -R_{\text{gas}} T \ln [P^{\text{SYS}}(x_{\text{PC}1}, x_{\text{PC}2}, T)]$ 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

$$rc_i = \frac{\lambda_i}{\sum_{i=1}^N \lambda_i} \quad (9)$$

In PCA, the larger the eigenvalue assigned to the PC components i becomes, the greater the contribution: $\text{SD}_1 > \text{SD}_2 > \dots$. 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 $rc_1 + rc_2$.

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% triisopropylsilane 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 Lys^[11] was introduced using Fmoc-Lys(Aloc)-OH (Aloc: allyloxycarbonyl), activated by *O*-benzotriazol-1-yl-*N,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

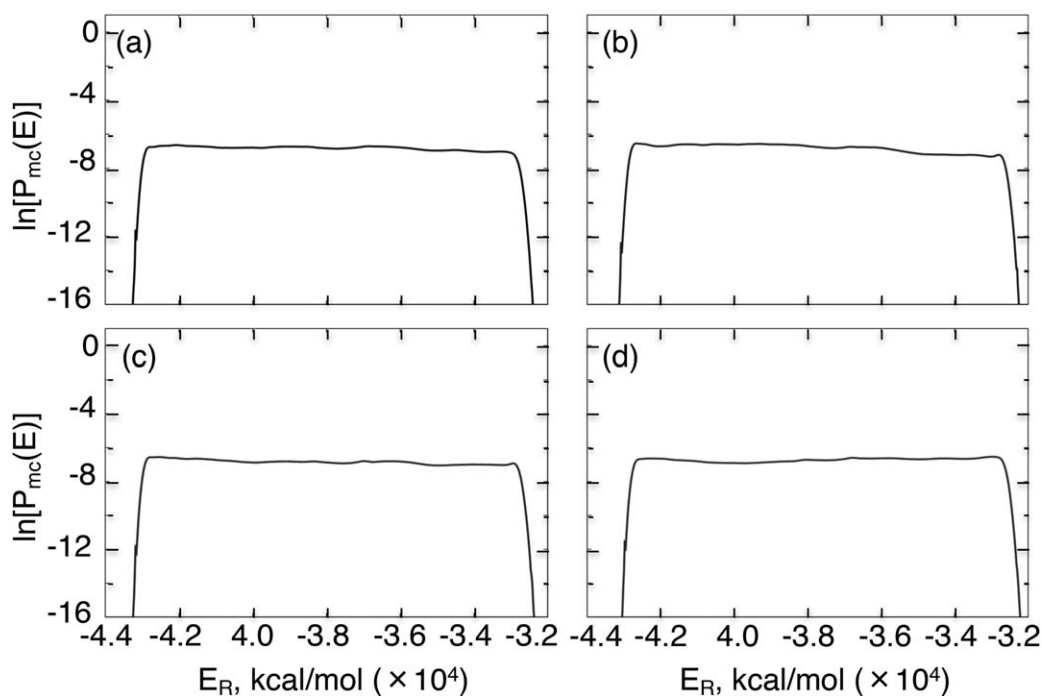


Figure 2. Flat energy distribution for the a) NonAc(H⁺), b) Ac(H⁺), c) NonAc, and d) Ac systems. Individual distributions ($P_{vmc}(E_R, v_i)$, $i=1, \dots, n_{vs}$) for the virtual states are integrated into the shown distribution $P_{mc}(E_R)$ using the method presented in an earlier report.^[33]

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 μ M. The peptide solutions were diluted to 40 μ M 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 $[\theta]$ (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 rc_1 and rc_2 from Q_{sum} were, respectively, 41.4% and 18.8%. Then, the contributions from the PC 1 and 2 axes ($rc_1 + rc_2$) were 60.2%. In Figure 3, we refer to the clusters as G_k^{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_1^{SYS} is assigned to the cluster of the global minimum PMF, which corresponds to a nearly complete helix (see structures in G_1^{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 $G_4^{NonAc(H+)}$. In other words, the CTD fragments are disordered.

All FELs involved not only the complete-helix cluster (G_1^{SYS}) but also partially helical ones. The tertiary structures are shown in $G_2^{NonAc(H+)}$, $G_6^{NonAc(H+)}$, $G_7^{NonAc(H+)}$, and $G_8^{NonAc(H+)}$ in Figure 4a, in $G_2^{Ac(H+)}$ and $G_4^{Ac(H+)}$ in Figure 4b, in G_2^{NonAc} , G_5^{NonAc} , G_6^{NonAc} , and G_7^{NonAc} in Figure 4c, and in G_4^A , G_5^A , and

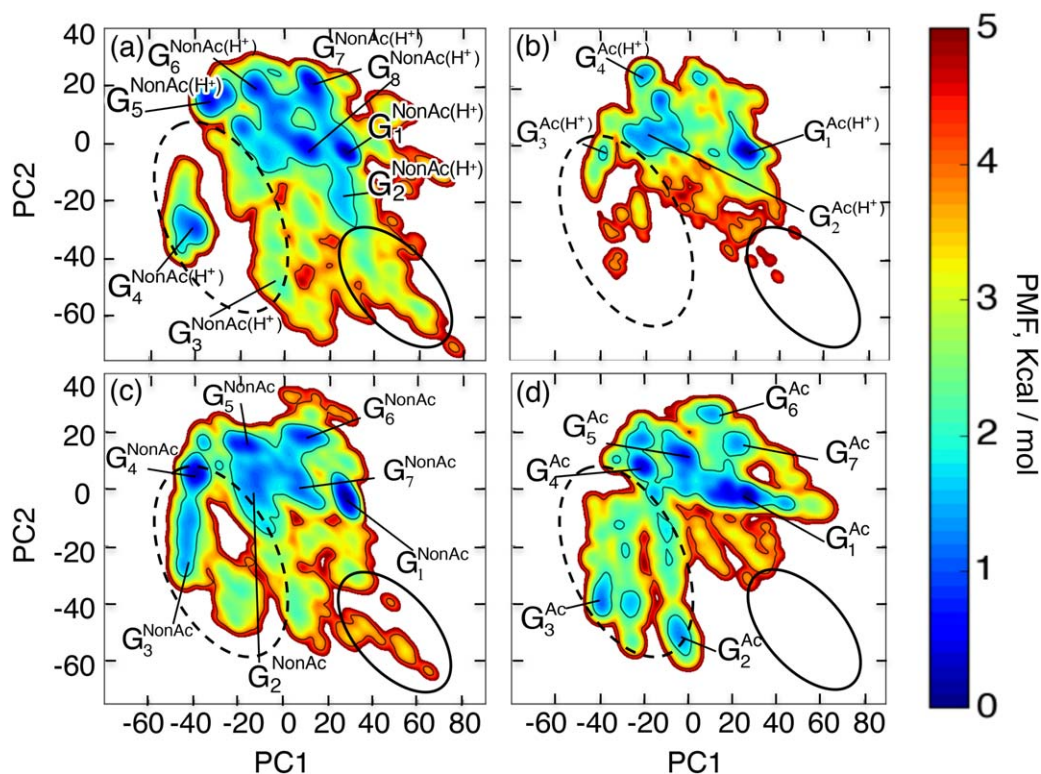


Figure 3. FEL, F^{SYS} , at 300 K for entire (17-residue) CTD fragments constructed in 2D PCA space: a) $F^{\text{NonAc(H}^+)}$, b) $F^{\text{Ac(H}^+)}$, c) F^{NonAc} , and d) F^{Ac} . We express a 2D site on F^{SYS} as r_m . The lowest PMF site, r_{min} , is set as $F^{\text{SYS}}(r_{\text{min}}) = 0.0$ kcal/mol. Clusters G_k^{SYS} are shown in FEL. See main text for details of the naming. PCA1 and PCA2 axes are computed from the entire ensemble (i.e., Q_{sum}). Consequently, the PCA axes for all the panels are common. [Color figure can be viewed at wileyonlinelibrary.com]

G_7^{Ac} in Figure 4d. The β -hairpins are also found as $G_4^{\text{NonAc(H}^+)}$ in Figure 4a, as $G_3^{\text{Ac(H}^+)}$ in Figure 4b (this is a distorted hairpin-like structure), as G_3^{NonAc} , G_4^{NonAc} , and G_5^{NonAc} in Figure 4c, and as G_3^{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_β , C_γ , and C_δ atoms of K372, K373, K381, K382, and K386; C_β atom of H380; C_β atom of L383; C_β atom of M384; and C_β atom of P385. The side-chain tip of the acetylated K382 in the $\text{Ac(H}^+)$ 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 $\text{NonAc(H}^+)$ 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.^[47] 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 $G_1^{\text{Ac(H}^+)}$, where the aK382 and L383 side-chains form a hydrophobic contact in the helix. Similarly, a conformation taken from G_1^{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 $G_2^{\text{Ac(H}^+)}$. 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_5^{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. If Lys382 is nonacetyl form, then repulsion interactions take place between Lys382 and the other positively charged residues in the helix because each sequence of the fragments includes six or seven positively

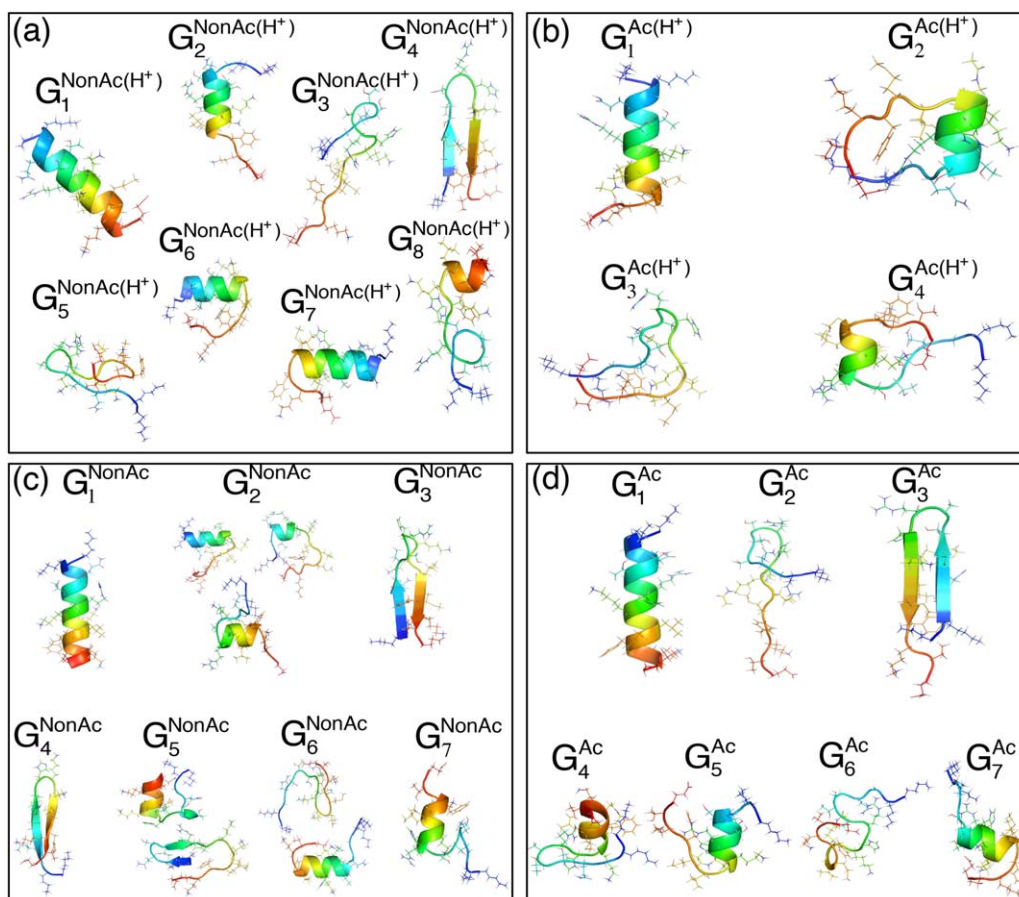


Figure 4. Structures in clusters G_k^{SYS} of FEL for entire CTD fragments: Tertiary structures picked from clusters G_k^{SYS} are colored by a rainbow (blue and red sides, respectively, show N and C-termini). [Color figure can be viewed at wileyonlinelibrary.com]

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_4^{NonAc} , 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_4^{NonAc} , 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 β -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_3^{Ac} , 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 $\text{Ac}(\text{H}^+)$ system might have the narrowest structural varieties among the four systems. To elucidate this feature quantitatively, we computed the standard deviation σ_{SYS} of the conformational distribution for $\text{C}\alpha$ atoms for each system as

$$\sigma_{\text{SYS}} = \left[\sum_i \left(\langle q_i^2 \rangle - \langle q_i \rangle^2 \right) \right]^{1/2}, \quad (10)$$

where the brackets are the ensemble average over conformations in each ensemble Q_{SYS} weighted at 300 K. The resultant values are: $\sigma_{\text{NonAc}(\text{H}^+)} = 37.6 \text{ \AA}$, $\sigma_{\text{Ac}(\text{H}^+)} = 30.6 \text{ \AA}$, $\sigma_{\text{NonAc}} = 36.5 \text{ \AA}$, and $\sigma_{\text{Ac}} = 36.9 \text{ \AA}$. These values of structural fluctuations are consistent with the radii of gyration in Table 3. Consequently, the $\text{Ac}(\text{H}^+)$ 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

	NonAc(H ⁺)	Ac(H ⁺)
Rg [Å]	8.6 ± 1.2	8.2 ± 0.8
	NonAc	Ac
Rg [Å]	8.6 ± 1.0	8.1 ± 1.9

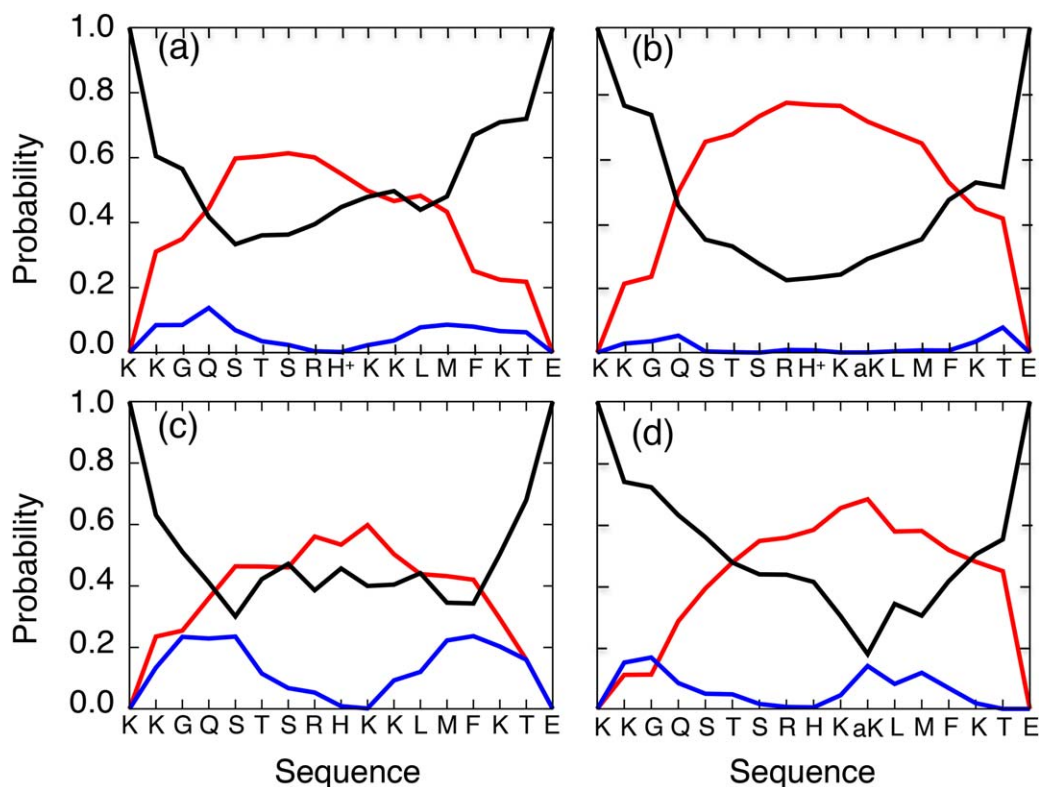


Figure 5. Secondary-structure contents along CTD sequence at 300 K: (a) NonAc(H⁺), (b) Ac(H⁺), (c) NonAc, and (d) Ac. Secondary structures are classified into three types: “helix” for α -helix, π -helix, or 3-10 helix; “sheet” for β -bridge or extended β strand, and “other” for else. Red, blue, and black lines, respectively, present contents of helix, sheet, and other. [Color figure can be viewed at wileyonlinelibrary.com]

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^[24,26] of a 14-residue p53 CTD fragment and a conventional MD simulation^[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⁺) 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 $Q_{\text{NonAc(H}^+)}$. 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.^[48] Our computational results demonstrated that both Q_{NonAc} and Q_{Ac} contain a helical fraction and that the helical

content of Q_{Ac} is larger than that of Q_{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.^[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.^[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^[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 ⁺)	39.1	5.1	55.8
Ac(H ⁺)	51.1	1.6	47.3
NonAc	36.3	12.4	51.2
Ac	41.2	5.9	53.0

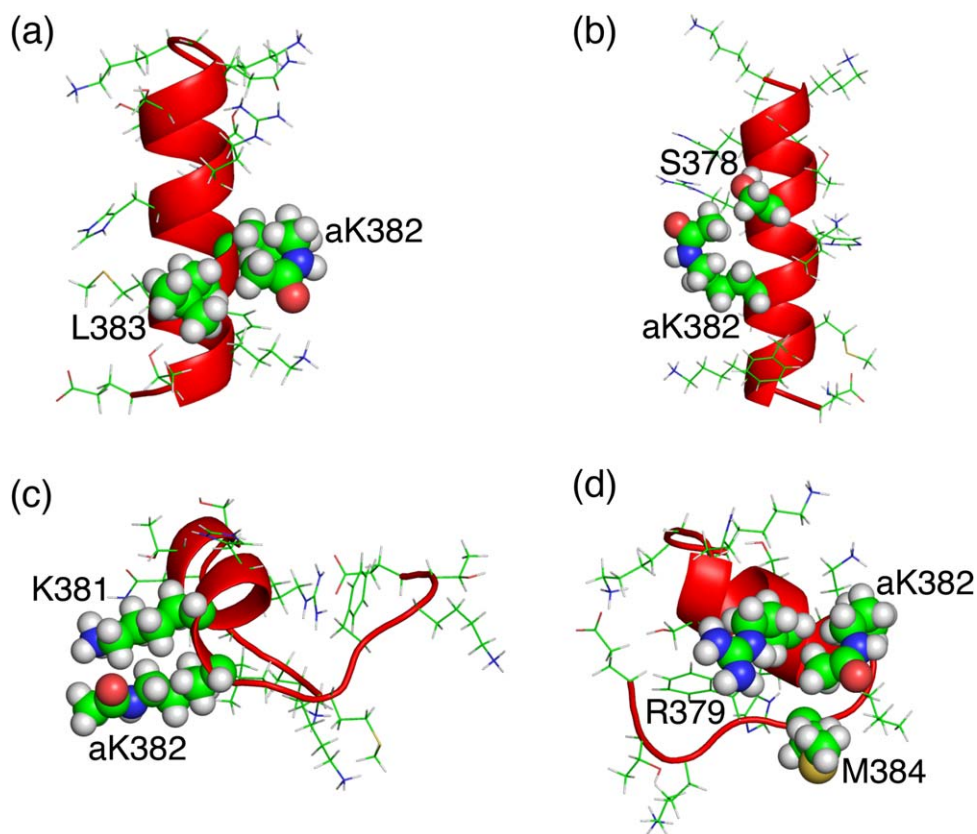


Figure 6. Tertiary structures taken from clusters a) $G_1^{Ac(H^+)}$, b) G_1^{Ac} , c) $G_2^{Ac(H^+)}$, and d) G_5^{Ac} . [Color figure can be viewed at wileyonlinelibrary.com]

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 $Q_{NonAc(H^+)}$ contains a smaller helix content than $Q_{Ac(H^+)}$. 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 $rc_1=73.5\%$ and $rc_2=14.4\%$; then $rc_1+rc_2=87.9\%$.

We refer to the clusters as S_k^{SYS} , where superscript SYS specifies the computed system and subscript k is a label assigned to clusters in Figure 9. Cluster S_1^{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

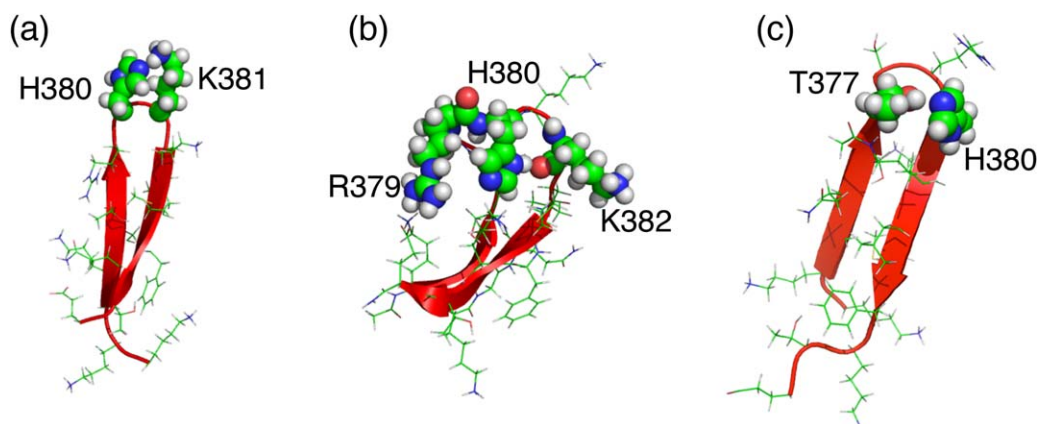


Figure 7. Tertiary structures taken from clusters a) G_4^{NonAc} , b) G_4^{NonAc} , and c) G_3^{Ac} . [Color figure can be viewed at wileyonlinelibrary.com]

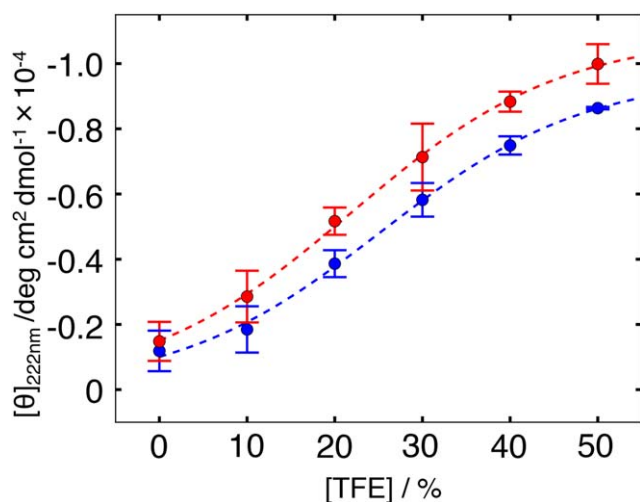


Figure 8. TFE-concentration dependence of mean molar ellipticity per residue, $[\theta]$, at 222 nm at 25°C for NonAc and Ac fragments. Blue and red points, respectively, present θ for the NonAc and Ac fragments. The smaller the value of θ at a TFE concentration, the greater the helix content at the concentration. Blue and red broken lines, respectively, stand for fitting lines of the points of NonAc and Ac by the sigmoidal function. [Color figure can be viewed at wileyonlinelibrary.com]

structures in each cluster. Again, this cluster corresponds to a helical cluster (see structures in S_1^{SYS} in Fig. 10) located at the same position in all panels.

Apparently, the ensemble $Q_{NonAc(H^+)}$ has the broadest distribution of the four ensembles. Many clusters are found in FEL (Fig. 9a). The tertiary structures taken from the clusters are diverse (Fig. 10a): The complete helix is $S_1^{NonAc(H^+)}$; partial helices are

$S_2^{NonAc(H^+)}$, $S_3^{NonAc(H^+)}$, $S_4^{NonAc(H^+)}$, $S_6^{NonAc(H^+)}$, $S_{10}^{NonAc(H^+)}$, $S_{11}^{NonAc(H^+)}$, and $S_{12}^{NonAc(H^+)}$; β hairpins are $S_7^{NonAc(H^+)}$ and $S_8^{NonAc(H^+)}$; and random-coiled structures are $S_5^{NonAc(H^+)}$ and $S_9^{NonAc(H^+)}$. Inter-cluster transitions can take place readily because free-energy barriers among the clusters are low.

Structural diversity decreases when K382 is acetylated and/or H380 is neutralized. In $Q_{Ac(H^+)}$, no β hairpin cluster exists, although three partial helix clusters ($S_1^{Ac(H^+)}$, $S_3^{Ac(H^+)}$, and $S_4^{Ac(H^+)}$) do (Fig. 9b). In Q_{NonAc} , only two helix clusters (S_1^{NonAc} and S_2^{NonAc}) and three β hairpin clusters (S_3^{NonAc} , S_4^{NonAc} , and S_5^{NonAc}) exist (Fig. 9c). In Q_{Ac} , two helix clusters (S_1^{Ac} and S_2^{Ac}) exist, but no hairpin cluster exists (Fig. 9d). In fact, the free-energy barriers among clusters in Figures 9b–9d are higher than those in Figure 9a. Therefore, inter-cluster transitions in Figures 9b–9d occur by passing narrower regions than those in Figure 9a. Figure 11 displays the experimentally determined complex structures and sampled conformations that are located near the bound form in the free-energy landscape (Fig. 9). Apparently, the sampled conformations closely resemble the experimental bound form.

The bound forms of the common binding region in the experimentally determined complex structures are also assigned to Figure 9. Bound forms S_{S100B} and S_{CBP} are mutually close in Figure 9 although S_{S100B} and S_{CBP} , respectively, denote a helix and a twisted conformation (See Fig. S1 of Supporting Information). The closeness of the two conformations results from the similarity of the $C\alpha$ atomic pair distances. Remember that the PCA space is generated based on the inter- $C\alpha$ atomic distances. Therefore, S_{CBP} can convert to S_{S100B} by minor rearrangements of the $C\alpha$ atomic pair distances.

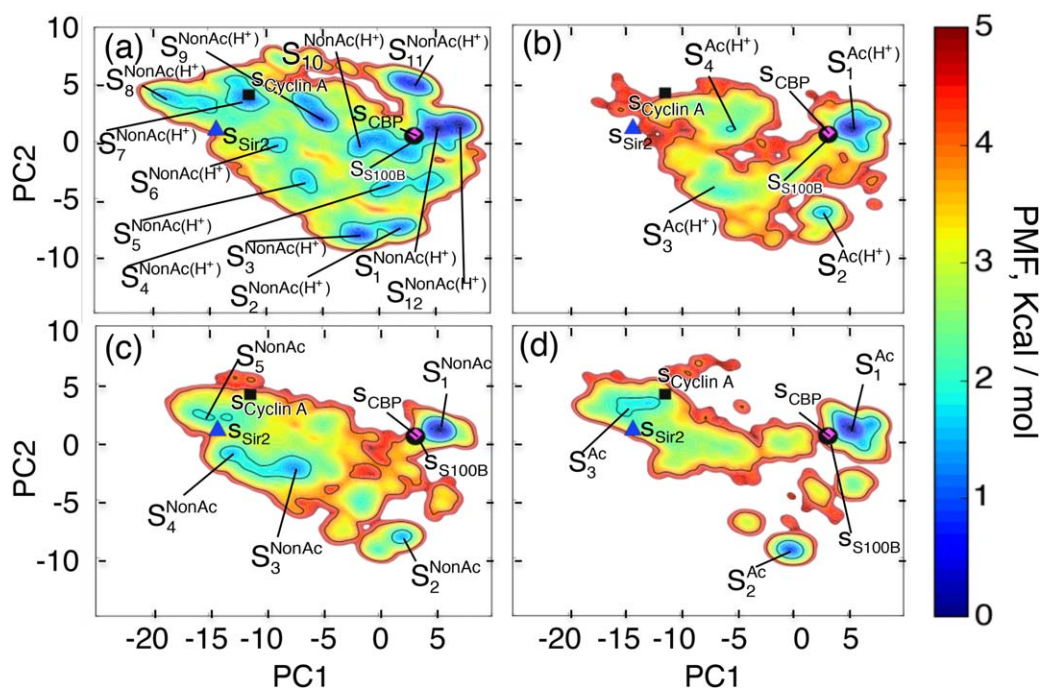


Figure 9. FEL, F^{SYS} , at 300 K for the common binding region (residues 380–386) constructed in 2D PCA space: a) $F^{NonAc(H^+)}$, b) $F^{Ac(H^+)}$, c) F^{NonAc} , and d) F^{Ac} . The lowest PMF site, r_{min} , is set as $F^{SYS}(r_{min}) = 0.0$ kcal/mol. Clusters S_k^{SYS} are shown in FEL. Black filled circle, blue triangle, black filled square, and magenta colored diamond, respectively, denote positions of the bound forms S_{S100B} , S_{Sir2} , $S_{CyclinA}$, 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_{sum}). Consequently, the PCA axes for all panels are common. [Color figure can be viewed at wileyonlinelibrary.com]

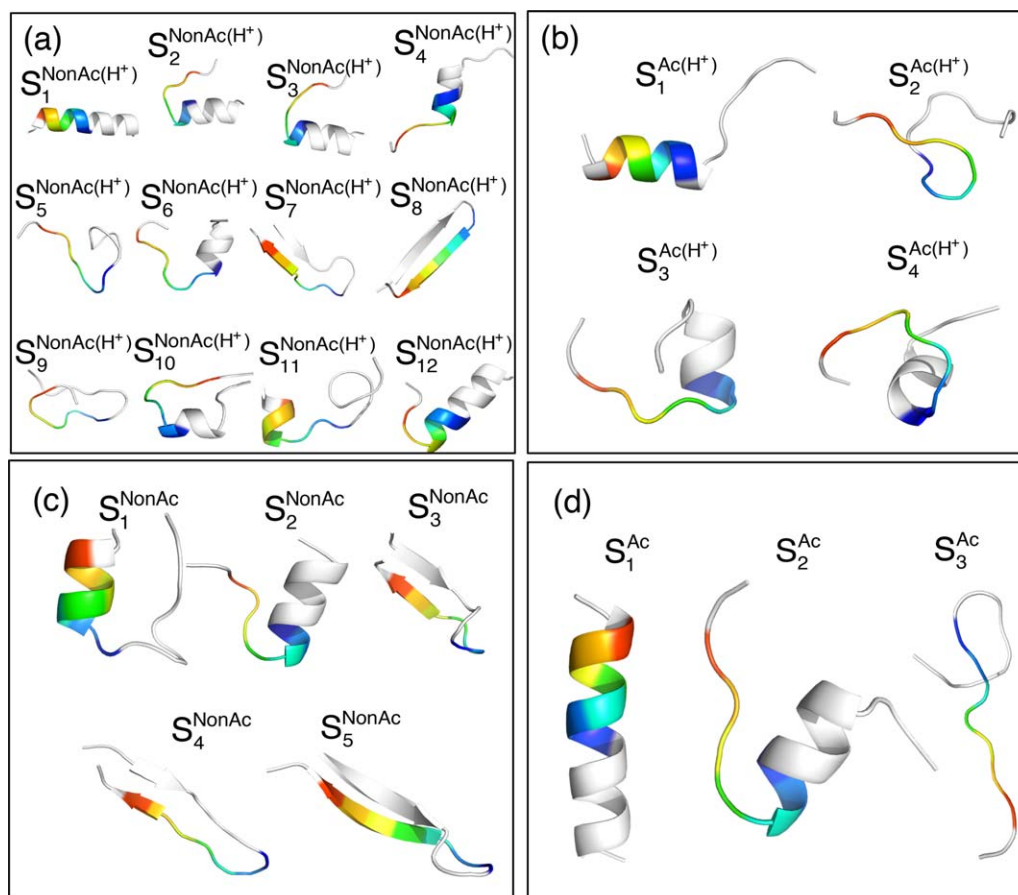


Figure 10. Structures in clusters of FEL for the common binding region: Tertiary structures picked from clusters S_k^{SYS} in FEL are colored by rainbow for the common binding region (blue and red sides are, respectively, the N and C-termini) and white for the other regions, which are not used for PCA. [Color figure can be viewed at wileyonlinelibrary.com]

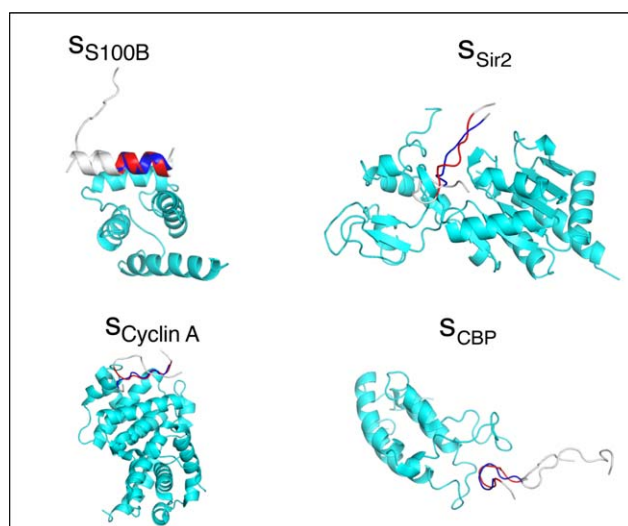


Figure 11. Experimentally determined complex structures and sampled conformations. The blue-colored region is the common-binding region of the experimental structure. Red is the sampled one. The sampled conformations are selected from the vicinity of the bound form in the free-energy landscape (Fig. 9), and are superimposed on the bound form. White-colored regions are outside the common-binding regions. Cyan-colored regions are partners. [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_{S100B} is located at a fringe of the most stable cluster $S_1^{NonAc(H^+)}$ (Fig. 9a). In other words, $Q_{NonAc(H^+)}$ prepares the bound form in advance. Furthermore, because the free-energy barriers from the other clusters to $S_1^{NonAc(H^+)}$ 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 $Q_{NonAc(H^+)}$ is smaller than that of $Q_{Ac(H^+)}$. Figure 9a suggests that $Q_{NonAc(H^+)}$ contains helical conformations sufficient for use for binding to S100B even if the helical content of $Q_{NonAc(H^+)}$ is less than that of $Q_{Ac(H^+)}$.

The population-selection mechanism might take place when the Ac fragment binds to CBP, where the conformations in the most stable cluster S_1^{Ac} can transition readily to the conformation S_{CBP} (Fig. 9d). However, the cluster S_1^{Ac} is not connected to the other clusters by low free-energy pathways. Consequently, the recruitment of conformations to S_1^{Ac} 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_{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.^[11,32] 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.^[52] 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 · multicanonical

How to cite this article: S. Iida, T. Mashimo, T. Kurosawa, 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, *Chem. Rev.* **2015**, *114*, 6661.
- [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.

- [10] T. Yamane, H. Okamura, Y. Nishimura, A. Kidera, M. Ikeguchi, *J. Am. Chem. Soc.* **2010**, *132*, 12653.
- [11] J. Higo, Y. Nishimura, H. Nakamura, *J. Am. Chem. Soc.* **2011**, *133*, 10448.
- [12] P. Tompa, M. Fuxreiter, *Trends Biochem. Sci.* **2008**, *33*, 2.
- [13] B. A. Shoemaker, J. J. Portman, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8868.
- [14] C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Yang, V. N. Uversky, A. K. Dunker, *BMC Genomics* **2008**, *9 Suppl 1*, S1.
- [15] E. Bochkareva, L. Kaustov, A. Ayed, G. Yi, Y. Lu, A. Pineda-Lucena, J. C. C. Liao, A. L. Okorokov, J. Milner, C. H. Arrowsmith, A. Bochkarev, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15412.
- [16] P. Di Lello, L. M. M. Jenkins, T. N. Jones, B. D. Nguyen, T. Hara, H. Yamaguchi, J. D. Dikeakos, E. Appella, P. Legault, J. G. Omichinski, *Mol. Cell* **2006**, *22*, 731.
- [17] P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine, N. P. Pavletich, *Science* **1996**, *274*, 948.
- [18] R. Rust, D. Baldisseri, D. Weber, *Nat. Struct. Mol.* **2000**, *7*, 570.
- [19] J. L. Avalos, I. Celic, S. Muhammad, M. S. Cosgrove, J. D. Boeke, C. Wolberger, *Mol. Cell* **2002**, *10*, 523.
- [20] S. Mujtaba, Y. He, L. Zeng, S. Yan, O. Plotnikova, Sachchidanand, R. Sanchez, N. J. Zeleznik-Le, Z. Ronai, M.-M. Zhou, *Mol. Cell* **2004**, *13*, 251.
- [21] E. D. Lowe, I. Tews, K. Y. Cheng, N. R. Brown, S. Gul, M. E. M. Noble, S. J. Gamblin, L. N. Johnson, *Biochemistry* **2002**, *41*, 15625.
- [22] J. van Dieck, D. P. Teufel, A. M. Jaulent, M. R. Fernandez-Fernandez, T. J. Rutherford, A. Wyslouch-Cieszyńska, A. R. Fersht, *J. Mol. Biol.* **2009**, *394*, 922.
- [23] W. J. Allen, D. G. S. Capelluto, C. V. Finkielstein, D. R. Bevan, *J. Phys. Chem. B* **2010**, *114*, 13201.
- [24] J. Chen, *J. Am. Chem. Soc.* **2009**, *131*, 2088.
- [25] C. McDowell, J. Chen, J. Chen, *J. Mol. Biol.* **2013**, *425*, 999.
- [26] I. Staneva, Y. Huang, Z. Liu, S. Wallin, *PLoS Comput. Biol.* **2012**, *8*, e1002682.
- [27] B. A. Berg, T. Neuhaus, *Phys. Rev. Lett.* **1992**, *68*, 9.
- [28] U. H. E. Hansmann, Y. Okamoto, F. Eisenmenger, *Chem. Phys. Lett.* **1996**, *259*, 321.
- [29] N. Nakajima, H. Nakamura, A. Kidera, *J. Phys. Chem. B* **1997**, *101*, 817.
- [30] T. Sugihara, J. Higo, H. Nakamura, *J. Phys. Soc. Japan* **2009**, *78*, 074003.
- [31] J. Ikebe, K. Umezawa, N. Kamiya, T. Sugihara, Y. Yonezawa, Y. Takano, H. Nakamura, J. Higo, *J. Comput. Chem.* **2011**, *32*, 1286.
- [32] K. Umezawa, J. Ikebe, M. Takano, H. Nakamura, J. Higo, *Biomolecules* **2012**, *2*, 104.
- [33] J. Higo, K. Umezawa, H. Nakamura, *J. Chem. Phys.* **2013**, *138*, 184106.
- [34] T. Terakawa, J. Higo, S. Takada, *Biophys. J.* **2014**, *107*, 721.
- [35] The UniProt Consortium *Nucleic Acids Res.* **2014**, *43*, D204.
- [36] T. Mashimo, Y. Fukunishi, N. Kamiya, Y. Takano, I. Fukuda, H. Nakamura, *J. Chem. Theory Comput.* **2013**, *9*, 5599.
- [37] J. Ryckaert, G. Ciccotti, H. Berendsen, *J. Comput. Phys.* **1977**, *327*, 341.
- [38] I. Fukuda, N. Kamiya, Y. Yonezawa, H. Nakamura, *J. Chem. Phys.* **2012**, *137*, 054314.
- [39] I. Fukuda, Y. Yonezawa, H. Nakamura, *J. Chem. Phys.* **2011**, *134*, 164107.
- [40] N. Kamiya, I. Fukuda, H. Nakamura, *Chem. Phys. Lett.* **2013**, *568–569*, 26.
- [41] L. V. Woodcock, *Chem. Phys. Lett.* **1971**, *10*, 257.
- [42] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
- [43] N. Kamiya, Y. S. Watanabe, S. Ono, J. Higo, *Chem. Phys. Lett.* **2005**, *401*, 312.
- [44] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- [45] P. Kollman, R. Dixon, W. Cornell, T. Fox, C. Chipot, A. Pohorille, In *The Development/Application of a “Minimalist” Organic/Biochemical Molecular Mechanic Force Field Using a Combination of Ab Initio Calculations and Experimental Data, Vol. 3*; W. van Gunsteren, P. Weiner, A. Wilkinson, Eds.; Springer: Netherlands, **1997**.
- [46] J. Ikebe, N. Kamiya, J. Ito, H. Shindo, **2007**, *16*, 1596.
- [47] R. P. Joosten, T. A. H. Te Beek, E. Krieger, M. L. Hekkelman, R. W. W. Hoof, R. Schneider, C. Sander, G. Vriend, *Nucleic Acids Res.* **2011**, *39*, 1.
- [48] O. D. Shahar, R. Gabizon, O. Feine, R. Alhadeff, A. Ganoth, L. Argaman, E. Shimshoni, A. Friedler, M. Goldberg, *PLoS One* **2013**, *8*, e78472.
- [49] P. Luo, R. L. Baldwin, *Biochemistry* **1997**, *36*, 8413.
- [50] N. Hirota, K. Mizuno, Y. Goto, *J. Mol. Biol.* **1998**, *275*, 365.
- [51] A. Micsonai, F. Wien, L. Kernya, Y. H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3095.
- [52] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, D. E. Shaw, *Plos One* **2012**, *7*, 1.

Received: 3 June 2016
Revised: 26 July 2016
Accepted: 31 August 2016
Published online on 13 October 2016