Molecular Mechanisms of Functional Modulation of Transcriptional Coactivator PC4 via Phosphorylation on Its Intrinsically Disordered Region

Qilin Xie, Kota Kasahara, Junichi Higo, and Takuya Takahashi*

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: To investigate the effects of phosphorylation on the function of the human positive cofactor 4 (PC4), an enhanced molecular dynamics (MD) simulation was performed. The simulation system consists of the N-terminal intrinsic disordered region (IDR) of PC4 and a complex that comprises the C-terminal acidic activation domain of a herpes simplex virion protein 16 (VP16ad) and a homodimer of the C-terminal structured core domain of PC4 (PC4ctd). An earlier report of an experimental study reported that the PC4−VP16ad interaction is modulated by incremental phosphorylation of the IDR. That report also proposed

a dynamic model where phosphorylated serine residues of a segment (SEAC) in the IDR contact positively charged residues (lysin and arginine) of another segment (K-rich) in the IDR. This contact formation induced by the phosphorylation results in variation of PC4−VP16ad interaction. However, this contact formation has not yet been measured directly because it is transiently formed and because the SEAC and K-rich segments are unstructured with high flexibility. We performed two simulations to mimic the incremental phosphorylation: the IDR was not phosphorylated in one simulation and only partially phosphorylated in the other. Our simulation results indicate that the phosphorylation weakens the IDR−VP16ad contact considerably with the induction of a compact structure in the IDR. This structure was stabilized by electrostatic interactions between the phosphorylated serine residues of a segment and lysine or arginine residues of another segment in the IDR, but the conformational fluctuation of this compact structure was considerably large. Consequently, the present study supports the experimentally proposed dynamic model. Results of this study can be important for computational elucidation of the functional modulation of PC4.
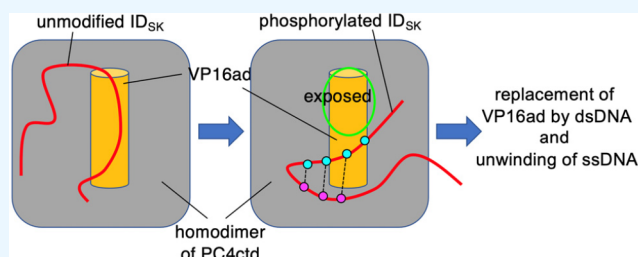
1. INTRODUCTION

Transcription has persisted as a central issue of molecular biology.[1,2] Infection of herpes simplex virus (HSV) to host cells (e.g., cells of yeast or mammals[3]) is related closely to transcription.[4] Herpes simplex virion protein 16 (VP16) (also known as Vmw65, ICP25, or α-TIF), a potent trans-activator,[5−7] is one element injected from HSV to the host cell. In infection, VP16 forms a ternary complex with HFC-1 and OCT-1, which are proteins that originate from the host cell. The ternary complex binds to a promoter region of the double-strand DNA.[4] Subsequently, a preinitiation complex (PIC) is formed by many general transcription factors (GTFs), mediator proteins, and RNA polymerase II, where the ternary complex is a part of PIC.

The C-terminal acidic activation domain of VP16 (VP16ad; residues 460−490) is intrinsically disordered; it is known as a protein that is able to bind to various proteins (a promiscuous property).[8] Addtionally, VP16 has been used to investigate transcription mechanisms.[8−10] Despite the intrinsically disordered property, VP16 adopts a helix conformation when it binds to the human positive cofactor 4 (PC4)[11] or to the Tfb1 subunit of a GTF, TFIIH.[12] This secondary structure

induction coupled with binding is called "coupled folding and binding".[13−17]

A report of a study of NMR has described that the C-terminal structured core region of PC4 (PC4ctd; residues 62−127) forms a homodimer (PC4ctd)$_2$ and binds to VP16ad. Figure S1a of the Supporting Information (SI) presents the NMR structure (PDB 2PHE). We designate this complex as VP16ad−(PC4ctd)$_2$ for this study. The N-terminal region (residues 1−61) of PC4 is not determined in the NMR complex; this region is intrinsically disordered in both the bound (complex) and unbound (free) states.[10] This disordered region, which is designated as PC4ntd, consists of SEAC, K-rich, and S-rich segments sequentially.[10] For the present study, we refer to the segment consisting of SEAC and K-rich segments as ID$_{SK}$. Figure S1b of the SI presents the sequence

of the complex following the UniProt residue numbering. The residue ordinal numbers in the PDB (PDB 2PHE) should be increased by one to the UniProt numbers.

Actually, PC4ntd interacts not only with VP16ad but also with double-stranded DNA (dsDNA) and single-stranded DNA (ssDNA).[10] Reportedly, the interaction surface of the PC4ctd homodimer to VP16ad overlaps with those of dsDNA and ssDNA. Furthermore, DNA unwinding occurs when dsDNA is binding to this surface.[8] Therefore, VP16ad, dsDNA, and ssDNA bind to PC4 exclusively. Additionally, it has been suggested that switching of the binding partner is modulated by the degree of phosphorylation of the SEAC segment.[10]

Phosphorylation is a chemical modification that is widely observed in transcription factors to regulate their activities.[18−22] In fact, phosphorylation of PC4 regulates its function in various transcription pathways.[23−28] The SEAC segment of PC4 has eight phosphorylatable serine sites: S9, S10, S11, S12, S13, S15, S17, and S19 (Figure S1b of the SI). Jonker et al. conducted extensive experiments with gradual phosphorylation of these serine residues by a kinase CKII.[10] Below, we summarize their findings obtained from experimentation. Given the VP16ad−(PC4ctd)$_2$ complex, the interaction between PC4 and VP16ad decreases concomitantly with the increasing number of phosphoserine residues (glutathione S-transferase (GST) pulldown assays). Similarly, given the dsDNA−(PC4ctd)$_2$ complex, the binding affinity of PC4 to dsDNA decreases, whereas the affinity to ssDNA increases with proceeding phosphorylation (electrophoretic mobility shift assays). Furthermore, the DNA unwinding activity of PC4 decreases (unwinding assays). A chemical shift perturbation (CSP) experiment has revealed a remarkable change of signals only in the ID$_{SK}$ region when the SEAC segment is phosphorylated.

Although controlling the number of phosphoserine residues experimentally and specifying the phosphorylated sites are difficult, Jonker et al. showed that the phosphorylation proceeds from the C-terminal side (S19) to the N-terminal side (S9) of the SEAC segment.[10] Based on these results, they proposed a scenario for the PC4's functional variation by gradual phosphorylation.[10] First, the unmodified PC4 (unphosphorylated PC4) interacts with VP16ad, the complex structure of which is presented in Figure S1a. The unmodified PC4 is then recruited to the promoter region. An important stage of the gradual phosphorylation is a period during which three serine residues (S15, S17, and S19) are phosphorylated from the C-terminal side of the SEAC segment (Figure S1 of the SI). In this stage, the PC4's binding affinity to VP16 becomes weaker in the complex;[10] also, VP16ad tends to be replaced by dsDNA in the complex because the dissociation constant ($K_d \approx 1 \mu$M) of VP16ad dissociating from the VP16ad−(PC4ctd)$_2$ complex is larger than that ($K_d \approx 50$ nM) of dsDNA dissociating from the dsDNA−(PC4ctd)$_2$ complex.[10,11] Although the DNA winding activity decreases gradually with the increase in the number of the phosphoserine residues, the activity persists when dsDNA binds to PC4, which means that dsDNA is replaced by ssDNA (or a DNA bubble). In the fully phosphorylated state of the SEAC segment, the ssDNA−(PC4ctd)$_2$ complex is mainly formed.

In the scenario presented by Jonker et al., they proposed a dynamic model for ID$_{SK}$, which can be a key to explain functional modulation by phosphorylation. The unmodified ID$_{SK}$ covers the VP16ad surface in the VP16ad−(PC4ctd)$_2$ complex. When S15, S17, and S19 are phosphorylated first, the

three phosphoserine residues interact attractively with lysine residues of the K-rich segment, forming a contact between the SEAC and K-rich segments (Figure S1c). Then, VP16ad is exposed to the solvent because the contact bends the ID$_{SK}$ segment, through which ID$_{SK}$ cannot cover a large surface of VP16ad. Then, the exposed VP16ad tends to be replaced by dsDNA in the complex. They presumed that further phosphorylation reinforces the interaction between the SEAC and K-rich segments. However, their model is difficult to detect directly by experimentation because PC4ntd is flexible and unstructured in either the bound or unbound state and because the bent conformation of ID$_{SK}$ is tentatively formed in the incremental phosphorylation process.

Molecular dynamics (MD) simulation is a sampling method that provides dynamic processes occurring in a biomolecular system with atomistic details. In particular, an enhanced sampling method (generalized-ensemble method) enables us to analyze larger conformational motions than a conventional MD method does.[29,30] Multidimensional virtual-system coupled molecular dynamics (mD-VcMD)[31] is an enhanced sampling method that we developed ourselves. To perform a mD-VcMD simulation, a user introduces a set of multiple reaction coordinates (RCs) in advance. Then the mD-VcMD enhances the sampling in the conformational space constructed by the multiple RCs. This method was applied to a molecular binding−unbinding process[32−36] and to motions of an intrinsic disordered region (IDR) of a protein.[34−36] We have proposed three variants of mD-VcMD to date:[31] original, subzone-based, and GA-guided mD-VcMD simulations.

For this study, we perform mD-VcMD of the VP16ad−(PC4ctd)$_2$ complex with ID$_{SK}$ to investigate two issues: variation of the ID$_{SK}$−VP16ad contact by phosphorylation of the SEAC segment and induction of a conformation in ID$_{SK}$ by phosphorylation. For these investigations, we prepared two models of the ID$_{SK}$ segment: an unmodified ID$_{SK}$ and a phosphorylated ID$_{SK}$ at three serine residues, namely, S15, S17, and S19. We obtained computational results indicating that the ID$_{SK}$−VP16ad contact is reduced considerably by phosphorylation and that phosphorylation induces fragile but compact conformations in ID$_{SK}$. These results support the dynamic model proposed by the Jonker et al.[10] Consequently, the current study presents, computationally, a mechanism for PC4 functional modulation induced by phosphorylation on ID$_{SK}$.

2. MATERIALS AND METHODS

2.1. The u-state and p-state Systems. The aim of this study is to analyze differences of the dynamic behaviors of ID$_{SK}$ between the unmodified and phosphorylated states in the presence of VP16ad−(PC4ctd)$_2$. For the analyses described herein, we set two systems: unmodified and phosphorylated. To sample the motions of ID$_{SK}$ effectively, we introduced two RCs to enhance the relative motions of ID$_{SK}$ toward VP16ad. Then we performed mD-VcMD simulations of the two systems and analyzed phosphorylation effects on the interaction between ID$_{SK}$ and the VP16ad−(PC4ctd)$_2$ complex. The mD-VcMD variant used for this study is the subzone-based VcMD.[31]

The computer model of the VP16ad−(PC4ctd)$_2$ complex was generated based on the NMR solution structure (PDB 2PHE) taken from the PDBj site (https://pdbj.org/). It is noteworthy that none of the entire intrinsically disordered (ID) segment (residues 1−61 consisting of the SEAC, K-rich, and S-rich segments) is involved in PC4ctd (residues 62−127)

of the NMR structure (Figure S1). The size of the periodic box should be markedly large if the two ID segments are involved in the system. The CSP experimental results indicated that the S-rich segment (residues 43−61), which is located at the root of the ID segment, is unaffected to any great degree by the phosphorylation (i.e., the S-rich segment is likely to behave as a random chain).[10] Therefore, we removed the S-rich segment from the system. The remaining ID segment, denoted as ID_{SK} (residues 1−42; the SEAC and K-rich segments), was treated as an isolated peptide, which moved freely in the periodic box during the simulation. Furthermore, to decrease computation costs, we put a single ID_{SK} segment in the periodic box. Later, we demonstrate that the single ID_{SK} segment causes different ID_{SK}−VP16ad interactions between the unmodified and phosphorylated systems.

Actually, VP16ad is α-helical in the NMR structure. Then, to refine the VP16ad conformation, we rebuild the conformation using MODELLER software[37] with the NMR structure of 2PHE for the reference conformation of modeling. The N-terminal of VP16ad was capped by an acetyl N-methyl group using PyMOL software.[38] Figure S2a depicts the resultant helical VP16ad conformation in the complex.

We prepared SEAC segments of ID_{SK} of two types: unmodified and phosphorylated SEAC segments, which are designated respectively as "unmodified-state" (u-state) and "phosphorylated-state" (p-state) SEAC segments. The phosphorylated residues are S15, S17, and S19, which are likely to be the first phosphorylated positions in the gradual phosphorylation process as described in section 1. Accordingly, entire systems with the u-state and p-state SEAC segments are designated respectively as a "u-state system" and a "p-state system".

The positioning of the ID_{SK} segment in the initial simulation conformation might affect the simulation results if ID_{SK} in the initial conformation involves a strong electrostatic attractive interaction to the VP16ad−$(PC4ctd)_2$ complex. If so, then ID_{SK} might not move from the initial position to any great degree. Therefore, simulation data might lack statistics. To avoid this artifact, the following procedures were taken. First, the initial conformation of the ID_{SK} segment was built using MODELLER. Then, this conformation was used for both the u-state and p-state SEAC segments. Next, we placed the modeled ID_{SK} so that the distance between the centroids of VP16ad and the SEAC region was about 45 Å and so that the intercentroid distance from VP16ad and the K-rich segment was also about 45 Å. Third, we put the VP16ad−$(PC4ctd)_2$ complex and the ID_{SK} segment in a periodic box filled with water molecules, where the box size was 90 Å× 90 Å × 90 Å (Figure S2b and c). Fourth, we introduced ions (Na^+ and Cl^-) so that the ionic concentration of the solution was 150 mM with neutralization of the net charge of the system. This procedure was done for both the u-state and p-state systems. ID_{SK} moved to a great degree with the adoption of various conformations during the simulation, which reproduced the intrinsically disordered nature of ID_{SK} in both the u-state and p-state systems.[10]

The total number of atoms was 71 441 (680 atoms in ID_{SK}, 1103 atoms for a single chain of PC4ctd, 446 atoms in VP16ad, 67 977 water molecules, 66 Na^+ ions, and 66 Cl^- ions) for the u-state system, and 71 615 (689 atoms in ID_{SK}, 1103 atoms for a single chain of PC4ctd, 446 atoms in VP16ad, 68 136 water molecules, 72 Na^+ ions, and 66 Cl^- ions) for the p-state system. After energy minimization using the steepest descent

method until the maximum atomic force was under 1000 kJ/mol/nm, we conducted an NPT simulation for 1.0 ns at 300 K with an integration time step of 2.0 fs. The resultant box sizes were respectively 89.5 Å × 89.5 Å × 89.5 Å and 89.6 Å × 89.6 Å × 89.6 Å for the u-state and p-state systems.

The force fields employed for the systems were AMBER99SB-ILDN[39] for polypeptides, Joung−Cheatham[40] parameters for ions, and the TIP3P[41] potential for water. The electrostatic interactions were calculated using the smooth particle-mesh Ewald method,[42] where the real-space cutoff length for the nonbonded potentials was set to 1.2 nm. All the MD simulations were performed using GROMACS 2018.8.[43] The geometry of moieties covalently bonded to hydrogen atoms was constrained using the LINCS method.[44] In fact, no structural restraint was applied to the system (the molecules were entirely flexible in the simulation) except for the hydrogen-related covalent bonds treated using the LINCS method explained above.

2.2. Two Reaction Coordinates (RCs). In mD-VcMD, the conformational motion is enhanced in a multidimensional RC space, which is defined in advance. For this study, we introduced two RCs into the system: $\lambda^{(h)}$ ($h = \alpha, \beta$). Consequently, "multi-dimensional (mD)" means "two-dimensional (2D)" herein. Figure S3 of the SI presents the two RCs: $\lambda^{(\alpha)}$ is defined as the distance from the centroid of the SEAC segment to that of VP16ad, and $\lambda^{(\beta)}$ is the intercentroid distance from the K-rich segment to VP16ad. Therefore, the moves of $\lambda^{(\alpha)}$ and $\lambda^{(\beta)}$ respectively represent the association−dissociation motions of the SEAC segment and the K-rich segment to VP16ad. To study the motions of ID_{SK} extensively, the upper and lower limits of the RCs should be sufficiently wide to cover both the unbound and bound conformations of ID_{SK} to VP16ad. Table S1 of the SI presents the lower and upper limits $[\lambda_1^{(h)}]_{min}$ and $[\lambda_{n_{vs}(h)}^{(h)}]_{max}$, respectively) for RCs, which are to be 5.0 and 46.9 Å, respectively, for both $h = \alpha$ and β; $n_{vs}(h)$ is explained later. As described in section 3, ID_{SK} moved widely in the space; a small probability was assigned to unbound conformations, and a large probability was assigned to bound conformations.

2.3. mD-VcMD. We briefly explain here the outline of mD-VcMD. Its methodological details are explained elsewhere.[31] First, the RC space is divided into small zones. Table S2 of the SI presents the actual zone partitioning along the two RCs, $\lambda^{(\alpha)}$ and $\lambda^{(\beta)}$. The number of zones along each RC is $n_{vs}(h)$ ($h = \alpha, \beta$) (Table S1). The upper and lower boundaries of the ith zone along $\lambda^{(h)}$ are denoted respectively as $[\lambda_i^{(h)}]_{min}$ and $[\lambda_i^{(h)}]_{max}$. A 2D zone, which is located in the i-th and j-th zones, respectively along $\lambda^{(\alpha)}$ and $\lambda^{(\beta)}$, is specified by 2D indices (i, j). The movable range for the 2D zone extends from $[\lambda_i^{(\alpha)}]_{min}$ to $[\lambda_i^{(\alpha)}]_{max}$ along $\lambda^{(\alpha)}$ and from $[\lambda_i^{(\beta)}]_{min}$ to $[\lambda_i^{(\beta)}]_{max}$ along $\lambda^{(\beta)}$.

In mD-VcMD, the conformation moves by a canonical MD at temperature T (300 K in this study) within a 2D zone (the "current zone") for a time period of $\Delta\tau$ ($\Delta\tau = 0.2$ ps (or 100 steps) in this study). Then, the conformation transitions from the current zone to another zone, which is one of the adjacent zones around the current zone in the 2D RC space. When the transition is achieved, the current zone is updated. Then, the conformation is confined in the new current zone. Repeating this procedure, the conformational space is sampled widely. The transition is controlled by a transition matrix B: an element $B_{(i, j)\rightarrow(k,l)}$ is the transition probability from a 2D zone

(i, j) to another 2D zone (k, l), which are mutually adjacent in the 2D RC space.

Because the actual simulation length is finite, none of the RC space from $[\lambda_1^{(h)}]_{\min}$ to $[\lambda_{n_{vs}(h)}^{(h)}]_{\max}$ (Table S1) is sampled by the finite simulation. Consequently, the mD-VcMD simulation is repeated iteratively. Thereby, the volume of the sampled conformation space is expanded. When iteration M has finished, quantity $Q_{cano}^M(i, j)$ is obtained from the first to M-th iterations, where $Q_{cano}^M(i, j)$ represents a probability distribution function of the system existing in zone (i, j). In other words, $Q_{cano}^M(i, j)$ is a system's canonical distribution (i.e., a thermally equilibrated distribution at the simulation temperature T), which is expressed in the 2D RC space spanned by the RC zone indices i and j when the convergence of $Q_{cano}^M(i, j)$ is obtained in iterations from 1 to M. The transition matrix B for the $(M + 1)$-th iteration is calculated using $Q_{cano}^M(i, j)$ according to a protocol introduced in our earlier studies.[31-36] Then, the iteration $M + 1$ is initiated.

To increase the sampling efficiency, we performed N_{run} (= 1000) runs in parallel in each iteration by starting the runs from different conformations. A conformational ensemble from the N_{run} trajectories can be regarded as a canonical ensemble.[45,46] Denoting the length of a run as L, if the L runs are initiated from various conformations, then the ensemble consisting of N_{run} trajectories covers a conformational space more widely than a long single run of length of $N_{run} \times L$ does.[45,46]

We have stored snapshots from the first to M-th iterations when the M-th iteration finished. The N_{run} initial conformations for the $(M + 1)$-th iteration were selected from those snapshots so that they are distributed as evenly as possible in the 2D RC space. An important exception is that runs for the first iteration start from a single conformation obtained from the NPT simulation that was done as described above.

When we judge that $Q_{cano}^M(i, j)$ has converged well, we calculate a statistical weight[31] that is assigned to each snapshot at 300 K. The ensemble of the weighted snapshots is used for computing various equilibrated quantities at 300 K. For each of the u-state and p-state systems, we performed 25 iterations, where the length of each of the N_{run} runs in an iteration was 10^6 steps. Because the simulation time step was 2 fs, a trajectory of 2 μs (= 2 fs $\times 10^6$ steps $\times$ 1000 runs) is obtained in an iteration for each system. Parameter L described above is 2 fs $\times 10^6$. We saved a snapshot once every 5000 steps of the simulation. Snapshots from the third to 25th iterations, consisting of 4.6×10^6 snapshots, were used for analyses (the first and second iterations were discarded). As described herein, this number 4.6×10^6 is designated as N_c. We obtained two ensembles of snapshots for the u-state and p-state systems, which are denoted respectively as $\zeta_{\text{u-state}}$ and $\zeta_{\text{p-state}}$.

2.4. Contacts between ID_{SK} and VP16ad. Presuming that $Q_{cano}^M(i, j)$ has converged at iteration M and that a statistical weight ω_k at 300 K, which is assigned to snapshot k, is calculated from $Q_{cano}^M(i, j)$ using a protocol in ref 31, then we can analyze the ID_{SK}−VP16ad contacts as follows. If the minimum inter-heavy atomic distance between two residues is smaller than 6.0 Å in a snapshot, then we judged that the residues are mutually contacting in the snapshot.

The contact ratio of residue i of ID_{SK} to VP16ad is defined as

$$C_{ID}(i) = \frac{\sum_{k=1}^{N_c} \omega_k \times \theta_k(i)}{\sum_{k=1}^{N_c} \rho_k} \tag{1}$$

One can recall that N_c represents the number of snapshots in $\zeta_{\text{u-state}}$ or $\zeta_{\text{p-state}}$. The term θ_k is defined as

$$\theta_k(i) = \begin{cases} 1 & \text{(if residue } i \text{ contacts to VP16ad in snapshot } k) \\ 0 & \text{(otherwise)} \end{cases} \tag{2}$$

Similarly, the contact ratio of residue j of VP16ad to ID_{SK} is defined as

$$C_{VP}(j) = \frac{\sum_{k=1}^{N_c} \omega_k \times \xi_k(j)}{\sum_{k=1}^{N_c} \rho_k} \tag{3}$$

where ξ_k is defined as

$$\xi_k(j) = \begin{cases} 1 & \text{(if residue } j \text{ contacts to } ID_{SK} \text{ in snapshot } k) \\ 0 & \text{(otherwise)} \end{cases} \tag{4}$$

Furthermore, we introduce the contact ratio for the pair of residues i of ID_{SK} and j of VP16ad as

$$C_{ID-VP}(i, j) = \frac{\sum_{k=1}^{N_c} \omega_k \times \eta_k(i, j)}{\sum_{k=1}^{N_c} \rho_k} \tag{5}$$

Also, $\eta_k(i, j)$ is defined as

$$\eta_k(i, j) = \begin{cases} 1 & \text{(if residues } i \text{ and } j \text{ contact in snapshot } k) \\ 0 & \text{(otherwise)} \end{cases} \tag{6}$$

A larger contact ratio $C_{ID}(i)$, $C_{VP}(j)$ or $C_{ID-VP}(i, j)$ is associated with more contact at equilibrium (300 K). If the ratio is 1, then the contact is always formed. If the ratio is 0, then no contact is formed at all.

Finally, we use the same expression to estimate the intra-ID_{SK} contact ratio between residues i and j of ID_{SK} as

$$C_{ID-ID}(i, j) = \frac{\sum_{k=1}^{N_c} \omega_k \times \eta_k(i, j)}{\sum_{k=1}^{N_c} \rho_k} \tag{7}$$

where

$$\xi_k(i, j) = \begin{cases} 1 & \text{(if residues } i \text{ and } j \text{ contact in snapshot } k) \\ 0 & \text{(otherwise)} \end{cases} \tag{8}$$

When analyzing the contacts, we have noted artifacts in some snapshots. Figure S4 presents an example of such a snapshot schematically, where ID_{SK} contacts both VP16ad and PC4ctd with the crossing of the periodic boundary. The fraction of those artificial snapshots against the all snapshots was 0.324, and we eliminated them from analyses by setting weight ω_k to 0 for those snapshots for both the u-state and p-state systems.

2.5. Principal Component Analysis (PCA). Because the system consists of many atoms, the structural distribution of the system is expressed in a high-dimensional conformational space, in principle. Actually, PCA has been used to visualize the distribution by projecting it in a low-dimensional

space.[47−49] Here, PCA was applied to the two conformational ensembles $\zeta_{\text{u-state}}$ and $\zeta_{\text{p-state}}$ of the snapshots. A general explanation for PCA is presented in the SI (section S1).

First, we select C_α atoms from the system, which specifies the relative positioning of ID_{SK} to VP16ad−(PC4ctd)$_2$. For this study, we select the C_α atoms from ID_{SK} (the group of these atoms is denoted as Atom$_{\text{ID}}$), those from VP16ad (denoted as Atom$_{\text{VP}}$), and those from PC4ctd that contact VP16ad in the NMR complex structure (denoted as Atom$_{\text{ctd}}$). The word "distance" used in this subsection is a $C_\alpha−C_\alpha$ atomic distance. Then, for each snapshot, we calculate all intra-Atom$_{\text{ID}}$ distances, all Atom$_{\text{ID}}$−Atom$_{\text{VP}}$ distances, and all Atom$_{\text{ID}}$−Atom$_{\text{ctd}}$ distances. We eliminate intra-Atom$_{\text{VP}}$, intra-Atom$_{\text{ctd}}$, and Atom$_{\text{VP}}$−Atom$_{\text{ctd}}$ distances because the VP16ad−(PC4ctd)$_2$ complex structure was maintained well during the mD-VcMD simulation. The variations of those distances are small; for that reason, the shape of the structural distribution is not influenced by the small variations. We denote the number of distances as N_d (= 3927). The relative position of ID_{SK} to VP16ad (and hence to VP16ad−(PC4ctd)$_2$) is described by the N_d distances.

Given a snapshot k, the computed distances are regarded as components of a vector defined in the N_d-dimensional space. We express this vector as $\boldsymbol{d}^k$: $\boldsymbol{d}^k = (d_1^k, d_2^k, ..., d_{N_d}^k)$, where d_i^k stands for the i-th distance of the snapshot, and $k = 1, ..., 2N_c$. After $\boldsymbol{d}^k$ is calculated for all snapshots in the combined ensemble ($\zeta_{\text{u-state}} + \zeta_{\text{p-state}}$), a variance−covariance matrix is computed using $2N_c$ snapshots stored in the combined ensemble. Equation S1 of the SI presents the actual formulation of the matrix. From the matrix, we obtained coordinate axes PC1 and PC2, which are convenient to describe the conformational distribution of the system. We designate the 2D space constructed by PC1 and PC2 as "2D PC space". The N_c snapshots in $\zeta_{\text{u-state}}$ or $\zeta_{\text{p-state}}$ are projected to the 2D PC space (eq S4 of the SI). Then PCA was applied using the Python library scikit-learn.[50]

To analyze the distribution of the conformation of ID_{SK}, we performed another PCA, where the intra-Atom$_{\text{ID}}$ distances are used to calculate the variance−covariance matrix. The ensemble average in eq S1 of the SI is taken over snapshots in the combined ensemble of $\zeta_{\text{u-state}}$ and $\zeta_{\text{p-state}}$. The other procedures are the same as those presented above.

A free-energy (potential of mean force) landscape was presented in 2D PC space as follows. After a density, $\rho(\text{PC1}, \text{PC2})$, of the ID_{SK} conformation at a position (PC1, PC2) in the 2D PC space was calculated, the density was converted to a potential of mean force, PMF(PC1, PC2), as

$$\text{PMF(PC1, PC2)} = -RT\ln[\rho(\text{PC1, PC2})] - c_0 \qquad (9)$$

where R is the gas constant and where c_0 is a constant introduced to reset PMF so that the lowest PMF is zero. The unit of PMF is kilocalories per mole.

2.6. Hydrogen Bonds and Salt bridges.

A hydrogen bond is detected with the criteria that the hydrogen−acceptor distance is less than 2.5 Å and that the donor−hydrogen−acceptor angle is greater than 120°. A salt bridge is detected with a criterion that the distance between the distal carbon atom of Asp or Glu and the nitrogen atom of Arg or the distal carbon atom of Lys is less than 6.0 Å.

3. RESULTS

The mD-VcMD simulation enhanced the intersegment motions between VP16ad and ID_{sk} because the two currently introduced RCs, $\lambda^{(\alpha)}$ and $\lambda^{(\beta)}$, are the VP16ad−ID_{SK} distances (Figure S3 of the SI). However, the motions related to the VP16ad−PC4ctd distance are not enhanced. Consequently, the VP16ad−(PC4ctd)$_2$ complex structure was maintained well during the simulation for both the u-state and p-state systems (data not shown). Although the phosphorylation of the SEAC segment decreases the binding affinity of VP16ad to the homodimer of (PC4ctd)$_2$,[10] it does not mean that VP16ad is released immediately from the homodimer.

As explained in section 2, we performed mD-VcMD up to iteration 25. To confirm the convergence of Q_{cano}^M (i, j), correlation between Q_{cano}^K $(K = 1, \cdots, 24)$ and Q_{cano}^{25} was calculated. Section S2 of the SI presents the method of computing the correlation. Figure S5 of the SI shows that the sampling is almost converged up to iteration 20. As shown in the analysis of section S3 of the SI, the ID_{SK} conformation converges to different regions between the u- and p-state systems in the conformational space (see Figure S6 and Table S3).

3.1. Decrement of ID_{SK}−VP16ad Contact by Phosphorylation.

We analyzed the ID_{SK}−VP16ad contact in the u-state and p-state systems. Figure 1a shows the contact ratio
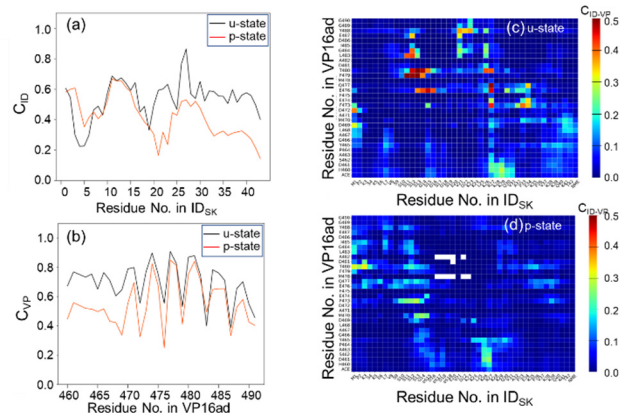
Figure 1. (a) Contact ratio C_{ID} (i) (eq 1) as a function of residue ordinal number i in ID_{SK} for both the u-state and p-state systems. (b) Contact ratio C_{VP} (j) (eq 3) as a function of residue ordinal number j of VP16ad. Inter-residue contact map between ID_{SK} and VP16ad for (c) the u-state system and (d) the p-state system. Amino acids pS15, pS17, and pS19 are phosphorylated serine residues. White regions were not sampled in the simulation because the free energy assigned to those regions is extremely high.

$C_{\text{ID}}(i)$ (eq 1) as a function of residue ordinal number i of ID_{SK}. In the u-state system, the K-rich segment of ID_{SK} contacted VP16ad, which is consistent with the CSP experiment.[8] It is noteworthy that remarkable contacts were apparent at K26 and R27 of the K-rich segment ($C_{\text{ID}}(i) \approx 0.8$; $i = 26$ and 27). In the p-state system, by contrast, the K-rich segment had considerably less contact to VP16ad than the K-rich segment in the u-state had. Figure 1b presents the contact ratio $C_{\text{VP}}(j)$ (eq 3). The periodic convexo-concave in this figure results from the helicity of VP16ad. The periodicity is less clear in its N-terminal side because the N-terminal of VP16ad is somewhat disordered in the simulation, which agrees with the NMR structure.[11] This figure shows that the contact ratio

$C_{VP}(j)$ for the p-state system decreased greatly in the N-terminal side of VP16ad. Figure 1a and b show that the ID_{SK}–VP16ad contact was decreased by the phosphorylation.

Then, to investigate details of the ID_{SK}–VP16ad contact, we calculated $C_{ID-VP}(i, j)$, which is the contact ratio between residue i of ID_{SK} and residue j of VP16ad (eq 5). The difference of $C_{ID-VP}(i, j)$ between the u-state and p-state systems is clear: remarkable contacts (we defined a remarkable contact as one with $C_{ID-VP}(i, j) \geq 0.4$) in the u-state system (Figure 1c) vanished in the p-state system (Figure 1d). Therefore, Figure 1 demonstrates that VP16ad in the p-state system was exposed to solvent more than VP16ad in the u-state system was. This finding supports Jonker's assumption.[10]

Figure 1c shows that the residues 11−16 of the SEAC segment contacted remarkably to E476, F479, T480, L483, and G484 of VP16ad. These ID_{SK}−VP16ad contacts constitute a mixture of hydrophobic and hydrophilic interactions because the contacting residues are a mixture of hydrophobic and hydrophilic residues. In addition, in the u-state system, K26 and R27 in the K-rich segment contacted remarkably to F473, E476, Q477, and T488. We concluded that these contacts were electrostatic because those residues are hydrophilic except for F473. Vanishing of the remarkable contact in the p-state system is examined later by analyzing the intra-ID_{SK} interactions.

Before analyzing the intra-ID_{SK} interaction, we checked the ID_{SK}−VP16ad interaction further. Tables 1a and 1b present

Table 1a. Hydrogen Bonds Observed Between ID_{SK} (u-state) and VP16ad

ID_{SK}	main/side	VP16ad	main/side	prob.[a]
ARG27	side chain	GLU476	side chain	0.482
ARG27	side chain	ASP476	side chain	0.269
SER15	side chain	GLU476	side chain	0.247
SER15	mainchain	GLU476	side chain	0.231
MET1	mainchain	ASP472	side chain	0.229

[a]Interacting residue pairs are arranged in descending order of probability.

Table 1b. Hydrogen Bonds Observed between ID_{SK} (p-state) and VP16ad

ID_{SK}	main/side	VP16ad	main/side	prob.[a]
ARG27	side chain	ASP469	side chain	0.196
LYS3	side chain	ASP481	side chain	0.185
ARG27	side chain	GLU476	side chain	0.135
ARG27	side chain	ASP461	side chain	0.102
S2P15	mainchain	GLU476	side chain	0.099

[a]Interacting residue pairs are arranged in descending order of probability.

hydrogen-bonding residue pairs between ID_{SK} and VP16ad, respectively, for both systems. Similarly, Tables 2a and 2b list the residue pairs with a salt bridge between ID_{SK} and VP16ad, respectively, for both systems. The section 2 presents the definition of the hydrogen bond and the salt bridge. Apparently, the hydrogen bonds in the u-state system (Table 1a) were more stable than those in the p-state system (Table 1b). However, the salt bridge stabilities were compatible between the two systems (Tables 2a and 2b). Because the phosphorylation of ID_{SK} induces the net charge of the three serine residues, one might expect that the phosphorylation

Table 2a. Salt Bridge Observed between ID_{SK} (u-state) and VP16ad

ID_{SK}	VP16ad	prob.[a]
ARG27	GLU476	0.269
ARG27	ASP469	0.111
ARG27	ASP461	0.093
LYS28	ASP481	0.074
LYS26	GLU476	0.071

[a]Interacting residue pairs are arranged in descending order of probability.

Table 2b. Salt Bridge Observed between ID_{SK} (p-state) and VP16ad

| ID_{SK}	VP16ad	Prob.[a]
LYS3	ASP481	0.239
ARG27	ASP469	0.114
LYS41	ASP481	0.085
LYS29	ASP469	0.082
LYS24	ASP469	0.078

[a]Interacting residue pairs are arranged in descending order of probability.

drastically alters the salt bridge interaction between ID_{SK} and VP16ad. However, the phosphoserine residues (pS15, pS17, and pS19) did not appear in Table 2b. We demonstrate later that those phosphoserine residues are related to the induction of interactions within ID_{SK}.

Although Figure 1c shows several remarkable contacts ($C_{ID-VP}(i, j) \geq 0.4$) between ID_{SK} and VP16ad in the u-state system, Tables 1a and 2a show that those contacts had a lower probability of hydrogen-bond (or salt bridge) formation (probability <0.4), except for the pair of ARG27−GLU476 in Table 1a. This apparent discrepancy is attributable to the stricter criteria for hydrogen-bond and salt bridge formation than the criterion used for calculating $C_{ID-VP}(i, j)$. Because the intrinsically disordered segment ID_{SK} is fluctuating greatly, the ID_{SK}−VP16ad interaction should be analyzed not only by the strict criteria for a hydrogen bond and salt bridge but also by the loosened criterion used for $C_{ID-VP}(i, j)$. Another reason for the discrepancy is that the contacts shown in Figure 1c involve hydrophobic contacts as well as electrostatic contacts, as discussed earlier.

3.2. Free-Energy Landscape in the 2D PC Space. The conformational ensembles of the systems at 300 K are summarized as free-energy landscapes (FEL), where the potential of mean force (PMF; eq 9) is assigned to each site of the 2D PC space constructed by the first and second principal axes (PC1 and PC2, respectively), as presented in section 2 in the text and section S1 of the SI. The resultant contribution ratios were $R_1 = 0.525$ for PC1 and $R_2 = 0.202$ for PC2. Therefore, 73% (= $R(1, 2)$; eq S6 of the SI) of the conformational variety was reproduced by PC1 and PC2. Remember that the PC axes were obtained from the combined ensemble ($\zeta_{u\text{-state}} + \zeta_{p\text{-state}}$); the definitions of PC1 and PC2 are common between the u-state and p-state systems. The coefficients of correlation were found to be 0.002 between PC1 and $\lambda^{(\alpha)}$, 0.003 between PC1 and $\lambda^{(\beta)}$, 0.010 between PC2 and $\lambda^{(\alpha)}$, and 0.011 between PC2 and $\lambda^{(\beta)}$. These small correlation coefficients indicate that the motion of ID_{SK} was comprised almost equally by motions along the PC1 and PC2 axes.

Figure 2a and b shows that ID_{SK} spread in similar 2D regions between the two systems, where the lowest free-energy (lowest
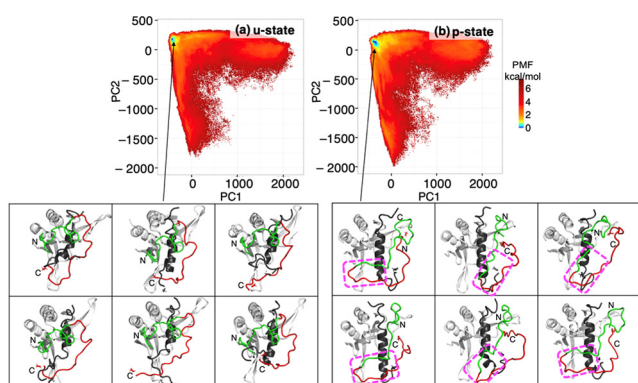

Figure 2. FEL for the (a) u-state and (b) p-state systems, which are presented in 2D PC space constructed by PC1 and PC2. The insets present typical conformations picked from the lowest free-energy basins, where "N" and "C" respectively denote the N- and C-termini of ID_{SK}. SEAC and K-rich segments are presented by green and red models, respectively. Gray and black models are respectively the $(PC4ctd)_2$ dimer and VP16ad. The magenta-colored frames are mentioned in the text.

PMF) basins overlap well. Consequently, one might pose the question: are the conformations in the lowest basins for the two systems similar? However, typical conformations picked from the basins were greatly different, as shown in the insets of Figures 2. Particularly, the molecular orientations of ID_{SK} were mutually different to a complete degree.

We defined the area for the lowest free-energy basin by 2D PC sites with PMF (PC1, PC2) ≤ 0.5 kcal/mol. Then, the ratios assigned to the lowest free-energy basin were, respectively, 0.104 and 0.157 for the u-state and p-state systems. Because the probability for the lowest free-energy basin was small, ID_{SK} was unstructured at 300 K in both systems. Later, we investigated more structural features of ID_{SK}.

3.3. Intra-ID_{SK} Contact Patterns in the p-state System.
Next, we check whether the ID_{SK} segment forms a specific structural feature. Figure 3 presents an intra-ID_{SK} contact map $C_{ID-ID}(i, j)$ (eq 7), which was computed from the conformations in the lowest free-energy basin (Figure 2); the
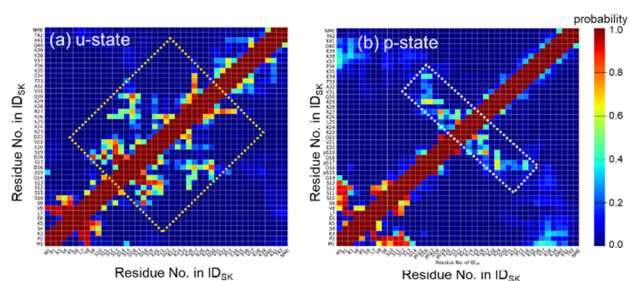

Figure 3. Intra-ID_{SK} contact map C_{ID-ID} (i, j) for conformations in the lowest free-energy basin (Figure 2) for (a) the u-state and (b) the p-state systems. The contents of the yellow and white rectangles in panels (a) and (b), respectively, are described in the text. Amino acids pS15, pS17, and pS19 are phosphorylated serine residues. White regions were not sampled in simulation because the free energy assigned to those regions is extremely high.

definition of the lowest free-energy basin was presented in the preceding subsection.

In the p-state system, a specific contact pattern can be seen, which is shown within a white broken-line rectangle in Figure 3b. We show the region involved in this contact pattern by the magenta-colored frames in the inset of Figure 2b. One may notice that this contact pattern is typically assigned to a hairpin structure. However, ID_{SK} does not involve inter-main chain hydrogen bonds. Thus, this hairpin-like conformation is not a hairpin defined in the protein secondary-structure classification.

In the u-state system, by contrast, we recognize various contact patterns in ID_{SK}, as shown by a yellow dashed rectangle in Figure 3a. These various contact patterns reflect the intrinsically disordered nature of ID_{SK} in the u-state system. One must recall that ID_{SK} is disordered in both the u-state and p-state systems, as shown in the experiment,[10] and that the fraction assigned to the lowest free-energy basins were small, as shown above.

To analyze the structural features of ID_{SK} further, we applied another PCA using the C_α atoms only in ID_{SK} (i.e., $Atom_{ID}$). Figure 4 portrays the resultant FELs for the two systems. This
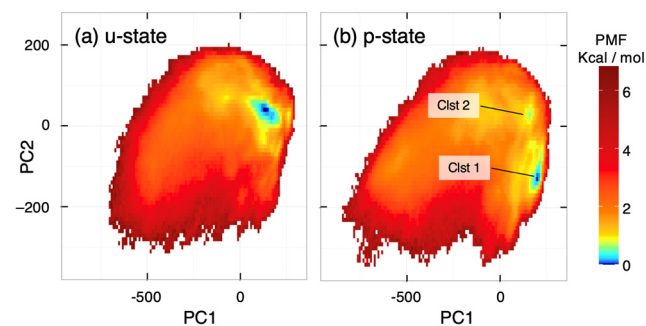
Figure 4. FELs for the (a) u-state and (b) p-state systems in 2D PC space constructed by PC1 and PC2. PCA was performed using C_α atoms in $Atom_{ID}$. Two clusters "Clst 1" and "Clst 2" are mentioned in the text.

figure shows that each system has a single major basin whereas each involves thermodynamically less important minor basins. We applied a clustering method (the single linkage clustering method[51]) to the ensemble of the p-state system and obtained two clusters: major and minor clusters, which are respectively labeled as Clst 1 and Clst 2 in Figure 4b. The fraction of Clst 2 to Clst 1 was 9%.

Two conformations of Figure 5a, which are examples selected from the major free-energy basin (Figure 4a) for the u-state system, were featureless without a characteristic structure. In contrast, the conformations from the major free-energy basin (Clst 1 of Figure 4b) of the p-state system had a feature: the phosphoserine residues in the SEAC segment interacted with lysine or arginine residues in the K-rich segment (Figures 5b). Those interactions are the origin of the specific contact pattern indicated by the white dashed rectangle in Figure 3b. The conformations from the minor basin (Clst 2) had a twisted form in ID_{SK}, as indicated by the magenta circle in Figure 5c. We note that the conformations from the major basin Clst 1 exhibited structural diversity: the interacting residue pairs are not unique. Moreover, the N- and C-termini of ID_{SK} fluctuated greatly. Similarly, the conformations from the minor basin Clst 2 were somewhat diverse.
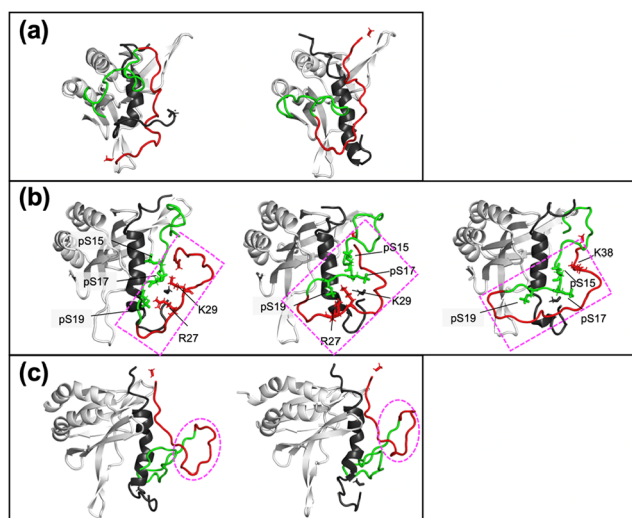
Figure 5. (a) Two conformations of the u-state system picked from the low free-energy basin of Figure 4a. (b) Three conformations from the major basin (Clst 1 of Figure 4b) of the p-state system. Side chains of three phosphoserine residues pS15, pS17, and pS19 are explicit; side chains of contact partners to the phosphoserine residues are also shown. Magenta rectangles represent the SEAC−K-rich contact region corresponding to the specific contact pattern shown by the white rectangle in Figure 3b. (c) Two conformations from the minor basin (Clst 2 of Figure 4b) of the p-state system. Magenta circles indicate the twisted conformation mentioned in the text. Green and red segments respectively represent the SEAC and K-rich segments. Gray and black models are respectively the $(PC4ctd)_2$ dimer and VP16ad.

Therefore, ID_{SK} had both the ordered nature and the disordered nature.

To analyze further the SEAC−K-rich contacts in Clst 1 and 2, we did the following analysis. We calculated an interatomic distance r_c, which is The distance from the side chain oxygen atom of serine or phosphoserine residues of the SEAC segment to the side chain carbon atom bound to the charged nitrogen atoms of lysin or arginine residues of the K-rich segment. Note that the side chain oxygen atom is connected to the phosphorus atom in phosphoserine. Then, we judged that a SEAC−K-rich contact was formed when $r_c \leq 5.5$ Å was satisfied. The SEAC−K-rich contact ratio for Clst 1 was 0.904, and that for Clst 2 was 0.281. Therefore, the association of the SEAC and K-rich segments was formed mainly in Clst 1. We also applied this procedure to the major cluster of the u-state system and found a ratio of 0.446.

The above analysis shows that the specific contact pattern between the SEAC and K-rich segments originated from the major basin, Clst 1, of the p-state system. As indicated by the magenta frame of Figures 2b and 5b, those contacts accompanied a relatively compact conformation in ID_{SK}, which we call an "intra-ID_{SK} compact conformation". We presume that the entity of the dynamic model proposed by Jonker et al.[10] is the intra-ID_{SK} compact conformation.

4. DISCUSSION

The experiments by Jonker et al. (i.e., CSP, GST pulldown assays, electrophoretic mobility shift assays, and unwinding assays) demonstrated that the phosphorylation of the SEAC segment changes the state of ID_{SK} considerably.[10] However, those experiments are unable to specify what happened in ID_{SK}

at an atomic level because ID_{SK} is intrinsically disordered during the process of phosphorylation. Especially, their CSP experiment did not measure the variation of the contact between ID_{SK} and VP16ad. Jonker et al. then proposed a dynamic model for ID_{SK}:[10] the phosphorylated SEAC segment induces an intra-ID_{SK} segment contact, which is stabilized by the electrostatic interactions between the phosphoserine residues of the SEAC segment and the lysine residues of the K-rich segments. This proposition is important for the functional modulation of PC4 because VP16ad in the VP16ad−$(PC4ctd)_2$ complex is exposed to the solvent by the formation of the SEAC−K-rich contact, in contrast to the unmodified ID_{SK} covering the VP16ad surface. The exposed VP16ad is replaced by dsDNA because the binding affinity of dsDNA to the homodimer $(PC4ctd)_2$ is stronger than that of VP16ad to the homodimer. They have also demonstrated by experimentation that dsDNA bound to $(PC4ctd)_2$ is unwound.

For computational study, we consider that investigation of the following two things from a single simulation is crucially import: one is to check the variation of the ID_{SK}−VP16ad contact by the phosphorylation, which was not measured in the work,[10] as mentioned above, and the other is to investigate formation of the intra-ID_{SK} segment contact. Our simulations for the u-state and p-state systems indicated that phosphorylation reduces the ID_{SK}−VP16ad contact considerably, which results in the exposure of VP16ad to the solvent (Figure 1c → d). Furthermore, we demonstrated that an intra-ID_{SK} compact conformation is induced by the phosphorylation (Figures 3 and 5), whereas the conformation exhibited structural diversity (conformations in Figure 2b and 5b). The SEAC−K-rich contact was stabilized by the phosphoserine−lysine and phosphoserine−arginine interactions. These results strongly support the proposition by Jonker et al.[10] The TOC graphic schematically presents the induction of the intra-ID_{SK} compact conformation and the subsequent processes.

For this study, the phosphorylated residues introduced were S15, S17, and S19. The SEAC segment, however, has eight phosphorylatable serine residues: S9, S10, S11, S12, S13, S15, S17, and S19. In fact, the three residues treated in this study are likely to be first phosphorylated in the gradual phosphorylation because the phosphorylation proceeds from the C-terminal side (S19) to the N-terminal side (S9) of the SEAC segment.[10] The current study computationally explains the early stage of phosphorylation. In PIC, the exposed VP16ad might be replaced by dsDNA more readily than VP16ad covered by unmodified ID_{SK} (i.e., the nonexposed VP16ad); then, dsDNA is unwound, as presumed by Jonker et al.[10] As the phosphorylation proceeds, the exposure of VP16ad increases because more residues are involved in intra-ID_{SK} compact conformation. Finally, when all eight serine residues are fully phosphorylated, the largest intra-ID_{SK} compact conformation might be formed. Although we did not study those events occurring in the later stage of the gradual phosphorylation, formation of the intra-ID_{SK} compact conformation formed in the early stage is inferred as a key to elucidating the transcription process related to the VP16−PC4 interaction. Furthermore, results of the present study suggest that the simulation can provide useful atomistic detailed information to understand the functional switch of PC4 by phosphorylation in various systems.[23−28]

We detected artificial snapshots in which ID_{SK} contacted both VP16ad and the $(PC4ctd)_2$ homodimer (Figure S4) crossing the periodic boundary and eliminated those snap-

shots. We presume that the effect of the artificial contact on the VP16ad−(PC4ctd)$_2$ complex structure is negligible because the helicity of VP16ad was maintained during the simulation.

The force field selection is important for an MD simulation. A simulation study by the Shaw group tested some force fields, which have been used widely for protein simulation, and proposed a force field, AMBER99SB-disp, which is applicable effectively to both ordered and disordered proteins.[52] This study also reported that AMBER99SB-ILDN provides appropriate results to be compared with experimental results when the simulation system is an ordered protein or relatively short disordered peptides.[52] We used AMBER99SB-ILDN in our previous study,[36] where a flexible drug molecule was binding to the deep binding pocket of a GPCR. The study demonstrated that the disordered N-terminal tail of GPCR captures the drug molecule and that this capture is essentially important to lead the drug to GPCR's binding pocket. Importantly, the N-terminal tail was always disordered to capture the ligand in the simulation. The present simulation study showed again that AMBER99SB-ILDN produces simulation data supportive to experimental results.[10]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c00364.

> Construction of the computed system, initial conformation of the simulation, definition of two reaction coordinates, scheme showing an artificial snapshot, parameters for 2D RC space constructed by $\lambda^{(\alpha)}$ and $\lambda^{(\beta)}$, setting of zones along $\lambda^{(\alpha)}$ and $\lambda^{(\beta)}$, principal component analysis, and convergence of $Q^M_{cano}(i, j)$ (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Takuya Takahashi − *College of Life Sciences, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan;* ⓞ orcid.org/0000-0002-6782-9206; Email: tkhs@sk.ritsumei.ac.jp

Authors

Qilin Xie − *Graduate School of Life Sciences, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan*

Kota Kasahara − *College of Life Sciences, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan;* ⓞ orcid.org/0000-0003-0207-6271

Junichi Higo − *Graduate School of Information Science, University of Hyogo, Kobe, Hyogo 650-0047, Japan; Research Organization of Science and Technology, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan;* ⓞ orcid.org/0000-0002-0408-1907

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.3c00364

Author Contributions

Research project was proposed by Q.X. and K.K. The computer program preparation was done by K.K. The MD simulation was performed by Q.X. Analyses were performed by Q.X. and J.H.. Manuscript writing was done by Q.X., J.H., and T.T. The entire process of this research was supervised by T.T.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ ABBREVIATIONS

PC4, human positive cofactor 4; MD, molecular dynamics; IDR, intrinsic disordered region; VP16, herpes simplex virion protein 16; VP16ad, activation domain of herpes simplex virion protein 16; PC4ctd, the C-terminal structured core domain of PC4; PC4ntd, the N-terminal structured core domain of PC4; PIC, preinitiation complex; GTFs, general transcription factors; dsDNA, double-stranded DNA; ssDNA, single-stranded DNA; SEAC, serin and acidic-rich region of PC4ntd; CSP, chemical shift perturbation; ID$_{SK}$, intrinsically disordered segments consisting of the SEAC and K-rich regions; u-state, unmodified-state; p-state, phosphorylated-state; mD-VcMD, multidimensional virtual-system coupled molecular dynamics; RCs, reaction coordinates; PCA, principal component analysis; FEL, free-energy landscape

■ REFERENCES

(1) Roeder, R. G. 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. *Nat. Struct. Mol. Biol.* **2019**, *26*, 783−791.

(2) Bylino, O. V.; Ibragimov, A. N.; Shidlovskii, Y. V. Evolution of regulated transcription. *Cells* **2020**, *9*, 1675.

(3) Cousens, D. J.; Greaves, R.; Goding, C. R.; O'Hare, P. The C-terminal 79 amino acids of the herpes simplex virus regulatory protein, Vmw65, efficiently activate transcription in yeast and mammalian cells in chimeric DNA-binding proteins. *EMBO J.* **1989**, *8*, 2337−2342.

(4) Fan, D.; Wang, M.; Cheng, A.; Jia, R.; Yang, Q.; Wu, Y.; Zhu, D.; Zhao, X.; Chen, S.; Liu, M.; Zhang, S.; Ou, X.; Mao, S.; Gao, Q.; Sun, D.; Wen, X.; Liu, Y.; Yu, Y.; Zhang, L.; Tian, B.; Pan, L.; Chen, X. The role of VP16 in the life cycle of Alphaherpesviruses. *Front. Microbiol.* **2020**, *11*, 1910.

(5) Triezenberg, S. J.; Kingsbury, R. C.; McKnight, S. L. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev.* **1988**, *2*, 718−729.

(6) Hirai, H.; Tani, T.; Kikyo, N. Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *Int. J. Dev. Biol.* **2010**, *54*, 1589−1596.

(7) Milbradt, A.; Kulkarni, M.; Yi, T.; Takeuchi, K.; Sun, Z.-Y. J.; Luna, R. E.; Selenko, P.; Näär, A. M.; Wagner, G. Structure of the

VP16 transactivator target in the Mediator. *Nat. Struct. Mol. Biol.* **2011**, *18*, 410−415.

(8) Jonker, H. R. A.; Wechselberger, R. W.; Boelens, R.; Kaptein, R.; Folkers, G. E. The intrinsically unstructured domain of PC4 modulates the activity of the structured core through inter- and intramolecular interactions. *Biochemistry* **2006**, *45*, 5067−5081.

(9) Uhlmann, T.; Boeing, S.; Lehmbacher, M.; Meisterernst, M. The VP16 activation domain establishes an active mediator lacking CDK8 in vivo. *J. Biol. Chem.* **2007**, *282*, 2163−2173.

(10) Jonker, H. R. A.; Wechselberger, R. W.; Pinkse, M.; Kaptein, R.; Folkers, G. E. Gradual phosphorylation regulates PC4 coactivator function. *FEBS J.* **2006**, *273*, 1430−1444.

(11) Jonker, H. R. A.; Wechselberger, R. W.; Boelens, R.; Folkers, G. E.; Kaptein, R. Structural properties of the promiscuous VP16 activation domain. *Biochemistry* **2005**, *44*, 827−839.

(12) Langlois, C.; Mas, C.; Di Lello, P.; Jenkins, L. M. M.; Legault, P.; Omichinski, J. G. NMR Structure of the complex between the Tfb1 subunit of TFIIH and the activation domain of VP16: Structural similarities between VP16 and p53. *J. Am. Chem. Soc.* **2008**, *130*, 10596−10604.

(13) Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321−331.

(14) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197−208.

(15) Sugase, K.; Dyson, H. J.; Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **2007**, *447*, 1021−1025.

(16) Nomura, M.; Uda-Tochio, H.; Murai, K.; Mori, N.; Nishimura, Y. The neural repressor NRSF/REST binds the PAH1 domain of the Sin3 corepressor by using its distinct short hydrophobic helix. *J. Mol. Biol.* **2005**, *354*, 903−915.

(17) Higo, J.; Nishimura, Y.; Nakamura, H. A free-energy landscape for coupled folding and binding of an intrinsically disordered protein in explicit solvent from detailed all-atom computations. *J. Am. Chem. Soc.* **2011**, *133*, 10448−10458.

(18) Whitmarsh, A. J.; Davis, R. J. Regulation of transcription factor function by Phosphorylation. *Cell. Mol. Life Sci.* **2000**, *57*, 1172−1183.

(19) Riedl, T.; Egly, J.-M. Phosphorylation in transcription: The CTD and more. *Gene Expr.* **2000**, *9*, 3−13.

(20) Kobor, M. S.; Greenblatt, J. Regulation of transcription elongation by phosphorylation. *Biochim. Biophys. Acta* **2002**, *1577*, 261−275.

(21) Meggio, F.; Pinna, L. A. One-thousand-and-one substrates of protein kinase CK2? *FASEB J.* **2003**, *17*, 349−368.

(22) Kasahara, K.; Shiina, M.; Higo, J.; Ogata, K.; Nakamura, H. Phosphorylation of an intrinsically disordered region of Ets1 shifts a multi-modal interaction ensemble to an out-inhibitory state. *Nucleic Acids Res.* **2018**, *46*, 2243−2251.

(23) Ge, H.; Zhao, Y.; Chait, B. T.; Roeder, R. G. Phosphorylation negatively regulates the function of coactivator PC4. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12691−12695.

(24) Kaiser, K.; Stelzer, G.; Meisterernst, M. The coactivator p15 (PC4) initiates transcriptional activation during TFIIATFIID-promoter complex formation. *EMBO J.* **1995**, *14*, 3520−3527.

(25) Rojas, D. A.; Urbina, F.; Solari, A.; Maldonado, E. The catalytic subunit of *Schizosaccharomyces pombe* CK2 (Cka1) negatively regulates RNA Polymerase II transcription through phosphorylation of positive cofactor 4 (PC4). *Int. J. Mol. Sci.* **2022**, *23*, 9499.

(26) Mustafi, P.; Hu, M.; Kumari, S.; Das, C.; Li, G.; Kundu, T. K. Phosphorylation-dependent association of human chromatin protein PC4 to linker histone H1 regulates genome organization and transcription. *Nucleic Acids Res.* **2022**, *50*, 6116−6136.

(27) Brzek, A.; Cichocka, M.; Dolata, J.; Juzwa, W.; Schümperli, D.; Raczynska, K. D. Positive cofactor 4 (PC4) contributes to the regulation of replication-dependent canonical histone gene expression. *BMC Mol. Biol.* **2018**, *19*, 9 DOI: 10.1186/s12867-018-0110-y.

(28) Dhanasekaran, K.; Kumari, S.; Boopathi, R.; Shima, H.; Swaminathan, A.; Bachu, M.; Ranga, U.; Igarashi, K.; Kundu, T. K.

Multifunctional human transcriptional coactivator protein PC4 is a substrate of Aurora kinases and activates the Aurora enzymes. *FEBS J.* **2016**, *283*, 968−985.

(29) Higo, J.; Ikebe, J.; Kamiya, N.; Nakamura, H. Enhanced and effective conformational sampling of protein molecular systems for their free energy landscapes. *Biophys. Rev.* **2012**, *4*, 27−44.

(30) Kasahara, K.; Terazawa, H.; Takahashi, T.; Higo, J. Studies on molecular dynamics of intrinsically disordered proteins and their fuzzy complexes: a mini-review. *Computat. Struct. Biotechnol. J.* **2019**, *17*, 712−720.

(31) Higo, J.; Kusaka, A.; Kasahara, K.; Kamiya, N.; Hayato, I.; Qilin, X.; Takahashi, T.; Fukuda, I.; Mori, K.; Hata, Y.; Fukunishi, Y. GA-guided mD-VcMD: A genetic-algorithm-guided method for multi-dimensional virtual-system coupled molecular dynamics. *Biophys. Physicobiol.* **2020**, *17*, 161−176.

(32) Higo, J.; Takashima, H.; Fukunishi, Y.; Yoshimori, A. Generalized-Ensemble Method Study: A helix-mimetic compound inhibits protein-protein interaction by long-range and short-range intermolecular interactions. *J. Comput. Chem.* **2021**, *42*, 956−969.

(33) Hayami, T.; Kamiya, N.; Kasahara, K.; Kawabata, T.; Kurita, J.; Fukunishi, Y.; Nishimura, Y.; Nakamura, H.; Higo, J. Difference of binding modes among three ligands to a receptor mSin3B corresponding to their inhibitory activities. *Sci. Rep.* **2021**, *11*, 6178.

(34) Higo, J.; Kawabata, T.; Kusaka, A.; Kasahara, K.; Kamiya, N.; Fukuda, I.; Mori, K.; Hata, Y.; Fukunishi, Y.; Nakamura, H. Molecular interaction mechanism of a 14−3-3 protein with a phosphorylated peptide elucidated by enhanced conformational sampling. *J. Chem. Inf. Model.* **2020**, *60*, 4867−4880.

(35) Higo, J.; Kasahara, K.; Wada, M.; Dasgupta, B.; Kamiya, N.; Hayami, T.; Fukuda, I.; Fukunishi, Y.; Nakamura, H. Free-energy landscape of molecular interactions between endothelin 1 and human endothelin type B receptor: Fly-casting mechanism. *Protein Eng. Des. Sel.* **2019**, *32*, 297−308.

(36) Higo, J.; Kasahara, K.; Bekker, G.-J.; Ma, B.; Sakuraba, S.; Iida, S.; Kamiya, N.; Fukuda, I.; Kono, H.; Fukunishi, Y.; Nakamura, H. Fly casting with ligand sliding and orientational selection supporting complex formation of a GPCR and a middle sized flexible molecule. *Sci. Rep.* **2022**, *12*, 13792.

(37) Marti-Renom, M. A.; Stuart, A.; Fiser, A.; Sánchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291−325.

(38) *The PyMol Molecular Graphics System*, ver. 2.0; Schrödinger, LLC, 2020. http://www.pymol.org/pymol.

(39) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct. Funct. Genet.* **2010**, *78*, 1950−1958.

(40) Joung, I. S.; Cheatham, T. E., III Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.

(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−936.

(42) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald potential. *J. Chem. Phys.* **1995**, *103*, 8577−8592.

(43) Kutzner, C.; Páll, S.; Fechner, M.; Esztermann, A.; de Groot, B. L.; Grubmüller, H. More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *J. Comput. Chem.* **2019**, *40*, 2418−2431.

(44) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(45) Higo, J.; Kamiya, N.; Sugihara, T.; Yonezawa, Y.; Nakamura, H. Verifying trivial parallelization of multicanonical molecular dynamics for conformational sampling of a polypeptide in explicit water. *Chem. Phys. Lett.* **2009**, *473*, 326−329.

(46) Ikebe, J.; Umezawa, K.; Kamiya, N.; Sugihara, T.; Yonezawa, T.; Takano, Y.; Nakamura, H.; Higo, J. Theory for trivial trajectory

parallelization of multicanonical molecular dynamics and application to a polypeptide in water. *J. Comput. Chem.* **2011**, *32*, 1286−1297.

(47) Pearson, K. On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **1901**, *2*, 559−572.

(48) Higo, J.; Galzitskaya, O. V.; Ono, S.; Nakamura, H. Energy landscape of a β-hairpin peptide in explicit water studied by multicanonical molecular dynamics. *Chem. Phys. Lett.* **2001**, *337*, 169−175.

(49) Higo, J.; Ito, N.; Kuroda, M.; Ono, S.; Nakajima, N.; Nakamura, H. Energy landscape of a peptide consisting of α-helix, 3_10-helix, β-turn, β-hairpin, and other disordered conformations. *Protein Sci.* **2001**, *10*, 1160−1171.

(50) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.

(51) Murtagh, F. A survey of recent advances in hierarchical clustering algorithms. *Comput. J.* **1983**, *26*, 354−359.

(52) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E4758−E4766.