Molecular interaction mechanism of 14-3-3ε protein with phosphorylated Myeloid leukemia factor 1 revealed by an enhanced conformational sampling

Junichi Higo,[1,] Takeshi Kawabata,[2] Ayumi Kusaka,[1] Kota Kasahara,[3] Narutoshi Kamiya,[1] Ikuo Fukuda,[1] Kentaro Mori,[1,4] Yutaka Hata,[1] Yoshifumi Fukunishi,[5] Haruki Nakamura.[2]*

[1]Graduate School of Simulation Studies, University of Hyogo, 7-1-28 Minatojima Minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

[2]Institute for Protein Research, Osaka University, 3-2 Yamada-oka, Suita, Osaka 565-0871, Japan

[3]College of Life Sciences, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan

[4]Maizuru College, 234 Shiroya, Maizuru, Kyoto, 625-8511 Japan

[5]Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 2-3-26, Aomi, Koto-ku, Tokyo, 135-0064, Japan

ABSTRACT: An enhanced conformational sampling, a genetic-algorithm-based multi-dimensional virtual-system coupled molecular dynamics (GA-based mD-VcMD), was applied to a peptide, phosphorylated Myeloid leukemia factor 1 (pMLF1), binding to a receptor protein, 14-3-3ε, in an explicit solvent. This method provides an equilibrated ensemble at room temperature, from which a density map of pMLF1 around 14-3-3ε can be calculated. At a low-density level, pMLF1 distributed everywhere around 14-3-3 ε involving pMLF1 free from or slightly contacting to 14-3-3ε. With increasing the density level, four clusters were identified on the 14-3-3ε surface, and free-energy barriers existed among some clusters. At a further high-density level, a single cluster remained, which was characterized by the inter-molecular contacts commonly observed in the crystal structure, and the molecular orientation of pMLF1 was ordered as in the crystal structure. Increasing the density level, the cluster mentioned above was split into two high-density spots (semi-native-like and native-like spots) connected by a narrow corridor. When pMLF1 passing the corridor, salt-bridge relay (switching) regarding the phosphorylated residue of pMLF1 was occurred. The current study supports an amino acid mutagenesis experiment, which was conducted to measure the complex-structure stability. We also performed a simulation of non-phosphorylated MLF1, and showed that the complex

structure was unstable, suggesting the important role of phosphorylation of MLF1 through activation of 14-3-3ε in a variety of human diseases.

INTRODUCTION

The 14-3-3 proteins are eukaryotic adapter proteins that binds short phosphorylated sequences (14-3-3 interaction sequences).[1-3] They bind to a large number of partner proteins, usually by recognition of a phosphoserine or phophothreonine motif. In human, there are seven 14-3-3 proteins: β, γ, ε, η, σ, τ, and ζ, which are quite similar to one another, among which the sequence identities are more than 60 %. Although several binding motif patterns have been proposed for the 14-3-3 interaction sequences, such as RSXpSXP or RXY/FXpSXP (pS is a phosphorylated serine, and denoted as Sep in three-letter code), the mechanism of their binding specificities is not clearly understood yet, except the salt bridges at the phosphorylated sites. The 14-3-3 interaction sequences that bind to the human 14-3-3ε and ζ proteins are listed, respectively, in Figs. S1A and S1B of Supporting Information. The 14-3-3 proteins are related to many human diseases. For example, they bind to myeloid leukemia factor 1 (MLF1), which has been reported to be overexpressed in acute myeloid leukemia.[4,5] Kimura et al reported that a mutation of nuclear distribution E homolog 1 (NDE1) had a significant association with schizophrenia, and the 14-3-3ε protein bound to the mutated site of NDE1.[6]

Molzan et al. solved an X-ray crystal structure (PDB ID: 3ual) of the human 14-3-3ε protein bound to a phosphorylated peptide (RSFpSEPFG) taken from myeloid leukemia factor 1 (MLF1).[7] They also measured dissociation constants K_D of alanine-mutated MLF1 peptides at the three residue positions -1, $+1$, and $+3$ (F33A, E35A, and F37A, respectively) relative to the phosphorylated-serine position (Sep 34) using the isothermal titration calorimetry (ITC), because these three amino-acid residues, Phe 33, Glu 35, and Phe 37, are not found in any other structurally known 14-3-3/target protein interaction. Interestingly, only the mutation F33A led to a significant decrease in affinity, which might be due to the loss of hydrophobic interactions by Phe 33 sandwiched between two hydrophobic residues Leu 223 and Leu 230 of the 14-3-3ε protein found in the X-ray crystal structure. In contrast, E35A and F37A exhibited almost no change in K_D. Phe 37 is located on the surface of 14-3-3ε in the complex forming a weak hydrophobic contact to the stem of Lys 40 of 14-3-3ε. This can be a reason for the negligibly small change of K_D by F37A. In the X-ray crystal structure, however, Glu 35 is located at a position that can form a salt bridge to Lys 50 or Lys 123 of 14-3-3ε. Therefore, it is difficult to explain the reason of the small change of K_D by E35A from the crystal structure.

Molzan et al. also pointed out that the residues at position -1 of six structurally-elucidated 14-3-3 interaction sequences, binding to 14-3-3 proteins, overlap well to one another, whereas the residues at position $+3$ are quite divergent. Figures S1B and S1C display the superimposed 29 known binding peptides presented in Figs. S1A and S1B using a software MATRAS.[8] To analyze the structural diversity quantitatively, we present Fig. S1E, which demonstrates clearly that the structural diversity increases as separating away from Sep 34. This suggests that a conserved region may play an important role in binding to the 14-3-3 proteins.

The superimposed structures (Figs. S1B and S1C) present that the situation for the position $+1$ is complicated. The main chain atoms for the position $+1$ are well conserved, however, the carboxyl group of Glu35 of MLF1 does not overlap with any other binding peptides. The diversity of the position $+1$ is similar to the position -1 (Fig. S1E). Furthermore, Glu35 forms two salt bridges to Lys 50 and Lys 123 of 14-3-3ε in the crystal structure. These observations

suggest that Glu35 may contribute to the stability of the complex structure considerably. However, as mentioned above, the mutation E35A results in the negligibly small changes of K_D.

Because the phosphorylated MLF1 segment used in the X-ray crystallography[7] and the ITC measurement was a short segment taken from the full length MKF1, it is likely that this segment is highly flexible in the free state, and that the encounter complex with 14-3-3ε is in a mixture of various conformations. Therefore, we presume that contribution of residues to the thermodynamic stability of the complex structure is determined not only by the most stable structure (i.e., the crystal structure) but also by semi-stable conformations (encounter complexes). We presume that the small contribution of Glu 35 to the stability may be understood by investigating the conformational flexibility, and that molecular dynamics (MD) simulation is useful to understand fully such a fragile system.

MD simulation can produce thermal motions of a molecular system at an atomic resolution. By using an atomistic model to express proteins and by taking solvent molecules into account explicitly, the obtained molecular motions become realistic. However, the sampling is usually limited around its initial conformation of simulation because of its heavy computational load. Thus, the conventional MD protocol is less effective to obtain a conformational ensemble equilibrated at a temperature (canonical ensemble), which covers the wide conformational space consisting of the unbound state, fragile encounter complexes, and the stable complex. The recent development in computer hardware is making a very-long simulation possible.[9,10] However, an equilibrated conformational ensemble at a room temperature is still beyond ability of these powerful computers as far as an ordinary MD simulation is performed, because statistical weight (thermodynamic weight) should be assigned to the sampled conformations (snapshots) that tend to be rather localized in a conformational space.

To increase sampling efficiency and to assign the statistical weight to the snapshots, generalized ensemble methods[11,12] such as the multicanonical sampling,[13-16] the replica exchange[17,18] method, and their variants[19-25] have been developed. These methods enhance sampling along a reaction coordinate, which is computable from the coordinates of the system's conformation such as the potential energy function. Although these methods are powerful to explore considerably stable states (major basins), they might overlook less stable states (minor basins) when these basins, which are separated to one another in the full-dimensional conformational space, overlap in the reaction-coordinate axis,[26,27] which may be called "degeneration of basins". As a result, conformational trapping may occur, and the minor basins may be overlooked. This oversight is not problematic when the minor basins are out of scope. However, an alternative approach is required when the minor basins act as bridges among the major basins or have key features for exerting molecular functions of the system.

Adaptive umbrella sampling[28,29] has a possibility to avoid such an oversight or trapping problem because one can set the reaction coordinate so that the major and minor basins are discriminated along the reaction-coordinate axis. However, this method requires fine tuning of a weight function (potential of mean force along the introduced reaction coordinate). Practically, the difficulty of the tuning may require a very long simulation (or increment of the number of iterative simulations) to treat a complicated system.

Recently, we introduced a new method, virtual-system coupled sampling method,[30] and combined it to generalized ensemble methods, multicanonical[31] and adaptive umbrella sampling.[26,32] These methods provide an equilibrated ensemble of the system.[33-37] However, they still require the fine tuning of the weight function to proceed iterations. Thus, based on the experiences from the above methods, we introduced a multi-dimensional virtual-system coupled

molecular dynamics (mD-VcMD) simulation.[38,39] This method is free from the fine tuning of the weight function because the weight function is not required to proceed iterations. We further introduced a sampling method, the genetic-algorithm-based mD-VcMD (GA-based mD-VcMD), as an extension of mD-VcMD (to be reported elsewhere). The GA-based method is also free from weight function. Besides, in the GA-based method, a concept of production run does not exist: Namely, all snapshots from all iterations are collected to construct the canonical ensemble. A simple protocol is available to assign a thermodynamic weight to snapshots in spite that the weight function is not used in the simulation. If a poorly-sampled region emerges in sampling, selective sampling can be performed focusing on the poorly-sampled region without breaking the proportion of the canonical ensemble.

In the present study, we performed the GA-based mD-VcMD simulations of a system consisting of 14-3-3ε protein and the MLF1 segment in an explicit solvent, where three reaction coordinates are introduced to control the molecular motions. Two types of MLF1, phosphorylated and non-phosphorylated MLF1, were examined. We obtained the contour maps of conformational distributions equilibrated at 300 K, which was comprised of a variety of states, i.e., unbound state, encounter-complex states, and the native-like complex state. The features of the distribution were considerably different between the two systems examined. Depending on the contour level to display the conformational distribution, various structural clusters were identified. We investigated the computed conformational distributions of Phe 33, Sep 34/Ser 34, Glu 35, and Phe 37 of MLF1 around the 14-3-3ε protein, and showed that the conformational distributions support the stability of the complex structure that was observed by experiments. Finally, we discuss a putative molecular binding mechanism between 14-3-3ε and MLF1.

METHODS

GA-based mD-VcMD.

We applied the GA-based mD-VcMD method to calculate density-contour maps (free-energy landscapes) of the molecular interactions between the ligand, MLF1, and the receptor, 14-3-3ε, at a room temperature (300 K) in a multidimensional (mD) reaction-coordinate (RC) space. Details of the GA-based mD-VcMD method is explained elsewhere. Here, we explain an outline for the method.

We introduce three RCs $\lambda^{(h)}$ $(h = \alpha, \beta, \gamma)$ in this study. Thus, "multi-dimensional" means three-dimensional (3D) here. An RC is defined by the inter-mass-center distance between two atom groups denoted by G_h^A and G_h^B. Superscripts A and B are introduced to express that the two atom groups are pairing to define a single RC $\lambda^{(h)}$. See Section 1 of Supporting Information (SI) for more details of the RC definition. The mD-VcMD controls the system's conformational motions at 300 K in the 3D space constructed by $\lambda^{(\alpha)}$, $\lambda^{(\beta)}$, and $\lambda^{(\gamma)}$ axes, and generates a conformational ensemble at equilibrated at 300 K that covers a wide 3D-RC space. The actual definition of RCs used for the present study is given in Table S1 of SI.

The GA-based mD-VcMD consists of iterative simulations, and produces a canonical ensemble with collecting snapshots from all the iterations. Importantly, the thermodynamic weight is assigned to each snapshot. If a quantity is defined by the atomic coordinates of the system, we can calculate a density-contour map of the quantity, using the canonical ensemble. The actual parameters used for mD-VcMD are given in Section 2 of SI.

Systems studied.

First, the X-ray crystal structure of the complex of 14-3-3ε with the phosphorylated MLF1 (denoted as pMLF1), is taken from the PDB data (ID: 3ual).[7] Residue ordinal numbers used in this study accord to those in the paper,[7] which are not the same as those in the PDB data. The

receptor 14-3-3ε forms a homodimer in the unbinding state, as shown in Fig. S3A schematically. The ligand pMLF1, which is MLF1 phosphorylated at serine 34 (Sep 34), binds to each of the two 14-3-3ε molecules (Fig. S3B), where Sep 34 of pMLF1 forms salt bridges to Arg 57 and Arg 130 of 14-3-3ε (Fig. S4). To reduce the computational load, in the present study, we used a single complex of 14-3-3ε with pMLF1 (Fig. S3C). The N- and C-termini of both 14-3-3ε and pMLF1 were capped by acetyl and N-methyl groups, respectively. The single complex taken from 3ual was immersed in a periodic water box ($86.2 \times 86.2 \times 90.8$ Å^3) under periodic boundary condition, and sodium and chlorine ions were introduced to neutralize the net charge of the entire system at a physiological ionic concentration. The resultant system consists of 65,776 atoms (3,700 atoms for 14-3-3ε; 137 atoms for pMLF1; 20,612 molecules for water; 46 chlorine ions and 57 for sodium ions).

Although the pMLF1 used for the experiment[7] was a segment of 14 residues long, the two N- and four C-terminal residues were not determined. Then, we treated the eight-residue determined segment in the present study: Ace-Arg-Ser-Phe-Sep^{34}-Glu-Pro-Phe-Gly-Nme, where Ace and Nme are the acetyl and N-methyl groups, and where the superscript "34" indicates the position of Sep 34. The receptor 14-3-3ε consists of 230 amino-acid residues (see PDB: 3ual). We refer to the current system consisting of the capped pMLF1 and 14-3-3ε as a pMLF1/14-3-3ε system, whether the two molecules form a complex or not.

We also computed a system of 14-3-3ε and the non-phosphorylated MLF1 (denoted as npMLF1), in which serine 34 is non-phosphorylated, and we refer to this system as an npMLF1/14-3-3ε system. To assess interactions between npMLF1 and 14-3-3ε, we performed the GA-based mD-VcMD of this system as well as the pMLF1/14-3-3ε system. First, we replaced Sep at the 34-th residue of pMLF1 by serine in the single complex of pMLF1/14-3-3ε. Next, this complex was immersed in the water box (the same box size as that for the pMLF1/14-3-3ε system), and ions were introduced to neutralize as having done for the pMLF1/14-3-3ε system. The resultant system consists of 65,771 atoms (3,700 atoms for 14-3-3ε; 134 atoms for npMLF1; 20,612 molecules for water; 46 chlorine ions and 55 for sodium ions). The amino-acid sequence of npMLF1 is: Ace-Arg-Ser-Phe-Ser^{34}-Glu-Pro-Phe-Gly-Nme, where the superscript "34" indicates the position of Ser 34.

The pMLF1/14-3-3ε complex structure generated above is the same as the crystal structure. On the other hand, a stable complex is not reported for the npMLF1/14-3-3ε system. However, the npMLF1/14-3-3ε complex structure generated above adopts the crystal structure of the pMLF1/14-3-3ε complex except for the sidechain of Ser 34. Thus, this structure is also called the crystal structure for convenience.

We aim to search the conformational space of both systems starting from randomized ligand's conformations of pMLF1 or npMLF1. The procedure for the randomization is explained in the next section.

Generation of initial conformations of GA-based mD-VcMD and setting RCs.

For both the pMLF1/14-3-3ε and npMLF1/14-3-3ε systems, after a short energy minimization of the above generated complex structure, a short constant-volume and constant-temperature (300 K) simulation (NVT simulation) was performed. Then, the system size was relaxed by a constant-pressure (1 atm) and constant-temperature simulation (NPT simulation). The resultant box size was $84.70 \times 84.67 \times 89.24$ Å^3 and $84.72 \times 84.69 \times 89.27$ Å^3 for the pMLF1/14-3-3ε and npMLF1/14-3-3ε systems, respectively. The resultant structures were still close to the crystal structure (data not shown). Therefore, we generated conformations where pMLF1 or npMLF1 is

far from the binding site of 14-3-3ε as explained below. The word "initial conformation" used in the present study represents a conformation in that pMLF1 or npMLF1 is apart from the pMLF1 binding site of 14-3-3ε.

Before explaining the procedure to generate the initial conformations, we explain three RCs used for the current GA-based mD-VcMD. The three RCs $\lambda^{(\alpha)}$, $\lambda^{(\beta)}$, and $\lambda^{(\gamma)}$ are defined as the inter-mass-center distances between the two atom groups, G_h^A and G_h^B $(h = \alpha, \beta, \gamma)$ shown in Fig. 1 and Table S1 of SI. The middle RC, $\lambda^{(\beta)}$, is the distance between the Sep 34/Ser 34 $(= G_\beta^B)$ of pMLF1 or npMLF1, and a helical segment $(= G_\beta^A)$ of 14-3-3ε near Sep 34/Ser 34 in the crystal structure. This distance controls the motion of Sep 34/Ser 34 and consequently the salt-bridge formation between 14-3-3ε and pMLF in the pMLF1/14-3-3ε system, whereas G_β^A does not involve the counter part of the salt bridge. The first and third RCs, $\lambda^{(\alpha)}$ and $\lambda^{(\gamma)}$, control the motions of the N- and C-termini of pMLF1 or npMLF1. In general, the smaller the RCs, the closer the conformation to the crystal structure. In fact, the RCs for the crystal structure are $\lambda^{(\alpha)} = 15.6$ Å, $\lambda^{(\beta)} = 7.9$ Å, and $\lambda^{(\gamma)} = 8.7$ Å.

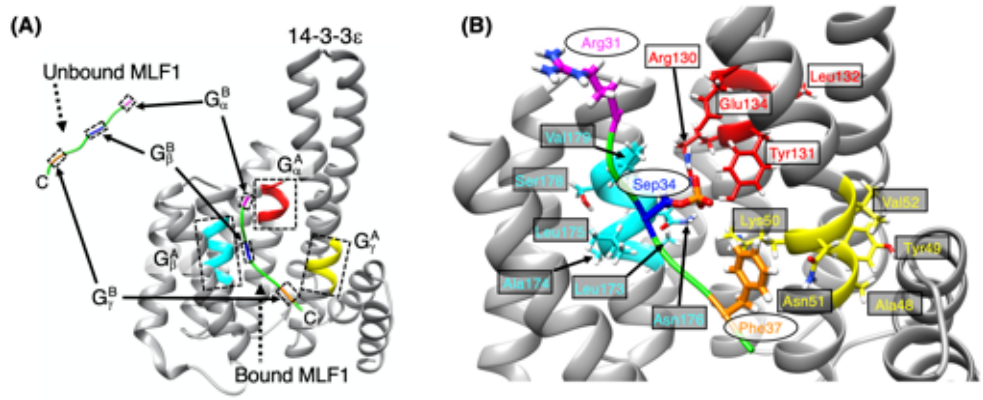

Figure 1. Atom groups to define three RCs $\lambda^{(h)}(h = \alpha, \beta, \gamma)$ for pMLF1/14-3-3ε system. (A) Receptor 14-3-3ε and ligand pMLF1 are shown in gray and green, respectively. Bound and unbound forms of pMLF1 are presented. $\lambda^{(\alpha)}$ is defined by mass-center distance between two atom groups G_α^A (red-colored fragment) and G_α^B (magenta), $\lambda^{(\beta)}$ by G_β^A (cyan) and G_β^B (blue), and $\lambda^{(\gamma)}$ by G_γ^A (yellow) and G_γ^B (golden yellow). Each atom group is surrounded by a rectangle. See Section 1 of SI for definition of atom groups. Character "C" is assigned to the C-terminal of pMLF1. (B) Sidechains involved in the atom groups are shown explicitly with the same coloring for panel (A). Residues for pMLF1 are shown in ovals and those for 14-3-3ε are done in rectangles. The npMLF1/14-3-3ε system is not shown. See Table S1 of SI for details of atom groups for both systems. Three residues Ala 133, Asn 176, and Phe 177 of 14-3-3ε are not shown because those are located at positions behind.

To generate the initial conformation, we applied weak forces along the three RCs, by which the RCs fell in the following ranges: 38 Å $\leq \lambda^{(h)} \leq$ 40 Å, starting from the crystal structure. For both systems, we performed 256 runs with different initial velocities of atoms, and used the final snapshots for the initial conformations of the GA-based mD-VcMD. In the resultant snapshots, pMLF1 or npMLF1 was away from the binding site of 14-3-3ε, and slightly contacting to or unbound from 14-3-3ε. Figure S5 of SI displays some of the initial conformations randomly picked from the 256 initial conformations.

Simulations.

Although the current system is a single heterodimer (Fig. S3C of SI), 14-3-3ε maintained its structure in preliminary NVT simulations (data not shown). Therefore, no structural restraints were introduced to keep the 14-3-3ε structure during the GA-based mD-VcMD simulations. As explained in the previous section, we performed 256 runs in parallel for the both systems starting from the randomized conformations generated in the above section.

The GA-based mD-VcMD simulation was performed using a computer program omegagene/myPresto[40] with the following condition: The SHAKE algorithm[41] to fix the covalent-bond lengths related to hydrogen atoms, the Berendsen thermostat to control temperature,[42] the zero-dipole summation method[43-45] for long-range electrostatic computations, a time-step of 2 fs ($\Delta t = 2$ fs), and simulation temperature of 300 K. An ensemble from the Berendsen thermostat can approximate a canonical distribution for a system of many atoms, whereas it generates a non-physical distribution for a small system.[46] To compute the original potential energy of the system, the Amber hybrid force fields[47] (mixture parameter $w = 0.75$) was used for pMLF1, npMLF1 and 14-3-3ε, the TIP3P model for water molecule,[48] and the Joung–Cheatham model for chloride and sodium ions.[49]

We performed 256 runs in parallel to raise the sampling efficiency further. This procedure, called a trivial trajectory-parallelization technique,[50,51] performs many independent runs in parallel from different initial conformations.

RESULTS

mD-VcMD simulation.

We repeated 30 and 14 iterations for the pMLF1/14-3-3ε and npMLF1/14-3-3ε systems, respectively. To sample carefully the electrostatic interactions between Sep 34 of pMLF1 and charged residues in 14-3-3ε, we did more iterations for pMLF1/14-3-3ε than npMLF1/14-3-3ε. Each of 256 runs in an iteration was performed for 2×10^6 steps (4 ns). Thus, the total simulation length was 30.720 μs (= $30 \times 256 \times 4$ ns) and 14.336 μs (= $14 \times 256 \times 4$ ns) for pMLF1/14-3-3ε and npMLF1/14-3-3ε, respectively. A snapshot was stored every 1×10^5 steps (200 ps), yielding 153,600 and 71,680 snapshots for pMLF1/14-3-3ε and npMLF1/14-3-3ε, respectively.

Contour maps in 3D-RC space.

We calculated a density function, denoted as $\rho_{3D-RC}(\boldsymbol{\lambda}) = \rho_{3D-RC}(\lambda^{(\alpha)}, \lambda^{(\beta)}, \lambda^{(\gamma)})$, in the 3D-RC space from the obtained ensemble: A probability of detecting the system in a RC volume of $d\boldsymbol{\lambda} = d\lambda^{(\alpha)} d\lambda^{(\beta)} d\lambda^{(\gamma)}$ is given by $\rho_{3D-RC}(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$. Figure 2 shows the two resultant contour maps for pMLF1/14-3-3ε (A) and npMLF1/14-3-3ε (B). The highest density region (i.e., the lowest free-energy region or the most thermodynamically stable region) was different between the two systems. In pMLF1/14-3-3ε, as shown later, the most stable region involved complex structures similar to the crystal structure. In npMLF1/14-3-3ε, the most stable region consists of conformations where npMLF1 and 14-3-3ε were unbound or slightly contacting. Later, we show that this difference in the most stable conformation was originated from presence or absence of the salt-bridges between MLF1 and 14-3-3ε.

Whereas the two systems presented different landscapes, two clusters existed in common. We named them C_1 and C_2 for pMLF1/14-3-3ε and C_1' and C_2' for npMLF1/14-3-3ε. The clusters C_1 and C_2 correspond roughly to C_1' and C_2', respectively. In either system, two clusters are connected by a corridor. The cluster C_1 is more stable than C_2, whereas C_1' is considerably less stable than C_2'. The cluster C_1' exists marginally as a cluster.

Although the two systems differ by one residue (Sep or Ser), the system's stability was different largely. Thus, phosphorylation of MLF1 strongly governs the interaction scheme with 14-3-3ε protein, and controls its physiological functions in cells relating to many different kinds of human diseases. In the next section, we further analyze the conformational distribution to uncover the role of Sep 34 in stabilization of the complex structure.

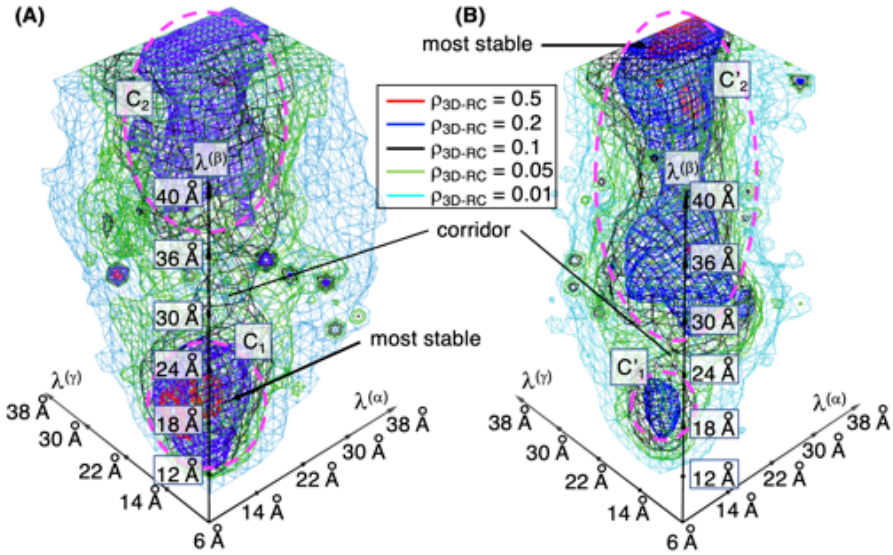

Figure 2. Distribution $\rho_{3D-RC}(\lambda^{(\alpha)},\lambda^{(\beta)},\lambda^{(\gamma)})$ of the systems in the 3D-RC space for (A) pMLF1/14-3-3ε and (B) npMLF1/14-3-3ε. Distribution is defined originally by $Q_{cano}(\boldsymbol{L})$, where $\boldsymbol{L} = [L^{(\alpha)}, L^{(\beta)}, L^{(\gamma)}]$ is a 3D index to specify a 3D zone (See Table S3 of SI). We converted $\boldsymbol{L}$ to $\boldsymbol{\lambda}$ as $\lambda^{(h)} = 0.5 \times \{[\zeta^{(h)}_{L^{(h)}}]_{min} + [\zeta^{(h)}_{L^{(h)}}]_{max}\}$ $(h = \alpha, \beta, \gamma)$, and expressed the distribution by $\rho_{3D-RC}(\lambda^{(\alpha)},\lambda^{(\beta)},\lambda^{(\gamma)})$, whose contour levels are shown in inset. The distribution is normalized as $\int \rho_{3D-RC}(\lambda^{(\alpha)},\lambda^{(\beta)},\lambda^{(\gamma)})\, d\lambda^{(\alpha)} d\lambda^{(\beta)} d\lambda^{(\gamma)} = 1$. Clusters are shown by broken-line circles with labels C_1, C_2, C'_1, and C'_2. Corridors connecting C_1 and C_2 as well as C'_1 and C'_2 are indicated by "corridor". Positions of the most stable (highest density) regions are indicated by "most stable". The RCs for the crystal structure are: $\lambda^{(\alpha)} = 15.6$ Å, $\lambda^{(\beta)} = 7.9$ Å, and $\lambda^{(\gamma)} = 8.7$ Å.

Here, in order to assess the conformational difference of pMLF1 (or npMLF1) between a snapshot and the crystal structure, we defined a measure $RMSD_{lig}$: The mainchain heavy-atomic root mean square deviation for residues 33–37 in pMLF1 (or npMLF1) after superposition of only 14-3-3ε between the snapshot and the crystal structure. Thus, $RMSD_{lig}$ reflects not only the internal motions of pMLF1 (or npMLF1) but also its translational and rotational motions relative to 14-3-3ε. The two residues Phe 33 and Phe 37 are specific in pMLF1 in comparison with other 14-3-3 interaction sequences, as mentioned in the Introduction–section.

Figures 3A-C illustrate conformations of the pMLF1/14-3-3ε system taken from clusters C_1 and C_2. Panel A displays the pMLF1 conformation with the smallest $RMSD_{lig}$ (0.69 Å). The sidechain positions of Sep 34 and Phe 37 coincided well with those in the crystal structure shown in yellow. Although Phe 33 was deviated from the crystal-structure position, this residue interacted with Leu 230 via a hydrophobic contact, as discussed later. Panel B is a conformation randomly picked from cluster C_1 ($RMSD_{lig} = 4.0$ Å). Whereas Phe 33 and Phe 37 were deviated from the crystal-structure positions, the position of Sep 34 agreed well with the crystal-structure position. Conformations in Fig. 3C are those from cluster C_2, where pMLF1 was far from the crystal-structure position, and weakly contacting to 14-3-3ε or isolated in solution.

Figures 3D-F display conformations of the npMLF1/14-3-3ε system from C'_1 and C'_2. Panel D is the smallest $RMSD_{lig}$ conformation in cluster C'_1 ($RMSD_{lig} = 1.60$ Å) and panel E is one randomly picked from cluster C'_1 ($RMSD_{lig} = 3.3$ Å). Three conformations in panel F are those randomly picked from cluster C'_2.

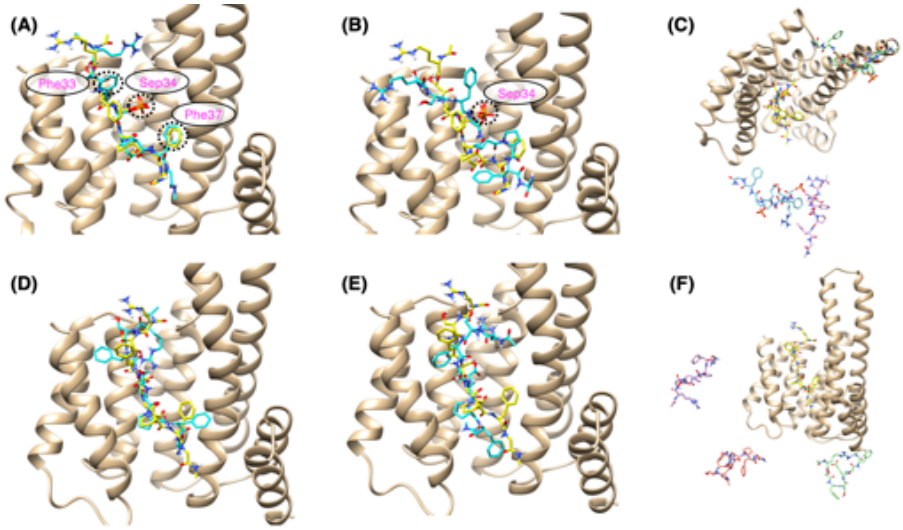

Figure 3. Panels (A)-(C) display conformations from the pMLF1/14-3-3ε system, and (D)-(F) do those from the npMLF1/14-3-3ε system. Yellow-colored MLF1 conformations in (A)-(C) and (D)-(F) are the equilibrated ones (NPT simulation) of the crustal structure. Residues labeled by ovals belongs to pMLF1. The pMLF1 in (A) is the smallest-$RMSD_{lig}$ conformation from cluster C_1, and one in (B) is a randomly picked conformation from cluster C_1. Residues Phe 33, Sep 34 and Phe 37 in (A) and Sep 34 in (B) are indicated by broken-line circles. Three conformations in (C) are those picked randomly from cluster C_2 of pMLF1. Panel (D) is the smallest $RMSD_{lig}$ conformation from cluster C_1', and (E) is one randomly picked conformation from cluster C_1'. Three conformations in (F) are those randomly picked from cluster C_2'.

Distribution of MLF1 around 14-3-3ε at 300 K.

As described in the Introduction–section and in Ref. 7, Phe 33, Sep 34, Glu 35, and Phe 37 of pMLF1 are characteristic residues thermodynamically, structurally, or in sequence comparison to other 14-3-3 interaction sequences (Fig. S1). Needless to say, Sep 34 is critical for the complex formation. Molzan *et al*. reported that Phe 33 contributed largely to the complex-structure stabilization in contrast to Glu 35 and Phe 37, which did not contribute to the stability. Therefore, we focused on the following four inter-residue distances: The first distance r^{Phe33} is the minimum distance from the ring center of Phe 33 of pMLF1 (or npMLF1) to the sidechain heavy atoms of Leu 223 and Leu 230 of 14-3-3ε. The second distance r^{Sep34} (or r^{Ser34}) is the minimum distance from the O□ atom of Sep 34 of pMLF1 (or Ser 34 of npMLF1) to the Cζ atoms of Arg 57 and Arg 130 of 14-3-3ε. We did not select the phosphate atom for Sep 34 in definition of r^{Sep34} because Ser 34 does not involve a phosphate atom. To compare behaviors of Sep 34 and Ser 34, the same atom should be selected. The third distance r^{Glu35} is the minimum distance from the Cγ atoms of Glu 35 of pMLF1 (or npMLF1) to the Nζ atoms of Lys 50 and Lys 123 of 14-3-3ε. The fourth distance r^{Phe37} is the distance from the ring center of Phe 37 of pMLF1 (or npMLF1) to the heavy atoms in the stem of Lys 50 of 14-3-3ε. Note that a lysine sidechain has a long hydrophobic stem.

After calculating a distance distribution function (or pair distance distribution function) $P(r^{\mu})$ ($\mu = Phe33, Sep\ 34/Ser\ 34, Glu35$, or $Phe37$), in which a probability to detect the distance in a window $[r^{\mu}, r^{\mu} + dr^{\mu}]$ is given by $P(r^{\mu})dr^{\mu}$, a radial distribution function (or radial density function) was defined as: $p(r^{\mu}) = P(r^{\mu})/4\pi r^{\mu 2}$ with normalization of $\int p(r^{\mu})dr^{\mu} = 1$. Figure 4 plots the radial distribution functions for the distances. A remarkable difference between the two systems is related to Sep/Ser 34: I.e., the sharp peak of $p(r^{Sep34})$ is seen in Fig. 4A, although $p(r^{Ser34})$ show a dull peak (Fig. 4B). This difference is because of presence or absence of the salt bridges associated with Sep 34 or Ser 34, as discussed later. According to the decrease

of the peak height from $p(r^{Sep34})$ to $p(r^{Ser34})$, the peaks of the other radial distribution functions decreased. This result is natural from Fig. 3B: Cluster C_1' is less stable than C_2' in the npMLF1/14-3-3ε system.

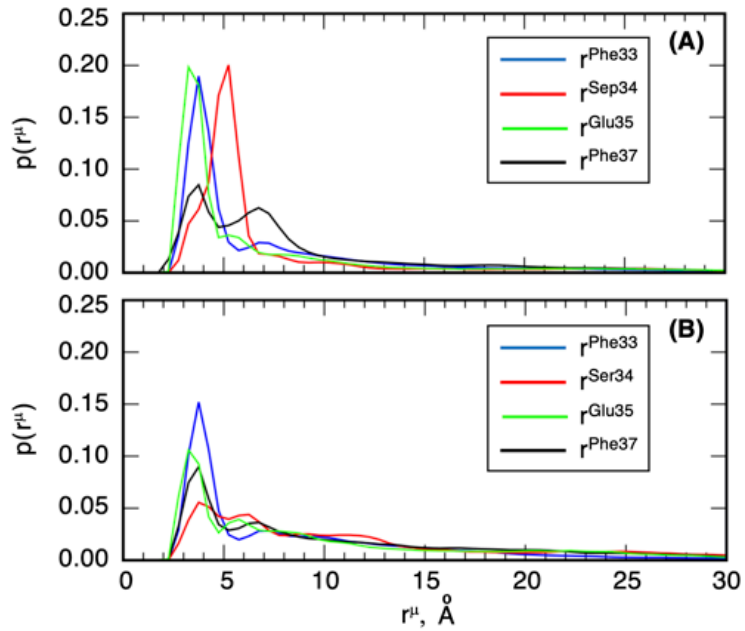

Figure 4. Radial distribution function $p(r^{\mu})$ for four distances: μ = Phe33, Sep/Ser34, Glu35, and Phe37. See the main text for definition of $p(r^{\mu})$. Panels (A) and (B) are for pMLF1/14-3-3ε and npMLF1/14-3-3ε, respectively.

Furthermore, in the pMLF1/14-3-3ε system, we note that $p(r^{Phe33})$ had a sharp peak whereas $p(r^{Phe37})$ had a dull peak (Fig. 4A). This result supports the experimental data that Phe 33 of pMLF1 contributes more to the complex stability than Phe 37 of pMLF1 does[A1]. Later, this point is discussed further.

The situation of Glu 35 of pMLF1 is problematic: $p(r^{Glu35})$ had a sharp peak at $r^{Glu35} \approx 3$ Å in the pMLF1/14-3-3ε system (Fig. 4A). Thus, Glu 35 seems to form a salt bridge to Lys 50 or Lys 123 of 14-3-3ε in the simulation. This result may indicate that Glu 35 contributes certainly to the complex stability. However, as mentioned in the Introduction–section, the mutation E35A of pMLF1 did not change the complex stability. Later, we analyze fluctuations of Glu 35 around 14-3-3ε to explain rationally this discrepancy.

Next, we analyze the spatial density of the Sep/Ser 34 position around 14-3-3ε. First, we divide the 3D real space into cubes, whose size ΔV is 2 Å × 2 Å × 2 Å. The cube position is defined by its center $\boldsymbol{r} = [x, y, z]$. Then, we sum up the statistical weights of snapshots whose Oγ atom of Sep 34 of pMLF1 (or Ser 34 of npMLF1) are detected in the cube. This number is interpreted as a density, and expressed as $\rho_{Sep34}^{O\gamma}(\boldsymbol{r})$ (or $\rho_{Ser34}^{O\gamma}(\boldsymbol{r})$). Last, we normalize the function as $\sum_i \rho_{\mu}^{O\gamma}(\boldsymbol{r}_i) = 1$, where the summation is taken over all cubes ($\forall i$). This normalization makes the density map comparable between the pMLF1/14-3-3ε and npMLF1/14-3-3ε systems. Other density functions appearing subsequently were calculated with the same method.

Figure 5A is the contour maps of $\rho_{Sep34}^{O\gamma}(\boldsymbol{r})$. Apparently, in the pMLF1/14-3-3ε system, the high-density spot was found around the Oγ-atomic position of Sep 34 in the crystal structure (red-colored contours in Fig. 5A). This figure also indicates that Sep 34 distributed almost entire surface of 14-3-3ε at a lower density level (dark green-colored contours). On the other hand, at a middle density level (the magenta-colored contours), the distribution was inhomogeneous. We found four large clusters $D_1, \ldots, D_4$ from the magenta-colored contours. The cluster D_1 involved the high-density spot mentioned above. Clusters D_1 and D_2 were connected smoothly without density discontinuity, or rather one may state that D_1 and D_2 form a single cluster. Contrarily, density discontinuity was found among clusters D_1, D_3, and D_4. This point is discussed again later.

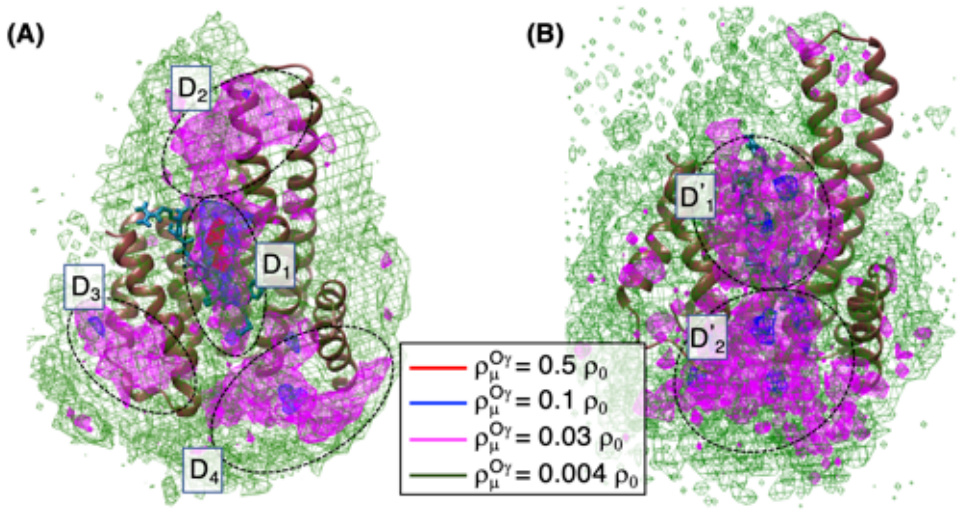

Figure 5. Density contour map $\rho_{\mu}^{O\gamma}(\boldsymbol{r})$ of the Oγ atomic position of residue 34 of MLF1: (A) $\rho_{Sep34}^{O\gamma}(\boldsymbol{r})$ and (B) $\rho_{Ser34}^{O\gamma}(\boldsymbol{r})$. Contour levels are shown in inset, where $\rho_0 = 0.01$. Shown molecular structures are the equilibrated ones (NPT simulation) of the crustal structure. Thus, they are close to the crystal structure. Clusters surrounded by broken-line circles, named D_i ($i = 1, \ldots, 4$) and D_i' ($i = 1,2$), are mentioned in text.

Figure 5B is the contour map of $\rho_{Ser34}^{O\gamma}(\boldsymbol{r})$ in the npMLF1/14-3-3ε system. No red-colored contours were found in this density map, which results from absence of a salt bridge at the 34-th residue. We found two large clusters, denoted as D_1' and D_2' in Fig. 5B, in the level of magenta-colored contours. The position of D_1' overlaps to that of D_1 in Fig. 5A. This means that npMLF1 has a weak tendency of staying around the pMLF1 position in the crystal structure.

Salt-bridge relay between pMLF1 and 14-3-3ε.

Now, we focus on the high-density spots involved in cluster D_1. Figure 6A shows that the contours of $\rho_{Sep34}^{O\gamma} = 0.5\ \rho_0$ are split into two spots, named E_1 and E_2 in the figure, and that E_1 involves the crystal-structure position of the Oγ atom of Sep 34 of pMLF1.

Figure 6B demonstrates a snapshot taken from a fringe of E_2, where Sep 34 formed a salt bridge to Arg 61 of 14-3-3ε, which is not observed in the crystal structure. Figure 6C displays a snapshot taken from the middle of E_2, where Sep 34 formed salt bridges to Arg 61 and Arg 57 of 14-3-3ε. Note that the salt bridge to Arg 57 is the one observed in the crystal structure. Also, we note that Arg 130 of 14-3-3ε did not participate in the salt bridge to Sep 34. Figure 6D displays a

snapshot picked from the corridor between E_1 and E_2, which shows that the salt bridge to Arg 61 was broken and the salt bridge to Arg 130 was not formed. Thus, the conformation at the corridor had only one salt bridge, which suggests that the conformation at the corridor is energetically less stable than that in E_2. Figure 6E shows a snapshot taken from E_1, in which the two salt bridges observed in the crystal structure were formed. We presume that appearance of the corridor between E_1 and E_2 is originated from decrease of a salt bridge. Thus, Figs. 6B-6E suggest that salt-bridge relay takes place when the conformation moves from E_2 to E_1.

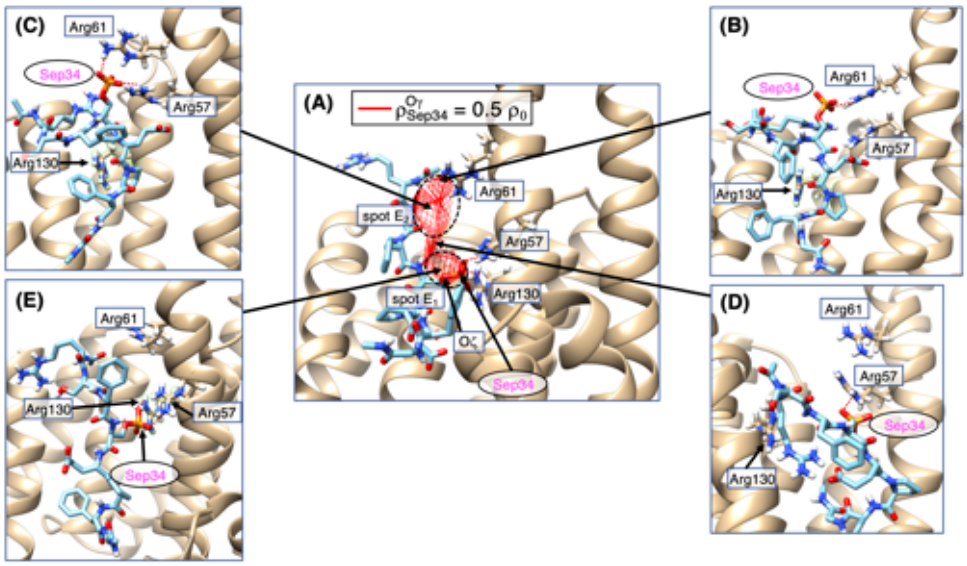

Figure 6. (A) Close-up of cluster D_1. Red-colored contours are those with $\rho^{O\gamma}_{Sep34} = 0.5\ \rho_0$ ($\rho_0 = 0.01$). Two high-density spots E_1 and E_2 are shown. (B) Snapshot taken from fringe of E_2. (C) Snapshot taken from middle of E_2, and (D) snapshot taken from corridor connecting E_1 and E_2. (E) Snapshot taken from E_1. Salt bridge is shown by brown-colored broken line. Residues labeled in ovals belong to pMLF1, and those in rectangles do to 14-3-3ε.

Spatial distribution of Phe 33 and Phe 37 of MKF1 around 14-3-3ε at 300 K.

Figure S6 shows that Phe 33 of pMLF1 is sandwiched by Leu 223 and Leu 230 of 14-3-3ε and that Phe 37 of pMLF1 contact to the hydrophobic sidechain stem of Lys 50 of 14-3-3ε. We calculated the density of the ring centers of Phe 33 and Phe 37, which are expressed by $\rho^{ring}_{Phe33}(\boldsymbol{r})$ and $\rho^{ring}_{Phe37}(\boldsymbol{r})$, respectively, of pMLF1 (Fig. 7A) or npMLF1 (Fig. 7B). The procedure to calculate the density function is the same as that for $\rho^{O\gamma}_{Sep34}(\boldsymbol{r})$. The ring-center position is the geometrical center of heavy atoms of the phenylalanine sidechain ring. For the pMLF1/14-3-3ε system, Phe 33 distributed in a region (named spot H in Fig. 7A) capable to contact to Leu 223 and/or Leu 230, and H is the largest spot in this figure. This large spot is the origin of the sharp peak of r^{Phe33} in Fig. 4A, indicating that Phe 33 forms the hydrophobic contacts readily to its interaction partners Leu 223 and Leu 230.

Two spots F_1 and F_2 for $\rho^{ring}_{Phe37}(\boldsymbol{r})$ in Fig. 7A are related to the two peaks of $p(r^{Phe37})$ in Fig. 4A. I.e., the spot F_1 is the origin of the peak at $r^{Phe37} \approx 4$ Å and F_2 is that at $r^{Phe37} \approx 7$ Å. In contrast to the large spot H for Phe 33, the both spots F_1 and F_2 for Phe 37 were small. This suggest that that Phe 37 less contributes to the stabilization of the complex structure.

In Fig. 7B, the red- and blue-colored contours were distributed widely and sparsely. Remember that Ser 34 in npMLF1/14-3-3ε distributed in clusters D_1' and D_2' in Fig. 5B. Figures 5B and 7B indicate that npMLF1 has no strong preferential spot. The three broken-line circles, named spots

H_1', H_2', and H_3' in Fig. 7B, were close to Leu 223 and Leu 230, although the size of the spots was small. The density in these three circles resulted in the peak of r^{Phe33} in Fig. 4B.

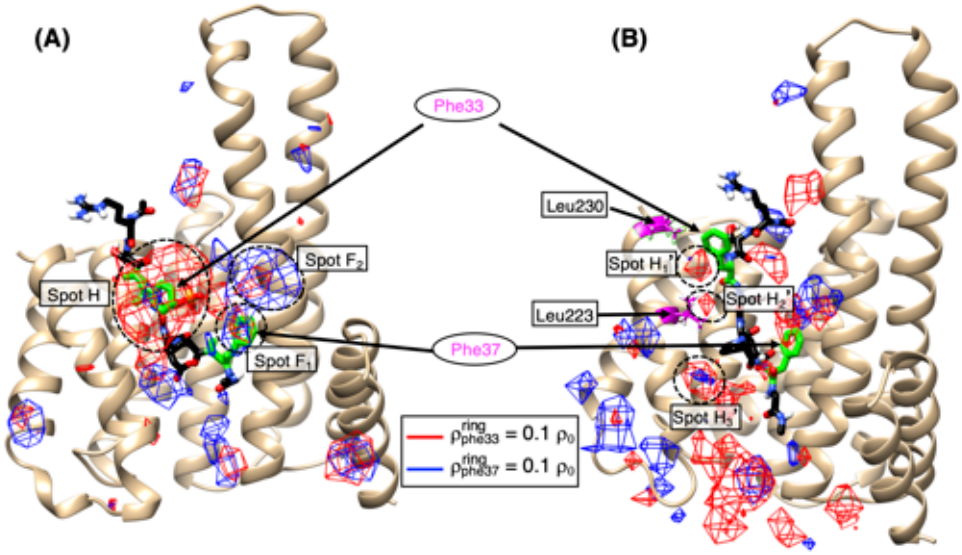

Figure 7. Spatial distributions of the ring-center positions of Phe 33 and Phe 37, denoted as ρ_{Phe33}^{ring} (red-colored contours) and ρ_{Phe37}^{ring} (blue-colored contours), respectively. The contour level is 0.1 ρ_0 ($\rho_0 = 0.01$) for both residues. Panels (A) and (B) are for pMLF1/14-3-3ε and npMLF1/14-3-3ε, respectively. Shown molecular structures are the equilibrated ones (NPT simulation) of the crustal structure, in which green-colored residues are Phe 33 and Phe 37. Residues labeled in ovals belong to pMLF1, and those in rectangles do to 14-3-3ε. In panel (A), broken-line circles, named spots H, F_1 and F_2, are mentioned in text. In panel (B), Leu 223 and Leu 230 are shown in magenta color, and three broken-line circles, named spots H_1', H_2', and H_3', are mentioned in text.

Spatial distribution of Glu 35.

We calculated the density functions $\rho_{Glu35}^{C\alpha}(\boldsymbol{r})$ and $\rho_{Glu35}^{C\gamma}(\boldsymbol{r})$ for the Cα and Cγ atomic positions of Glu 35 of pMLF1, respectively, for the pMLF1/14-3-3ε system. The procedure to calculate the density is the same as that for $\rho_{Sep34}^{O\gamma}(\boldsymbol{r})$. Figure 8A shows high density regions. The shown molecular structure is the equilibrated structure by the NPT simulation, in which Glu 35 forms salt bridges to Lys 50 and Lys 123 of 14-3-3ε. The contours indicate that both the Cα and Cγ atoms shifted from the crystal-structure positions toward the solvent, and that the mainchain of Glu 35 fluctuated largely as indicated by the broken-line arrow along the magenta-colored contours in Fig. 8A. One may consider that this positional shift may hinder Glu 35 of pMLF1 to form salt bridges to Lys 50 and Lys 123 of 14-3-3ε. On the other hand, Figure 4A indicated that Glu 35 forms the salt bridge to Lys 50 and/or Lys 123. The black contours in Fig. S7A demonstrates that Glu 35 could take a position (broken-line circle of Fig. S7A), from which Glu 35 was able to form the salt bridges to both Lys 50 and Lys 123. However, the density assigned to this position was small ($\rho_{Glu35}^{C\gamma}(\boldsymbol{r}) = 0.05\,\rho_0$). Thus, it seems that Fig. 8A and Fig. 4A are contradictory to each other.

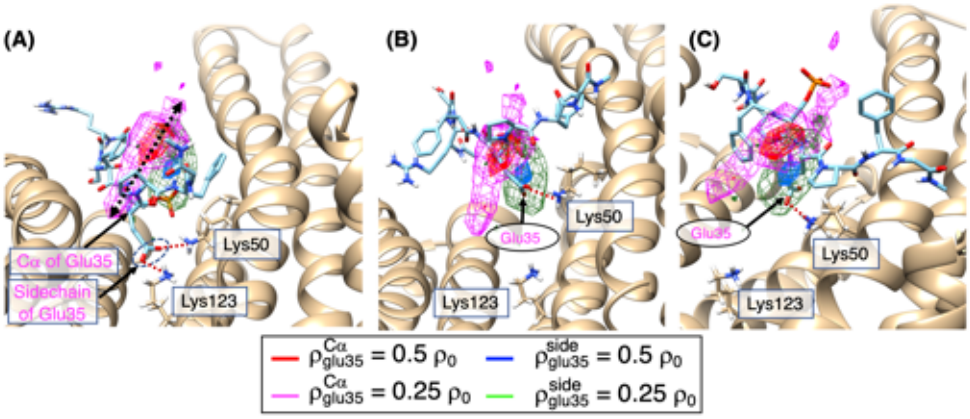

Figure 8. (A) Density contour maps of $\rho^{C\alpha}_{Glu35}(\boldsymbol{r})$ and $\rho^{C\gamma}_{Glu35}(\boldsymbol{r})$ for the pMLF1/14-3-3ε system. Shown molecular structure is the equilibrated one (NPT simulation), where Cα atom and sidechain tip of Glu 35 of pMLF1 are indicated by "Cα of Glu35" and "Sidechain of Glu35", respectively. Contour levels are shown in inset ($\rho_0 = 0.01$). Panels (B) and (C) display snapshots picked randomly from conformations that are involved in the contour regions of panel (A). Residues labeled in ovals belong to pMLF1, and those in rectangles do to 14-3-3ε. Salt bridge is shown by brown-colored broken line.

To elucidate this discrepancy, we checked snapshots whose Cα and Cγ atoms were involved in the contours of $\rho^{C\alpha}_{Glu35}(\boldsymbol{r}) = 0.25\rho_0$ (magenta contours in Fig. 8A) and $\rho^{C\gamma}_{Glu35}(\boldsymbol{r}) = 0.25\rho_0$ (green contours of Fig. 8A), respectively, and found that a majority of the snapshots in these contour regions formed a salt bridge only to Lys 50 of 14-3-3ε. Figures 8B and 8C display two snapshots picked randomly from the contour regions. To clarify this observation, we calculated $p(r^{Glu35-Lys50})$ and $p(r^{Glu35-Lys123})$: The former is a radial distribution function for the distance $r^{Glu35-Lys50}$ from the Cγ atom of Glu 35 of pMLF1 to the Nζ atom of Lys 50 of 14-3-3ε, and the latter is a radial distribution function for the distance $r^{Glu35-Lys123}$ from the Cγ atom of Glu 35 to the Nζ atom of Lys 123 of 14-3-3ε. Remember that these distances were used to define r^{Glu35} in Fig. 4. These radial distribution functions indicate that the salt bridge between Glu 35 and Lys 123 was rarely formed in the thermal fluctuations of Glu 35 (Fig. S7B).

Furthermore, we calculated salt-bridge-formation probabilities $Q_{Glu35-Lys50}$ and $Q_{Glu35-Lys123}$, which are defined as $Q_{Glu35-Lys50} = \sum_i^{r^{Glu35-Lys50} \leq 5.5\ \text{Å}} w_i$ and $Q_{Glu35-Lys123} = \sum_i^{r^{Glu35-Lys123} \leq 5.5\ \text{Å}} w_i$, respectively. The summation for $Q_{Glu35-Lys50}$ is taken over snapshots with $r^{Glu35-Lys50} \leq 5.5$ Å and that for $Q_{Glu35-Lys123}$ is done over those with $r^{Glu35-Lys123} \leq 5.5$ Å. One may consider that the value of 5.5 Å is larger than a limit distance for a salt bridge. However, the Cγ atom, not Oε atoms, was used to calculate $r^{Glu35-Lys50}$ and $r^{Glu35-Lys123}$, and Glu 35 has large positional fluctuations as shown above: Remember that a snapshot was saved once every 200 ps. Thus, we increased the limit distance. The resultant fraction $Q_{Glu35-Lys123}/Q_{Glu35-Lys50}$ was 0.18×10^{-5}, which again indicated that the salt bridge between Glu 35 and Lys 50 was dominant.

We note that Lys 50 of 14-3-3ε is located on the surface of 14-3-3ε. Thus, the sidechain stem of Lys 50 can modulate its sidechain-tip position to maintain the salt bridge to Glu 35 of pMLF1. In contrast, Lys 123 is somewhat buried in 14-3-3ε. Therefore, the sidechain tip of Lys 123 cannot move together with the positional shift of Glu 35, and consequently the salt bridge between Glu 35 of pMLF1 and Lys 123 of 14-3-3ε is not formed.

We presume that the positional shift and large fluctuations of Glu 35 of pMLF1 weaken the contribution of Glu 35 to the complex-structure stability. Remember that the sidechain stem of Lys 50 of 14-3-3ε is the hydrophobic interaction partner of Phe 37 in the crystal structure. Therefore, the positional fluctuations of Glu 35 of pMLF1 affect the sidechain motions of Lys 50 of 14-3-3ε, and hence hinder the hydrophobic interaction between Phe 37 of pMLF1 and Lys 50 of 14-3-3ε. I.e., the small contribution of Glu 35 and Phe 37 to the complex-structure stability may correlate to each other.

Orientation of pMLF1 around 14-3-3ε.

In this section, we focus on the molecular orientation of pMLF1. We defined an orientation vector $\boldsymbol{e}_i$ of pMLF1 for a snapshot i by a unit vector ($|\boldsymbol{e}_i| = 1$) pointing from the Cα atom of Phe 33 to the Cα atom of Phe 37 of the snapshot. The orientation vector $\boldsymbol{e}_{ref}$ for the reference conformation (the crystal structure) was also calculated in the same way: $|\boldsymbol{e}_{ref}| = 1$.

We analyzed correlation between the orientation of pMLF1 and the position of Sep 34. We divided again the 3D real space into cubes of 2.0 Å × 2.0 Å × 2.0 Å, whose positions are assigned to the cube centers $\boldsymbol{r}$. A snapshot was assigned to a cube when the Oγ atom of Sep 34 of the snapshot is involved in the cube. This procedure is the same as that used for defining $\rho_{Sep34}^{O\gamma}(\boldsymbol{r})$. Last, we took average of orientation vectors for each cube:

$$\overline{\boldsymbol{v}}(\boldsymbol{r}) = \frac{\sum_i \delta_{cube}(\boldsymbol{r};i) w_i \boldsymbol{e}_i}{\sum_i \delta_{cube}(\boldsymbol{r};i) w_i}, \tag{1}$$

where w_i is the thermodynamic weight ($w_i \geq 0$) assigned to the i-th snapshot. The $\delta_{cube}(\boldsymbol{r};i)$ is a delta function defined as:

$$\delta_{cube}(\boldsymbol{r};i) = \begin{cases} 1 & \text{(if O}\gamma\text{ atom of snapshot } i \text{ is in cube at } \boldsymbol{r}) \\ 0 & \text{(else)} \end{cases}. \tag{2}$$

The quantity $\sum_i \delta_{cube}(\boldsymbol{r};i) w_i$ in Eq. 1 is equivalent to $\rho_{Sep34}^{O\gamma}(\boldsymbol{r})$, the density of the O$\gamma$ atom of Sep 34 in the cube at $\boldsymbol{r}$. Note that $\overline{\boldsymbol{v}}(\boldsymbol{r})$ is no longer a unit vector. The vector $|\overline{\boldsymbol{v}}(\boldsymbol{r})|$ indicates a degree of orientational ordering of snapshots detected in cube $\boldsymbol{r}$: $|\overline{\boldsymbol{v}}(\boldsymbol{r})|$ takes the maximum (= 1.0) when all $\boldsymbol{e}_i$ have the same orientations. If each $\boldsymbol{e}_i$ has uncorrelated orientation, then $|\overline{\boldsymbol{v}}(\boldsymbol{r})|$ becomes small.

Next, we calculated a scalar product of $\overline{\boldsymbol{v}}(\boldsymbol{r})$ and $\boldsymbol{e}_{ref}$, and assigned it to the cube $\boldsymbol{r}$.

$$SC(\boldsymbol{r}) = \overline{\boldsymbol{v}}(\boldsymbol{r}) \cdot \boldsymbol{e}_{ref}, \tag{3}$$

which takes a value from -1 to 1. Green and orange arrows in Fig. 9 are the vectors $\overline{\boldsymbol{v}}(\boldsymbol{r})$ that satisfied the following two conditions:

$$\rho_{Sep34}^{O\gamma}(\boldsymbol{r}) \geq 0.03\rho_0, \tag{4}$$

and

$$|\overline{\boldsymbol{v}}(\boldsymbol{r})| \geq 0.5. \tag{5}$$

Here, the orientation of pMLF1 in the crystal structure is shown by a black arrow. Equation 4 is to pick cubes involved in clusters $D_1,\ldots,D_4$ in Fig. 5, and Eq. 5 is to display relatively well ordered $\overline{\boldsymbol{v}}(\boldsymbol{r})$. Apparently, only cluster D_1 exhibited remarkable orientational ordering of pMLF1, and its majority was occupied by $SC(\boldsymbol{r}) \geq 0.5$ (green-colored arrows). In contrast, cubes with $SC(\boldsymbol{r}) < 0.5$ was rare and sparse. We conclude that clusters D_2, D_3, and D_4 are occupied by less ordered orientational vectors, and that pMLF1 ordered only in cluster D_1. Thus, the orientation of pMLF1 is ordered as reaching the high-density region (i.e., low free-energy region), which is advantageous for binding to 14-3-3ε.

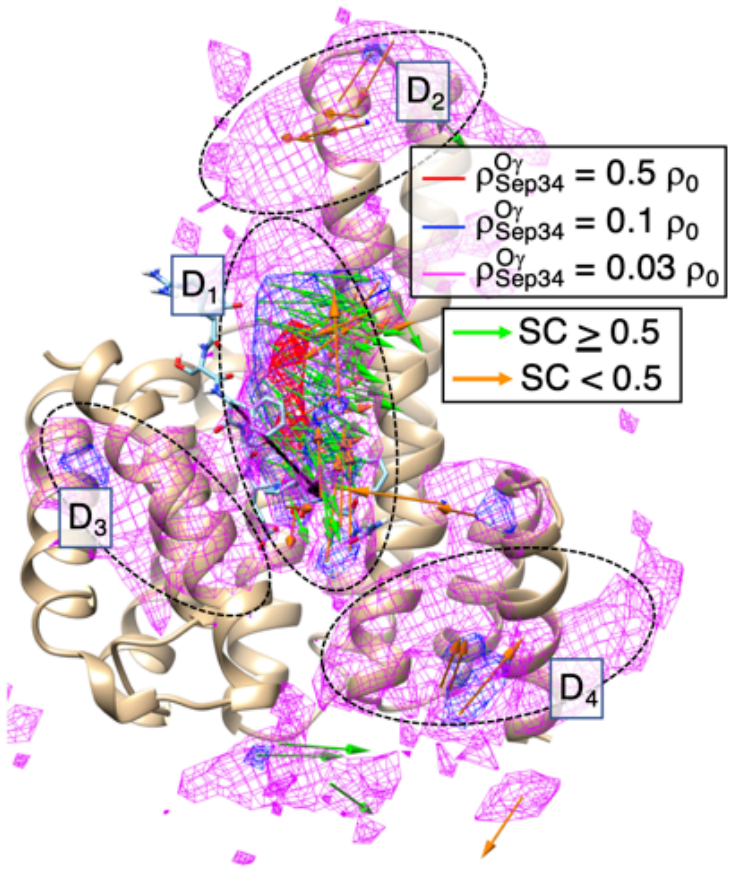

Figure 9. Averaged vectors $\overline{\boldsymbol{v}}(\boldsymbol{r})$ in cubes $\boldsymbol{r}$ where $\rho_{Sep34}(\boldsymbol{r}) \geq 0.03\rho_0$ ($\rho_0 = 0.01$) and $|\overline{\boldsymbol{v}}(\boldsymbol{r})| \geq 0.5$. Red-, blue-, and magenta-colored contours as well as clusters $D_1,\ldots,D_4$ are the same as those in Fig. 5, although 14-3-3ε is viewed in different orientation from Fig. 5 to show position-dependency of $\overline{\boldsymbol{v}}(\boldsymbol{r})$. Green-colored arrows are vectors with $SC(\boldsymbol{r}) \geq 0.5$, and orange-colored ones are those with $SC(\boldsymbol{r}) < 0.5$. Vectors are multiplied by 10.0 to show the vectors clearly: arrow $= 10.0 \times \overline{\boldsymbol{v}}(\boldsymbol{r})$. Shown structure is the equilibrated one (NPT simulation) of the crustal structure of pMLF1/14-3-3ε. Black arrow points from the Cα atom of Phe 33 to the Cα atom of Phe 37, which shows orientation of pMLF1 in the crystal structure.

Site dependency of residue interactions between pMLF1 and 14-3-3ε.

Remember that we calculated the four distances r^{μ} ($\mu = Phe33, Sep34, Glu35, Phe37$) for each snapshot for Fig. 4. Here, we calculated the thermal average of the distance in each cube for the pMLF1/14-3-3ε system:

$$\bar{r}^{\mu}(\boldsymbol{r}) = \frac{\sum_i \delta_{cube}(\boldsymbol{r};i) w_i r^{\mu}(i)}{\sum_i \delta_{cube}(\boldsymbol{r};i) w_i}, \tag{6}$$

Where $r^{\mu}(i)$ is r^{μ} of snapshot i. The delta function δ_{cube} (Eq. 2) works so that only snapshots, whose Oγ atom of Sep 33 are detected in the cube at $\boldsymbol{r}$, contribute to $\bar{r}^{\mu}(\boldsymbol{r})$. The $\bar{r}^{\mu}(\boldsymbol{r})$ quantifies space dependency of the contacts of the four residues of pMLF1 to their contact partners of 14-3-3ε found in the crystal structure as a function of $\boldsymbol{r}$ (i.e., the Oγ atom position of Sep 33 of pMLF1).

Figure 10A displays the overall spatial patterns of $\bar{r}^{\mu}(\boldsymbol{r})$. The red-colored contours indicate regions with $\bar{r}^{Sep34}(\boldsymbol{r}) \leq 5.5$ Å. One may consider that the value of 5.5 Å is larger than a limit distance for salt bridge. However, the contours are assigned to the Oγ atom of Sep 34 not to three O atoms of the phosphate group. Therefore, we increased the limit distance. We also set $\bar{r}^{Glu35}(\boldsymbol{r}) \leq 5.5$ Å as described before. The limits for Phe 33 and Phe 37 of pMLF1 were set to 7 Å: $\bar{r}^{\mu}(\boldsymbol{r}) \leq 7$ Å, which is larger than a value for a hydrophobic contact. This is because $r^{\mu}(i)$ was calculated using the ring center (not individual atoms in the ring) of Phe 33 and Phe 37 of pMLF1.

Figure 10A shows the large spatial region for Phe 33 of pMLF1 (blue-colored contour). This region indicates that Phe 33 could form the hydrophobic contacts to Leu 223 and Leu 230 of 14-3-3ε in the wide spatial region: When the Oγ atom of Sep 34 was inside of the blue-colored contour region, pMLF1 could form a hydrophobic contact to its interaction partners. This result may raise a possibility that these hydrophobic contacts formable in the wide spatial region acts as attractive interactions between pMLF1 and 14-3-3ε. We have no definite answer to this argument. Our previous work of mD-VcMD,[39] however, suggests that a long N-terminal tail of a protein may capture its ligand before the ligand binds to the most probable binding site of the receptor (fly-casting mechanism).[52-55] Thus, the hydrophobic contacts with respect to Phe 37 of pMLF1 may lead to the most stable complex structure.

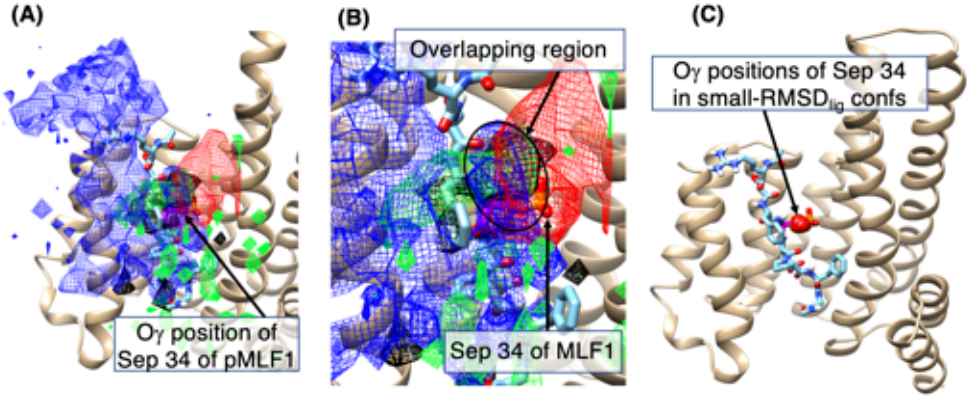

Figure 10. (A) Regions of cubes that satisfy $\bar{r}^{\mu}(\boldsymbol{r}) \leq r_{lim}$, where $\boldsymbol{r}$ is the center of cube. Red-colored contour regions are those for $\mu = Sep34$ of pMLF1 and $r_{lim} = 5.5$ Å, blue-colored ones are those for $\mu = Phe33$ of pMLF1 and $r_{lim} = 7.0$ Å, and black-colored ones are those for $\mu = Glu35$ of pMLF1 and $r_{lim} = 5.5$ Å, and green-colored ones are those for $\mu = Phe37$ of pMLF1 and $r_{lim} = 7.0$ Å. Black arrow with words "Oγ position of Sep 34 of pMLF1" indicates its position in the crystal structure. (B) Closeup around Sep 34 of pMLF1 in the crystal structure. Four contours overlap around the black circle. (C) The Oγ-atomic positions (red-colored spheres indicated by words "Oγ positions of Sep 34 in small-$RMSD_{lig}$ confs") of snapshots with $RMSD_{lig} \leq 1.0$ Å. Shown molecular structure is the equilibrated one (NPT simulation) of the crustal structure.

Figure 10A also shows a relatively large spatial region for Sep 34 of pMLF1 by the red contours. This region overlaps well to the high-density spots E_1 and E_2 (Fig. 6). This result is natural because Sep 34 can form the salt bridges when Sep 34 approaches the binding partners.

On the contrary, the green contours for Phe 37 of pMLF1 were sparse around the surface of 14-3-3ε, which suggests that the hydrophobic interaction by Phe 37 with 14-3-3ε is weaker than

that by Phe 33 (blue contours) in stabilizing the complex structure. Black contours for Glu 35 of pMLF1 were localized. This means that Glu 35 could form the salt bridge only when Glu 35 reached the localized region.

Figure 10B is the closeup of Fig. 10A focusing on Sep 34 of pMLF1. Apparently, the four contours overlapped in a narrow region around the Sep 34 position of the crystal structure, which is indicated by a black-line circle with words "Overlapping region" in the figure. We emphasize that this narrow region involves the Oγ atomic position of Sep 34 in the crystal structure. This suggests that the simultaneous formation of the four important intermolecular interactions, regarding Phe 33, Sep 34, Glu 35, and Phe 37 of pMLF1, are possible only when the Oγ atom is near the position of the crystal structure. As shown in Fig. 9, Glu 35 shifted somewhat from the crystal-structure position towards solvent. Interestingly, the black-contours regarding Glu 35 are the consequence from the shift of Glu 35. In other word, the pMLF conformations with the Glu-35 shift belong to the crystal-structure-like complex.

To check this conjecture further, we picked snapshots with $RMSD_{lig} \leq 1.0$ Å, and displayed the Oγ-atomic positions of those small-$RMSD_{lig}$ snapshots in the 3D real space. In fact, Fig. 10C demonstrates that the Oγ-atomic positions (red-colored spheres in Fig. 10C) of those snapshots overlapped well to its crystal-structure position. This result also supports that simultaneous satisfaction of the four intermolecular interactions leads to the crystal-structure-like complex form.

DISCUSSION

The current study could propose a binding process of pMLF1 to 14-3-3ε (Fig. 11). Before pMLF1 binds to 14-3-3ε, pMLF1 exists everywhere around 14-3-3ε at a low concentration ($\rho_{Sep34} = 0.004\rho_0$; dark green-colored contours of Fig. 11). It is presumable from the magenta-colored contours ($\rho_{Sep34} = 0.03\rho_0$) that pMLF1 can contact to four regions of the 14-3-3ε surfaces denoted by $D_1, \ldots, D_4$ more stably or frequently than to the other regions, which results in formation of an encounter complex. If pMLF1 binds to D_1 first, then pMLF1 may form the complex structure similar to the crystal structure quickly by adjusting the conformation.

Because the sum of the regions $D_2 + D_3 + D_4$ is larger than D_1, it is likely that pMLF1 binds more frequently to one of D_2, D_3, and D_4 than to D_1. Remember that free-energy barriers exist among three clusters D_1, D_3, and D_4. Therefore, if pMLF1 contacts to D_3 or D_4 first, then the complex should jump over the free-energy barriers to reach D_1. Thus, the motion from D_3 to D_1 or from D_4 to D_1 may take a long time. Otherwise, the encounter complex generated in D_3 or D_4 may be dissociated sooner or later.

If the encounter complex is formed in D_2 first, then no energy jump is required to move from D_2 to D_1 because the two clusters are connected smoothly without a free-energy barrier. Remember that pMLF1 has no orientational tendency in D_2. When pMLF1 moves to D_1, Sep 34 of pMLF1 meets the spot E_2, and binds to Arg 61 and/or Arg 57 of 14-3-3ε. Simultaneously, the molecular orientation of pMLF1 tends to be aligned as in the crystal structure. At this stage, neither the hydrophobic contacts regarding Phe 33 and Phe 37 of pMLF1 nor the salt bridge regarding Glu 35 of pMLF1 have yet been formed. When pMLF1 goes through the corridor from spot E_2 to E_1, the salt-bridge relay takes place: The salt bridge between Sep 34 of pMLF1 and Arg 61 of 14-3-3ε is broken, and one between Sep 34 of pMLF1 and Arg 130 of 14-3-3ε is formed finally. When Sep 34 fluctuates in spot E_1, which involves the crystal-structure position (the white-circle position in Fig. 11), Phe 33 and Phe 37 form the hydrophobic contacts to 14-3-

3ε. Simultaneously, Glu 35 forms the salt bridge to Lys 50 whereas the salt bridge to Lys 123 are not formed.

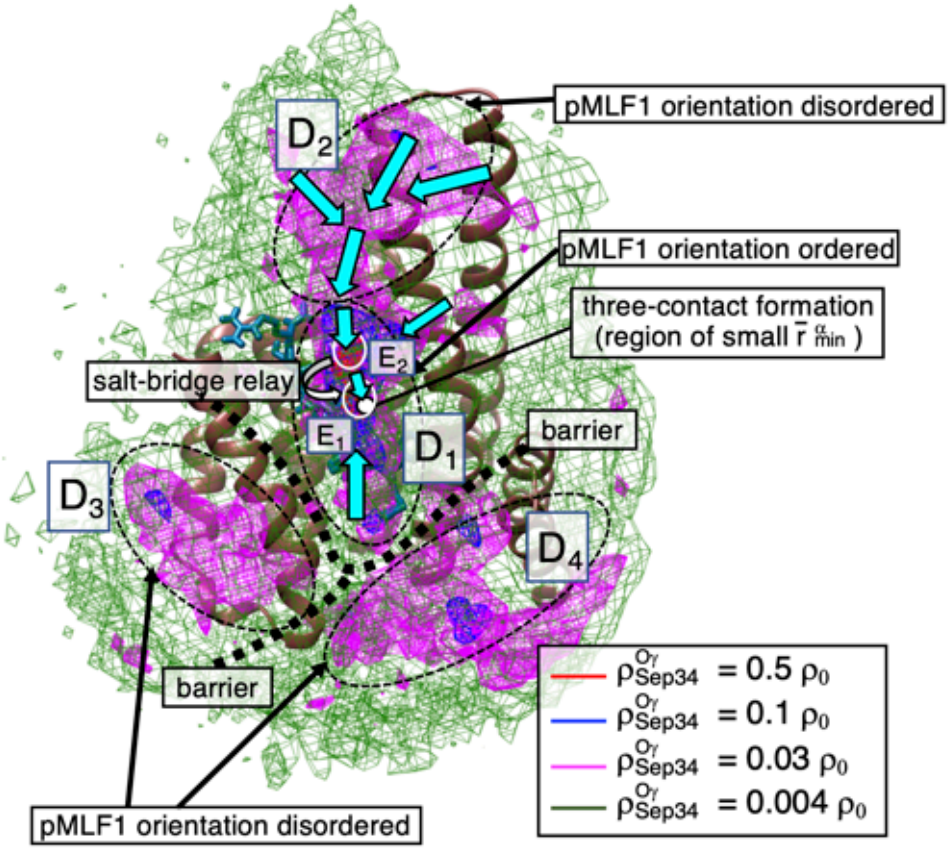

Figure 11. Scheme of binding process of pMLF1 to 14-3-3ε. Contour maps represent $\rho_{Sep34}(\boldsymbol{r})$ at four density levels, which are the same as those for Fig. 5. Clusters D_1,…, D_4 are also from Fig. 5. Free-energy barriers are shown by thick broken lines. Cyan-colored arrows represent conformational motions of pMLF1 from cluster D_2 to D_1. Molecular orientation of pMLF1 is ordered well in D_1, whereas it is less ordered in D_2, D_3, and D_4. Position where salt-bridge relay takes place is shown by curved arrow. White open circles are high-density spots E_1 and E_2. White filled circle is the position where four intermolecular contacts, quantified by $\bar{r}^{\mu}(\boldsymbol{r})$ ($\mu = Phe33, Sep34, Glu35, Phe37$), are formed simultaneously.

CONCLUSIONS

The equilibrated distribution functions derived from the current GA-based mD-VcMD provided important aspects to understand the binding process of pMLF1 to 14-3-3ε: The encounter-complex formation in clusters $D_1, \ldots, D_4$, free-energy barriers among D_1, D_3, and D_4, the pMLF1 orientation ordering in the high-density region, salt-bridge relay to reach the complex structure similar to the crystal structure, and the narrow spot essentially important for formation of the crystal structure. These aspects were naturally derived from the current simulation without a gap in logic, and can be useful to proceed other experimental and computational study.

In our simulation, Glu 35 shifted toward solvent somewhat, and exhibited relatively large positional fluctuations even in the high-density spot (Fig. 8). These positional shift and large fluctuations were induced by the absence of the salt bridge to Lys 123 of 14-3-3ε, which exists in the crystal structure (PDB ID: 3ual). We noted that our simulation study supports the experimental data that the mutation E35A does not alter K_D for the pMLF1/14-3-3ε complex.[7] Our simulation also suggests that the large fluctuations of Glu 35 affects the hydrophobic

interaction of Phe 37 with the sidechain stem of Lys 50. Mutagenesis K50A of 14-3-3ε may provide insightful information to our suggestion.

The current study showed that the GA-based mD-VcMD can treat a very flexible ligand molecule without restraining the molecular flexibility. We also note that this method can search a wide conformational space, where the ligand covers not only the vicinity of the binding site of the receptor observed in the crystal structure, but also regions distant from the binding site. It is because this method sets RCs to search the wide 3D real space and assigns the thermodynamic weight to all sampled snapshots. Although we do not know the best selection of the reaction coordinates a priori, degeneration of basins is in general resolved by increasing the space dimensionality. This is why the multiple reaction coordinates was useful to sample the highly flexible chain, pMLF1, binding to the receptor in the present study. In contrast, increment of the space dimensionality makes sampling data sparse. Thus, a powerful and simple sampling method is required to handle the multiple RCs. The present study has demonstrated that the GA-based mD-VcMD is adequate for treating the 3D-RC space.

Recently, we performed multicanonical molecular dynamics (McMD) of an intrinsically disorder peptide, p53 C-terminal domain (p53CTD), binding to its receptor protein (S100B) to investigate inter-molecular interactions.[37] Interestingly, this study suggested that the degree of disorder of p53CTD may increase when binding to S100B. We named this emerging state upon binding an *extra–disorder* state. In theory, the GA-based mD-VcMD has an ability to investigate the extra-disorder state throughout because pMLF1 is a highly flexible peptide. In fact, as shown in Fig. 3C, free pMLF1 snapshots were sampled. However, the upper limits of 40 Å for RCs (Table S2) are still short to sample pMLF1 conformations entirely free from 14-3-3ε: I.e., a pMLF1 snapshot, which was un-touching 14-3-3ε, could touch by rotating the snapshot (data not shown). As a future study, we set the RC limits to values large enough to generate the free-state snapshots. Furthermore, we note that this increment of RCs leads the GA-based mD-VcMD to a method that can calculate K_D directly by calculating the probabilities for both the bound and free states.

The GA-based mD-VcMD simulation has shown that the density contour map is not simple. For instance, the current study suggests that molecular modification that can change the corridor width between spot E_1 and E_2 alters the kinetic constants and thus dissociation constant for the complex formation/dissociation. The current simulation also suggests a reason why E35A did not alter K_D. These insights may open a new approach for drug design. We emphasize that the large-scale conformational ensemble equilibrated at a room temperature, which covers the free, encounter-complex, and the most-stable states, are useful to widen the research area for drug discovery.

ACKNOWLEDGMENT

The authors thank Prof. Norio Ozaki in Nagoya University Graduate School of Medicine for his valuable discussion about the functional roles of the 14-3-3 proteins that interact with the phosphorylated sequences. J. H. was supported by JSPS KAKENHI Grant No. 16K05517 and by the Development of core technologies for innovative drug development based upon IT from Japan Agency for Medical Research and Development, AMED. J. H. and K. K. were supported by the HPCI System Research Project (Project IDs: hp190017, hp190018, hp190027, hp200025, hp200063, and hp200090). The mD-VcMD was performed on the TSUBAME3.0 supercomputers at the Tokyo Institute of Technology. K. K. was also supported by JSPS

KAKENHI Grant No. 16K18526. H. N. was supported by a Grant-in-Aid for Scientific Research on Innovative Areas (24118008) and a Grant-in-Aid for Challenging Exploratory Research (16K14711) from JSPS. N. K. wad supported by JSPS KAKENHI Grant (Number JP20H03229). It was performed in part under the Cooperative Research Program of the Institute for Protein Research, Osaka University, CR-19-05 and CR-20-05.

REFERENCES

(1) Yaffe, M. B.; Rittinger, K.; Volinia, S.; Caron, P. R.; Aitken, A.; Leffers, H.; Gamblin, S. J.; Smerdon, S. J.; Cantley, L. C. The Structural Basis for 14-3-3:Phosphopeptide Binding Specificity. *Cell* **1997**, *91*, 961–971.

(2) Yaffe, M. B. How Do 14-3-3 Proteins Work? --Gatekeeper Phosphorylation and the Molecular Anvil Hypothesis. *FEBS Lett*. **2002**, *513*, 53–57.

(3) Johnson, C.; Crowther, S.; Stafford, M. J.; Campbell, D. G.; Toth, R.; MacKintosh, C. Bioinformatic and Experimental Survey of 14-3-3-Binding Sites. *Biochem. J*. **2010**, *427*, 69–78.

(4) Matsumoto, N.; Yoneda-Kato, N.; Iguchi, T.; Kishimoto, Y.; Kyo, T. ; Sawada, H.; Tatsumi, E.; Fukuhara, S. Elevated MLF1 Expression Correlates with Malignant Progression from Myelodysplastic Syndrome. *Leukemia* **2000**, *14*, 1757–1765.

(5) Winteringham, L. N.; Endersby, R.; Kobelke, S.; McCulloch, R. K.; Williams, J. H.; Stillitano, J.; Cornwall, S. M.; Ingley, E.; Klinken, S.P. Myeloid Leukemia Factor 1 Associates with a Novel Heterogeneous Nuclear Ribonucleoprotein U-Like Molecule. *J. Biol. Chem*. **2006**, *281*, 38791–38800.

(6) Kimura, H.; Tsuboi, D.; Wang, C.; Kushima, I.; Koide, T.; Ikeda, M.; Iwayama, Y.; Toyota, T.; Yamamoto, N.; Kunimoto, S.; Nakamura, Y.; Yoshimi, A.; Banno, M.; Xing, J.; Takasaki, Y.; Yoshida, M.; Aleksic, B.; Uno, Y.; Okada, T.; Iidaka, T.; Inada, T.; Suzuki, M.; Ujike, H.; Kunugi, H.; Kato, T.; Yoshikawa, T.; Iwata, N.; Kaibuchi, K.; Ozaki. N. Identification of Rare, Single-Nucleotide Mutations in NDE1 and Their Contributions to Schizophrenia Susceptibility. *Schizophr Bull*. 2015, *41*, 744–753.

(7) Molzan, M.; Weyand, M.; Rose, R.; Ottmann, C. Structural Insights of the MLF1/14-3-3 Interaction. *FEBS J*. **2012**, *279*, 563–571.

(8) Kawabata, T.; Nishikawa. K. Protein Structure Comparison Using the Markov Transition Model of Evolution. *Proteins* **2000**, *41*, 108–122.

(9) Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M. Ø.; Xu. H.; Trbovic, N.; Friesner, R. A.; Palmer, A. G., III; Shaw, D. E. Microsecond Molecular Dynamics Simulation Shows Effect of Slow Loop Dynamics on Backbone Amide Order Parameters of Proteins. *J. Phys. Chem. B* **2008**, *112*, 6155–6158.

(10) Narumi, T.; Yasuoka, K.; Taiji, M.; Höfinger, S. Current Performance Gains from Utilizing the GPU or the ASIC MDGRAPE-3 within an Enhanced Poisson Boltzmann Approach. *J Comput. Chem*. **2009**, *30*, 2351-2357.

(11) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Generalized-Ensemble Algorithms for Molecular Simulations of Biopolymers. *Biopolymers* **2001**, *60*, 96–123.

(12) Higo, J.; Ikebe, J.; Kamiya, N.; Nakamura, H. Enhanced and Effective Conformational Sampling of Protein Molecular Systems for Their Free Energy Landscapes. *Biophysical Rev*. **2012**, *4*, 27–44.

(13) Berg, B. A.; Neuhaus, T. Multicanonical Ensemble: A New Approach to Simulate First-Order Phase Transitions. *Phys. Rev. Lett*. **1992**, *68*, 9–12.

(14) Hansmann, U. H. E.; Okamoto, Y. Prediction of Peptide Conformation by Multicanonical Algorithm: New Approach to the Multiple-Minima Problem. *J. Comput. Chem*. **1993**, *14*, 1333–1338.

(15) Kidera, A. Enhanced Conformational Sampling in Monte Carlo Simulations of Proteins: Application to a Constrained Peptide. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 9886–9889.

(16) Nakajima, N.; Nakamura, H.; Kidera, A. Multicanonical Ensemble Generated by Molecular Dynamics Simulation for Enhanced Conformational Sampling of Peptides. *J. Phys. Chem. B* **1997**, *101*, 817–824.

(17) Hukushima, K.; Nemoto, K.. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Jpn*. **1996**, *65*, 1604–1608.

(18) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett*. **1999**, *314*, 141–151.

(19) Lee, J. New Monte Carlo Algorithm: Entropic Sampling. *Phys. Rev. Lett*. **1993**, *71*, 211–214.

(20) Wang, F.; Landau, D. P. Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett*. **2001**, *86*, 2050–2053.

(21) Darve, E.; Pohorille, A. Calculating Free Energies Using Average Force. *J. Chem. Phys*. **2001**, *115*, 9169–9183.

(22) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12566.

(23) Iba, Y.; Chikenji, G.; Kikuchi, M. Simulation of Lattice Polymers with Multi-Self-Overlap Ensemble. *Phys. Soc. Jpn*. **1998**, *67*, 3327–3330.

(24) Swendsen, R. H.; Wang, J. S. Replica Monte Carlo Simulation of Spin Glasses. *Phys. Rev. Lett*. **1986**, *57*, 2607-2609.

(25) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys*. **2002**, *116*, 9058–9067.

(26) Higo, J.; Dasgupta, B., Mashimo T.; Kasahara, K.; Fukunishi, Y., Nakamura, H. Virtual-System-Coupled Adaptive Umbrella Sampling to Compute Free-Energy Landscape for Flexible Molecular Docking. *J. Comput. Chem.* **2015**, *36*, 1489-1501.

(27) Higo, J.; Kasahara, K.; Dasgupta, B.; Nakamura, H. Enhancement of Canonical Sampling by Virtual-State Transitions. *J. Chem. Phys.* **2017**, 146, 044104.

(28) Paine, G. H.; Scheraga, H. A. Prediction of the Native Conformation of a Polypeptide by a Statistical-Mechanical Procedure. I. Backbone Structure of Enkephalin. *Biopolymers* **1987**, *26*, 1391–1436.

(29) Mezei, M. Adaptive Umbrella Sampling: Self-Consistent Determination of the Non-Boltzmann Bias. *J. Comput. Phys.* **1987**, *68*, 237–248.

(30) Higo, J.; Nakamura, H. Virtual States Introduced for Overcoming Entropic Barriers in Conformational Space. *Biophysics* **2012**, *8*, 139–144.

(31) Higo, J.; Umezawa, K.; Nakamura, H. A Virtual-System Coupled Multicanonical Molecular Dynamics Simulation: Principle and Its Application to Free-Energy Landscape of Protein-Protein Interaction with an All-Atom Model in Explicit Solvent. *J. Chem. Phys.* **2013**, *138*, 184106.

(32) Dasgupta, B.; Higo J.; Nakamura, H. Faster Binding Free-Energy Landscape Calculation by Virtual-State Coupled Adaptive Umbrella Sampling. *Biophys. J.* **2016**, *110*, 55a.

(33) Terakawa, T.; Higo, J.; Takada, S. Multi-Scale Ensemble Modeling of Modular Proteins with Intrinsically Disordered Linker Regions: Application to p53. *Biophys. J.* **2014**, *107*, 721–729.

(34) Okuda, M.; Higo, J.; Komatsu, T.; Konuma, T.; Sugase, K., Nishimura, Y. Dynamics of the Extended String-Like Interaction of TFIIE with the p62 Subunit of TFIIH. *Biophys. J.* **2016**, 111, 950–962.

(35) Dasgupta, B.; Nakamura, H.; Higo J. Flexible Binding Simulation by a Novel and Improved Version of Virtual-System Coupled Adaptive Umbrella Sampling. *Chem. Phys. Lett.* **2016**, *662*, 327–332.

(36) Iida, S.; Mashimo, T.; Kurosawa, T.; Hojo, H.; Muta, H.; Goto, Y.; Fukunishi, Y.; Nakamura, H.; Higo, J. Variation of Free-Energy Landscape of the p53 C-terminal Domain Induced by Acetylation: Enhanced Conformational Sampling. *J. Comput. Chem.* **2016**, *37*, 2687–2700.

(37) Iida, S.; Kawabata T.; Kasahara, K.; Nakamura, H.; Higo, J. Multimodal Structural Distribution of the p53 C-Terminal Domain Upon Binding to S100B via a Generalised Ensemble Method: From Disorder to Extra-Disorder. *J. Chem. Theory Comput.* **2019**, *15*, 2597–2607.

(38) Hayami, T.; Higo, J.; Nakamura, H.; Kasahara, K. Multidimensional Virtual-System Coupled Canonical Molecular Dynamics to Compute Free-Energy Landscapes of Peptide Multimer Assembly. *J Comput. Chem.* **2019**, *40*, 2453–2463.

(39) 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.

(40) Kasahara, K.; Ma, B.; Goto, K.; Dasgupta, B.; Higo, J.; Fukuda, I.; Mashimo, T.; Akiyama, Y.; Nakamura, H. myPresto/Omegagene: a GPU-Accelerated Molecular Dynamics Simulator Tailored for Enhanced Conformational Sampling Methods with a Non-Ewald Electrostatic Scheme. *Biophys. Physicobiol.* **2016**, *13*, 209–216.

(41) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(42) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(43) Kamiya, N.; Fukuda, I.; Nakamura, H. Application of Zero-Dipole Summation Method to Molecular Dynamics Simulations of a Membrane Protein System. *Chem. Phys. Lett.* **2013**, *568–569*, 26–32.

(44) Fukuda, I.; Kamiya, N.; Yonezawa, Y.; Nakamura, H. Simple and Accurate Scheme to Compute Electrostatic Interaction: Zero-Dipole Summation Technique for Molecular System and Application to Bulk Water. *J. Chem. Phys.* **2012**, *137*, 054314.

(45) Fukuda, I.; Yonezawa, Y.; Nakamura, H. Molecular Dynamics Scheme for Precise Estimation of Electrostatic Interaction via Zero-Dipole Summation Principle. *J. Chem. Phys.* **2011**, *134*, 164107.

(46) Morishita, T. Fluctuation Formulas in Molecular-Dynamics Simulations with the Weak Coupling Heat Bath. *J Chem Phys* **2000**, *113*, 2976–2982.

(47) Kamiya, N.; Watanabe, Y. S.; Ono, S.; Higo, J. AMBER-based hybrid force field for conformational sampling of polypeptides. *Chem. Phys. Lett.* **2005**, *401*, 312–317.

(48) 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*, 926–935.

(49) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

(50) 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.

(51) 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.

(52) Arai, M. Unified Understanding of Folding and Binding Mechanisms of Globular and Intrinsically Disordered Proteins. *Biophys. Rev.* **2018**, *10*, 163–181.

(53) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. Speeding Molecular Recognition by Using the Folding Funnel: The Fly-Casting Mechanism. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8868–8873.

(54) Sugase, K.; Dyson, H. J.; Wright, P. E. Mechanism of Coupled Folding and Binding of an Intrinsically Disordered Protein. *Nature* **2007**, *447*, 1021–1025.

(55) Kofuku, Y.; Yoshiura, C.; Ueda, T.; Terasawa, H.; Hirai, T.; Tominaga, S.; Hirose, M.; Maeda, Y.; Takahashi, H.; Terashima, Y.; Matsushima, K.; Shimada, I. Structural Basis of the Interaction between Chemokine Stromal Cell-derived Factor-1/CXCL12 and Its G-protein-coupled Receptor CXCR4. *J. Biol. Chem.* **2009**, *284*, 35240–35250.

Supporting Information

Molecular interaction mechanism of 14-3-3 ε protein with phosphorylated Myeloid leukemia factor 1 revealed by an enhanced conformational sampling

Junichi Higo,[1,*] Takeshi Kawabata,[2] Ayumi Kusaka, [1] Kota Kasahara,[3] Narutoshi Kamiya, [1] Ikuo Fukuda, [1] Kentaro Mori, [1,4] Yutaka Hata, [1] Yoshifumi Fukunishi,[5] Haruki Nakamura. [2]

[1]Graduate School of Simulation Studies, University of Hyogo, 7-1-28 Minatojima Minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

[2]Institute for Protein Research, Osaka University, 3-2 Yamada-oka, Suita, Osaka 565-0871, Japan

[3]College of Life Sciences, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan

[4]Maizuru College, 234 Shiroya, Maizuru, Kyoto, 625-8511 Japan

[5]Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 2-3-26, Aomi, Koto-ku, Tokyo, 135-0064, Japan.

[*] Electronic mail: higo@protein.osaka-u.ac.jp

Figure S1

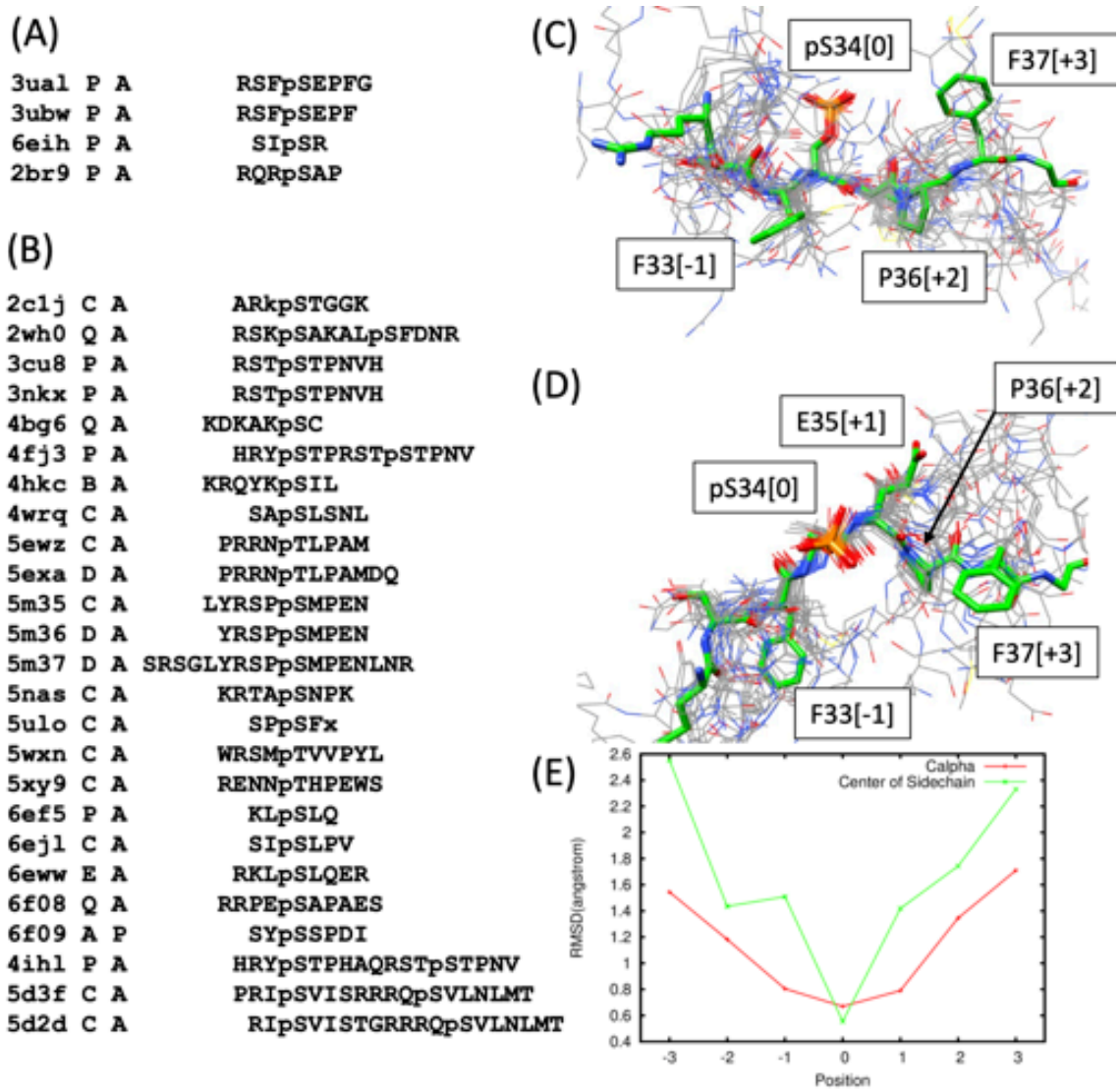

Figure S1. (A) Four structurally known phosphorylated peptides bound to human 14-3-3 ε protein. PDB IDs, chain IDs of binding peptide, chain IDs of 14-3-3 protein and amino acid sequences of binding peptides are shown. (B) 25 structurally known phosphorylated peptides bound to human 14-3-3ζ protein. Sequence identity between ε and ζ is 67 %. (C) A side view of the superimposed 3D structures of binding peptides to 14-3-3ε or 14-3-3ζ proteins. They were superimposed by their bound 3D structures of 14-3-3 proteins on A chain of PDB ID:3ual, using the program MATRAS[1]. The peptide structure of P chain (i.e., pMLF1) of PDB ID:3ual is shown in green. (D) A top view of the superimposed 3D structures of binding peptides. (E) RMSD plots of the superimposed 28 peptides with the P chain of PDB ID:3ual. Red line shows RMSDs of the corresponding Cα atoms, and the green line those of center of sidechain atoms. The residue numbering accords to UniProt (M.MLF1_HUMAN; https://www.uniprot.org/uniprot/P58340).

1. Atom group and reaction coordinate (RC)

Consider two atom groups G_h^A and G_h^B $(h = \alpha, \beta, \gamma)$ in a molecular system. The RC, $\lambda^{(h)}$, is defined by the distance between centers of mass of G_h^A and G_h^B (Fig. S2). Superscripts A and B

indicate simply that two atom groups are pairing to define $\lambda^{(h)}$, and then, one can exchange the superscripts as: $G_h^A \rightarrow G_h^B$ and $G_h^B \rightarrow G_h^A$ without changing the value of $\lambda^{(h)}$.

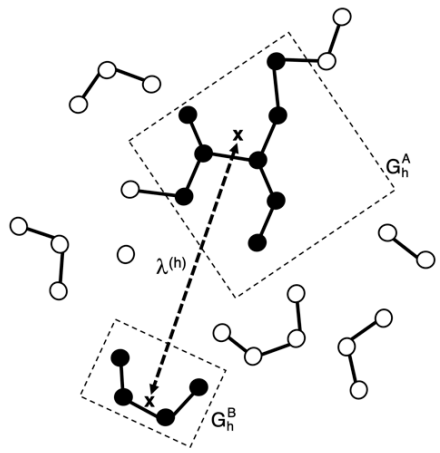

Figure S2. Two atom groups, G_h^A and G_h^B, are indicated by two rectangles, where in-group atoms are shown by black filled circles. Center of mass of each atom group is presented by a cross. The distance between the two centers of mass is $\lambda^{(h)}$ (broken-line with arrows).

2. Parameters used for mD-VcMD

Methodological detains for the GA-based mD-VcMD is explained elsewhere. Here, we explain parameters which control simulations. The atom groups defining three RCs are listed in Table S1. Also see Fig. 1 of main text, which graphically shows the atom groups and RCs. Table S2 lists parameters $n_{vs}(h)$, $[\zeta_1^{(h)}]_{min}$, $[\zeta_{n_{vs}(h)}^{(h)}]_{max}$, and $\Delta\lambda^{(h)}$. Table S3 lists actual setting of zones: $\{[\lambda_k^{(h)}]_{min}, [\lambda_k^{(h)}]_{max};\ k = 1, \ldots, n_{vs}(h)\}$. An interzone transition was attempted once every $20\ ps$: The IVT transition interval was set to 1×10^4 steps.

Table S1. Atom groups to define three RCs.

	Atom Group[a)]	
RC	G_h^A	G_h^B
$\lambda^{(\alpha)}$	Arg 130-Glu 134[b)]	Arg 31
$\lambda^{(\beta)}$	Leu 173-Val 179	Sep 34/Ser 34[c)]
$\lambda^{(\gamma)}$	Ala 48-Val 52	Phe 37

[a] Two atom groups are expressed as G_h^A or G_h^B ($h = \alpha, \beta, \gamma$) to define $\lambda^{(h)}$: $\lambda^{(h)}$ is defined by the distance between the centers of mass of G_h^A and G_h^B. Superscripts A and B are assigned to 14-3-3ε and pMLF1 (or npMLF1), respectively.

[b] "$X\ n_1$-$Y\ n_2$" represents all mainchain heavy and Cβ atoms from amino-acid X at the n_1-th residue to amino-acid Y at the n_2-th residue.

[c] Sep 34 and Ser 34 are the phosphorylated serine of pMLF1 and the serine of npMLF1 at the 34-th residue, respectively.

Table S2. Parameters[a] for RC-space division

h	$n_{vs}(h)$	$[\zeta_1^{(h)}]_{min}$	$[\zeta_{n_{vs}(h)}^{(h)}]_{max}$	$\Delta\lambda^{(h)}$
α	16	6 Å	40.0 Å	4 Å
β	22	6 Å	40.5 Å	3 Å
γ	16	6 Å	40.0 Å	4 Å

TableS3. Setting of zones.

Zone No. [a]	zones [b]				
k	$[\lambda_k^{(\alpha)}]_{min}$	$[\lambda_k^{(\alpha)}]_{max}$	$[\lambda_k^{(\beta)}]_{min}$	$[\lambda_k^{(\beta)}]_{max}$	$[\lambda_k^{(\gamma)}]_{min}$ $[\lambda_k^{(\gamma)}]_{max}$ [c]
1	6.0	10.0	6.0	9.0	
2	8.0	12.0	7.5	10.5	
3	10.0	14.0	9.0	12.0	

4	12.0	16.0	10.5	13.5
5	14.0	18.0	12.0	15.0
6	16.0	20.0	13.5	16.5
7	18.0	22.0	15.0	18.0
8	20.0	24.0	16.5	19.5
9	22.0	26.0	18.0	21.0
10	24.0	28.0	19.5	22.5
11	26.0	30.0	21.0	24.0
12	28.0	32.0	22.5	25.5
13	30.0	34.0	24.0	27.0
14	32.0	36.0	25.5	28.5
15	34.0	38.0	27.0	30.0
16	36.0	40.0	28.5	31.5
17			30.0	33.0
18			31.5	34.5
19			33.0	36.0
20			34.5	37.5
21			36.0	39.0
22			37.5	40.5

[a] Number of zones n_{vs} is 16 for $\lambda^{(\alpha)}$ and $\lambda^{(\beta)}$, and 22 for $\lambda^{(\gamma)}$.

[b] Unit of zones is Å. The meaning of $[\lambda_k^{(h)}]_{min}$ and $[\lambda_k^{(h)}]_{max}$ $(h = \alpha, \beta, \gamma;\ k = 1, \ldots, n_{vs}(\lambda^{(h)})$ is given in Section 2 of SI.

[c] Parameters for $\lambda^{(\gamma)}$ are not shown because those are exactly the same as for $\lambda^{(\alpha)}$.

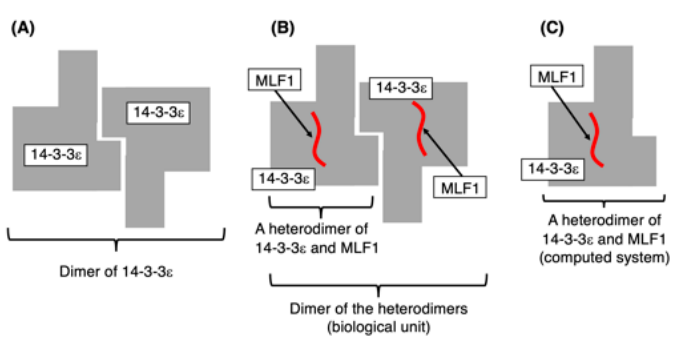

Figure S3

Figure S3. Shaded polygon and red-colored string are receptor 14-3-3ε and ligand MLF1, respectively. (A) Dimer of 14-3-3ε formed in the

unbound state of 14-3-3ε. (B) Dimer of heterodimers [MLF1/14-3-3ε]$_2$ formed when MLF1 binds to 14-3-3ε. (C) Computed system.

Figure S4

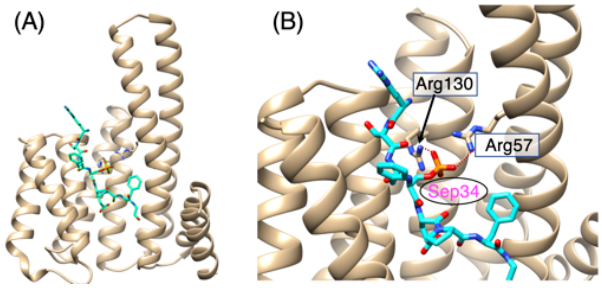

Figure S4. (A) Overview of MLF1/14-3-3ε complex. Ligand pMLF1, phosphorylated at serine 34 (Sep 34), is shown by cyan-colored model, and 14-3-3ε by light yellow-colored model. (B) Close up of two residues Arg 57 and Arg 130 of 14-3-3ε and Sep 34, which form salt-bridges (brown-colored broken lines).

Figure S5

Figure S5. The initial conformations of GA-based mD-VcMD for (A) pMLF1/14-3-3ε and (B) npMLF1/14-3-3ε systems. Eight conformations picked randomly from 256 initial ones are shown. Red-colored model in panel (A) is pMLF1 in the native complex structure. Note that this red-colored model is not used for the initial conformation of the GA-based mD-VcMD.

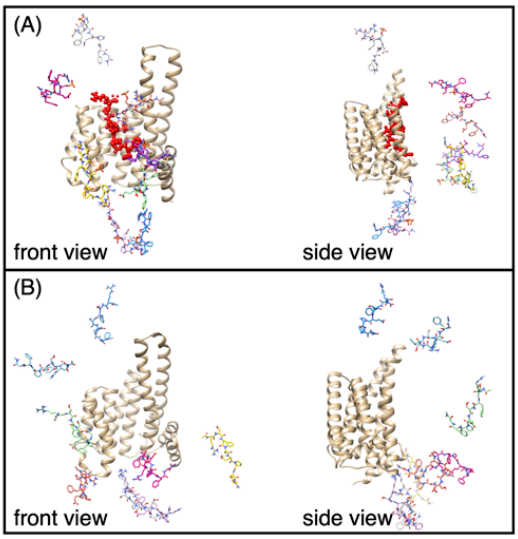

Figure S6

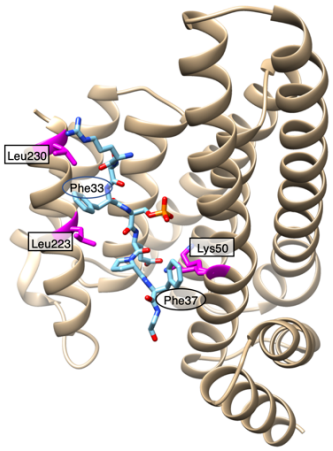

Figure S6. Phe 33 and Phe 37 belong to pMLF1. Lys 50, Leu 223, and Leu 230 of 14-3-3ε are shown by magenta. In the crystal structure, Phe 33 of pMLF1 contacts to Leu 223 and Leu 230 of 14-3-3ε. Phe 37 contacts to the stem of Lys 50 of 14-3-3ε. Two hetero sidechains are shown for Lys 50.

Figure S7

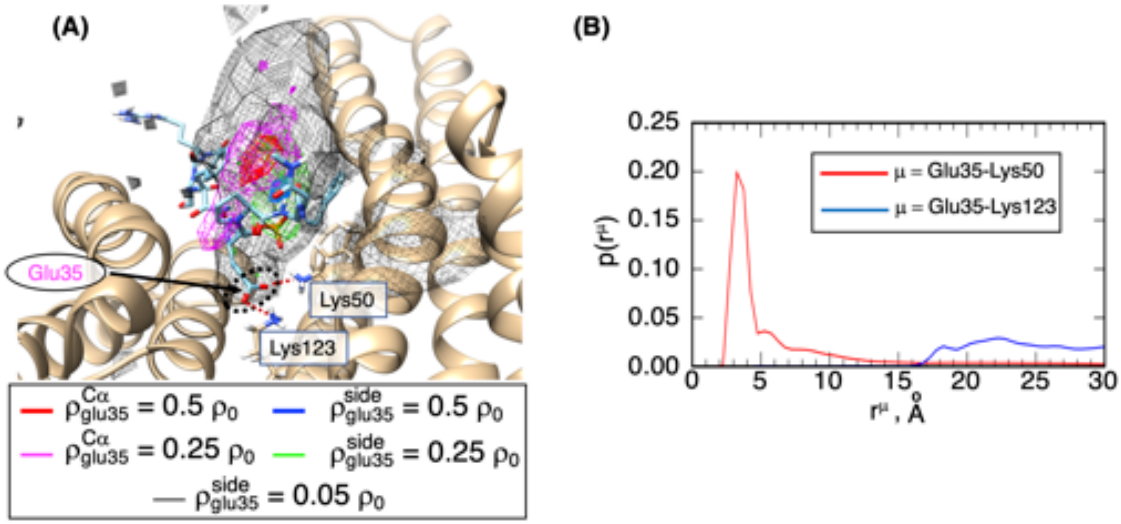

Figure S7. (A) Density contour maps of $\rho^{C\alpha}_{Glu35}(\boldsymbol{r})$ and $\rho^{C\gamma}_{Glu35}(\boldsymbol{r})$ for pMLF1/14-3-3ε. Shown molecular structure is the equilibrated one (NPT simulation), focusing on Glu 35 of pMLF1. Contour levels are shown in inset ($\rho_0 = 0.01$). Black-colored contours ($\rho^{C\gamma}_{Glu35}(\boldsymbol{r}) = 0.05\rho_0$) cover the sidechain-tip position of Glu 35 of the equilibrated structure, where Glu 35 forms salt bridges (brown-colored broken line) to Lys 50 and Lys 123 of 14-3-3ε. (B) Radial distribution function, $p(r^{Glu35-Lys50})$, for distance $r^{Glu35-Lys50}$ from the Cγ atom of Glu 35 of pMLF1 to the Nζ atom of Lys 50 of 14-3-3ε, and radial distribution function $p(r^{Glu35-Lys123})$ for distance $r^{Glu35-Lys123}$ from the Cγ atom of Glu 35 to the Nζ atom of Lys 123 of 14-3-3ε.

References

[1] Kawabata, T.; Nishikawa, K. Protein structure comparison using the Markov transition model of evolution. *Proteins* **2000**, *41*, 108–122.