Supporting Information:
Explicit Characterization of the Free Energy Landscape of pKID–KIX Coupled Folding and Binding

Song-Ho Chong, Haeri Im, and Sihyun Ham*

Department of Chemistry, The Research Institute of Natural Sciences, Sookmyung Women’s University, Cheongpa-ro 47-gil 100, Yongsan-Ku, Seoul 04310, Korea

E-mail: sihyun@sookmyung.ac.kr
Phone: +82 2 710 9410. Fax: +82 2 2077 7321
METHODS

Free Energy Relations and Equations

We outline the derivation of the relations and equations discussed in the main text that involve various free energies. The starting point is the potential energy part of the partition function for a solute-solvent system

\[ Z_{\text{tot}} = \int dr_u \int dr_v e^{-\beta E_{\text{tot}}(r_u, r_v)} \]  

(S1)

Here, \( r_u \) and \( r_v \) abbreviate all the coordinates of atoms constituting the solute and solvent, respectively, and \( E_{\text{tot}} = E_u + E_{uv} + E_v \) is the total potential comprising the solute potential \( (E_u) \), the solute-solvent interaction potential \( (E_{uv}) \), and the solvent potential \( (E_v) \). By introducing the solvation free energy \( G_{\text{solv}}(r_u) \) for a specific solute configuration \( r_u \)

\[ e^{-\beta G_{\text{solv}}(r_u)} = (1/Z_v) \int dr_v e^{-\beta [E_{uv}(r_u, r_v) + E_v(r_v)]} \]  

(S2)

where \( Z_v = \int dr_v e^{-\beta E_v(r_v)} \), one obtains

\[ Z \equiv (Z_{\text{tot}}/Z_v) = \int dr_u e^{-\beta f(r_u)} \]  

(S3)

that solely involves the solute configuration \( r_u \). Here enters \( f = E_u + G_{\text{solv}} \) that defines the free energy landscape. \( f \) also determines the thermodynamics of the system through \( F = -k_B T \log Z \).

We next outline the derivation of the equation of motion which indicates that the solute dynamics is also dictated by \( f \). The derivation proceeds as follows. (i) The starting point is Newton’s equations of motion for the solute variables \( A_u = \{r_u, p_u\} \) and solvent variables \( A_v = \{r_v, p_v\} \), where \( p_u \) and \( p_v \) are the solute and solvent momenta, respectively. (ii) A projection operator \( \mathcal{P}_u \) is introduced that projects any dynamical variable onto the space
spanned by $g_a = \delta(A_u - a)$, where $a$ denotes a specific realization of $A_u$. $P_u$ has an effect of projecting out the solvent degrees of freedom from the equation of motion. (iii) We apply the projection-operator formalism developed by Mori and Fujisaka.\textsuperscript{S2} (The orthonormality of $g_a$ is assumed in the original work, but this assumption can easily be removed.) (iv) We assume the time-scale separation between the solute and solvent dynamics. These lead to the following Langevin-type equation for the solute dynamics (with $r_{u,i}$ denoting the position of a solute atom $i$, $m_i$ its mass, and the dot the time derivative):

$$m_i \ddot{r}_{u,i} = -\partial f/\partial r_{u,i} - \gamma_0 \dot{r}_{u,i} + R_{u,i}. \tag{S4}$$

Here, the bare friction coefficient $\gamma_0$ and the random force $R_{u,i}$ are connected by the fluctuation-dissipation relation. Equation S4 clearly indicates that the solute dynamics is also dictated by $f$. It also guarantees that, after long times, the solute configuration $r_u$ is populated proportional to $e^{-\beta f(r_u)}$, which is in consistent with Equation S3. From here on, the solute configuration shall be denoted as $r$ for notational simplicity.

The relation between $f(r)$ and the free energy profile $F(Q) = -k_B T \log Z(Q)$ can be derived as follows. Here, $Z(Q)$ is the restricted partition function defined only with those solute configurations $\{r\}$ satisfying $Q = Q(r)$:

$$Z(Q) = \int_{Q=Q(r)} d\mathbf{r} e^{-\beta f(\mathbf{r})} \tag{S5}$$

(In the main text, $F(Q)$ is defined as $-k_B T \log P(Q)$ with $P(Q) = Z(Q)/Z$ since this definition is more familiar in literature; $F(Q)$’s from these two definitions differ only by an unimportant overall constant.) Let us introduce a probability distribution $P_Q(\mathbf{r})$ defined such that $P_Q(\mathbf{r}) = e^{-\beta f(\mathbf{r})}/Z(Q)$ for those configurations satisfying $Q = Q(\mathbf{r})$ and $P_Q(\mathbf{r}) = 0$ otherwise. The solute configurational entropy $S_{\text{config}}(Q)$ associated with the configurations
conforming to $Q = Q(r)$ can then be defined as

$$S_{\text{config}}(Q) = -k_B \int_{Q=Q(r)} d\mathbf{r} P_Q(\mathbf{r}) \log P_Q(\mathbf{r})$$  \hspace{1cm} (S6)

Using $P_Q(\mathbf{r}) = e^{-\beta f(r)}/Z(Q)$, we get $F(Q) = f(Q) - TS_{\text{config}}(Q)$, where $f(Q) = \langle f(\mathbf{r}) \rangle_Q$ is the average taken with $P_Q(\mathbf{r})$.

The relation between $F(Q)$ and the thermodynamic free energy $F_X$ for state $X$ ($= f$ or $u$) can be derived in a similar manner. We shall characterize state $X$ as the certain region of the reaction coordinate: $Q > Q_f$ for the folded state and $Q < Q_u$ for the unfolded state with cutoffs $Q_f$ and $Q_u$. Then, $F_X$ is given by $F_X = -k_B T \log Z_X$ with $Z_X = \int_X dQ Z(Q)$, where $X$ in the $Q$-integration refers either to $Q > Q_f$ or $Q < Q_u$. Let us introduce a probability distribution $P_X(Q)$ defined such that $P_X(Q) = Z(Q)/Z_X$ when $Q$ belongs to state $X$ and $P_X(Q) = 0$ otherwise. The configurational entropy $S_{\text{config},X}^{\text{distr}}$, associated with the presence of various $Q$’s in state $X$, can then be defined as

$$S_{\text{config},X}^{\text{distr}} = -k_B \int_X dQ P_X(Q) \log P_X(Q)$$  \hspace{1cm} (S7)

Using $P_X(Q) = Z(Q)/Z_X$, we get $F_X = \langle F(Q) \rangle_X - TS_{\text{config},X}^{\text{distr}}$, where $\langle \ldots \rangle_Q$ is the average taken with $P_X(Q)$. Thus, the thermodynamic $S_{\text{config},X}$ comprises $\langle S_{\text{config}}(Q) \rangle_X$ and $S_{\text{config},X}^{\text{distr}}$.

**Molecular Dynamics Simulations**

We investigated the 34-residue form of the pKID region (residues 116–149) and the 87-residue construct of the KIX domain (residues 586–672). We first carried out molecular dynamics simulation for the free pKID and the equilibrium simulations for the pKID–KIX complex: the former simulation is done to obtain a disordered pKID configuration to be used in the spontaneous binding simulations, and the latter simulations to investigate the nature of the native intermolecular contacts. (All the simulations reported in the present study are summarized in Table S1.) The initial structures for these simulations were obtained from
the corresponding parts of the NMR complex structure (PDB entry 2LXT\textsuperscript{S4}). We then performed spontaneous pKID–KIX binding simulations using the disordered pKID structure taken from the free pKID simulation and the KIX structure from the NMR complex structure, which were placed with a $\sim$60 Å center-of-mass distance. No artificial attractive force was applied between pKID and KIX. All the simulations were done at $T = 300$ K and $P = 1$ bar using the pmemd.cuda module in AMBER16\textsuperscript{S5} with the CHARMM22* force field\textsuperscript{S6} and the TIP3P water model.\textsuperscript{S7} The CHAMBER utility\textsuperscript{S8} was employed to use CHARMM22*.

The following common procedures were applied to conduct our simulations. The system was filled with water molecules and neutralizing counter ions in a cubic box with a buffer size of 15 Å. The energy minimization of the system was first carried out with harmonic restraints of a 500 kcal/(mol Å$^2$) force constant (1000/4000 steps of steepest descent/conjugate gradient minimizations), followed by the one without restraints (5000/5000 steps of steepest descent/conjugate gradient minimizations). We then increased the system temperature from 0 K to 300 K with a 20 ps simulation at constant volume, followed by a 200 ps simulation at $P = 1$ bar with the Berendsen’s method.\textsuperscript{S9} These equilibration steps were repeated with different seeds for initial velocities to perform independent production runs. The first 5 ns part of the production runs was done at constant pressure, and the rest at constant volume. Lennard-Jones interactions were truncated at 12 Å, and Coulomb interactions were treated with the particle mesh Ewald method.\textsuperscript{S10} We used the SHAKE algorithm\textsuperscript{S11} to perform the simulations with a 2 fs time step. A production run of 1 $\mu$s length was performed for the free pKID; six independent equilibrium simulations of 1 $\mu$s length for the pKID–KIX complex; and ten independent simulations of 1.5 to 10 $\mu$s length for the spontaneous pKID–KIX binding. More details of our simulations are summarized in Table S1.

**Structural Analyses**

We computed the fraction of native intermolecular contacts ($Q_{\text{int}}$) and that of native intrapKID contacts ($Q_{\text{pKID}}$) following the procedure detailed in Ref. S3 based on a list of native
contact pairs. Native contact pairs refer to those heavy atom contact pairs – \( i \) atom in residue \( \theta_i \) and \( j \) atom in residue \( \theta_j \) separated by less than 4.5 Å and satisfying \( |\theta_i - \theta_j| > 3 \) in the case of intramolecular contacts – found in the native structure, which is usually taken from an X-ray or NMR study. However, if we define native contact pairs for the pKID–KIX complex using the NMR structure (PDB entry 2LXT\textsuperscript{S4}), we obtain, on average, \( Q_{\text{int}} = 0.48 \) and \( Q_{\text{pKID}} = 0.74 \) from our equilibrium pKID–KIX complex simulations. Since the typical fraction of native contacts exceeds \( \sim 0.9 \) for the simulated folded structures,\textsuperscript{S3} this implies that the simulated pKID–KIX structures largely deviate from the NMR structure. Indeed, we observe large (>5 Å) \( \text{C}^{\alpha} \) root-mean-square fluctuations to the NMR structure as shown in Figure S11. In this connection, we notice that another peptide – the mixed-lineage leukemia (MLL) peptide – is actually bound in the NMR complex structure (see Figure S11), which is removed in our simulations. An inspection of Figure S11, with the knowledge of the pKID and MLL binding sites on KIX, indicates that the removal of MLL in our simulations significantly affects the KIX and pKID structures. This is in agreement with the presence of allosteric communication between pKID and MLL through KIX.\textsuperscript{S4} It is therefore inappropriate to adopt the NMR complex structure in defining the native contact pairs, and we instead used the simulated complex structures for this purpose. Rather than choosing a single representative structure, native contact pairs were defined based on well populated contact pairs during the equilibrium complex simulations: with a choice of 50/60/70 % population, we obtain, on average, \( Q_{\text{int}} = 0.87/0.93/0.97 \) and \( Q_{\text{pKID}} = 0.93/0.93/0.98 \). The results reported in the main text refer to \( Q_{\text{int}} \) and \( Q_{\text{pKID}} \) defined with 70% population: we confirmed that similar results were obtained with choices of 50% and 60% population.

The alpha helical contents were computed based on the DSSP algorithm.\textsuperscript{S12} The angle between the pKID \( \alpha_A \) and \( \alpha_B \) helices was calculated based on \( \text{C}^{\alpha} \) atoms of Arg-125, Arg-131 and Arg-135. A heavy atom contact is considered formed between two residues if the minimum heavy-atom distance is <4.5 Å. Typical contacts between hydrophobic residues fulfill such a criterion. A hydrogen bond is considered formed between two residues if the
minimum O-O, O-N or N-N distance is <3.5 Å. The APBS (Adaptive Poisson-Boltzmann Solver) plugin\textsuperscript{S13} implemented in PyMOL\textsuperscript{S14} was used for computing the surface electrostatic potential displayed in Figure 4a of the main text.

**Computational Procedure of Solvation Free Energy**

For each of the simulated solute configurations, we applied the 3D-RISM (three-dimensional reference interaction site model) theory\textsuperscript{S15,S16} to compute the solvation free energy $G_{\text{solv}}$. This is an integral-equation theory for obtaining the 3D distribution function $g_\gamma(r)$ of the water site $\gamma$ at position $r$ around the solute. In this theory, the distribution function is obtained by self-consistently solving the 3D-RISM equation

$$h_\gamma(r) = \sum_\gamma' \int dr' \chi_{\gamma\gamma'}(|r-r'|) c_{\gamma'}(r')$$

(S8)

and the approximate closure relation

$$h_\gamma(r) = \begin{cases} \exp[d_\gamma(r)] - 1 & \text{for } d_\gamma(r) \leq 0 \\ d_\gamma(r) & \text{for } d_\gamma(r) > 0 \end{cases}$$

(S9)

in which $d_\gamma(r) = -u_\gamma(r)/(k_B T) + h_\gamma(r) - c_\gamma(r)$. Here $h_\gamma(r) = g_\gamma(r) - 1$ and $c_\gamma(r)$ are the total and direct correlation functions, respectively; $\chi_{\gamma\gamma'}(r)$ denotes the site-site solvent susceptibility function which can be obtained either from simulations or integral-equation calculations; and $u_\gamma(r)$ is the solute-solvent interaction potential for a given solute configuration. We used the same numerical procedure as described in Ref. S16 to solve the above equations. Solvation free energy can then be computed from the following analytical expression:

$$G_{\text{solv}} = \rho k_B T \sum_\gamma \int dr \left[ \frac{1}{2} h_\gamma(r)^2 \Theta(-h_\gamma(r)) - c_\gamma(r) - \frac{1}{2} h_\gamma(r)c_\gamma(r) \right]$$

(S10)

Here, $\rho$ is the average solvent number density, and $\Theta$ is the Heaviside step function.
Construction of the Free Energy Landscape

The free energy \( f = E_u + G_{\text{solv}} \) that defines the landscape is given by the gas-phase potential energy \( (E_u) \), which can be calculated from the force field parameters adopted in the simulations, and the solvation free energy \( (G_{\text{solv}}) \), which can be computed from the 3D-RISM theory. The free energy landscape \( f(Q_{\text{int}}, Q_{\text{pKID}}) \) presented in Figure 3b of the main text was constructed from the successful 10 \( \mu \)s spontaneous pKID–KIX binding simulation as follows. We first computed \( Q_{\text{int}} \), \( Q_{\text{pKID}} \), and \( f \) along the trajectory using \( 10^6 \) configurations saved with a 10 ps interval. The results shown in Figure 2c (\( Q_{\text{int}} \) and \( Q_{\text{pKID}} \) vs. time) and Figure 3a (\( f \) vs. time) of the main text are running averages over 1 ns (100 configurations). By dividing the configuration space spanned by \( 0 \leq Q_{\text{int}} \leq 1 \) and \( 0 \leq Q_{\text{pKID}} \leq 1 \) into 2500 small areas after discretizing each of \( Q_{\text{int}} \) and \( Q_{\text{pKID}} \) into 50 bins, we collected simulated configurations that belong to each small area and then computed the average \( f \) value from those configurations; the resulting \( f(Q_{\text{int}}, Q_{\text{pKID}}) \) is displayed as the surface colored yellow in Figure 3b of the main text. The free energy landscape \( f(n_{\text{contact}}, Q_{\text{pKID}}) \) shown in Figure 5a of the main text was constructed in a similar manner (the configuration space spanned by \( 0 \leq n_{\text{contact}} \leq 400 \) and \( 0 \leq Q_{\text{pKID}} \leq 1 \) is divided into \( 50 \times 50 = 2500 \) small areas) by combining all the configurations (\( 1.65 \times 10^6 \) configurations) taken from the nine unsuccessful binding simulation trajectories.
References

(S1) Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press: New York, 2006.

(S2) Mori, H.; Fujisaka, H. On nonlinear dynamics of fluctuations. *Prog Theor Phys* 1973, 49, 764–775.

(S3) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc Natl Acad Sci USA* 2013, 110, 17874–17879.

(S4) Brüschweiler, S.; Konrat, R.; Tollinger, M. Allosteric communication in the KIX domain proceeds through dynamic repacking of the hydrophobic core. *ACS Chem Biol* 2013, 8, 1600–1610.

(S5) Case, D. A.; Ben-Shalom, I.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. AMBER16 (University of California, San Francisco). 2016.

(S6) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys J* 2011, 100, L47–L49.

(S7) 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.
(S8) Crowley, M. F.; Williamson, M. J.; Walker, R. C. CHAMBER: Comprehensive support for CHARMM force fields within the AMBER software. *Int J Quantum Chem* **2009**, *109*, 3767–3772.

(S9) 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.

(S10) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N$-log($N$) method for Ewald sums in large systems. *J Chem Phys* **1993**, *98*, 10089–10092.

(S11) 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.

(S12) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.

(S13) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **2001**, *98*, 10037–10041.

(S14) Schrödinger LLC, The PyMOL molecular graphics system, version 1.8. 2015.

(S15) Kovalenko, A. Three-dimensional RISM theory for molecular liquids and solid-solid interfaces. Molecular Theory of Solvation. Dordrecht, 2003; pp 169–275.

(S16) Imai, T.; Harano, Y.; Kinoshita, M.; Kovalenko, A.; Hirata, F. A theoretical analysis on hydration thermodynamics of proteins. *J Chem Phys* **2006**, *125*, 024911.
Figure S1: Free energy landscape constructed from 12 individual folding pathways in the ∼400 µs simulation trajectory of HP-35 and comparison with the average $f(Q)$. 
Figure S2: Fraction of native intra-pKID contacts ($Q_{pKID}$) versus the simulation time from an equilibration simulation of free pKID, initiated from the folded structure taken from the NMR complex structure. The simulated structure at 912 ns is displayed, which was used as the initial pKID structure in our spontaneous pKID–KIX binding simulations.
Figure S3: Fraction of native intermolecular contacts $Q_{\text{int}}$ (left panel) and fraction of native intra-pKID contacts $Q_{\text{pKID}}$ (right panel) versus the simulation time from the successful 10 µs spontaneous pKID–KIX binding trajectory.
Figure S4: (a,b) Alpha helical contents of the pKID $\alpha_B$ helix (a) and the angle between the pKID $\alpha_A$ and $\alpha_B$ helices (b) versus the simulation time from the successful 10 $\mu$s spontaneous pKID–KIX binding trajectory. The top panels show the results up to 3 $\mu$s in which the time regions are colored as in Figure 2c of the main text to highlight the various stages of the binding process. The bottom panels display the results for the entire time range.
Figure S5: Free energy $f = f_{pKID} + \Delta f_{int}$ versus the simulation time from the successful 10 $\mu$s spontaneous pKID–KIX binding trajectory.
Figure S6: Steps to construct the folding landscape of pKID based on the free energy landscape $f(Q_{\text{int}}, Q_{pKID})$ shown in Figure 3b of the main text. In the left middle panel, the dashed lines denote the linear fits to the respective curves of the same color. The left bottom panel is obtained using the polar coordinate $r = 1 - Q_{pKID}$: the linear fits were used for those $Q_{pKID}$ values for which the simulation data are unavailable.
Figure S7: Comparison of the free energy landscape \( f(Q_{\text{int}}, Q_{\text{pKID}}) \) projected onto the \( Q_{\text{pKID}} \) axis for the initial diffusive stage (red squares) and the final specific complex stage (cyan squares) taken from Figure 3c of the main text with the result from the equilibration simulation of the free pKID (brown circles) and from the equilibrium pKID–KIX complex simulations (green circles). To obtain \( f = f_{\text{pKID}} + \Delta f \) for the equilibration simulation of the free pKID, from which only \( f_{\text{pKID}} \) can be computed, we added a constant \( \Delta f \) estimated from \( f = f_{\text{pKID}} + \Delta f \) of the initial diffusive stage at \( Q_{\text{pKID}} = 0.6 \).
Figure S8: Free energy landscape of HP-35 (colored orange; taken from Figure 1a of the main text) versus pKID (colored red for before binding, magenta for the initial folding stage, and cyan for the final specific complex stage; taken from Figure 3c of the main text). Red and orange dashed lines denote the linear fits to the corresponding curves of the same color. For a direct comparison, the free energy relative to its $Q = 0$ value (estimated from the respective linear fits) is plotted.
Figure S9: Presence (indicated by a dot) of the hydrophobic contacts (left panel), of the contacts involving the phosphorylated Ser-133 (middle panel), and of the contacts associated with the local surface electrostatic complementarity (right panel) versus the simulation time from the successful 10 µs spontaneous pKID–KIX binding trajectory.
Figure S10: (a–g) Unsuccessful pKID–KIX binding trajectories (labeled a to g) projected onto the free energy landscape $f(n_{\text{contact}}, Q_{\text{pKID}})$ along with their final structures. The time regions from 0 to 1000 ns and from 1000 to 1500 ns are colored red and orange, respectively.
Figure S11: (a) NMR complex structure of PDB entry 2LXT that consists of pKID (colored orange), KIX (green), and MLL (cyan). The MLL peptide is removed when we prepare for the starting configurations of our simulations. (b,c) Cα root-mean-square fluctuations (RMSF) relative to the NMR complex structure for pKID (b) and KIX (c). Averages (filled circles) and standard errors (error bars) were computed from the six independent equilibrium simulations of the pKID–KIX complex. Locations of the alpha helices (horizontal bars) and contacting residues (squares) are indicated in each panel. The removal of MLL in our equilibrium pKID–KIX complex simulations affects those residues in KIX which are in contact with MLL in the NMR complex structure (marked with cyan squares). This in turn affects the RMSF values of the pKID residues.
Table S1: Summary of the simulations reported in the present work

| simulation type                                      | waters | ions | box size [Å³] | # of traj. | length |
|-----------------------------------------------------|--------|------|---------------|-----------|--------|
| (i) Equilibration simulation of free pKID           |        |      |               |           |        |
| (to obtain disordered pKID structure)               |        |      |               |           |        |
|                                                     | 16,644 | none | 80.7 × 80.4 × 80.9 | 1          | 1.0 µs |
| (ii) Equilibrium pKID–KIX complex simulation        |        |      |               |           |        |
| (to analyze native intermolecular contacts)         |        |      |               |           |        |
|                                                     | 23,168 | 3 Cl⁻ | 90.7 × 90.4 × 90.7 | 6          | 1.0 µs |
| (iii) Spontaneous pKID–KIX binding simulation       |        |      |               |           |        |
|                                                     | 46,717 | 3 Cl⁻ | 113.8 × 113.6 × 114.0 | 1          | 10.0 µs|
|                                                     |        |      |               | 2          | 3.0 µs |
|                                                     |        |      |               | 7          | 1.5 µs |

| total simulation time                                |        |      |               |           | 33.5 µs|

*a Initial size of the simulation box.  b Number of independent production runs.
Table S2: Population (%) of residue contacts at the pKID–KIX binding interface

| pKID (helix ID) | KIX (helix ID) | heavy atom contact | hydrogen bond/salt bridge |
|----------------|---------------|--------------------|--------------------------|
| L128 (α_A)     | Y658 (α_3)    | 84.5 ± 1.7         |                          |
| R131           | Y658 (α_3)    | 96.5 ± 0.6         |                          |
| Y134 (α_B)     | A654 (α_3)    | 98.3 ± 0.5         |                          |
| Y134 (α_B)     | E655 (α_3)    | 95.4 ± 0.9         |                          |
| Y134 (α_B)     | Y658 (α_3)    | 97.2 ± 0.6         |                          |
| I137 (α_B)     | A654 (α_3)    | 96.6 ± 1.0         |                          |
| I137 (α_B)     | I657 (α_3)    | 97.6 ± 0.6         |                          |
| I137 (α_B)     | Y658 (α_3)    | 93.4 ± 0.8         |                          |
| L138 (α_B)     | A654 (α_3)    | 85.9 ± 3.3         |                          |
| L141 (α_B)     | L603 (α_1)    | 78.0 ± 14.2        |                          |
| L141 (α_B)     | Y650 (α_3)    | 83.6 ± 13.7        |                          |
| L141 (α_B)     | L653 (α_3)    | 80.5 ± 14.7        |                          |
| L141 (α_B)     | A654 (α_3)    | 78.0 ± 14.3        |                          |
| R124 (α_A)     | E655 (α_3)    | 61.2 ± 10.4        | 57.7 ± 10.7              |
| R125 (α_A)     | E648 (α_3)    | 81.3 ± 5.6         | 63.2 ± 9.2               |
| R125 (α_A)     | H651 (α_3)    | 94.2 ± 4.3         | 33.9 ± 12.1              |
| pS133          | Y658 (α_3)    | 96.8 ± 1.1         | 49.4 ± 16.2              |
| pS133          | K662 (α_3)    | 90.4 ± 3.9         | 89.3 ± 4.1               |
| pS133          | R669          | 46.5 ± 14.2        | 45.9 ± 14.2              |
| pS133          | R671          | 71.5 ± 6.0         | 71.5 ± 6.0               |
| Y134 (α_B)     | H651 (α_3)    | 98.9 ± 0.7         | 60.0 ± 12.4              |
| D140 (α_B)     | K606 (α_1)    | 87.8 ± 4.1         | 67.5 ± 6.9               |
| D144 (α_B)     | R646 (α_3)    | 35.7 ± 7.4         | 27.4 ± 6.0               |

*a Average ± standard error computed from the six independent 1 µs equilibrium simulations of the pKID–KIX complex.