

## Supporting Information for

### Temperature-Dependent Fold-Switching Mechanism of the Circadian Clock Protein KaiB

Ning Zhang<sup>1,8,†</sup>, Damini Sood<sup>1,8,§</sup>, Spencer C. Guo<sup>2,8</sup>, Nanhao Chen<sup>3</sup>, Adam Antoszewski<sup>2,‡</sup>, Tegan Marianchuk<sup>4</sup>, Supratim Dey<sup>1,#</sup>, Yunxian Xiao<sup>1</sup>, Lu Hong<sup>4,¶</sup>, Xiangda Peng<sup>5</sup>, Michael Baxa<sup>5</sup>, Carrie Partch<sup>6</sup>, Lee-Ping Wang<sup>3,\*</sup>, Tobin R. Sosnick<sup>5,\*</sup>, Aaron R. Dinner<sup>2,\*</sup>, Andy LiWang<sup>1,7,\*</sup>

<sup>1</sup>Department of Chemistry & Biochemistry, University of California, Merced, CA, 95343, United States

<sup>2</sup>Department of Chemistry and James Franck Institute, University of Chicago, Chicago, IL, 60637

<sup>3</sup>Department of Chemistry, University of California, Davis, CA, 95616

<sup>4</sup>Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL, 60637

<sup>5</sup>Department of Biochemistry & Molecular Biology, University of Chicago, Chicago, IL, 60637

<sup>6</sup>Department of Chemistry & Biochemistry, University of California, Santa Cruz, CA, 95064

<sup>7</sup>Center for Cellular and Biomolecular Machines, University of California, Merced, CA, 95343

<sup>8</sup>Equal contribution.

\*Corresponding authors.

<sup>†</sup>Current address: Qingdao Institute of Bioenergy and Process, Chinese Academy of Sciences, Qingdao 266101, China

<sup>§</sup>Department of Medicine, Duke University Medical Center, Durham, NC 27710

<sup>‡</sup>Current address: D. E. Shaw Research, New York, NY 10036

<sup>#</sup>Current address: Feinberg School of Medicine, Northwestern University, Chicago, IL, 60611

<sup>¶</sup>Current address: AbCellera, Vancouver, BC V5Y 0A1, Canada

**Email:** aliwang@ucmerced.edu

#### **This PDF file includes:**

Supporting text  
Figures S1 to S16  
Tables S1 to S4  
Datasets S1 to S17  
SI References

## Supporting Text

### EXPERIMENTAL DETAILS

#### Measuring residue-specific $\Delta G$ values for $\text{gsKaiB}^{\text{D91R}} \rightleftharpoons \text{fsKaiB}^{\text{D91R}}$ .

$\Delta G$  values for the  $\text{gsKaiB}^{\text{D91R}} \rightleftharpoons \text{fsKaiB}^{\text{D91R}}$  reaction can be estimated from integrated HSQC peak volumes on equilibrated samples as follows:

$$\Delta G = -RT \ln K_{\text{eq}}$$

with

$$K_{\text{eq}} = p_{\text{fs,eq}}/p_{\text{gs,eq}},$$

and

$$p_{\text{gs,eq}} = V_{\text{gs,eq}}/(V_{\text{gs,eq}} + V_{\text{fs,eq}}),$$

$$p_{\text{fs,eq}} = V_{\text{fs,eq}}/(V_{\text{gs,eq}} + V_{\text{fs,eq}}),$$

where,  $p_{\text{gs,eq}}$  and  $p_{\text{fs,eq}}$  are the residue-specific fractional equilibrium populations of  $\text{gsKaiB}^{\text{D91R}}$  and  $\text{fsKaiB}^{\text{D91R}}$ , respectively, and  $V_{\text{gs,eq}}$  and  $V_{\text{fs,eq}}$  are the corresponding HSQC peak volumes of  $\text{gsKaiB}^{\text{D91R}}$  and  $\text{fsKaiB}^{\text{D91R}}$  under equilibrium conditions.

#### Determination of residue-specific $\Delta G^{\ddagger}_{\text{gs} \rightarrow \text{fs}}$ and $\Delta G^{\ddagger}_{\text{gs} \leftarrow \text{fs}}$ values from temperature-jump experiments on $\text{KaiB}^{\text{D91R}}$ (and $\text{KaiB}^{\text{G89A}}$ ).

The observed rates of  $^{15}\text{N}$ -enriched  $\text{KaiB}^{\text{D91R}}$  fold switching,  $k_{\text{obs}}$ , upon a jump in sample temperature were obtained by fitting changes in fractional populations,  $p_{\text{gs}}$  and  $p_{\text{fs}}$ , determined from volume integration of HSQC peaks to the following equations:

$$\begin{aligned} p_{\text{gs}} &= 1 - (A e^{-k_{\text{obs}} t} + B) \\ p_{\text{fs}} &= A e^{-k_{\text{obs}} t} + B \end{aligned}$$

Residue-specific forward and reverse rate constants,  $k_{\text{gs} \rightarrow \text{fs}}$  and  $k_{\text{gs} \leftarrow \text{fs}}$ , of the  $\text{gsKaiB}^{\text{D91R}} \rightleftharpoons \text{fsKaiB}^{\text{D91R}}$  reaction were determined for each temperature as follows:

$$\begin{aligned} k_{\text{gs} \rightarrow \text{fs}} &= k_{\text{obs}} p_{\text{fs,eq}} \\ k_{\text{gs} \leftarrow \text{fs}} &= k_{\text{obs}} p_{\text{gs,eq}}. \end{aligned}$$

These rate constants allowed us to determine the activation free energies,  $\Delta G^{\ddagger}_{\text{gs} \rightarrow \text{fs}}$  and  $\Delta G^{\ddagger}_{\text{gs} \leftarrow \text{fs}}$ , using transition-state theory:

$$\begin{aligned} \Delta G^{\ddagger}_{\text{gs} \rightarrow \text{fs}} &= -RT \ln(h k_{\text{gs} \rightarrow \text{fs}} / k_{\text{B}} T) \\ \Delta G^{\ddagger}_{\text{gs} \leftarrow \text{fs}} &= -RT \ln(h k_{\text{gs} \leftarrow \text{fs}} / k_{\text{B}} T), \end{aligned}$$

where  $h$  is Planck's constant.

#### The exchange behavior of $\text{KaiB}^{\text{HDX}}$ is in a mixed EX1-EX2 regime.

One method to identify the extent to which a residue is in the EX1 or EX2 regime is to examine its HDX rate as a function of pH, taking advantage of the known linear dependence of  $\log(k_{\text{chem}})$  on pH. Since  $k_{\text{obs}}$  is independent of  $k_{\text{chem}}$  under EX1 conditions,  $k_{\text{obs}}$  will not be affected by pH and hence, the plot of  $\log(k_{\text{obs}})$  versus pH will have zero slope, provided the stability of the protein does not change over the chosen pH range. In contrast, the slope should be one under EX2 conditions where  $k_{\text{obs}}$  is proportional to  $k_{\text{chem}}$ . An assumption underlying this approach is that the stability of the protein does not change over the pH

values tested; we confirmed that this is the case by thermal melting studies tracking far-UV circular dichroism for KaiB<sup>HDX</sup> (Fig. S16).

We observed that the slopes in the  $\log_{10}(k_{\text{obs}})$  versus pH plot for most sites are between 0.2 and 1. Because the slopes in the EX1 and EX2 limits are 0 and 1, respectively, the exchange behavior falls in a mixed EX1-EX2 regime and is dependent on the particular residue (Fig. S17). That said, we expect that the exchange is closer to the EX1 limit. In addition, elevated pH and urea concentrations amplify the probability of the hydrogen-deuterium exchange (HDX) data being in the EX1 regime because these two factors increase the intrinsic exchange rate and decrease the refolding rate, respectively, thereby increasing the likelihood that they satisfy the EX1 condition  $k_{\text{chem}} \gg k_{\text{cl}}$ . Specifically, the chemical exchange rate ( $k_{\text{chem}}$ ) is accelerated 10-fold at pH 6.5 versus pH 5.5, while the folding rate ( $k_{\text{cl}}$ ) is reduced at higher urea concentrations. In the EX1 limit,  $k_{\text{obs}} = k_{\text{unf}}$  (i.e.,  $k_{\text{op}}$ ), and hence  $k_{\text{obs}}$  should be independent of  $k_{\text{chem}}$ , whereas in the EX2 limit,  $k_{\text{obs}} = k_{\text{chem}}/\text{PF}$ , and thus dependent on  $k_{\text{chem}}$ . These behaviors, along with  $k_{\text{chem}}$ 's inherent increase by 10x for every pH unit, imply that the slope of the  $\log_{10}(k_{\text{obs}}/k_{\text{chem}})$  plot should be negative in the EX1 limit but zero in the EX2 limit. Since the observed slopes are predominantly negative, the data better support exchange being in the EX1 limit. Overall, the pH dependence of the rates suggests that the exchange for these sites is closer to the EX1 limit, especially with added denaturant. This in turn implies that the  $k_{\text{obs}}$  values mostly reflect the unfolding rate of a large-scale opening event.

## COMPUTATIONAL DETAILS

### Systems studied with Upside.

We used CHARMM-GUI version 3.7 (1) to prepare systems for Upside simulations. The X-ray crystal structure for tetrameric wildtype (WT) ground state KaiB was retrieved from the Protein Data Bank (PDB ID 2QKE (2)). Monomeric gsKaiB was isolated by modeling only chain B, since all its residues were resolved. The NMR structure of a fold-switched KaiB mutant was also retrieved from the Protein Data Bank (PDB ID 5JYT (3)), and CHARMM-GUI was used to mutate the sequence back to the WT sequence. Both gsKaiB and fsKaiB structures were truncated and mutated to the KaiB<sup>D91R</sup> sequence or the KaiB<sup>HDX</sup> sequence (Table S4). All simulations were performed and analyzed with a combination of Upside (4, 5) and MDTraj 1.9.9 (6). All structures were visualized in VMD (7). The source code, parameters, and usage examples for Upside can be obtained from <https://github.com/sosnicklab/upside2-md/>. Code used to analyze simulations is available at <https://github.com/dinner-group/KaiB-fold-switching>.

### Parameterization of Upside proline isomerization potential.

To allow for the explicit isomerization of the prolines during the unbiased dynamical Galerkin approximation (DGA) simulations, we employed a two-well proline isomerization potential. The minima are at  $\omega$  dihedral angle values of 180° (*trans*) and 0° (*cis*) and are separated by a 2.5–5.0  $k_B T$  barrier. This barrier is set much lower than the experimental value of 76–92 kJ/mol (8–10) to enable adequate sampling of the proline isomerization states during unbiased simulations. Despite this choice, we expect to obtain meaningful pathways because we adjust the difference in energy between the *cis* and *trans* energy minima ( $\Delta E_{\text{cis/trans}} = E_{\text{cis}} - E_{\text{trans}}$ ) to match literature values (11, 12). To set  $\Delta E_{\text{cis/trans}}$ , we ran a series of simulations of five-amino acid fragments, each with at least one proline that was allowed to isomerize. The literature values for  $\Delta E_{\text{cis/trans}}$  are dependent on the preceding amino acid (11, 12); in KaiB, the seven prolines are preceded by one of four unique amino acids (Xaa = Ala, Leu, Thr, Pro). For all small peptide simulations, a five amino acid segment was taken from PDB ID 5JYT and mutated and capped using CHARMM-GUI to the sequence H<sub>3</sub>CCO-

Ala-Xaa-Pro-Ala-Lys-NH<sub>2</sub>. For each peptide, we ran eight independent unbiased simulations of length 20 million Upside time units initialized from each isomerization state (*cis* or *trans*), where  $\Delta E_{cis/trans}$  ranged from  $-1.0$  to  $2.7 k_B T$ . The proline dihedral angle was calculated, and the proline was considered *cis* if  $\omega \in [-90^\circ, 90^\circ]$ .

#### Upside temperature calibration.

To investigate the role of temperature in the fold-switching dynamics of KaiB, we needed to set a simulation temperature, as the Upside temperature scale is uncertain to within 10–20° (13). We ran 8 sets of temperature replica-exchange (T-REMD) simulations of KaiB<sup>D91R</sup>, each with 48 replicas. We initialized 8 sets of T-REMD simulations from the gsKaiB, fsKaiB, and intermediate structures characterized by the secondary structure collective variables (**Collective variables**). Prolines were allowed to isomerize under a double-well potential with a barrier of  $2.5 k_B T$  (**Parameterization of proline isomerization potential**), to allow more rapid equilibration of the *cis/trans* states. Each simulation was 3 million Upside time units long at temperatures ranging between  $T = 0.78$  and  $T = 1.02$  (in units of  $k_B$ ), with exchanges attempted every 10 time units. Coordinates were saved every 50 time units. The first 500,000 frames of each trajectory were discarded as burn-in samples. After reweighting all 384 trajectories with the multistate Bennett acceptance ratio (MBAR) method (14), we determined the melting temperature of fsKaiB to be  $T = 0.92$  (**Fig. S18**), the point where the number of KaiB hydrogen bonds as a function of simulation temperature decreased and the radius of gyration increased dramatically.

#### Initialization of Upside simulations.

To achieve good coverage of the fold-switching pathways for the DGA simulations, we did the following: as described in **Upside temperature calibration**, we performed  $8 \times 48$  simulations using T-REMD to adequately sample the conformational space of gsKaiB, fsKaiB, and unfolded intermediates. We drew 32 structures evenly spaced by 1000 frames starting from the 10,000<sup>th</sup> frame from each T-REMD trajectory. From each of the  $8 \times 48 \times 32$  starting structures, we launched 2 independent unbiased simulations of length 200,000 time units, one initialized with P63, P70, P71, and P72 in the *cis* state, and one initialized with the prolines in the *trans* state. In these simulations, all prolines were subject to a double-well potential with a barrier of  $5 k_B T$ . We generated data sets in this way at Upside temperatures of 0.87, 0.89, 0.91, which are near but below the melting temperature (see the **Upside temperature calibration**). We saved configurations every 200 time units.

#### Collective variables.

To characterize the structures in Upside simulations, we define a number of collective variables based on differences in fractions of contacts in the gsKaiB and fsKaiB states. We compute the fraction of contacts between non-hydrogen atoms using the formula in ref. (15):

$$Q_s = (1/N) \sum_{ij} \{1 + \exp[\beta(r_{ij} - \lambda s_{ij}^0)]\}^{-1},$$

where the sum runs over pairs of non-hydrogen atoms in the reference structure  $s$  within 0.45 nm of each other ( $s = \text{gsKaiB}$  or  $\text{fsKaiB}$  as constructed as described in **Systems studied with Upside**),  $r_{ij}$  is the distance in the structure of interest,  $s_{ij}$  is the distance in reference structure  $s$ , and  $N$  is a normalization factor; only  $ij$  pairs from residues separated by three or more sequence positions were included. We set the  $\beta$  and  $\lambda$  parameters to be 50 nm<sup>-1</sup> and 1.8, respectively.  $\beta$  determines the “softness” of the switching function, determining how quickly the contact function decays from 1 to 0 after the reference distance is passed.  $\lambda$  is the tolerance for the reference distance, providing slack to allow atoms at near-native distance to still be considered contacts.

As noted in the main text, we also considered subsets of the contacts in the gsKaiB and fsKaiB states. These included fractions of contacts for the N-terminal half and the C-terminal half denoted  $Q_N$ ,  $Q_{gs/fs,C}$ . We also considered fractions of contacts for individual fold-

switching secondary structure elements, following a similar naming convention. While the definition above accurately captured the melting/repositioning of  $\alpha$ -helices, we found it too sensitive to register shifts of  $\beta$ -sheets. Thus, for  $\beta$ -strands, we modify the definition of  $Q_s$  and instead measure the number of contacts to neighboring  $\beta$ -strands, normalized by the number of contacts between those strands in the experimental structures. For  $Q_{fs,\beta3}$  we measured contacts between  $\beta3$  and  $\beta1$  in fsKaiB. For  $Q_{fs,\beta4}$  we measured contacts between  $\beta4$  and  $\beta3$  in fsKaiB. For  $Q_{gs,\beta4}$  we measured contacts between  $\beta4$  and  $\beta1$  in gsKaiB. See **Table S2** for definitions of the secondary structures and **Fig. 1** for their arrangements. Because it is possible for non-native conformations to have more contacts between specified  $\beta$ -strands than do the experimental structures,  $Q$  for  $\beta$ -strands ranges from 0 to approximately 1.3. This means that the typical ranges for  $Q_{fs,\alpha2} - Q_{gs,\beta3}$  and  $Q_{fs,\alpha3} - Q_{gs,\beta4}$  are between  $-1.3$  and  $+1$ , while the typical ranges for  $Q_{fs,\beta3} - Q_{gs,\alpha2}$  and  $Q_{fs,\beta4} - Q_{gs,\alpha3}$  are between  $-1$  and  $+1.3$ .

### DGA weights and basis set construction.

Because the unbiased trajectories used to compute kinetic statistics in DGA are not sampled from a stationary, equilibrium distribution in general, typically one must reweight the sampled data to the correct equilibrium distribution. Mathematically, given a sampling distribution  $\mu$  that we want to reweight to an equilibrium distribution  $\pi$  we have for any function  $f$

$$\langle f \rangle_\pi = \langle wf \rangle_\mu / \langle w \rangle_\mu$$

where  $w = \pi/\mu$  are the weights and bracket subscripts indicate the distributions used for the averages. Typical approaches for computing the weights (16, 17) yielded poor agreement with the equilibrium distribution sampled with T-REMD, which we expect to be reasonably accurate. Therefore, we instead chose to set the weight of each frame in our unbiased dataset to the MBAR weight of the initial frame in the trajectory, assuming that the MBAR weights are stationary over the trajectory. This is not strictly true because the T-REMD simulations are of finite duration, and we initialize isomerizing prolines with equal weights even though we expect their weights to be different at equilibrium. In practice, we found this procedure yielded potentials of mean force from DGA (**Fig. 4**) in good agreement with the T-REMD results (**Fig. S19**).

To generate the basis for computing the probability of reaching the fsKaiB state before the gsKaiB state, we first used PyEMMA (18) to compute molecular features from our existing unbiased trajectory data. We computed contacts between residue pairs more than three residues away in the set 3, 9, 11, 13, 16, 23, 33, 38, 41, 43, 45, 47, 51, 56, 59, 63, 65, 67, 70, 72, 75, 78, 81, 84, 87, 90, and 93. Contacts were counted when two residues had any non-hydrogen atom distance between them less than 0.45 nm. We then reduced the dimensionality of the binary contact maps to five dimensions using the integrated variational approach to conformational dynamics (IVAC) (19), integrating over lag times from 1 to 200 frames (200 to 40,000 time steps). We combined the five IVAC coordinates with  $\cos(\omega)$  for P63, P70, P71, and P72, to emphasize the importance of the proline isomerization state, and then clustered the resulting nine-dimensional data into 500 clusters using mini-batch  $k$ -means in scikit-learn (20). Each cluster defined a basis function that was one on the cluster and zero otherwise. The results were only mildly sensitive to the size of the basis set (the number of clusters).

To obey the homogeneous boundary conditions of the DGA operator equations, we defined our gsKaiB and fsKaiB states in terms of our previously described CVs using the potentials of mean force from T-REMD and distributions of CVs from unbiased simulations. The cutoffs for the CVs we used are listed in **Table S5**. For calculating the probability of reaching the fsKaiB state before the gsKaiB state ( $q$ ), we used a function that was one in the fsKaiB state and zero elsewhere as the DGA “guess” function (16, 21), and we used an improved algorithm that incorporated memory (22) with a lag time of 100 frames (20,000 time steps) and one memory kernel.

### Reactive flux through unfolded state.

To separate reactive trajectories (that is, trajectories which fold-switch) that pass through a completely unfolded (i.e., N-terminal half unfolded) state from all reactive trajectories, we follow the procedure from ref. (23), which treats reactive trajectories in open domains. Following their notation, we are interested in dynamics of a Markov process  $X_t$  in some domain  $S$  comprises an “open” domain  $O \subset S$  and an artificial state  $\omega$  that closes the dynamics. Here, we define the completely unfolded state  $\omega$  as containing structures where  $Q_N < 0.16$ ,  $|Q_{fs,\beta3} - Q_{gs,\alpha2}| < 0.5$ ,  $|Q_{fs,\beta4} - Q_{gs,\alpha3}| < 0.5$ , and  $|Q_{fs,\alpha3} - Q_{gs,\beta4}| < 0.5$ . The cutoffs on the secondary structure variables are to exclude states with significant C-terminal secondary structure, but our tests show that removing them does not significantly alter the calculations.

We would like to study dynamics of transitions between two disjoint states  $A \subset O$  and  $B \subset O$  that stay in  $O$  the whole time (i.e., they avoid  $\omega$ ). In our case,  $A$  represents the gsKaiB state,  $B$  represents the fsKaiB state, and  $\omega$  represents a completely unfolded state. By computing the ratio of total reactive flux from paths that avoid the unfolded state  $\omega$ ,  $k_{gs \rightarrow fs}^O$ , to the total reactive flux disregarding the unfolded state  $\omega$ ,  $k_{gs \rightarrow fs}$ , we can obtain an estimate of the fraction of reactive trajectories which do or do not involve complete unfolding as  $1 -$

$$\frac{k_{gs \rightarrow fs}^O}{k_{gs \rightarrow fs}} \text{ or } \frac{k_{gs \rightarrow fs}^O}{k_{gs \rightarrow fs}}, \text{ respectively.}$$

To compute the reactive flux, we require the probability to first reach fsKaiB over gsKaiB,  $q$ , as well as the probability of having last come from gsKaiB over fsKaiB,  $\bar{q}$  (16, 23, 24). We can estimate  $q$  and  $\bar{q}$  for the paths that avoid the unfolded state using a modified equation

$$\begin{aligned} \mathcal{S}_{(A \cup B \cup \omega)^c}^\tau q(x) &= q(x) \text{ if } x \in O \setminus (A \cup B) \\ q(x) &= 0 \text{ if } x \in A \cup \omega \\ q(x) &= 1 \text{ if } x \in B \end{aligned}$$

where  $\mathcal{S}_S^\tau f(x) = E[f(X_{\min\{\tau, T^+(0)\}}) \mid X_0 = x]$  is the “stopped” transition operator and  $T^+(0) = \min\{s > 0 \mid X_s \in S^c\}$  is the first exit time from a set  $S$  (22). One can define an equation similarly for  $\bar{q}$  (see ref. (23) for further technical details). We then compute the total reactive flux following ref. (16) with a lag time of 200 frames (40,000 time steps) and with either  $q$  or  $1 - \bar{q}$  as the choice of reaction coordinate.

### Hydrogen-deuterium exchange simulations.

HDX simulations were performed on the KaiB<sup>HDX</sup> mutant starting from the fsKaiB structure. Prolines were fixed in their initial isomerization state (i.e., *cis* for P63, P70, and P72) during the entire simulation. We ran T-REMD simulations with 48 replicas spaced linearly between  $T = 0.78$  and  $T = 1.02$ , each of length 2 million time steps and with attempts at Monte Carlo swaps every 10 frames. Following ref. (13) we compute the free energy of exchange for each residue  $i$  as

$$\Delta G_{\text{HDX},i} = RT \ln \left( \frac{\langle w \cdot \text{PS}_i \rangle}{1 - \langle w \cdot \text{PS}_i \rangle} \right)$$

where  $\text{PS}_i$  is a “protection state” value that roughly represents the competency of a given backbone hydrogen for exchange (between 0 and 1), and  $w$  is the weight assigned by MBAR. Angle brackets denote an ensemble average over all simulation frames and replicas.

Because Upside simulations of HDX effectively measure local stability, they correspond to the EX2 regime. This requires that  $k_{\text{cl}} \gg k_{\text{chem}}$  where  $k_{\text{cl}}$  and  $k_{\text{chem}}$  are the closing and intrinsic exchange rates, respectively.

### All-atom simulations.

Compared to the Upside simulations, all-atom molecular dynamics simulations are computationally much more costly, as each time step represents only 2 fs of physical time,

compared to approximately 30 ps in Upside (a 15,000-fold speed difference), even before accounting for the reduced number of degrees of freedom in the latter. On the other hand, the all-atom simulations provide a route to estimating quantitative free-energy differences. We simulated the temperature dependence of the free-energy difference between gsKaiB and fsKaiB as:

$$\begin{aligned}\Delta G_{FS-GS}(T) &= \Delta G_{FS-GS}(T_0) + \Delta\Delta G_{FS-GS}(T - T_0) \\ \Delta\Delta G_{FS-GS}(T - T_0) &= \Delta G_{FS}(T - T_0) - \Delta G_{GS}(T - T_0)\end{aligned}$$

where  $\Delta G_{FS-GS}(T)$  denotes the free energy of fold switching as a function of temperature,  $T_0$  is an arbitrary reference temperature, and  $\Delta\Delta G_{FS-GS}(T - T_0)$  represents the temperature-dependent shift in the free energy of fold switching. We focused on computing only  $\Delta\Delta G_{FS-GS}(T - T_0)$ , which is computationally much less costly because one only needs to simulate the free energy dependence on temperature for individual structures. To compute one such term,  $\Delta G_{GS}(T - T_0)$ , for example, one requires the Boltzmann factor that gives the relative probability of a structure at two temperatures, *i.e.*,

$$\frac{P(T_i)}{P(T_j)} = e^{-E\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)},$$

which can be easily computed from the potential energies along the simulation trajectories. The Boltzmann factors can be used to compute the relative free energies of all temperature points using the MBAR method.

The structures of gsKaiB and fsKaiB used in the Upside simulations were also used for the all-atom simulations. The protonation states of both gsKaiB and fsKaiB were determined by H++ (25) at pH 7. The behavior of the protein systems was described by the AMBER-FB15 force field and the solvent was described by the TIP3P-FB water model. Previous studies have shown that this combination of force field and water model yields accurate melting curves for an  $\alpha$ -helical peptide and  $\beta$  hairpin (26). After solvating the proteins with 7210 waters in a  $70 \times 70 \times 60 \text{ \AA}^3$  box, eight  $\text{Cl}^-$  ions were added to balance the charge of the system. The preparation of the molecular dynamics simulations was done by *tLeap* in Amber16 (27). Because the temperature dependence of free energy is proportional to the system size, we ensured that the protein sequence, and number of solvent molecules and ions were the same in both the gsKaiB and fsKaiB simulations.

A stepwise minimization process was performed to relax each system. First, the energy was minimized while subjecting the protein atoms to a harmonic energy restraint to their original positions. The side chains of the protein were optimized next, followed by an energy minimization without restraints. After the energy minimization process, a series of NVT simulations were carried using Langevin dynamics to heat the system to the targeted temperatures over 200 ps, followed by an NPT simulation in which the density was equilibrated at 1.0 atm using the Berendsen barostat over 200 ps. Finally, production molecular dynamics simulations were carried out in the NPT ensemble. The simulations used the particle mesh Ewald (PME) treatment of electrostatic interactions with a real-space cutoff of 10  $\text{\AA}$ , as well as a cutoff of 10  $\text{\AA}$  for van der Waals interactions. A time step of 2 fs was used in the simulations and the SHAKE algorithm (28) was used to constrain the bonds to hydrogen atoms.

### Modeling the kinetics of fsKaiB<sup>D91R</sup>-CI complex formation.

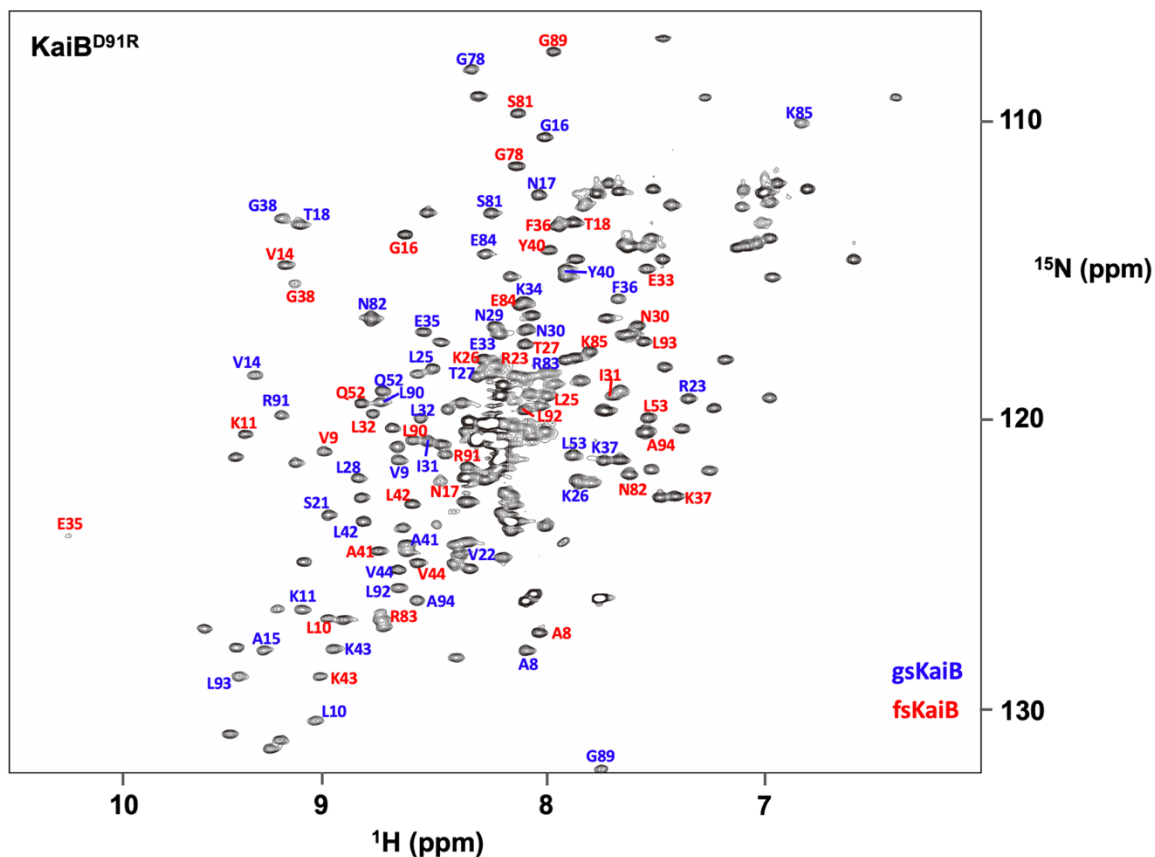
We applied the following boundary conditions:

$$[\text{gsKaiB}^{\text{D91R}}] + [\text{fsKaiB}^{\text{D91R}}] + [\text{fsKaiB}^{\text{D91R}}\text{-CI}] = 1 \text{ } \mu\text{M}$$

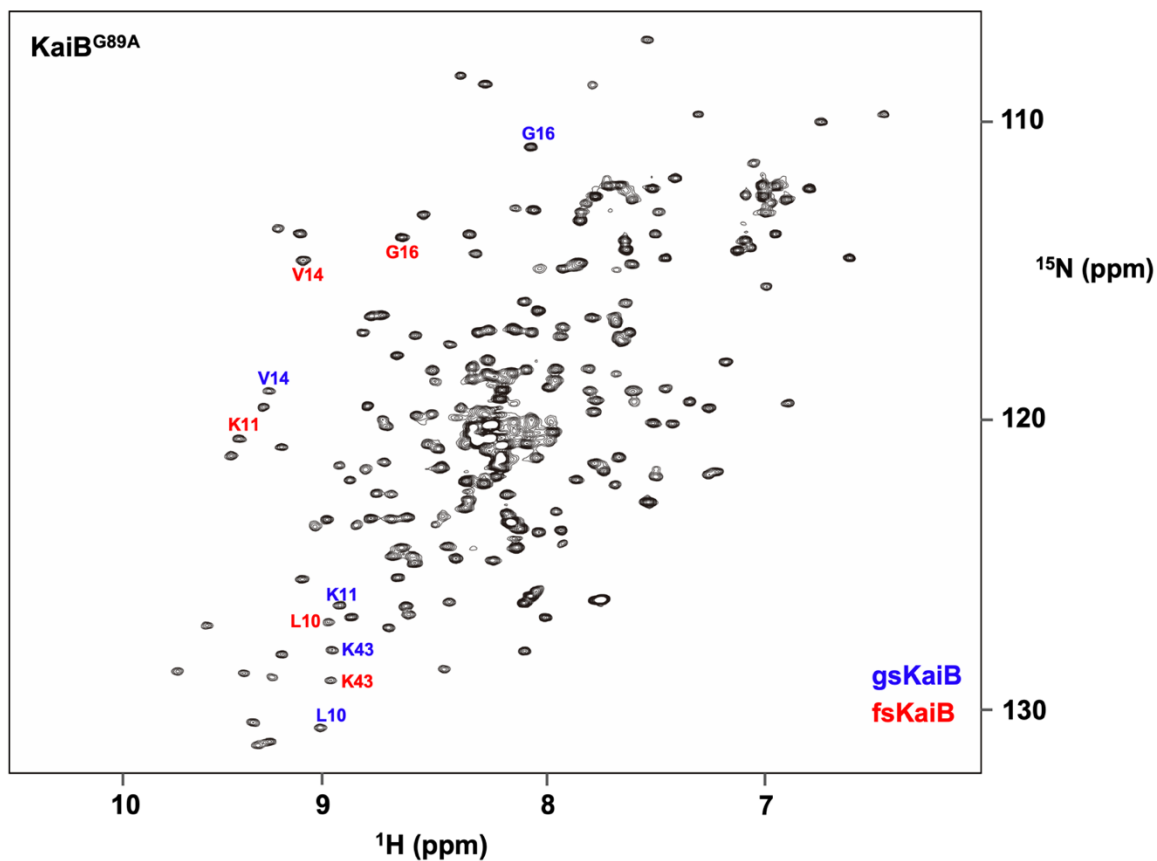
$$[\text{CI}] + [\text{fsKaiB}^{\text{D91R}}\text{-CI}] = 10 \text{ } \mu\text{M}.$$

The molar excess of CI ensures that the kinetics of formation of the fsKaiB<sup>D91R</sup>-CI complex depends primarily on the properties of KaiB<sup>D91R</sup>. We set  $k_{\text{on}}$  to  $10^4 \text{ h}^{-1} \mu\text{M}^{-1}$  so that fsKaiB<sup>D91R</sup>-CI complexation is much faster than other rates. Regarding  $k_{\text{off}}$ , we previously determined it to be  $0.03 \text{ h}^{-1}$  at 30 °C for the fsKaiB<sup>D91R</sup>-CI complex (29) and use that value here under the simplifying assumption that it is independent of temperature.

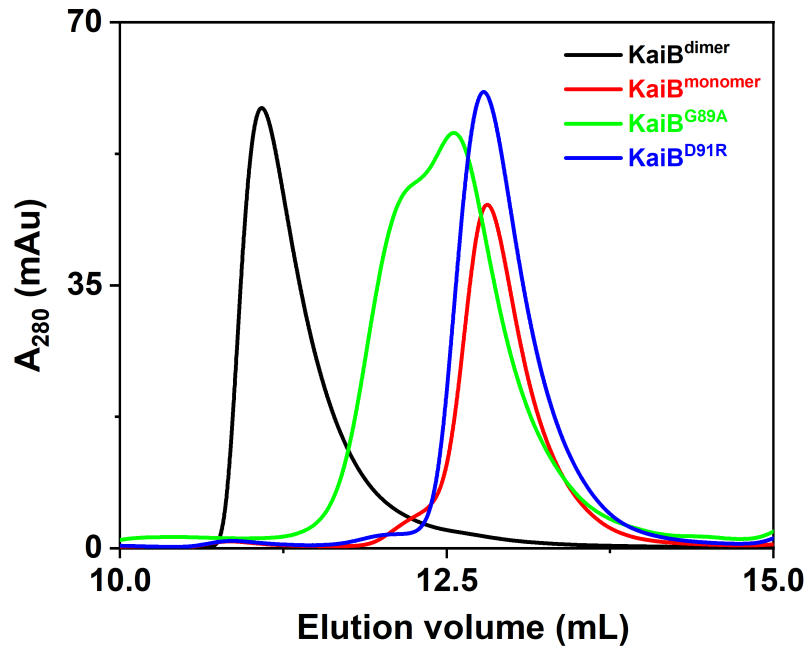




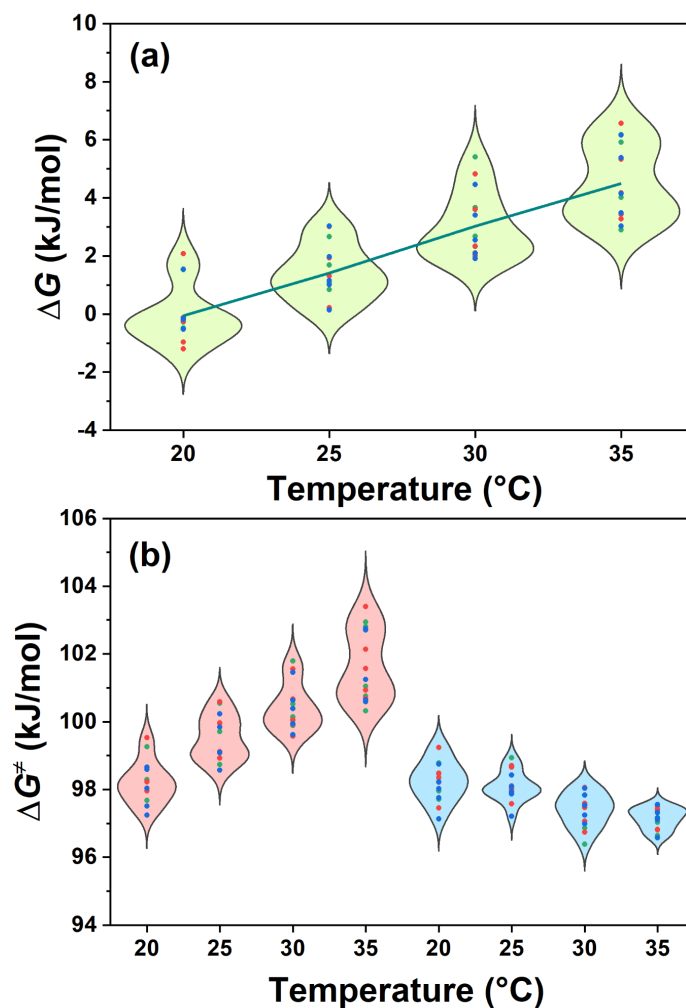
**Fig. S1. Assigned peaks in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of KaiB<sup>D91R</sup>.** The spectrum was acquired at a sample temperature of 25 °C. Assigned peaks for gsKaiB and fsKaiB are indicated with blue and red labels, respectively. Chemical-shift assignments were determined by Chang et al., 2015 (29).



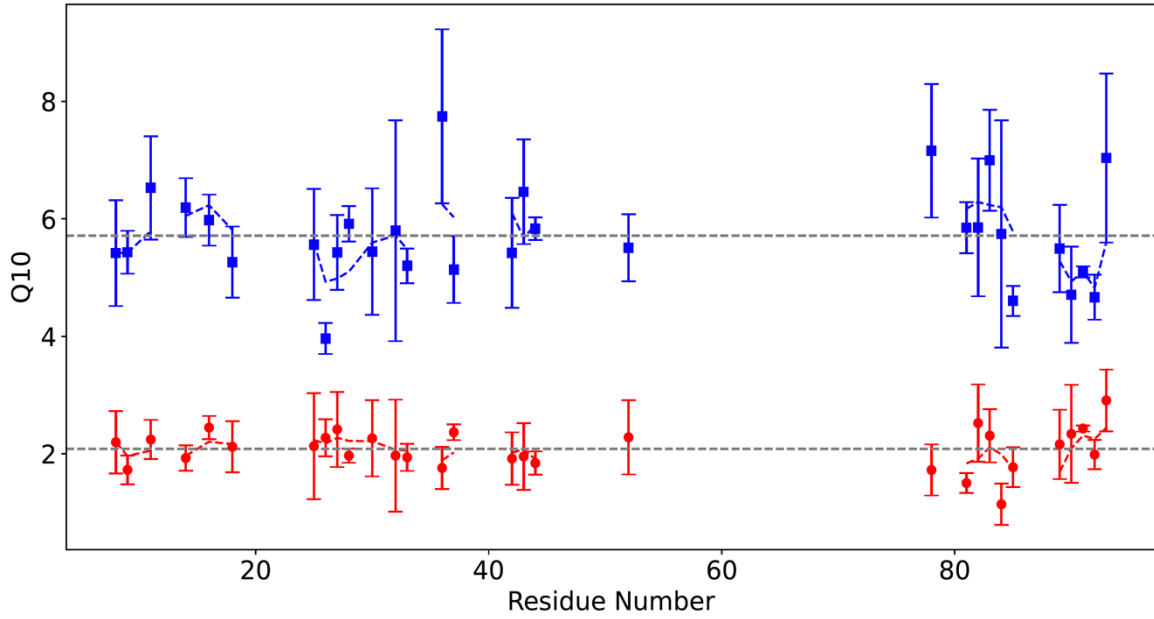
**Fig. S2. Assigned peaks in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of KaiB<sup>G89A</sup>.** The spectrum was acquired at a sample temperature of 25 °C. Assigned peaks for gsKaiB and fsKaiB are indicated with blue and red labels, respectively. Chemical-shift assignments were determined by Chang et al., 2015 (29).



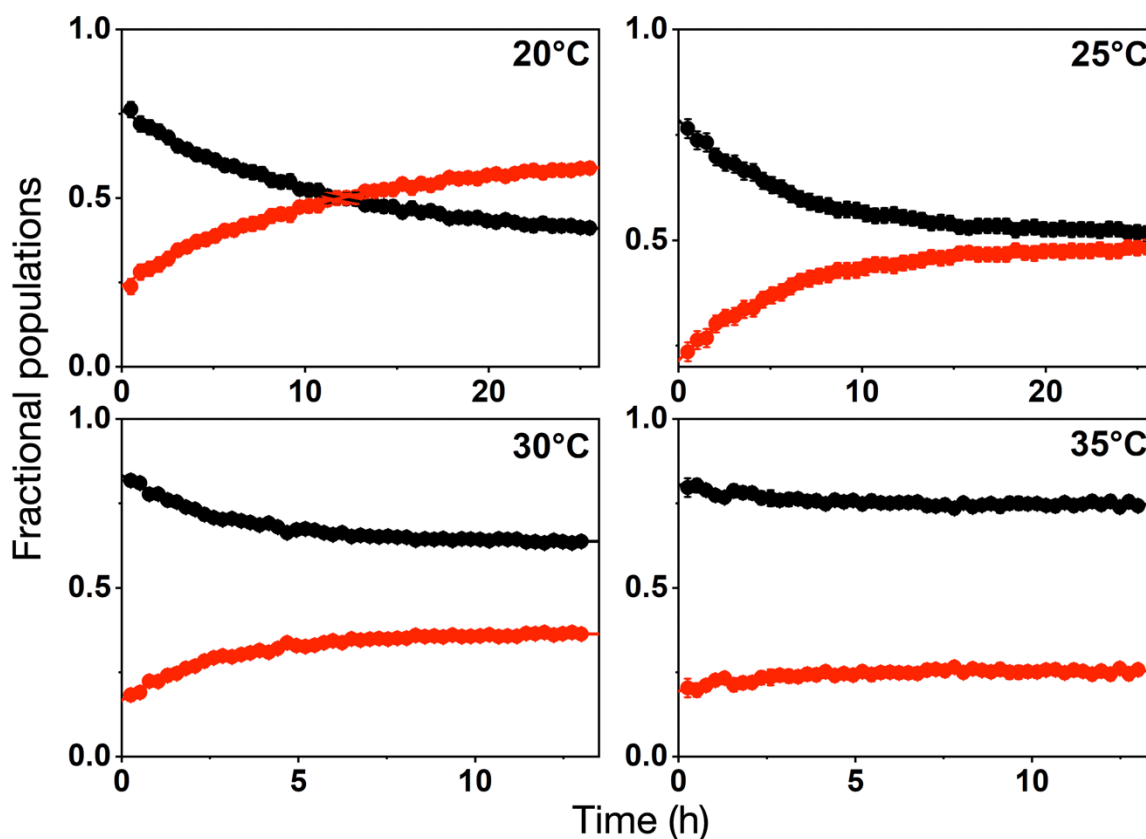
**Fig. S3. Size-exclusion chromatograms of KaiB constructs at room temperature.** KaiB<sup>dimer</sup> (KaiB 1-94 Y8A Y94A (black)), KaiB<sup>monomer</sup> (KaiB 1-94 Y8A G89A D91R Y94A (red)), KaiB<sup>G89A</sup> (KaiB 1-94 Y8A G89A Y94A (green)), and KaiB<sup>D91R</sup> (KaiB 1-94 Y8A, D91R, Y94A (blue)). KaiB<sup>monomer</sup> and KaiB<sup>dimer</sup> were previously determined to be monomeric and dimeric, respectively (29).



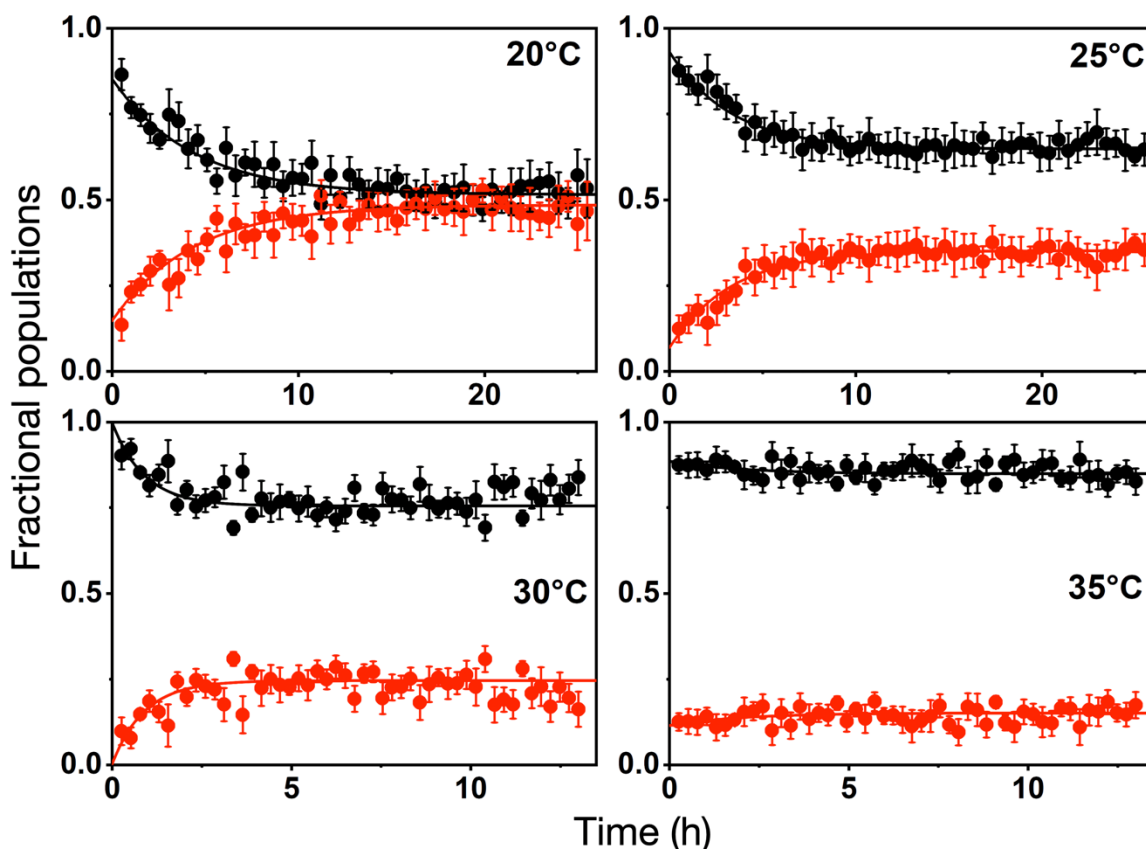
**Fig. S4. Temperature dependence of the  $gsKaiB^{G89A} \rightleftharpoons fsKaiB^{G89A}$  reaction.** (a) Residue-specific  $\Delta G$  values for  $gsKaiB \rightleftharpoons fsKaiB$  are plotted as a function of temperature. Red, green, and blue points are used to distinguish data from three replicates and are superimposed on distribution diagrams for each temperature sampled. The blue line is a fit to mean values as a function of temperature. (b)  $\Delta G^\ddagger_{gs \rightarrow fs}$  (pink) and  $\Delta G^\ddagger_{gs \leftarrow fs}$  (blue) values as a function of temperature. Red, green, and blue data points represent different replicate data sets.



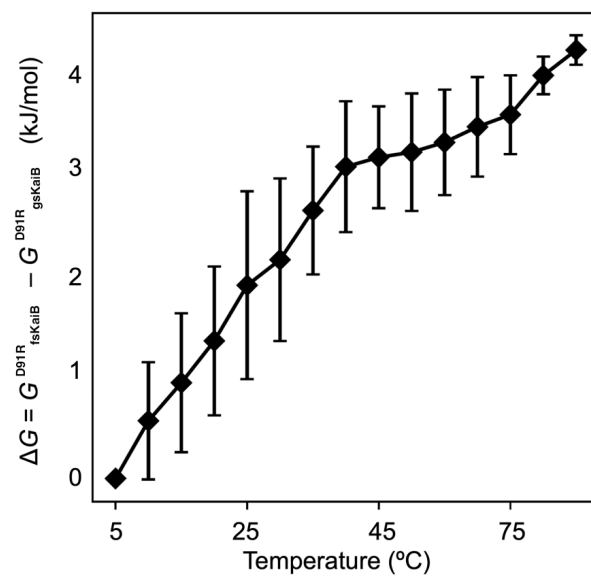
**Fig. S5. Residue-specific Q10 values derived from the kinetics of fold switching measured at 25 and 35 °C for KaiB<sup>D91R</sup>.** The red and blue points represent the means of three replicates for the gsKaiB → fsKaiB and gsKaiB ← fsKaiB directions, respectively, and the standard deviations are represented as error bars. The gray horizontal dashed lines are plotted at the global means for each direction whereas the red and blue dashed lines represent local means with a window of three residues.



**Fig. S6. The kinetics of KaiB<sup>D91R</sup> fold switching upon a high-to-low temperature jump.** A freshly prepared sample of <sup>15</sup>N-enriched KaiB<sup>D91R</sup> was incubated at 40 °C for at least 24 h before inserting it into a 14.1 T NMR spectrometer that was set at either 20 °C, 25 °C, 30 °C, or 35 °C. <sup>15</sup>N, <sup>1</sup>H HSQC spectra were collected at regular intervals after a dead time of approximately four minutes. Black and red data points represent fractional populations of *gsKaiB<sup>D91R</sup>* and *fsKaiB<sup>D91R</sup>* folds, respectively. HSQC peak volumes were determined by nmrPipe and nmrDraw. Each data point represents a mean of fractional populations across residues with resolved and assigned peaks.

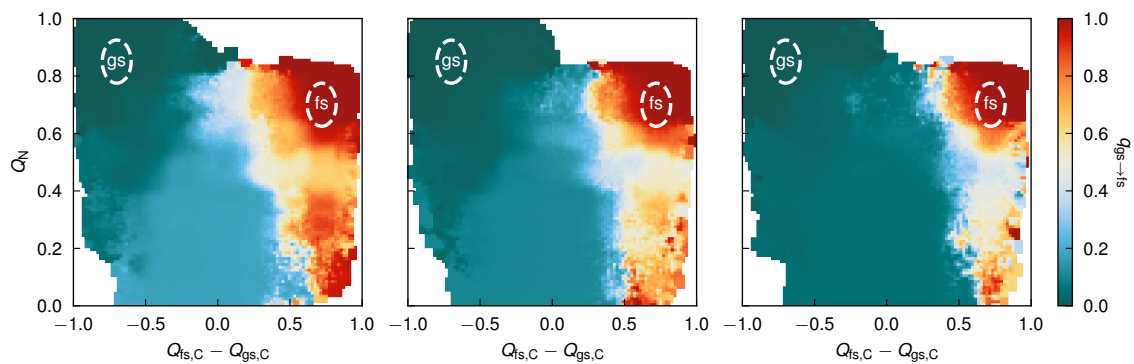


**Fig. S7. The kinetics of KaiB<sup>G89A</sup> fold switching upon a high-to-low temperature jump.** A freshly prepared sample of <sup>15</sup>N-enriched KaiB<sup>G89A</sup> was incubated at 40 °C for at least 24 h before inserting it into a 14.1 T NMR spectrometer that was set at either 20 °C, 25 °C, 30 °C, or 35 °C. <sup>15</sup>N, <sup>1</sup>H HSQC spectra were collected at regular intervals after a dead time of approximately four minutes. Black and red data points represent fractional populations of gsKaiB<sup>G89A</sup> and fsKaiB<sup>G89A</sup> folds, respectively. HSQC peak volumes were determined by nmrPipe and nmrDraw. Each data point represents a mean of fractional populations across residues with resolved and assigned peaks.

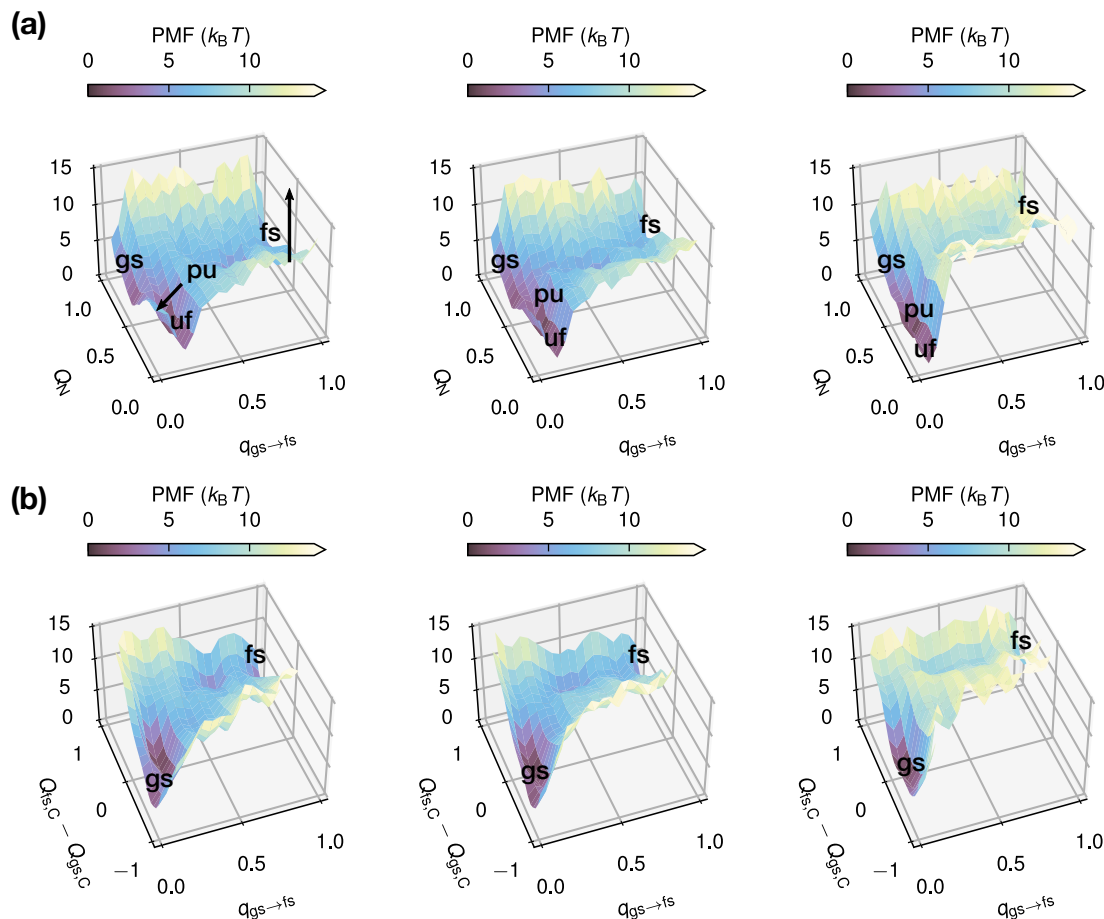


**Fig. S8. Temperature dependence of the free-energy difference,  $\Delta G = G^{\text{D91R}}_{\text{fsKaiB}} - G^{\text{D91R}}_{\text{gsKaiB}}$ , from all-atom molecular dynamics simulations.** Error bars were computed from repeating the analysis twice, once using only the first half of the trajectory and a second time using only the second half.

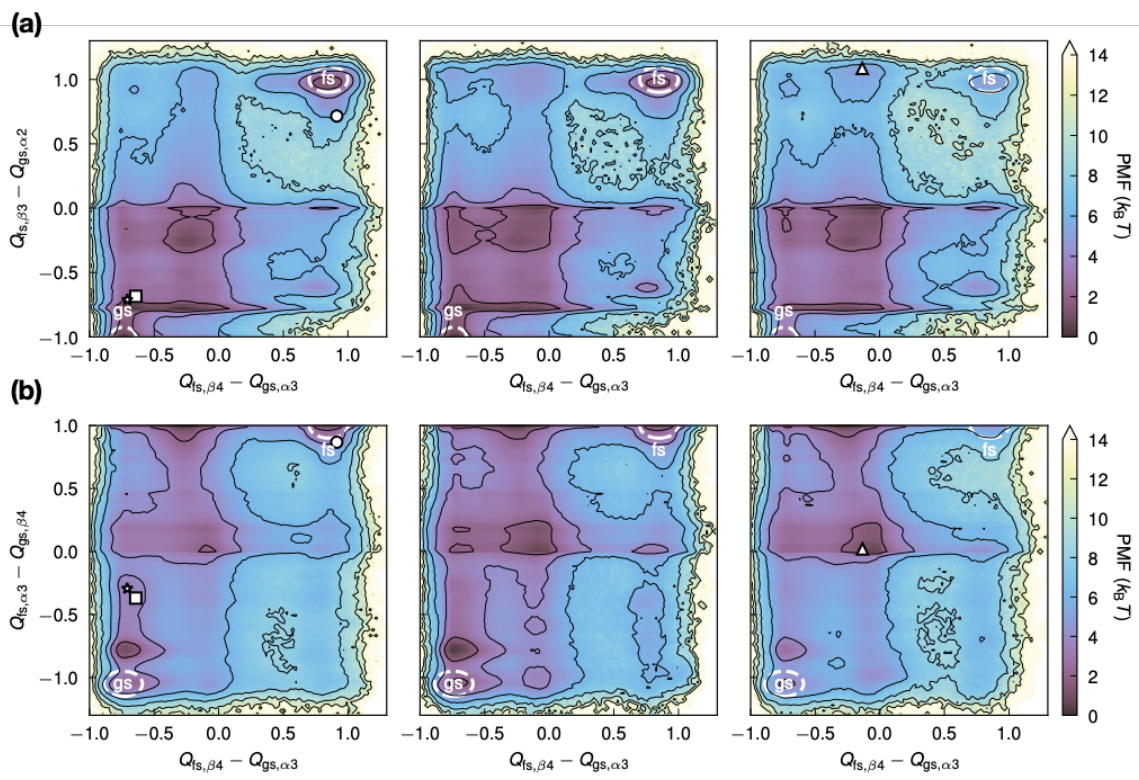




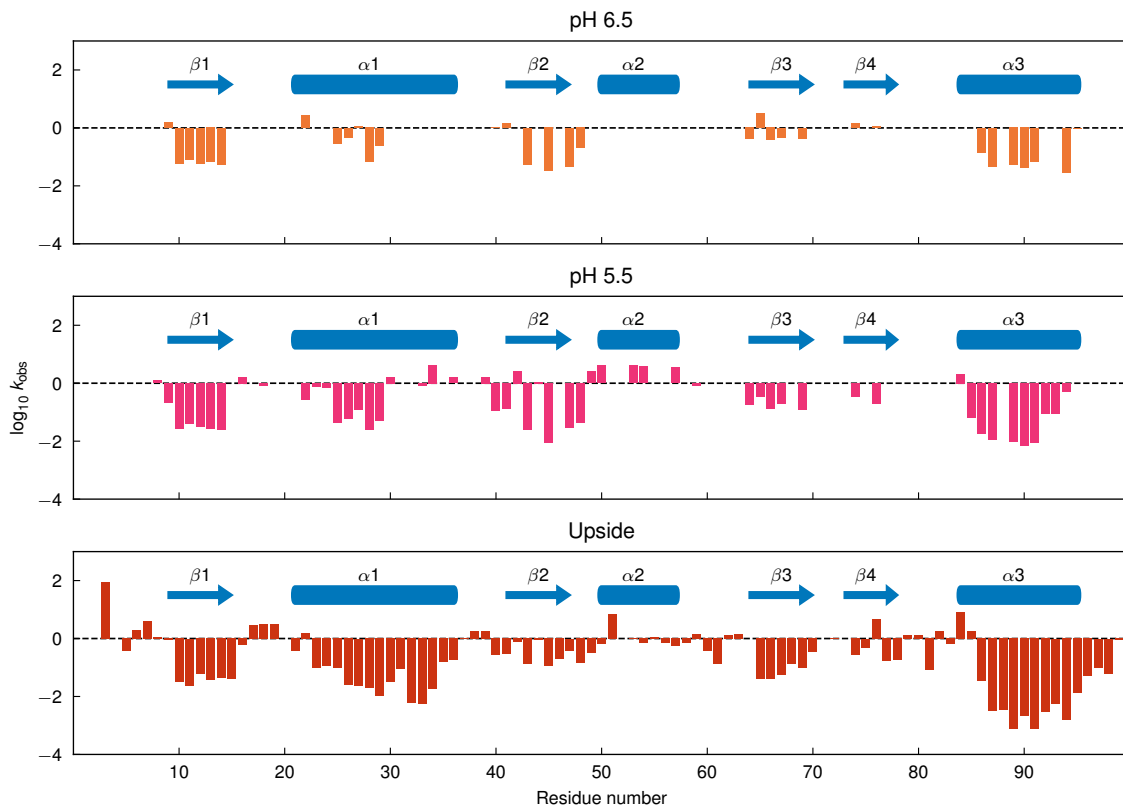
**Fig. S9. Fold switching progress for an Upside model of KaiB<sup>D91R</sup>.** We show the probability of reaching fsKaiB before gsKaiB as a function of the fraction of N-terminal contacts ( $Q_N$ ) and the difference in the fractions of gsKaiB and fsKaiB C-terminal contacts ( $Q_{fs,C} - Q_{gs,C}$ ). From left to right, the Upside temperature is increasing.



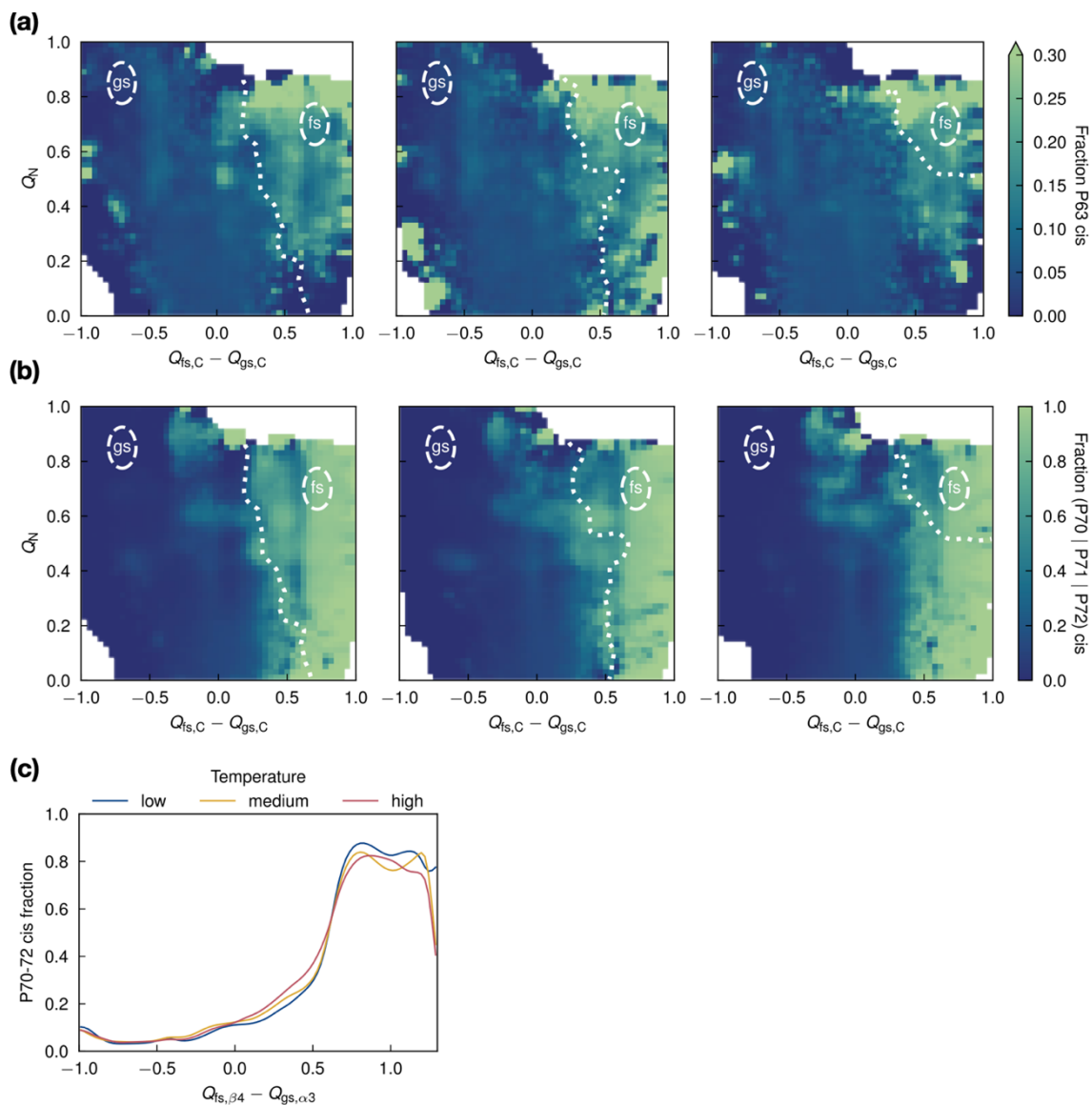
**Fig. S10. Free energy profiles of fold switching for an Upside model of KaiB<sup>D91R</sup>.** Potentials of mean force as a function of the probability of proceeding to fsKaiB over gsKaiB,  $q$ , and **(a)** the fraction of native contacts in the N-terminal half of the protein ( $Q_N$ ), and **(b)** the difference in the fractions of gsKaiB and fsKaiB C-terminal contacts between the gsKaiB and fsKaiB states ( $Q_{fs,C} - Q_{gs,C}$ ). The “pu” (partially unfolded) and “uf” (unfolded) states correspond approximately to the metastable intermediates at (0.0, 0.4) and (0.0, 0.0) in the plots in **Fig. 4**. The arrows depict the movement of the semi-folded and unfolded states towards gsKaiB and the increasing free energy of fsKaiB as temperature increases. From left to right, the Upside temperature is increasing.



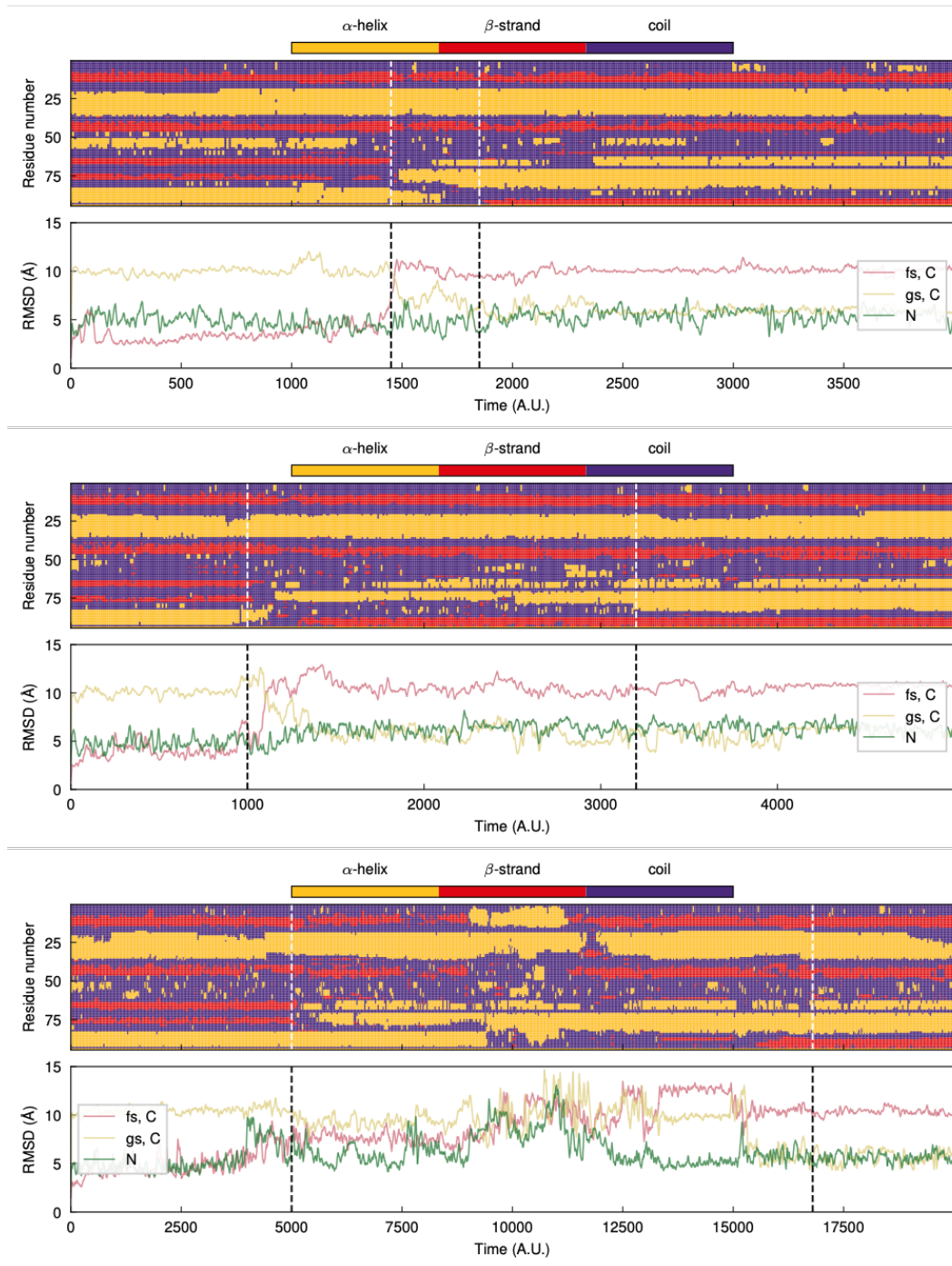
**Fig. S11. Free-energy profiles of the fold-switching secondary structure for an Upside model of KaiB<sup>D91R</sup>.** Potentials of mean force as a function of (a)  $Q_{fs,\beta4} - Q_{gs,\alpha3}$  and  $Q_{fs,\beta3} - Q_{gs,\alpha2}$  and (b)  $Q_{fs,\beta4} - Q_{gs,\alpha3}$  and  $Q_{fs,\alpha3} - Q_{gs,\beta4}$ . Approximate location of the intermediate structures shown in Fig. 4a are marked with the corresponding symbols. From left to right, the Upside temperature is increasing. Contour lines are drawn every  $2 k_B T$ .



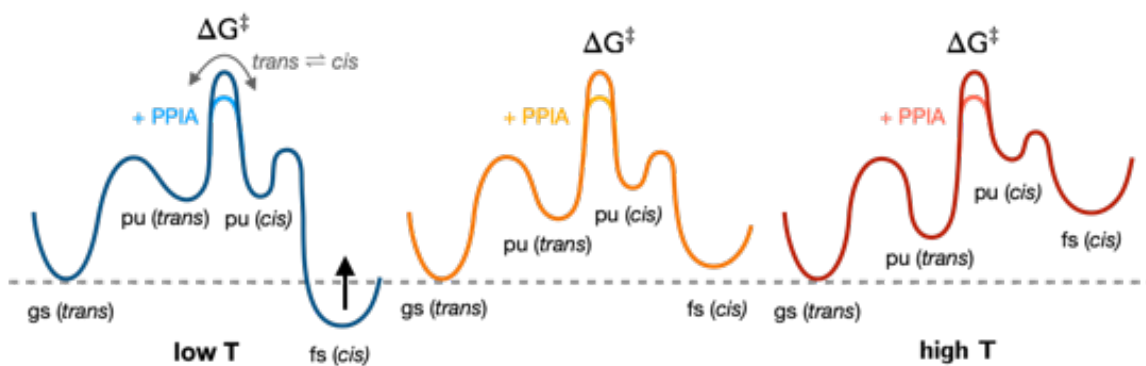
**Fig. S12. Observed and computed hydrogen-deuterium exchange rates of KaiB<sup>HDX</sup>.** From top to bottom, log<sub>10</sub> of  $k_{\text{obs}}$  rates (in h<sup>-1</sup>) for experiments at pH 6.5 and pH 5.5, and from Upside simulations. Upside simulations correspond to  $T = 0.91$  and are computed by converting stabilities (free energies) using the experimentally measured  $k_{\text{chem}}$  values (see supplementary discussion) at pH 5.5 with the relation  $k_{\text{obs}} = k_{\text{chem}}/\text{PF}$ , where  $\text{PF} = 1 + \exp(-\Delta G_{\text{HDX}}/RT)$ , assuming a temperature conversion of  $0.85 \leftrightarrow 298$  K.



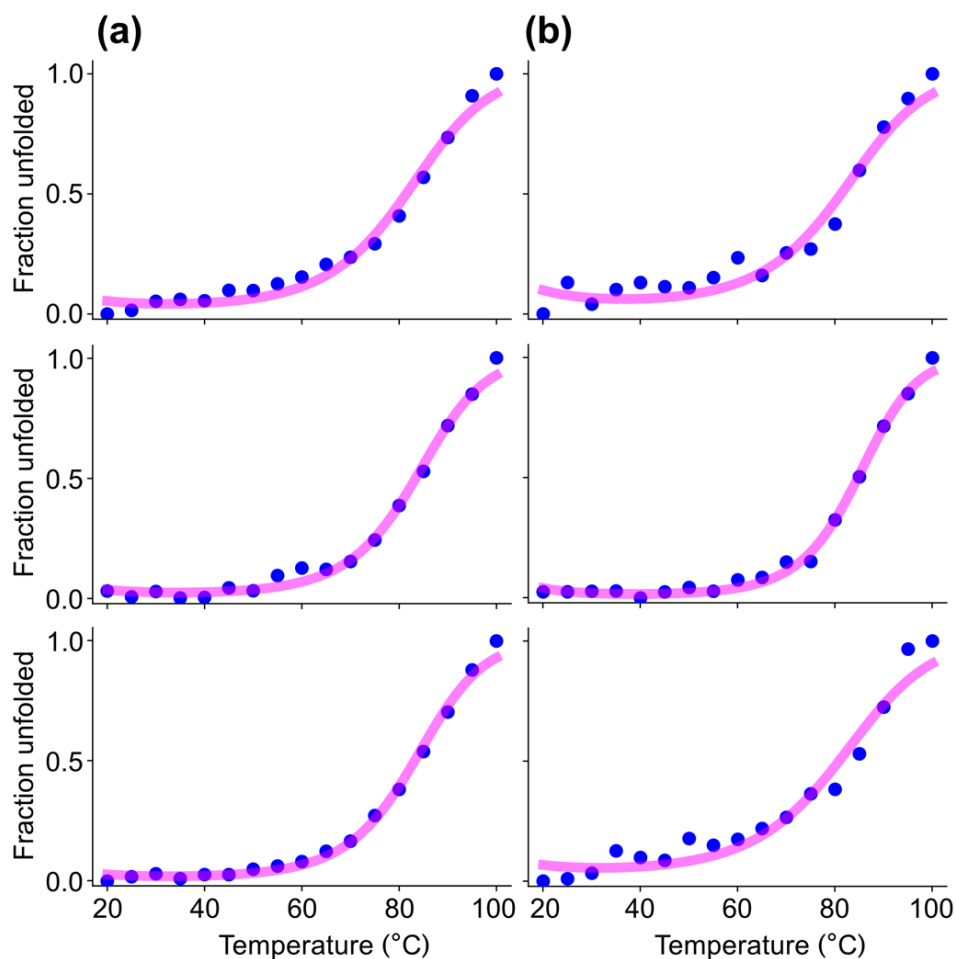
**Fig. S13. Role of isomerization of prolines in Upside simulations of KaiB<sup>D91R</sup> fold switching.** Average fraction of (a) P63 and (b) any of P70–P72 in *cis* as a function of the fraction of N-terminal contacts ( $Q_N$ ) and the difference in the fractions of gsKaiB and fsKaiB C-terminal contacts ( $Q_{fs,C} - Q_{gs,C}$ ). From left to right, the Upside temperature is increasing. The approximate location of the  $q = 0.5$  surface is marked as a dashed line. (c) Average fraction of any of P70–P72 in *cis* as a function of  $Q_{fs,\beta4} - Q_{gs,\alpha3}$ .



**Fig. S14. Proline isomerization is sufficient to induce fold-switching in an Upside model of KaiB<sup>D91R</sup>.** Upside simulations starting from fsKaiB with P63, P70, and P72 fixed in *trans*. (upper panels) Secondary structure assigned via DSSP (30) and (lower panels) RMSD to either the C-terminal half of fsKaiB, the C-terminal half of gsKaiB, or the N-terminal half (Table S2) during three successful fold-switching simulations from fsKaiB to gsKaiB other than the one shown in the main text. RMSDs are plotted as a moving average over 20 frames. The dotted lines indicate the approximate start and end times of the fold-switching event.

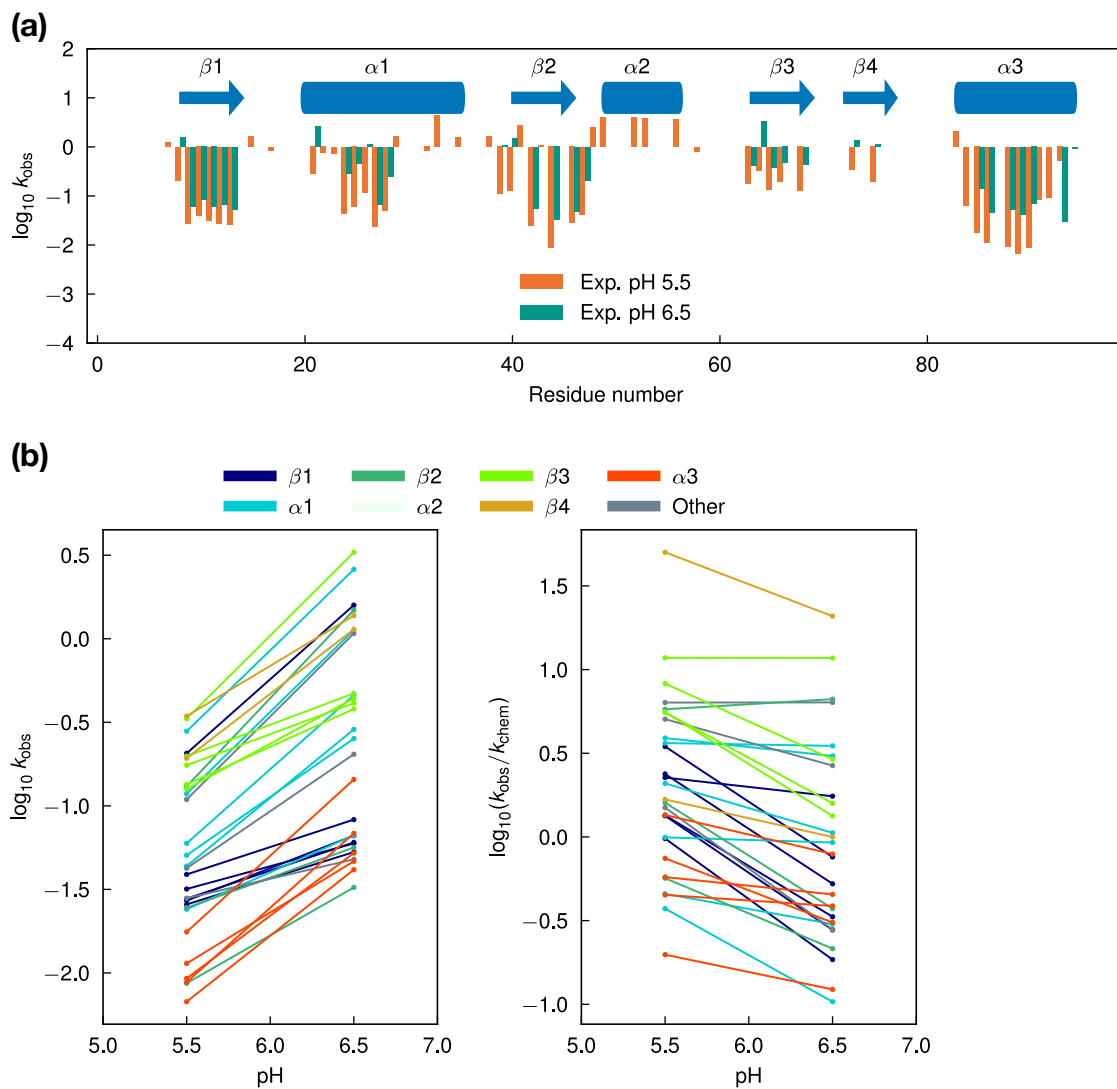


**Fig. S15. Temperature dependence of free-energy landscape for KaiB<sup>D91R</sup> fold switching.** The partially unfolded intermediates that undergo proline isomerization are labeled “pu.”

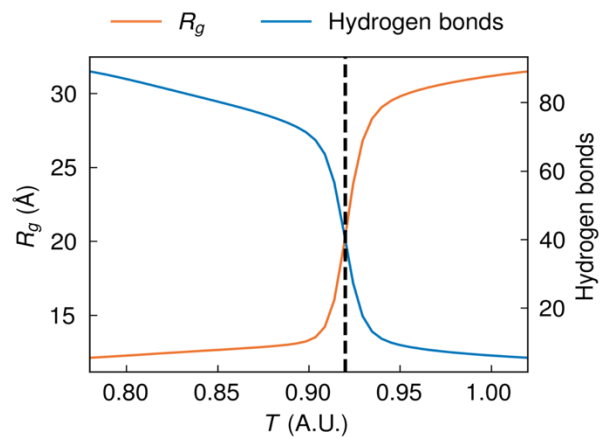


**Fig. S16. Thermal denaturation profiles for KaiB<sup>HDX</sup> at (a) pH 5.5 and (b) pH 6.5.** Three replicate experiments were carried out for both pH values on a Jasco J-1500 CD spectrometer. Fits (pink curves) to the data (blue points) yielded melting points of  $83 \pm 1$  °C and  $82 \pm 1$  °C for pH 5.5 and pH 6.5, respectively. The accuracy of these melting points is approximate as full denaturation of KaiB<sup>HDX</sup> was not achieved over this range of temperatures. The concentrations of KaiB<sup>HDX</sup> were 14  $\mu$ M and 12  $\mu$ M at pH 5.5 and pH 6.5, respectively, in 20 mM sodium phosphate and 100 mM NaCl, titrated with HCl to adjust pH values. Measurements were done at a wavelength of 220 nm at 5 °C intervals from 20 °C to 100 °C (17 data points total) with a temperature ramp of 1 °C /min. Each data point represents the average value of seven readings at 220 nm.

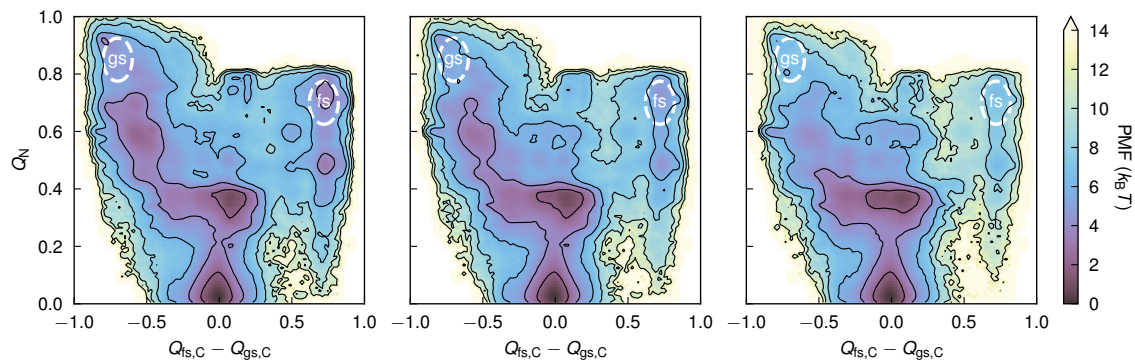




**Fig. S17. HDX rates for KaiB<sup>HDX</sup> as a function of pH.** (a)  $\log_{10}$  of  $k_{\text{obs}}$  (in  $\text{h}^{-1}$ ) for residues measured at pH 5.5 and 6.5. (b) Left,  $\log_{10} k_{\text{obs}}$ , and right,  $\log_{10}(k_{\text{obs}}/k_{\text{chem}})$  at same pH values and colored by fsKaiB secondary structure. In the plot of  $k_{\text{obs}}$  and the plot of  $k_{\text{obs}}/k_{\text{chem}}$ , a slope of 1 and 0 respectively, can be attributable to a 10-fold increase in  $k_{\text{obs}}$  with a one unit increase in pH matching a similar 10-fold increase in  $k_{\text{chem}}$ , and are the expected slopes in the EX2 limit.



**Fig. S18. Melting curve for Upside model of KaiB<sup>D91R</sup>.** The black dashed line indicates the estimated melting temperature of  $T = 0.92$ .  $R_g$  is the radius of gyration.



**Fig. S19. Free-energy profile of KaiB<sup>D91R</sup> from T-REMD Upside simulations.** Potential of mean force as a function of the fraction of N-terminal contacts ( $Q_N$ ) and the difference in the fractions of gsKaiB and fsKaiB C-terminal contacts ( $Q_{fs,C} - Q_{gs,C}$ ). From left to right, the temperature is increasing. Contour lines are drawn every  $2 k_B T$ .

**Table S1. Solvent accessible surface area (ASA) of TeKaiB computed via GETAREA (curie.utmb.edu).**

State	PDB ID	Polar ASA (Å <sup>2</sup> )	Nonpolar ASA (Å <sup>2</sup> )
gs	1VGL, chain A	2243	3355
fs	5JWO, chain B	1775	3490

**Table S2. Residue definitions, in terms of amino acid numbers, for the secondary structure elements and N/C-termini in KaiB<sup>D91R</sup>.**

Name	gs	fs
N-terminal half	1–50	1–50
C-terminal half	51–94	51–94
$\beta_{3gs} \rightleftharpoons \alpha_{2fs}$	59–61	50–57
$\alpha_{2gs} \rightleftharpoons \beta_{3fs}$	62–67	64–69
$\alpha_{3gs} \rightleftharpoons \beta_{4fs}$	72–81	73–77
$\beta_{4gs} \rightleftharpoons \alpha_{3fs}$	88–94	84–94

**Table S3. Equilibrium population fraction of the *cis* isomerization state for selected prolines in Upside simulations of KaiB<sup>D91R</sup>.**

% cis	P63	P70	P71	P72
<b>gsKaiB*</b>	0.0	0.1	0.0	5.6
<b>fsKaiB*</b>	28.7	8.2	89.9	3.8

\*The gsKaiB and fsKaiB states were defined as in **Table S4**.

**Table S4. List of KaiB variants.**

<b>KaiB<sup>G89A</sup> for NMR temperature-jump experiments and SEC</b>	
Variant	FLAG- <i>TeKaiB</i> -1-94-Y8A-Y94A-G89A-FLAG
Construct	pET-28b-SUMO-FLAG- <i>TeKaiB</i> -1-94-Y8A-Y94A-G89A-FLAG
Protein sequence	DYKDDDDKMAPLRKTAVLKLYVAGNTPNSVRALKTLNNILEKEFKGVYALKVIDVLKNPQ LAEEKILATPTLAKVLPPPVRRIIGDLSNREKVLIALDLLADYKDDDDK
<b>KaiB<sup>D91R</sup> for NMR temperature-jump experiments and SEC</b>	
Variant	FLAG- <i>TeKaiB</i> -1-94-Y8A-Y94A-D91R-FLAG
Construct	pET-28b-SUMO-FLAG- <i>TeKaiB</i> -1-94-Y8A-Y94A-D91R-FLAG
Protein sequence	DYKDDDDKMAPLRKTAVLKLYVAGNTPNSVRALKTLNNILEKEFKGVYALKVIDVLKNPQ LAEEKILATPTLAKVLPPPVRRIIGDLSNREKVLIGLRLLADYKDDDDK
<b>KaiB<sup>Dimer</sup> for SEC</b>	
Variant	FLAG- <i>TeKaiB</i> -1-94-Y8A-Y94A -FLAG
Construct	pET-28b-SUMO-FLAG- <i>TeKaiB</i> -1-94-Y8A-Y94A -FLAG
Protein sequence	DYKDDDDKMAPLRKTAVLKLYVAGNTPNSVRALKTLNNILEKEFKGVYALKVIDVLKNPQ LAEEKILATPTLAKVLPPPVRRIIGDLSNREKVLIGLDLLADYKDDDDK
<b>KaiB<sup>Monomer</sup> for SEC</b>	
Variant	FLAG- <i>TeKaiB</i> -1-94-Y8A-Y94A-G89A-D91R-FLAG
Construct	pET-28b-SUMO-FLAG- <i>TeKaiB</i> -1-94-Y8A-Y94A-G89A-D91R-FLAG
Protein sequence	DYKDDDDKMAPLRKTAVLKLYVAGNTPNSVRALKTLNNILEKEFKGVYALKVIDVLKNPQ LAEEKILATPTLAKVLPPPVRRIIGDLSNREKVLIALRLLADYKDDDDK

<b>KaiB<sup>HDX</sup> for Hydrogen-deuterium exchange (HDX) experiments</b>	
Variant	<i>TeKaiB</i> -1-99-Y8A-Y94A-P71A-G89A-D91R
Construct	pET-28b-SUMO- <i>TeKaiB</i> -1-99-Y8A-Y94A-P71A-G89A-D91R
Protein sequence	MAPLRKTAVLKLYVAGNTPNSVRALKTLNNILEKEFKGVYALKVIDVLKNPQLAEEDKILA TPTLAKVLPPPVRRIIGDLSNREKVLIGLRLLA

**Table S5. State definitions of gsKaiB and fsKaiB for DGA calculations.**

<b>CV</b>	<b>gs cutoff</b>	<b>fs cutoff</b>
$Q_{fs,\alpha 2} - Q_{gs,\beta 3}$	$< -0.50$	—
$Q_{fs,\beta 3} - Q_{gs,\alpha 2}$	$< -0.75$	$> 0.85$
$Q_{fs,\beta 4} - Q_{gs,\alpha 3}$	$< -0.55$	$> 0.65$
$Q_{fs,\alpha 3} - Q_{gs,\beta 4}$	$< -0.90$	$> 0.80$
$Q_N$	$> 0.70$	$> 0.65$
$Q_s$ ( $s = \text{gs or fs}$ )	$> 0.62$	$> 0.61$
$Q_{s,C}$ ( $s = \text{gs or fs}$ )	$> 0.65$	$> 0.75$
RMSD to $s$ (nm)	$< 0.45$	$< 0.35$



## Supporting Datasets

### Datasets S1-S8

- Fractional populations of gsKaiB and fsKaiB determined from  $^{15}\text{N}$ ,  $^1\text{H}$  HSQC spectra as a function of time after jumping the sample from 4 °C to 20, 25, 30, or 35 °C. The fractional populations are means of values from residues with resolved and assigned peaks in the gsKaiB and fsKaiB states. Each dataset contains three replicates on separate sheets.
- Column A: time (h).
- Column B: fractional population of fsKaiB.
- Column C: fractional population of gsKaiB.
- Dataset S1: Temperature jump from 4 → 20 °C for KaiB<sup>D91R</sup>.
- Dataset S2: Temperature jump from 4 → 20 °C for KaiB<sup>G89A</sup>.
- Dataset S3: Temperature jump from 4 → 25 °C for KaiB<sup>D91R</sup>.
- Dataset S4: Temperature jump from 4 → 25 °C for KaiB<sup>G89A</sup>.
- Dataset S5: Temperature jump from 4 → 30 °C for KaiB<sup>D91R</sup>.
- Dataset S6: Temperature jump from 4 → 30 °C for KaiB<sup>G89A</sup>.
- Dataset S7: Temperature jump from 4 → 35 °C for KaiB<sup>D91R</sup>.
- Dataset S8: Temperature jump from 4 → 35 °C for KaiB<sup>G89A</sup>.

### Datasets S9 and S10

- Fractional populations of gsKaiB and fsKaiB determined from  $^{15}\text{N}$ ,  $^1\text{H}$  HSQC spectra as a function of time after jumping the sample from 40 °C to 20, 25, 30, or 35 °C. The fractional populations are means of values from residues with resolved and assigned peaks in the gsKaiB and fsKaiB states. Within each dataset individual sheets contain data for a separate temperature jump.
- Column A: time (h).
- Column B: fractional population of gsKaiB.
- Column C: fractional population of fsKaiB.
- Dataset S9: Temperature jumps for KaiB<sup>D91R</sup>.
- Dataset S10: Temperature jumps for KaiB<sup>G89A</sup>.

### Dataset S11

- First replicate of KaiB<sup>D91R</sup> + 0, 5, 10, 25, 50, or 100  $\mu\text{M}$  PPIA.
- Column A: time (h).
- Columns B – AO: residue-specific fractional populations of fsKaiB or gsKaiB.
- Fractional populations of fsKaiB and gsKaiB at each concentration of PPIA are on separate sheets.

### Dataset S12

- Second replicate of KaiB<sup>D91R</sup> + 0, 5, 10, 25, 50, or 100  $\mu\text{M}$  PPIA.
- Column A: time (h).
- Columns B – AO: residue-specific fractional populations of fsKaiB or gsKaiB.
- Fractional populations of fsKaiB and gsKaiB at each concentration of PPIA are on separate sheets.

**Dataset S13**

- KaiB<sup>D91R</sup> + 0  $\mu$ M PPIA + 55  $\mu$ M cyclosporinA (1st and second replicates).
- KaiB<sup>D91R</sup> + 50  $\mu$ M PPIA + 55  $\mu$ M cyclosporinA (1st and second replicates).
- Column A: time (h).
- Columns B – A0: residue-specific fractional populations of fsKaiB or gsKaiB.
- Fractional populations of fsKaiB and gsKaiB at each concentration of PPIA are on separate sheets.

**Datasets S14 – S17**

- Residue-specific rates of hydrogen-deuterium exchange determined by NMR at 12 °C and separate urea concentrations.
- Column A: Residue name.
- Column B: Observed rate of hydrogen-deuterium exchange in h<sup>-1</sup>.
- Dataset S14: pH 5.50, first replicate.
- Dataset S15: pH 6.52, first replicate.
- Dataset S16: pH 5.50, second replicate.
- Dataset S17: pH 6.52, second replicate.
- Each spreadsheet has separate sheets for the different urea concentrations.

## SI References

1. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
2. R. Pattanayek, *et al.*, Structural model of the circadian clock KaiB–KaiC complex and mechanism for modulation of KaiC phosphorylation. *EMBO J* **27**, 1767–1778 (2008).
3. R. Tseng, *et al.*, Structural basis of the day-night transition in a bacterial circadian clock. *Science* **355**, 1174–1180 (2017).
4. J. M. Jumper, N. F. Faruk, K. F. Freed, T. R. Sosnick, Trajectory-based training enables protein simulations with accurate folding and Boltzmann ensembles in cpu-hours. *PLOS Comput. Biol.* **14**, e1006578 (2018).
5. J. M. Jumper, N. F. Faruk, K. F. Freed, T. R. Sosnick, Accurate calculation of side chain packing and free energy with applications to protein molecular dynamics. *PLOS Comput. Biol.* **14**, e1006342 (2018).
6. R. T. McGibbon, *et al.*, MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528–1532 (2015).
7. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28 (1996).
8. C. L. Gustafson, *et al.*, A Slow Conformational Switch in the BMAL1 Transactivation Domain Modulates Circadian Rhythms. *Mol. Cell* **66**, 447–457.e7 (2017).
9. G. Schoetz, O. Trapp, V. Schurig, Determination of the cis-trans isomerization barrier of several L-peptidyl-L-proline dipeptides by dynamic capillary electrophoresis and computer simulation. *Electrophoresis* **22**, 2409–2415 (2001).
10. M. Shibukawa, A. Miyake, S. Eda, S. Saito, Determination of the cis–trans Isomerization Barriers of L-Alanyl-L-proline in Aqueous Solutions and at Water/Hydrophobic Interfaces by On-Line Temperature-Jump Relaxation HPLC and Dynamic On-Column Reaction HPLC. *Anal. Chem.* **87**, 9280–9287 (2015).
11. G. Fischer, Chemical aspects of peptide bond isomerisation. *Chem. Soc. Rev.* **29**, 119–127 (2000).
12. U. Reimer, *et al.*, Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *J. Mol. Biol.* **279**, 449–460 (1998).
13. X. Peng, *et al.*, Prediction and Validation of a Protein’s Free Energy Surface Using Hydrogen Exchange and (Importantly) Its Denaturant Dependence. *J. Chem. Theory Comput.* **18**, 550–561 (2022).
14. M. R. Shirts, J. D. Chodera, Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, 124105 (2008).
15. R. B. Best, G. Hummer, W. A. Eaton, Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci.* **110**, 17874–17879 (2013).

16. J. Strahan, *et al.*, Long-Time-Scale Predictions from Short-Trajectory Data: A Benchmark Analysis of the Trp-Cage Miniprotein. *J. Chem. Theory Comput.* **17**, 2948–2963 (2021).
17. H. Wu, *et al.*, Variational Koopman models: Slow collective variables and molecular kinetics from short off-equilibrium simulations. *J. Chem. Phys.* **146**, 154104 (2017).
18. M. K. Scherer, *et al.*, PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).
19. C. Lorpaiboon, E. H. Thiede, R. J. Webber, J. Weare, A. R. Dinner, Integrated Variational Approach to Conformational Dynamics: A Robust Strategy for Identifying Eigenfunctions of Dynamical Operators. *J. Phys. Chem. B* **124**, 9354–9364 (2020).
20. F. Pedregosa, *et al.*, Scikit-learn: Machine Learning in Python. *Mach. Learn. PYTHON*.
21. E. H. Thiede, D. Giannakis, A. R. Dinner, J. Weare, Galerkin approximation of dynamical quantities using trajectory data. *J. Chem. Phys.* **150**, 244111 (2019).
22. C. Lorpaiboon, S. C. Guo, J. Strahan, J. Weare, A. R. Dinner, Accurate estimates of dynamical statistics using memory. *J. Chem. Phys.* **160**, 084108 (2024).
23. P. Miron, F. J. Beron-Vera, L. Helfmann, P. Koltai, Transition paths of marine debris and the stability of the garbage patches. *Chaos Interdiscip. J. Nonlinear Sci.* **31**, 033101 (2021).
24. W. E., E. Vanden-Eijnden, Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. *Annu. Rev. Phys. Chem.* **61**, 391–420 (2010).
25. J. C. Gordon, *et al.*, H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33**, W368–W371 (2005).
26. L.-P. Wang, *et al.*, Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **121**, 4023–4039 (2017).
27. D. A. Case, *et al.*, AmberTools. *J. Chem. Inf. Model.* **63**, 6183–6191 (2023).
28. H. C. Andersen, Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **52**, 24–34 (1983).
29. Y.-G. Chang, *et al.*, A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science* **349**, 324–328 (2015).
30. W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
31. R. Fraczkiwicz, W. Braun, Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules. *J. Comput. Chem.* **19**.