

Supplementary Material

Zippering and Unzipping of Adenylate Kinase: Atomistic Insights into the Ensemble of Open ↔ Closed Transitions

Oliver Beckstein^{*,a,c,1}, Elizabeth J. Denning^{b,1}, Juan R. Perilla^b, Thomas B. Woolf^{*,c,b}

^aDepartment of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

^bDepartment of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

^cDepartment of Physiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Abstract

In the Supplementary Material we give more details on the definition of the geometry of AKeco and provide a discussion about the computation of FRET distance probability distributions from PMFs. We also provide more details on the DIMS method. We note the parameters used for DIMS and explain how we arrived at a robust parameter set, which should be of use for new users of the method. We also discuss the use of the Onsager-Machlup score to assess trajectory diversity. DIMS-MD simulations with protonated (charge-neutral) acidic residues in the salt bridge zipper are presented. The zipper mechanism is shown to be of electrostatic origin. Additional figures to the main text are presented: histograms of the duration of simulated DIMS transitions, RMSD for matches of X-ray structures versus DIMS transitions, the angle-projected potential energy error analysis of the PMF, an analysis of the conservation of charged residues in almost 1000 AdK sequences, plots for all salt bridges discussed in the main paper, and an analysis of the secondary structure stability in the hinge regions that shows intermittent unwinding of some of the hinges during transitions.

Geometry of AdK

The LID and NMP domain rotate on a number of hinge residues relative to the CORE domain¹. Here we define the NMP domain to consist of residues 30–59 in the AKeco sequence and the LID of residues 122–159; these numbers correspond to residues 30–59 and 120–157 in the AK consensus sequence² and are similar to previous definitions^{2,3}. A simplified description of this movement is afforded by the tuple of the two angles θ_{NMP} and θ_{LID} . The NMP-CORE angle θ_{NMP} is formed by the centres of geometry of the backbone and C_{β} atoms in residues 115–125 (CORE/LID), 90–100 (CORE), and 35–55 (NMP). θ_{LID} is defined equivalently as the angle between 179–185 (CORE), 115–125 (CORE/hinge/LID), and 125–153 (LID).

In general, it is very difficult to find reaction coordinates to construct a low dimensional free energy surface. Although Kubitzki and de Groot⁴ point out that the LID-opening motion in a AdK transition observed in their simulations can not be described by a simple hinge bending motion domain angles still appear to be reasonable coordinates as they seem to capture at least a large proportion of the conformational change and have the additional advantage that they can be intuitively interpreted.

*Correspondence: OB (phone: +44 1865 613304; fax: +44 1865 613238) or TBW (phone: +1 (410) 614-2643; fax: +1 (410) 614-4436)

Email addresses: oliver.beckstein@bioch.ox.ac.uk (Oliver Beckstein), edennin3@jhu.edu (Elizabeth J. Denning), jrperilla@jhu.edu (Juan R. Perilla), twoolf@jhmi.edu (Thomas B. Woolf)

¹These authors contributed equally to this work.

Calculating FRET distance distributions from the PMF?

Formally, the probability distribution $p(d)$ of FRET distances d is related to the PMF \mathcal{W} via integration of the PMF-derived probability over d -contours in angle space,

$$p(d) = \frac{\int_0^{2\pi} d\theta_{\text{NMP}} \int_0^{2\pi} d\theta_{\text{LID}} \delta[\hat{d}(\theta_{\text{NMP}}, \theta_{\text{LID}}) - d] e^{-\mathcal{W}(\theta_{\text{NMP}}, \theta_{\text{LID}})/kT}}{\int_0^{2\pi} d\theta_{\text{NMP}} \int_0^{2\pi} d\theta_{\text{LID}} e^{-\mathcal{W}(\theta_{\text{NMP}}, \theta_{\text{LID}})/kT}}. \quad (\text{S1})$$

Calculation of $p(d)$ requires \mathcal{W} to be computed for the whole space spanned by $0 \leq \theta < 2\pi$ or the PMF must be bounded in the computed region, i.e. having barriers of many kT confining the system. In other words, if the PMF does not include all likely states that are compatible with a distance d then the estimates for $p(d)$ will have a possibly large systematic error. Our PMF already covers a much larger region of configuration space than is typically computed but it still hints at additional accessible states, especially along θ_{LID} . Hence FRET distributions computed from this PMF would not be consistent or comparable to experiment. We should point out that PMFs derived from a single reaction coordinate potentially face the same problem unless one can be assured that sampling is complete perpendicular to that reaction coordinate. Multi-dimensional PMFs that have only been sampled near one transition path face this problem even more so because in that case sampling has, by definition, only extended to the vicinity of the path. At least for AdK we show that the free energy landscape is rather open and so the enzyme can easily explore regions far from obvious transition paths, making any approach based on Equation S1 difficult.

DIMS Method

The dynamic importance sampling calculation in this work employ the soft-ratcheting algorithm¹⁰ in conjunction with the root mean square deviation progress variable ρ , the RMSD of the protein structure from the target conformation. For each new MD step attempt the change in progress is computed as $\Delta\rho = \rho(t + \Delta t) - \rho(t)$. Note that $\Delta\rho < 0$ indicates a step *towards* the target. One can define $\phi := -\rho$ so that the progress variable actually increases with progress towards the system but this interpretational transformation is irrelevant for the algorithm as described.

Soft-ratcheting accepts a MD step with probability

$$p(\Delta\phi) = \begin{cases} 1 & \text{if } \phi > 0 \\ \exp\left(-\left|\frac{\Delta\phi}{\Delta\phi_0}\right|^2\right) & \text{if } \phi \leq 0 \end{cases}; \quad (\text{S2})$$

the probability solely depends on the change in progress $\Delta\phi = \phi(t + \Delta t) - \phi(t)$ of the proposed step. The ‘softness’ parameter $\Delta\phi_0$ can be used to tune the system’s ability to explore alternative pathways that may initially lead away from B but may help to escape local traps in the free energy landscape. The procedure effectively implements a Maxwell’s daemon which only allows those rare fluctuations to occur that move a system ‘uphill’ across a barrier.

Determining the optimal softness parameter $\Delta\phi_0$ was tested on the closed to open transition. Figure S10 shows how different $\Delta\phi_0$ values affect the transition, in particular the acceptance ratio, which we define as the number of accepted DIMS steps divided by the total number of steps (accepted and rejected) for blocks of a ‘significant’ number of DIMS steps. ‘Significant’ means that the progress variable has not changed more than a given resolution threshold (here we used 0.1 Å) or a large number of steps (such as 50,000). Figure S10A shows that a scale parameter of 10^{-4} Å causes the acceptance ratio to remain high. These parameter values are too weak because they allow the transitions to diverge with RMSD values away from the target (i.e. open) structure. For $\Delta\phi_0$ values of 5×10^{-5} Å and 1×10^{-5} Å the RMSD distance toward the target structure does not decrease below approximately 2.7 Å and 2 Å, i.e. the bias is not quite strong enough to complete the transition. An optimal acceptance ratio was achieved using a scale parameter bias value of 1×10^{-6} Å. The trajectories converged between 0.5 Å and 1 Å and still maintained diversity between different transitions, as judged by the Onsager-Machlup score (see below). Scaling parameter values smaller than 1×10^{-6} were explored (data not shown) but not used because while the transitions completed the trajectory diversity decreased. Figure S10B shows that attempts to move towards the target structure can vary considerably in length. Typically, once the *first* step in an initial attempt moves towards the target, the transition rapidly progresses and completes. This analysis hints at the considerable diversity in system behaviour based on the initial random assignment of velocities alone.

A single soft-ratcheting transition typically completes within 3-5 h of CPU time on a modest single Intel Xeon processor and takes between 85 ps and 135 ps simulated time with the

exact distribution of times depending somewhat on the direction of the transition and if the endpoints were fixed or chosen from an ensemble of structures (Figure S11). On occasions initial attempts to move towards the target structure are unsuccessful and moves are rejected for millions of attempts even though eventually the transition begins and rapidly completes (Figure S10). To improve the trajectory yield we terminate any simulation in which more than 10,000 subsequent moves are rejected. Move rejections are computationally inexpensive because variation is provided by generating new random velocities from a Gaussian distribution around $T = 300$ K. These are used with the current forces to integrate another trial MD step. Because forces are not recomputed, much larger rejection cut-offs can be used in principle although not deemed necessary in the present case as the set of parameters described here yielded a 75% overall completion rate of transitions.

Onsager-Machlup score

To assess trajectory diversity, the cumulative Onsager-Machlup (OM) score was computed for each trajectory using the OMSCORE CORREL facility in CHARMM. The step-score is the Onsager-Machlup action for the given time step, $s(t)$,

$$s(t) := \sum_{i=1}^{N_{\text{atom}}} \left(\frac{\mathbf{x}_i(t) - \mathbf{x}_i(t - \Delta t)}{\Delta t} - \frac{\mathbf{F}_i}{m_i \eta} \right)^2 \quad (\text{S3})$$

The cumulative OM score is

$$S(t) := \sum_{t'=0}^t \Delta t s(t') \quad (\text{S4})$$

and hence the normalized cumulative score is

$$\hat{S}(t) := \frac{S(t)}{S(0)} \quad (\text{S5})$$

The OM score of a trajectory of length t_{traj} is the cumulative OM score of the last frame, $S_{\text{OM}} = S(t_{\text{traj}})$. The lower the OM score, the more likely is it that the transition would have occurred without the bias. It is not an absolute probability but it can be used to rank-order trajectories relative to each other. The results in Figure S12 indicate that we generate a relatively diverse family of transitions.

Electrostatic nature of the salt bridge zipper

At pH 7 all charged residues are in their default charged state as calculated by PROPKA¹¹. In order to investigate the sensitivity of the observed salt bridge zipper to the strength of the ionic interaction, 80 DIMS simulations were performed in which the the four acidic zipper residues D33, D54, D158, and E170 were modelled in their protonated (low pH) and hence neutral state. Although the range of predicted pK_a value of 2.5 to 4.5 of these residues makes it unlikely that exactly this charge configuration would be observable in experiment, these simulations can still serve as an extreme case in which the interaction strength has been reduced to the minimum in a physiological plausible manner.

The salt bridge existence probability along the $\Delta\rho$ progress measure in Figure S26 shows that contacts that existed in the closed state broke much sooner along the transition than in the fully ionized state (Figure 7 in the main paper). While the four salt bridges existed for at least half the transition in more than 50% of the trajectories at pH 7, this is only the case for D33–R156 in the protonated simulations. For the three other salt bridges no interactions persist and they move apart in sync with the opening of the cleft between NMP and LID domain. K57–E170 appears to be most affected and shows virtually no ionic bond in the protonated state.

These simulations demonstrate that electrostatic interactions are the driving force for the zipper mechanism and suggest that salt bridges in the zipper could be differentially affected by changes in pH or ionic strength of the solution. Variations in the force field and/or solvation model, which would probably result in somewhat different interaction strengths, are likely to produce behaviour between the two extremes studied here.

References

- Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, et al. (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* 450: 913–916. doi:10.1038/nature06407.
- Vonrhein C, Schlauderer GJ, Schulz GE (1995) Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. *Structure* 3: 483–490.
- Lu Q, Wang J (2008) Single molecule conformational dynamics of adenylate kinase: Energy landscape, structural correlations, and transition state ensembles. *J Am Chem Soc* 130: 4772–4783. doi:10.1021/ja0780481.
- Kubitzki MB, de Groot BL (2008) The atomistic mechanism of conformational transition in adenylate kinase: a TEE-REX molecular dynamics study. *Structure* 16: 1175–1182. doi:10.1016/j.str.2008.04.013.
- Wolf TB (1998) Path corrected functionals of stochastic trajectories: Towards relative free energy and reaction coordinate calculations. *Chem Phys Lett* 289: 433–441.
- Zuckerman DM, Woolf TB (1999) Dynamic reaction paths and rates through importance-sampled stochastic dynamics. *J Chem Phys* 111: 9475–9484.
- Zuckerman DM, Woolf TB (2001) Efficient dynamic importance sampling of rare events in one dimension. *Phys Rev E* 63: 016702. doi:10.1103/PhysRevE.63.016702.
- Jang H, Woolf TB (2006) Multiple pathways in conformational transitions of the alanine dipeptide: An application of dynamic importance sampling. *J Comp Chem* 27: 1136–1141.
- Arora K, Brooks CL III (2007) Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc Natl Acad Sci USA* 104: 18496–18501.
- Zuckerman DM, Woolf TB (2006) Rapid determination of multiple reaction pathways in molecular systems: The soft-ratcheting algorithm. *ArXiv:physics/0209098*.
- Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Struct Funct Genet* 61: 704–21. doi:10.1002/prot.20660.
- Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22: 2577–2637. doi:10.1002/bip.360221211.
- Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, et al. (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature* 450: 838–844. doi:10.1038/nature06410.
- Hanson JA, Duderstadt K, Watkins LP, Bhattacharyya S, Brokaw J, et al. (2007) Illuminating the mechanistic roles of enzyme conformational dynamics. *Proc Natl Acad Sci USA* 104: 18055–18060. doi:10.1073/pnas.0708600104.
- Sinev MA, Sineva EV, Ittah V, Haas E (1996) Domain closure in adenylate kinase. *Biochemistry* 35: 6425–6437. doi:10.1021/bi952687j.
- Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 18: 6097–6100.

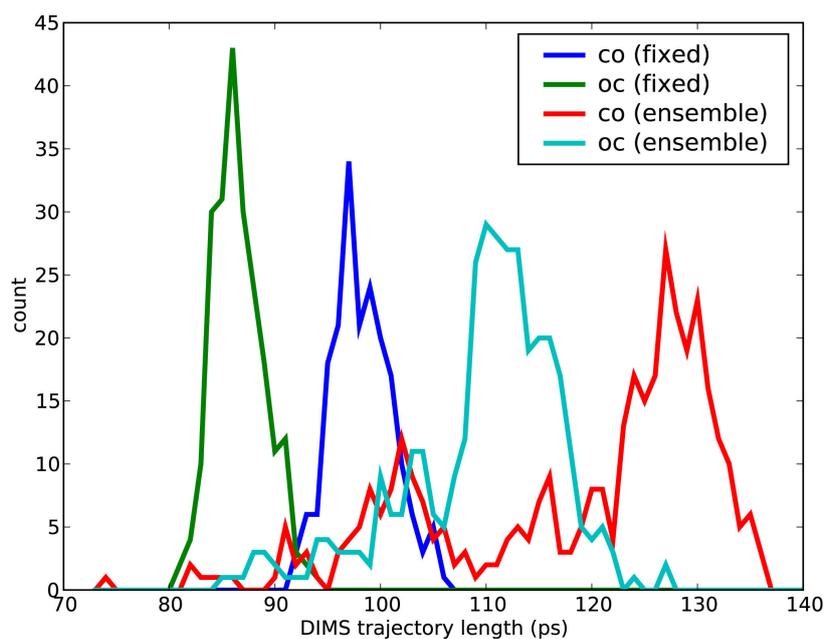


Figure S11: Transition lengths. The total time length for the DIMS transitions are plotted as a histogram of actual path times and colored according to direction and endpoint sampling. The label *oc* denotes trajectories starting from the open state and moving towards the closed state whereas *co* is a transition in the opposite direction. *Fixed* transitions used identical endpoint structures between all transitions whereas endpoints in *ensemble* transitions were randomly drawn from equilibrium MD simulations at the endpoints.

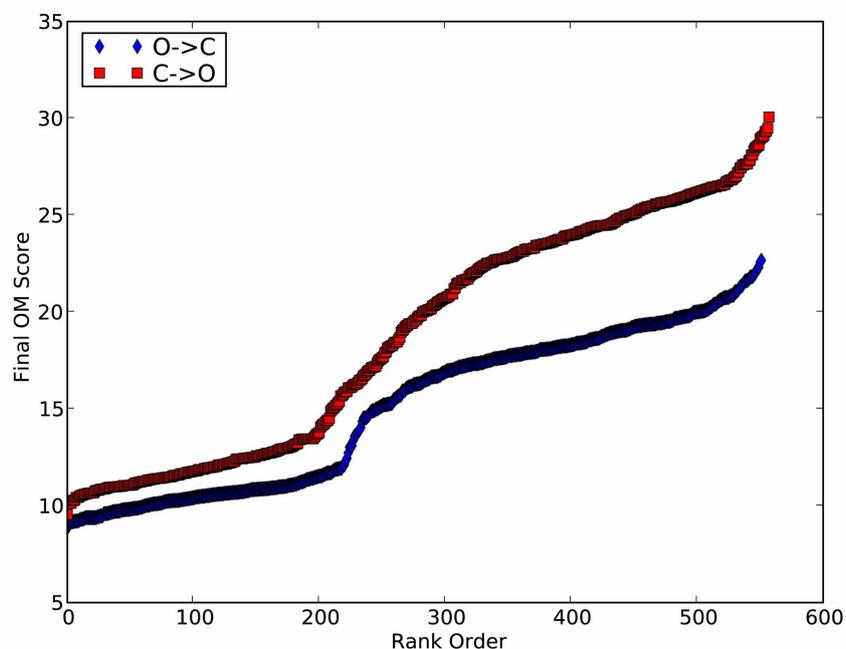


Figure S12: Onsager-Machlup (OM) scores for DIMS trajectories. Blue diamonds represent the open→closed transition and the red square the closed→open transition. Trajectories are rank ordered by OM score. This shows that not all trajectories are equally to likely occur and DIMS also samples some paths off from the most likely ones.

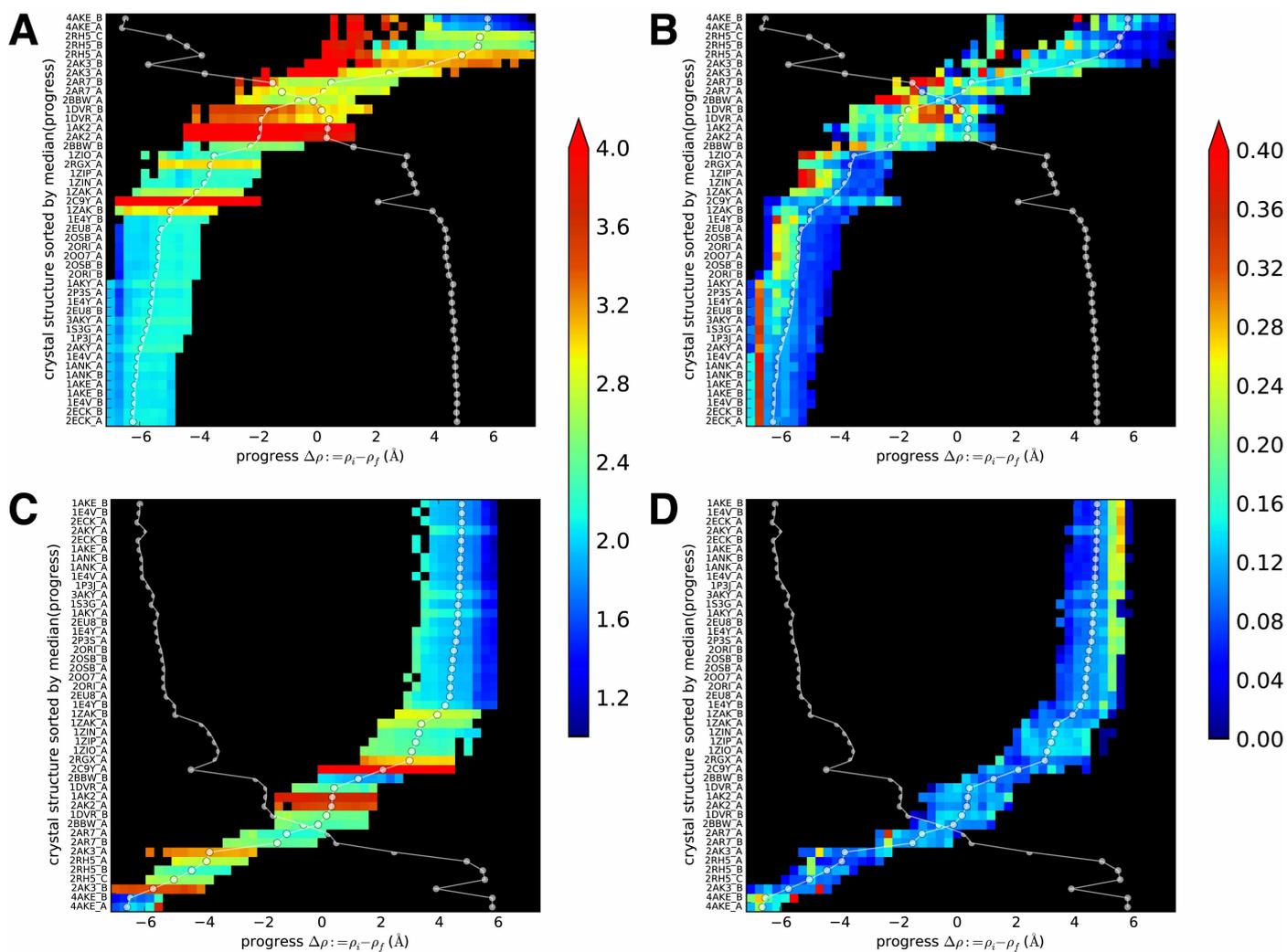


Figure S13: The RMSD and standard deviation of matches between X-ray structures and DIMS trajectory frames along the transition. Data from 330 DIMS transitions between endpoint ensembles. A: Average RMSD (closed \rightarrow open). B: Standard deviation (closed \rightarrow open). C: Average RMSD (open \rightarrow closed) D: Standard deviation (open \rightarrow closed). All distances in Å.

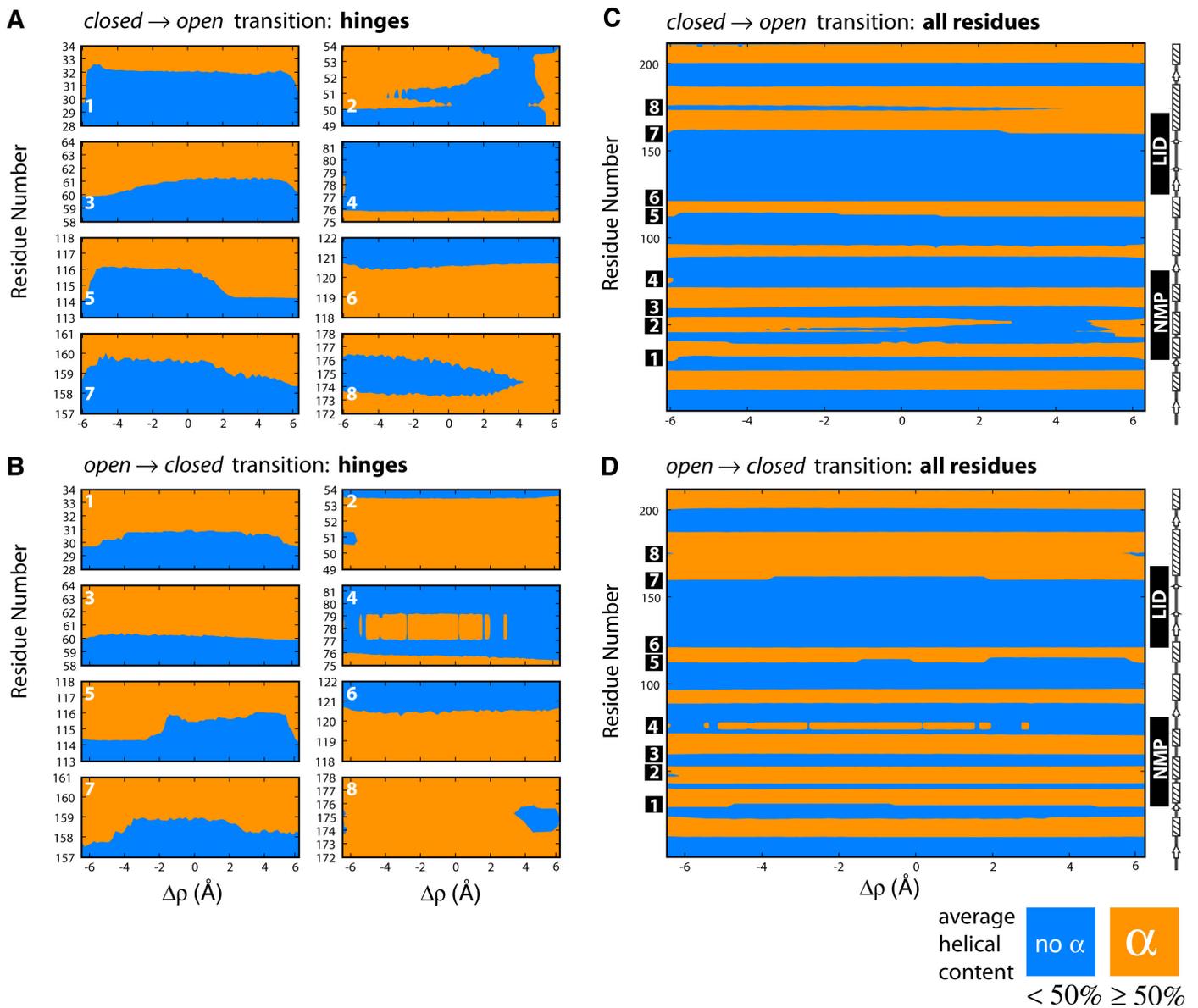


Figure S14: Loss of helical structure during DIMS transitions. The secondary structure was determined with DSSP¹² from DIMS transitions along the progress measure $\Delta\rho = \rho^A - \rho^B$ where A is the initial state and B the target. Loss of helical structure is quantified by assigning 1 to a frame in which DSSP reports any helix and 0 for any non-helix conformation; thus the helix indicator function I_h is averaged in bins along $\Delta\rho$. A residue is colored orange if $\langle I_h \rangle > 0.5$ (“helical”) and blue otherwise (“non-helical”). A, B: Details of the hinge regions 1–8 identified by Henzler-Wildman et al.¹. C, D: All residues in AdK. A, C: closed \rightarrow open transitions. B, D: open \rightarrow closed. Transient loss of helical structure occurs for hinges 1, 2, 3, 5, and 7. The position of the NMP (residues 30–59) and LID domain (122–159) are indicated by black bars. Black boxes with white numbers indicate the eight hinge regions. The secondary structure of AdK is shown for all residues on the right hand side, with α -helices denoted by rectangles and β -strands as arrows.

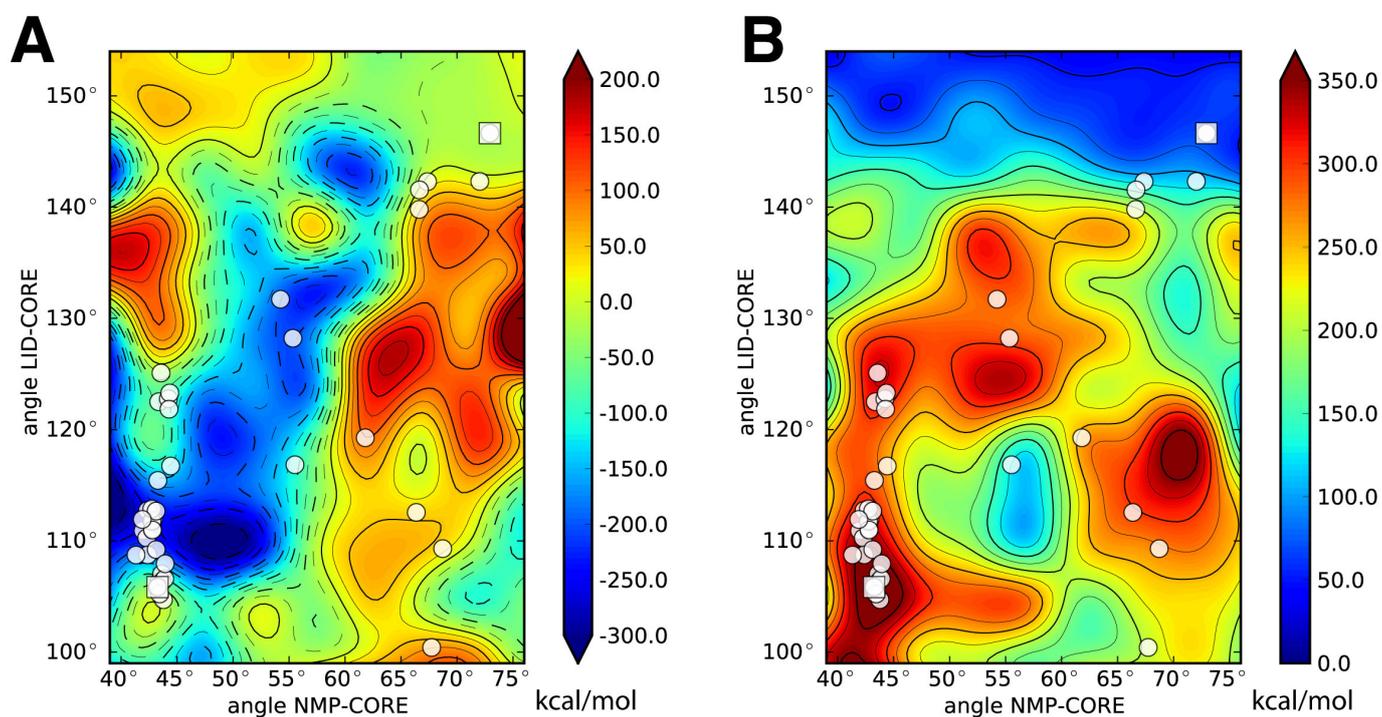


Figure S15: Angle-projected potential energy landscape of AdK. A: The potential energy in the CHARMM22 force field was recalculated from umbrella sampling simulations without the harmonic umbrella restraints. The closed state (1AKE, square in lower left corner) is taken as the reference state at $\Delta E = 0$. B: The standard deviation in each bin is large, on the order of 200 kcal/mol, which precludes calculating a meaningful entropy map from the PMF and the energy via $\mathcal{W} = \Delta E - T \Delta S$. All energies are in kcal/mol.

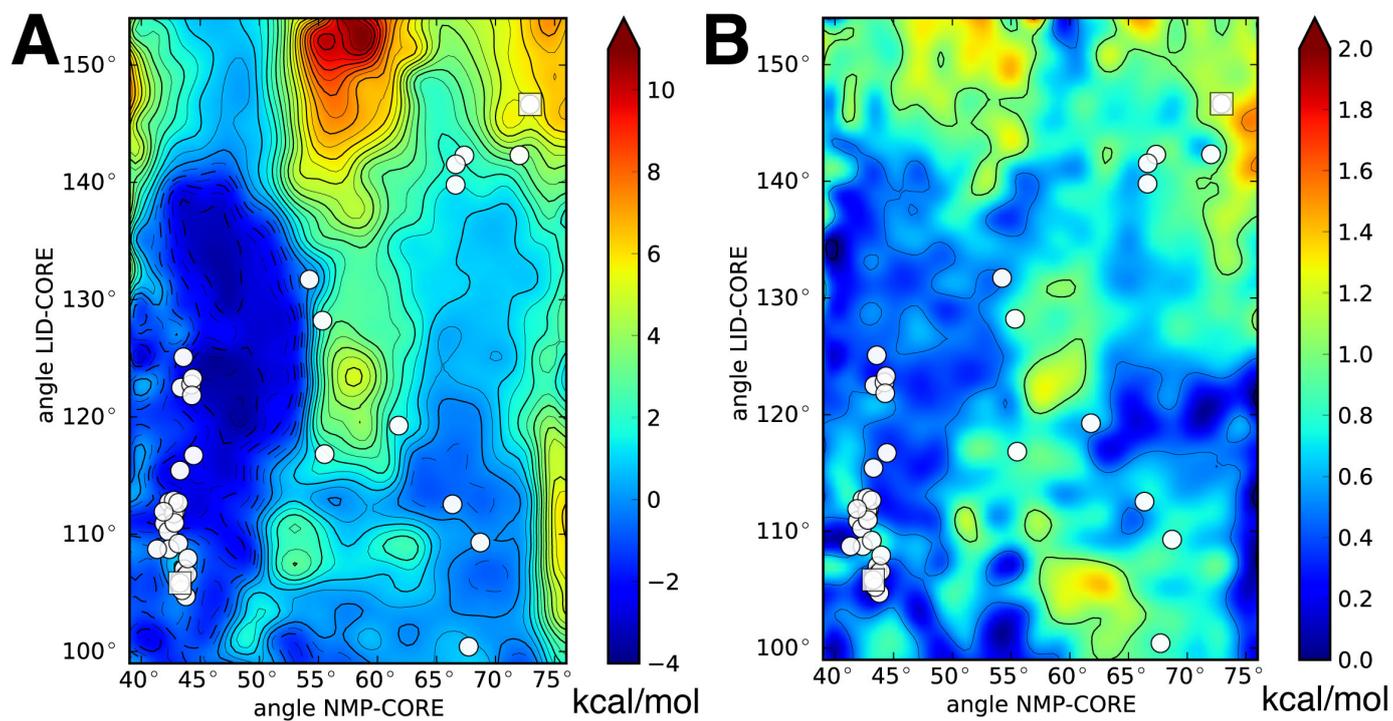


Figure S16: Potential of mean force error analysis. The PMF was calculated from three blocks of 1000 frames (100 ps) each, with the first 2000 (200 ps) frames discarded as equilibration. A: The PMF is the average of these three blocks. B: The error estimate is the standard deviation of the block average. All energies are in kcal/mol.

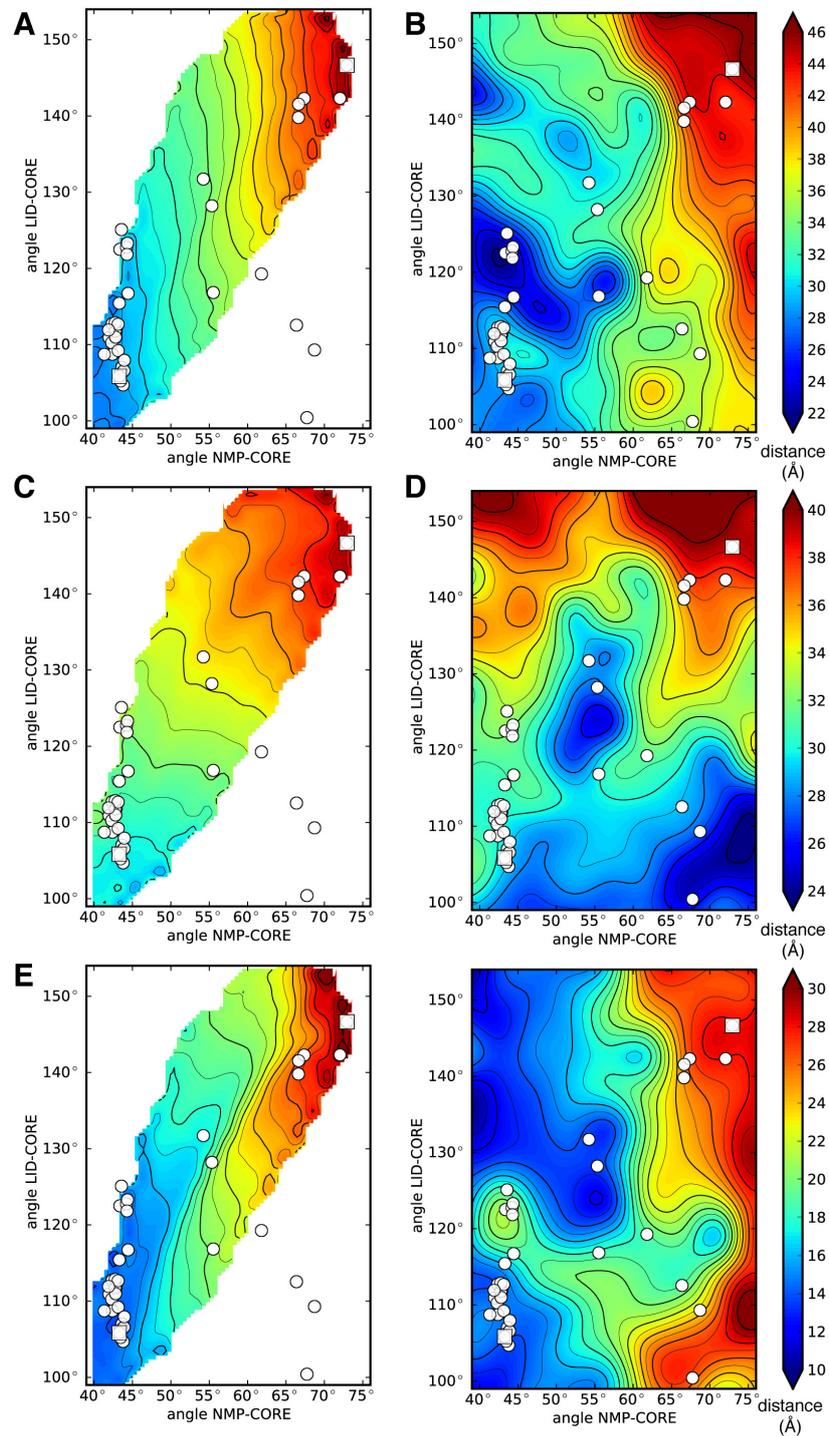


Figure S17: Projection of FRET distances (residue C_{α} - C_{α}) onto angle coordinates. A, B: I52-K145 (NMP-LID), corresponding to Y52-K145 in *A. aeolicus*¹³; C, D: A127-A194 (LID-CORE)¹⁴; E, F: A55-V169 (NMP-LID/CORE)¹⁵. Data are shown for DIMS transitions (left: A, C, E) and umbrella sampled simulations (right: B, D, F).

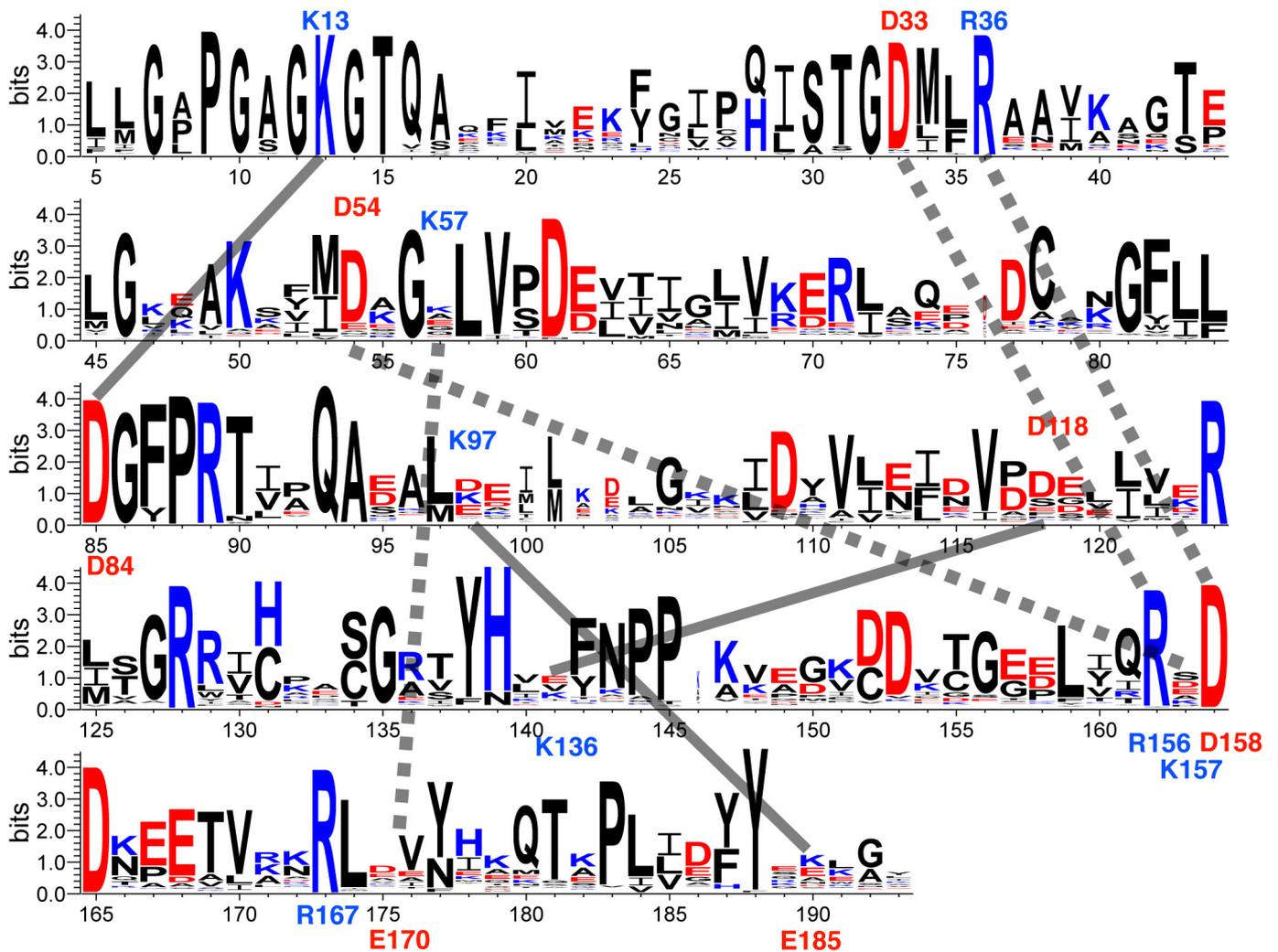


Figure S18: Conservation of salt bridges in AdKs. The sequence logo¹⁶ of the AdK family (PFAM PF00406, with LID-less variants and distant members removed) shows that many basic (blue) and acidic (red) residues are highly conserved. Charged residues involved in salt bridges discussed in the text are shown with their AKeco numbering. Persistent salt bridges are drawn as heavy continuous lines. Four salt bridges that are successively broken during the closed→open transition are indicated with broken lines. They form a salt bridge ‘zipper’ as described in the text.

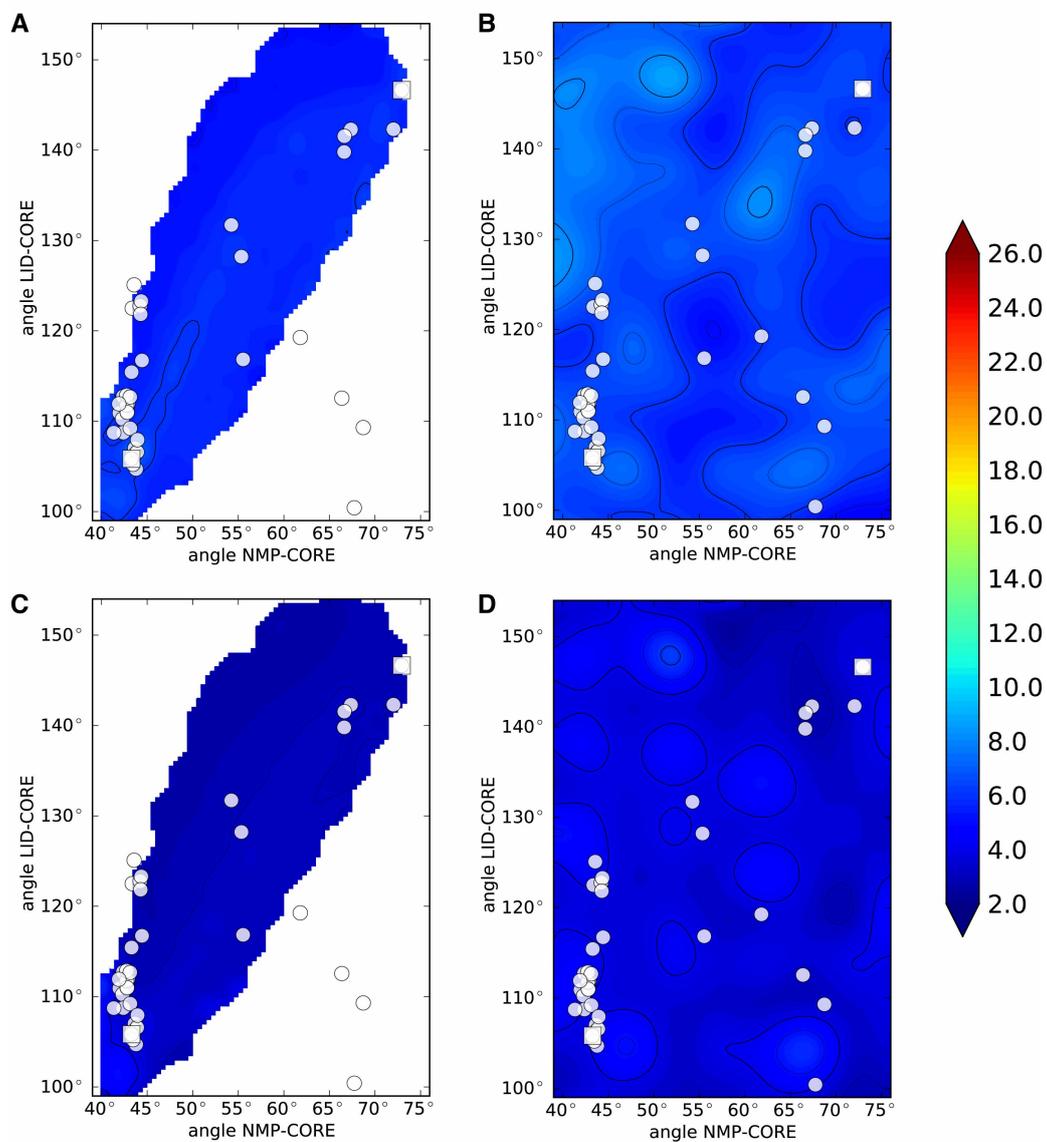


Figure S19: Salt bridge formation of CORE-CORE residue pairs as the transition progresses (in Å). A: K97-N190 (DIMS) B: K97-N190 (PMF) C: K97-E185 (DIMS) D: K97-E185 (PMF).

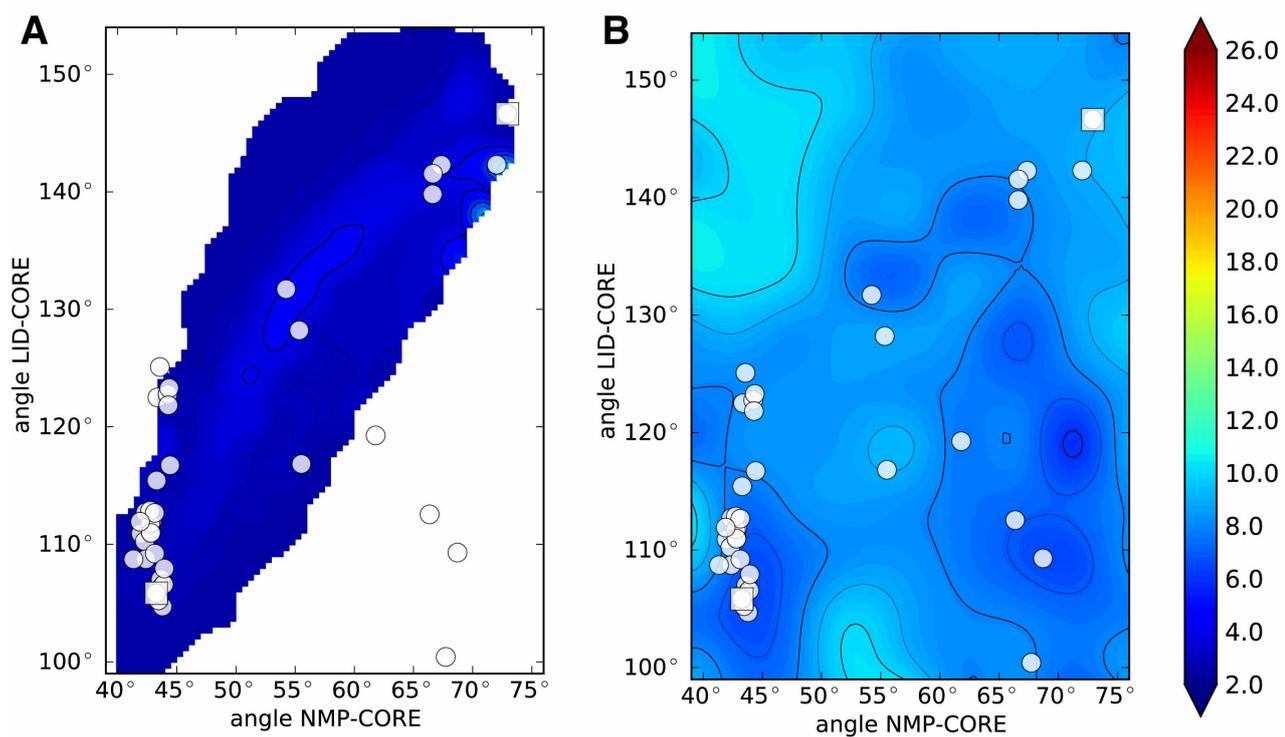


Figure S20: Salt bridge formation of an NMP-NMP residue pair as the transition progresses (in Å). A: E44-K47 (DIMS) B: E44-K47 (PMF)

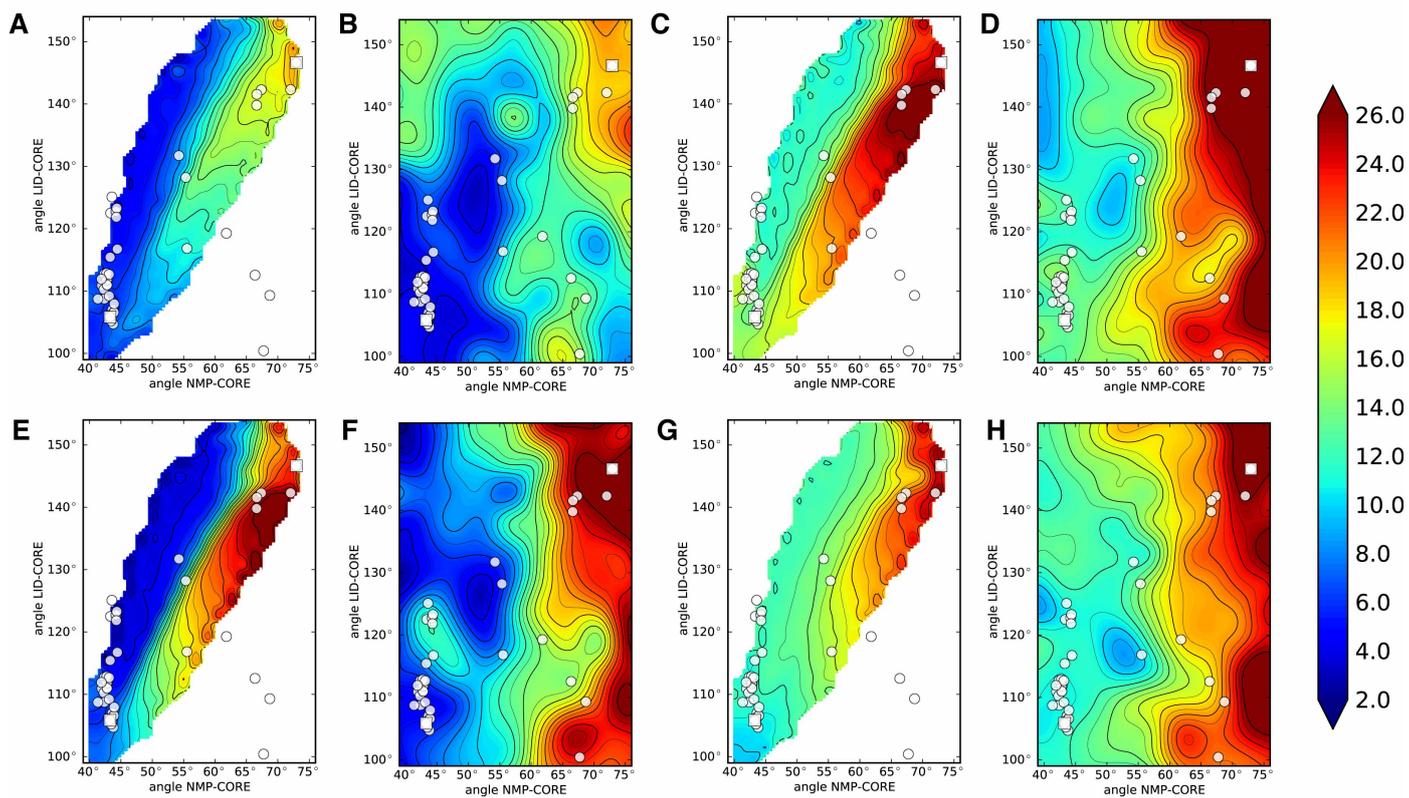


Figure S21: Salt bridge formation of NMP-LID residue pairs as the transition progresses (in Å). A: D33-R156 (DIMS) B: D33-R156 (PMF) C: D54-R156 (DIMS) D: D54-R156 (PMF) E: D54-K157 (DIMS) F: D54-K157 (PMF) G: K57-D158 (DIMS) H: K57-D158 (PMF)

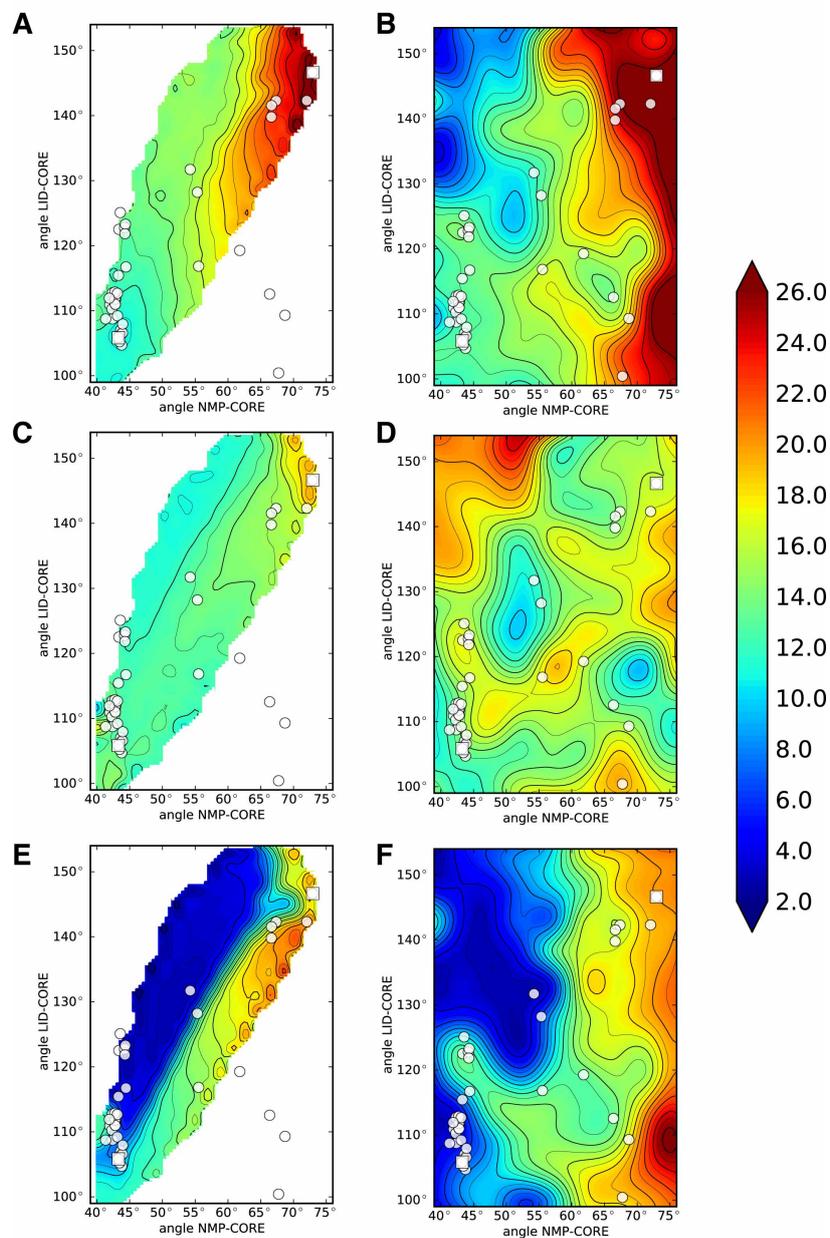


Figure S22: Salt bridge formation of NMP-CORE residue pairs as the transition progresses (in Å). A: D54-R167 (DIMS) B: D54-R167 (PMF) C: R36-E170 (DIMS) D: R36-E170 (PMF) E: K57-E170 (DIMS) F: K57-E170 (PMF)

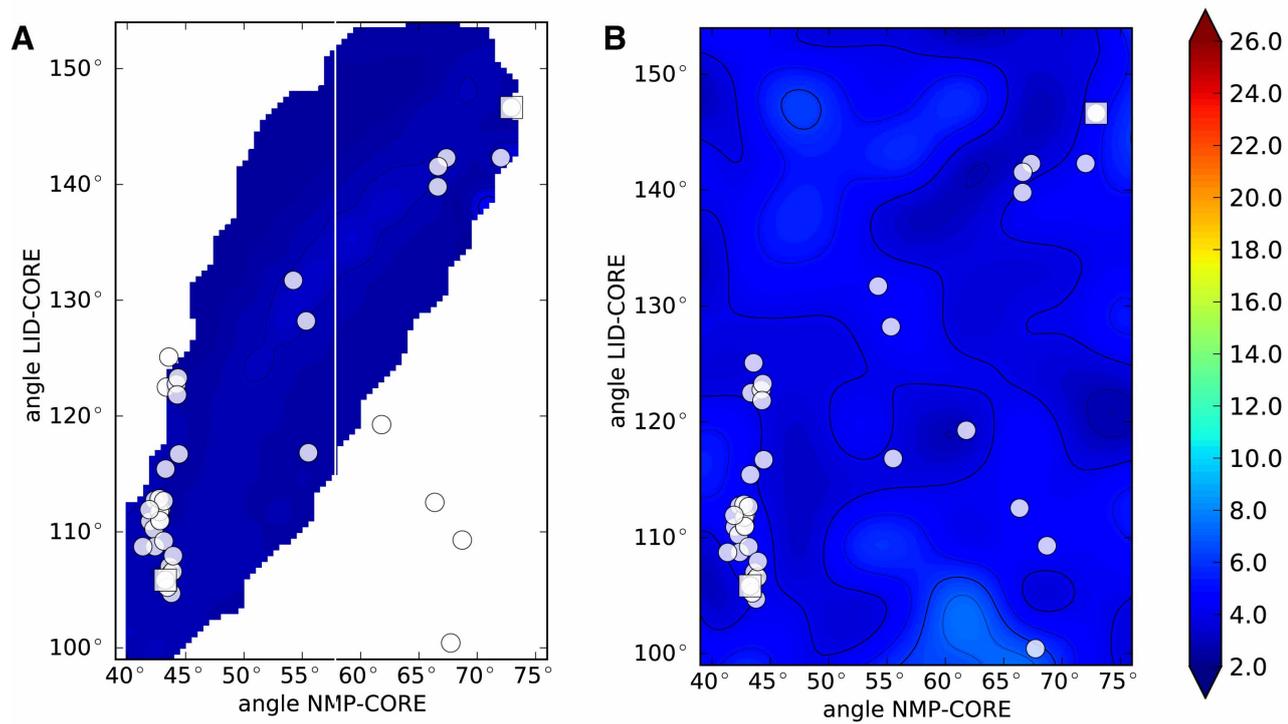


Figure S23: Salt bridge formation of a LID-CORE residue pair as the transition progresses (in Å). A: D118-K136 (DIMS) B: D118-K136 (PMF)

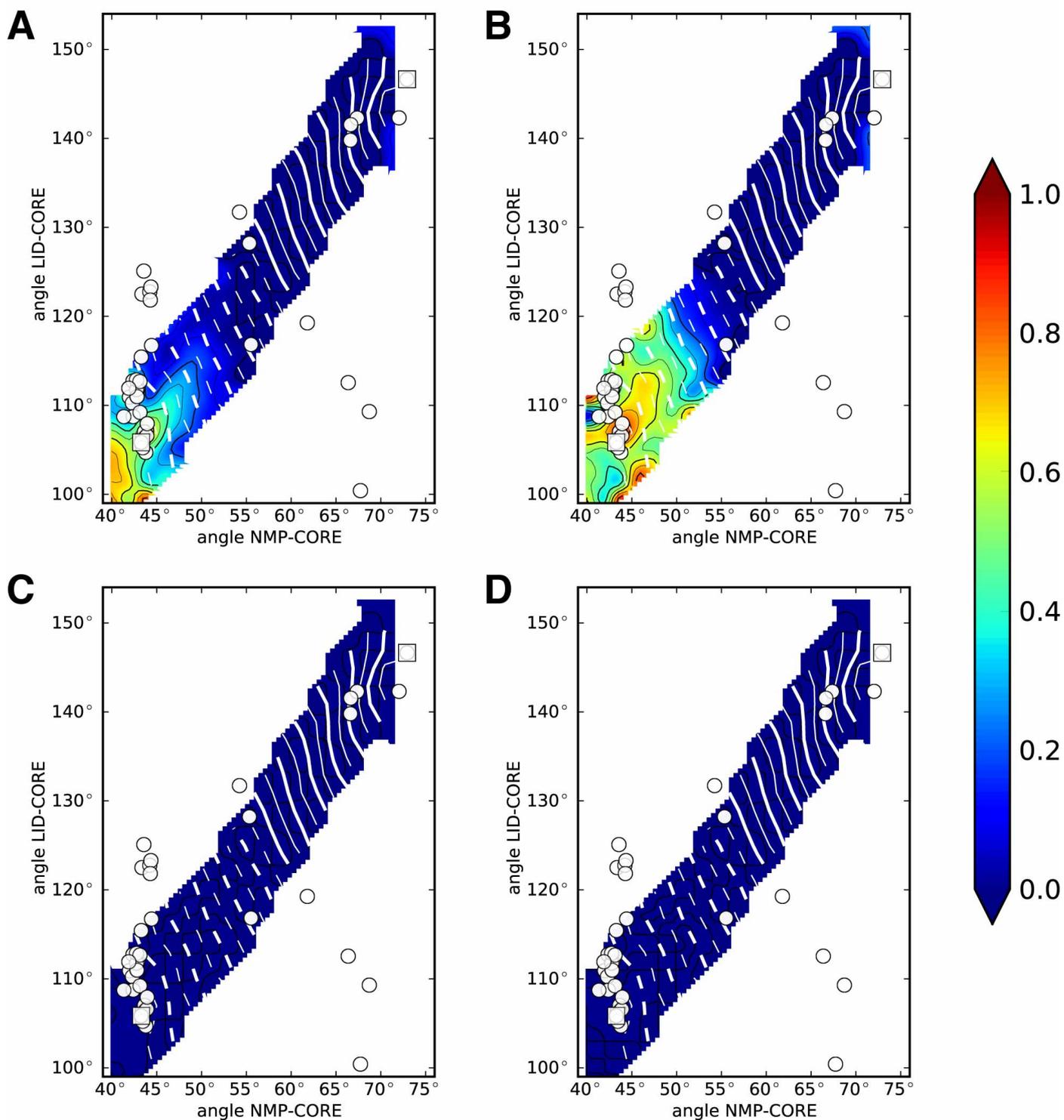


Figure S24: Salt bridge existence probability projected on NMP and LID angles along the open→closed DIMS transition. A: D33-R156; B: R36-D158; C: D54-K157; D: K57-E170. The contour lines for the $\Delta\rho$ coordinate are shown as white lines; the heavy white lines go from -5 \AA (dashed) to $+5\text{ \AA}$ (solid).

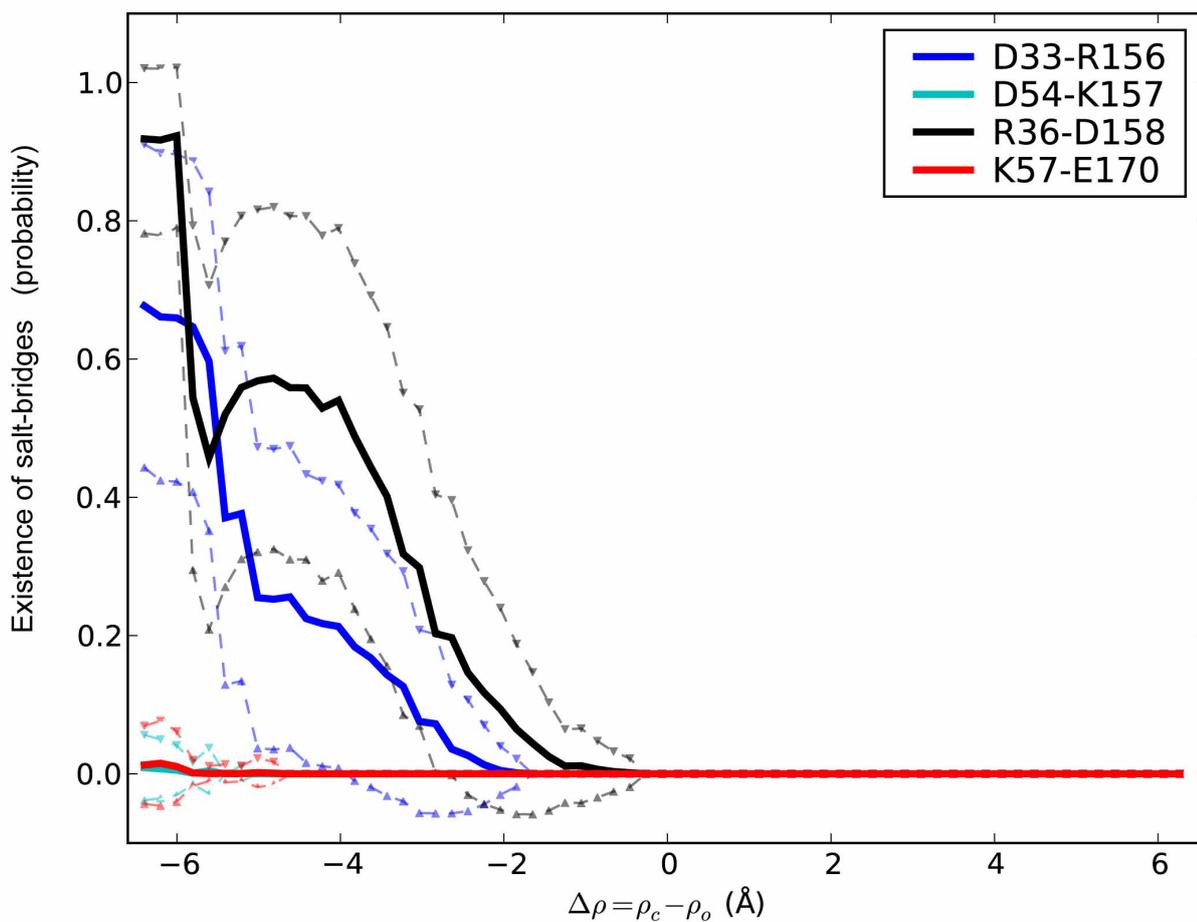


Figure S25: Salt bridge existence probability projected on $\Delta\rho$ from open→closed transition. The heavy lines are the mean of the data, dashed lines indicate one standard deviation of the average. The projection is calculated from the 2D data in Figure S24 as the average value over stripes of constant $\Delta\rho$.

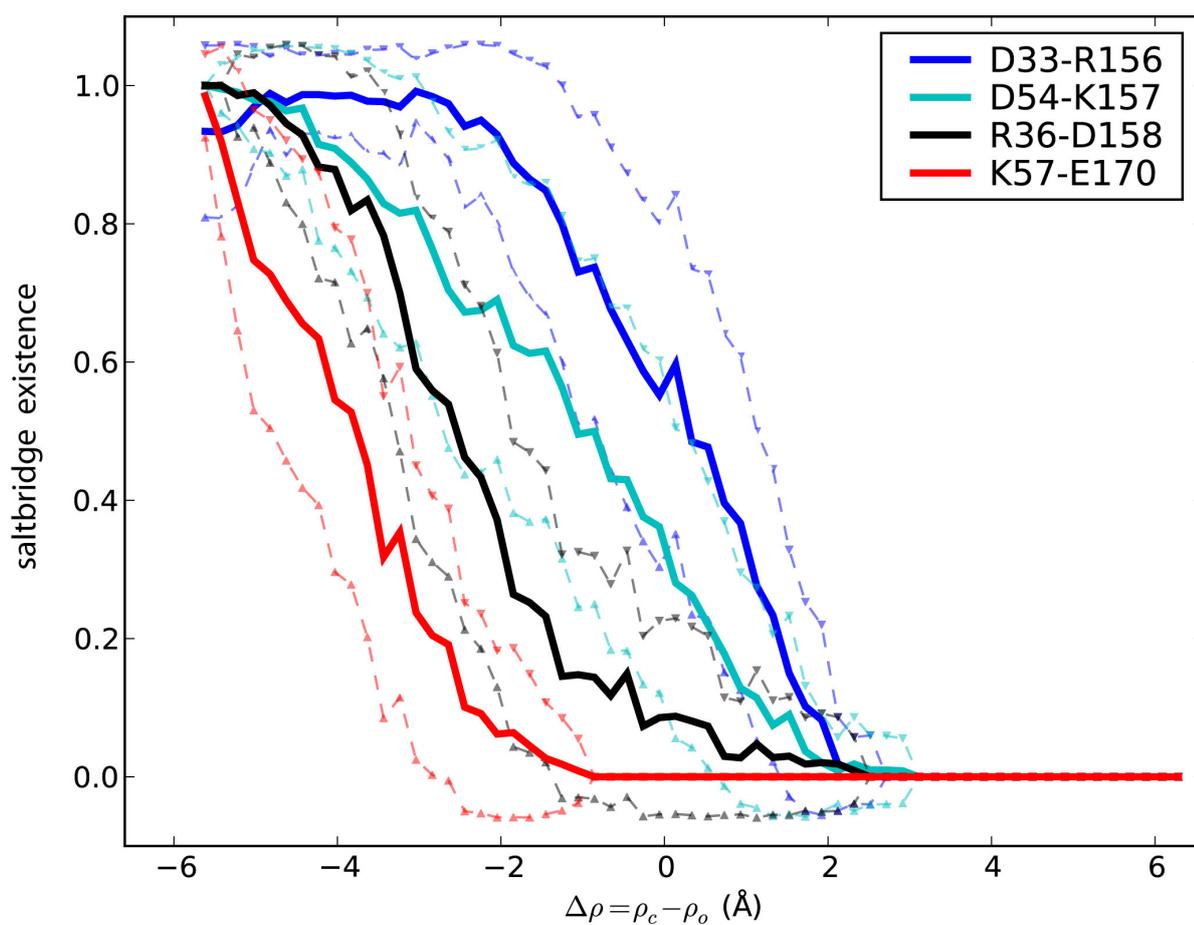


Figure S26: Salt bridge existence in the case of neutral acidic zipper residues. The existence probability was projected on $\Delta\rho$ from open→closed transition in which D33, D54, D158, and E170 were protonated. The heavy lines are the mean of the data, dashed lines indicate one standard deviation of the average.