Supplementary Information

Structural Basis for Reduced Dynamics of Three Engineered HNH Endonuclease Lys-to-Ala Mutants of the Cas9 Enzyme

Jimin Wang[1,*], Erin Skeens[2], Pablo R. Arantes[3], Federica Maschietto[4], Brandon Allen[4], Gregory W. Kyro[4], George P. Lisi[2,*], Giulia Palermo[3,*], Victor S. Batista[4,*]

[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520-8114, USA. [2]Department of Molecular and Cell Biology and Biochemistry, Brown University, Providence, RI 02912, USA. [3]Department of Bioengineering and Department of Chemistry, University of Californian Riverside, Riverside, CA, USA. [4]Department of Chemistry, Yale University, New Haven, CT 06511-8499, USA.

Supporting Information contains one supporting section and five supporting figures.

Computing eigenvector centrality from Kabsch-Sander hydrogen-bond energy

There have been significant efforts in the development of computational methods to measure and predict changes upon perturbation of allosteric sites. Traditionally, the correlation is computed from α-carbon displacements, using measures such as generalized correlation coefficients based on mutual information. For a system, typically represented as a network of nodes, the corresponding graph is described as an adjacency matrix, with elements defining the magnitude of the generalized correlation between nodes. Although correlation of α-carbon displacements has been shown to accurately capture collective protein motions associated with large conformational changes mediating the communication between allosteric and catalytic sites, it often fails to match NMR experiments. NMR α-carbon, N, H chemical shifts are directly related to modification of electrostatic interactions such as hydrogen bonds, which are primary determinants in protein structure, stability, and specificity. By using backbone electrostatics in place of α-carbon displacements, we achieve higher correspondence between experimental NMR and computational data, providing a molecular-level understanding of the experimental observations.

We define the electrostatic energy between a residue *i* and a neighboring residue *j* as given by the Kabsch-Sander (KS) energy computed between a backbone *i-th* CO group and *j-th* NH group. KS hydrogen bonds are defined, assuming partial charges of (i) carbonyl oxygen = -0.42 e, carbonyl carbon = +0.42 e, amide hydrogen = +0.20 e, amide nitrogen = -0.20 e, according to the following equation:

$$\boldsymbol{E} = 0.42e * 0.20e * 33.2\frac{kcal}{mol * nm} * \left(\frac{1}{r_{ON}} + \frac{1}{r_{CH}} - \frac{1}{r_{OH}} - \frac{1}{r_{CN}}\right), \quad (1)$$

where a hydrogen-bond is identified only if *E* is less than -0.5 kcal/mol. We denote the KS adjacency matrix as a non-symmetric matrix whose elements reflect the energetics between each pair of residues, accounted for as either a donor or an acceptor, along rows and columns, respectively. We then define the acceptor (donor) degree centrality matrix, constructed from degree centrality vectors computed for each frame of the MD trajectory, averaging over rows (columns) of each matrix. The resulting matrices indicate the total electrostatic contributions of each acceptor (donor) residue with every other residue in the protein. We use linearized generalized correlation coefficients to quantify how donor-acceptor pairs are perturbed across the trajectory, which provides a reliable measure of linear correlations without suffering from the flaws of noncollinearity associated with Pearson correlation. The derived correlation matrices describe how any two donors/acceptors are correlated. We apply eigenvector centrality to then assess the importance of each residue (node), which has been previously applied to proteins to rank the nodes according to both the number of connections and relevance in terms of information flow. The eigen decomposition of the eigenvector centrality matrices yields donor and acceptor centrality vectors, which identify the acceptor (donor) residues whose electrostatics are most largely perturbed across the dynamics. The acceptor eigenvector centrality is then subtracted from the donor centrality for the wt and each

mutant, followed by normalization of this resulting centrality. This centrality for the mutants is then subtracted from that of the wt, to produce the difference in centrality metric that is presented in Fig. 3.

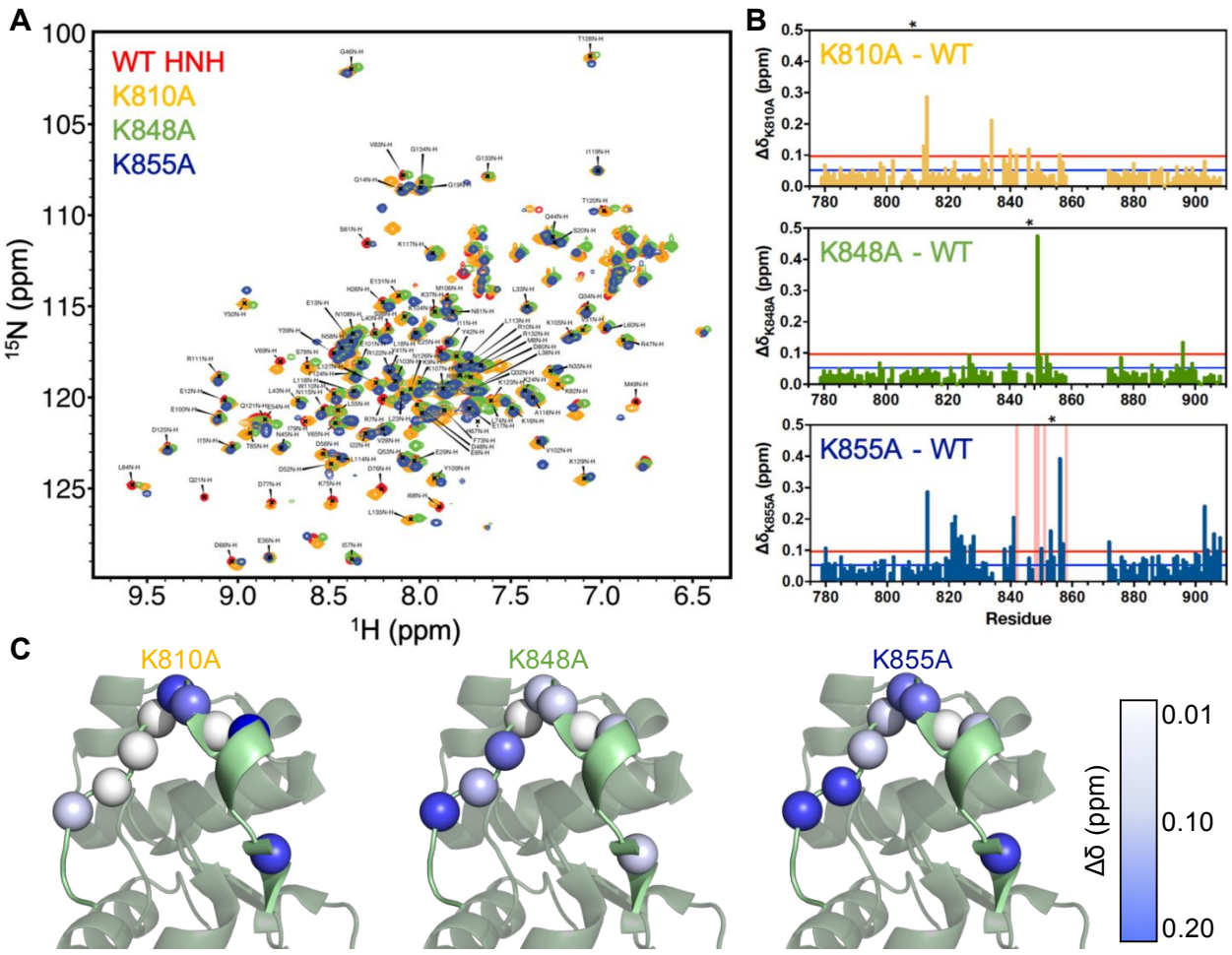


Figure S1. Structural perturbations caused by Lys-to-Ala mutations in the HNH domain of CRISPR Cas9 detected by NMR spectroscopy. (A) ^{1}H-^{15}N HSQC spectral overlay of the WT (red), K810A (yellow), K848A (green) and K855A (blue) mutants of the isolated HNH domain. (B) Combined chemical shift perturbations (Δδ) of the Lys-to-Ala HNH mutants, relative to WT HNH. The site of mutation is denoted by an asterisk above each plot. Blue lines represent the 10% trimmed mean of all chemical shifts and red lines signify 1.5σ above the 10% trimmed mean. Pink bars denote sites of observed line broadening. Chemical shift perturbations of the Y836-containing loop (E827-D837) are shown in (C), where sphere color represents the magnitude of the chemical shift (Δδ) for each residue as denoted in the legend.

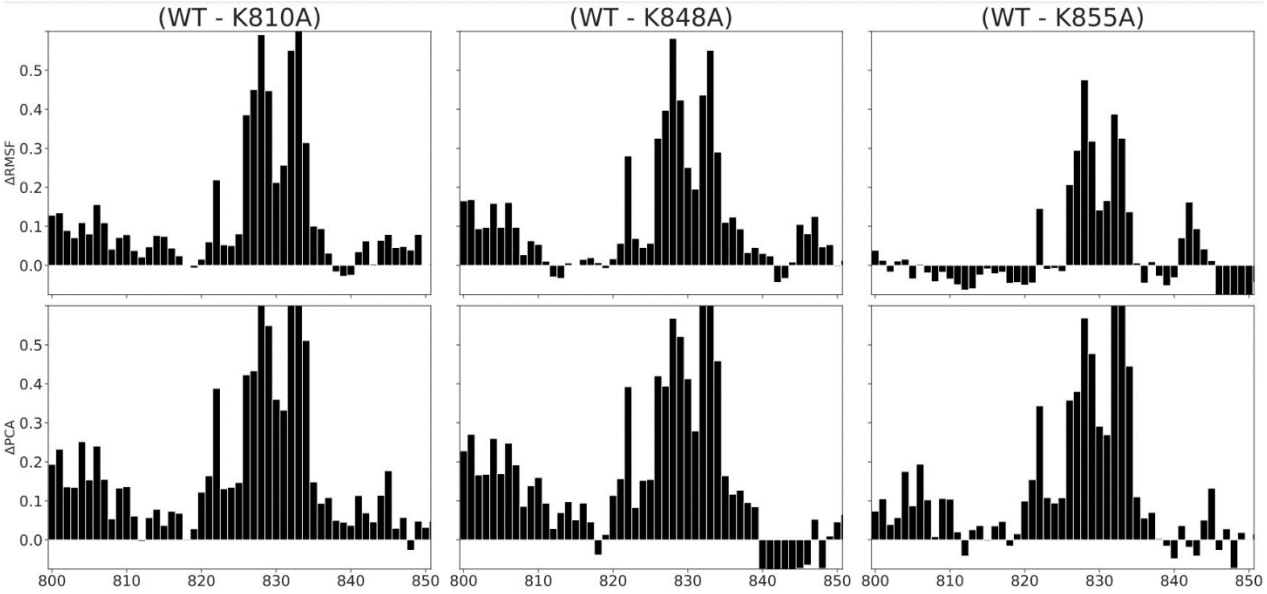


Figure S2. ΔRMSF and ΔPCA between wt and each of the three mutants. ΔPCA is the difference in variance of residues P800-D850 projected onto the first principal component. Much of the variance is due to this single essential mode, highlighted by the similarity between the ΔRMSF and ΔPCA plots.

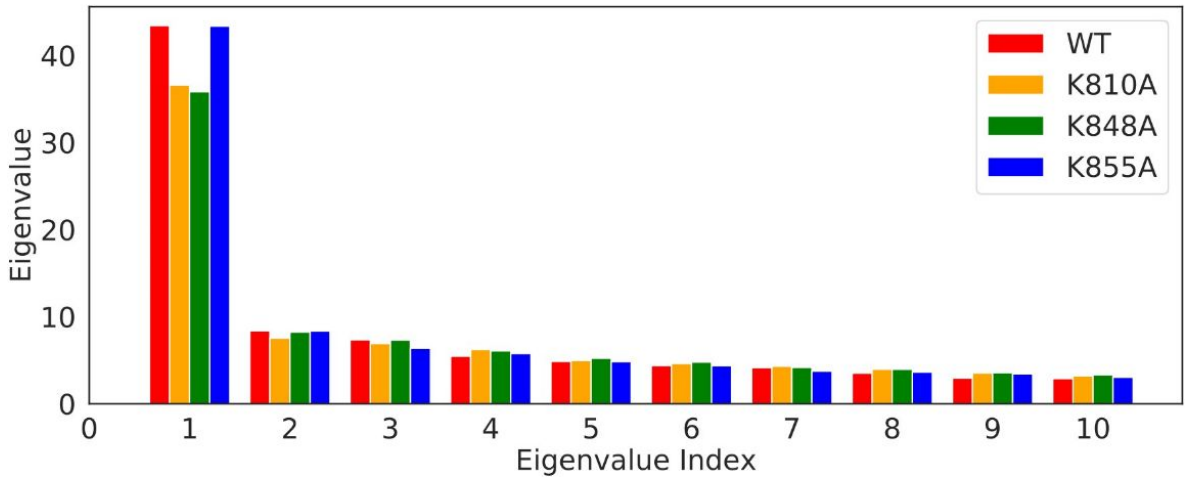


Figure S3. Largest 10 eigenvalues obtained from the adjacency matrix used to compute eigenvector centrality from the α-carbon coordinates for the wt (red), K810A (orange), K848A (green), and K855A (blue) mutants (in the limit of $\lambda \rightarrow \infty$).

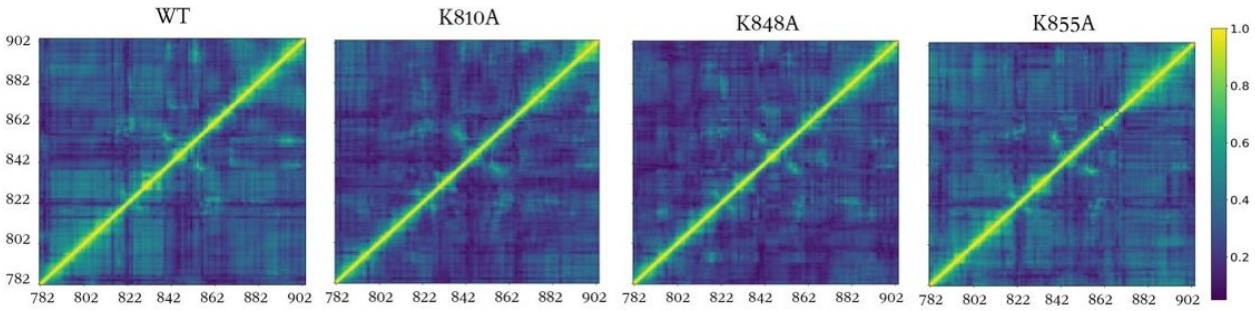
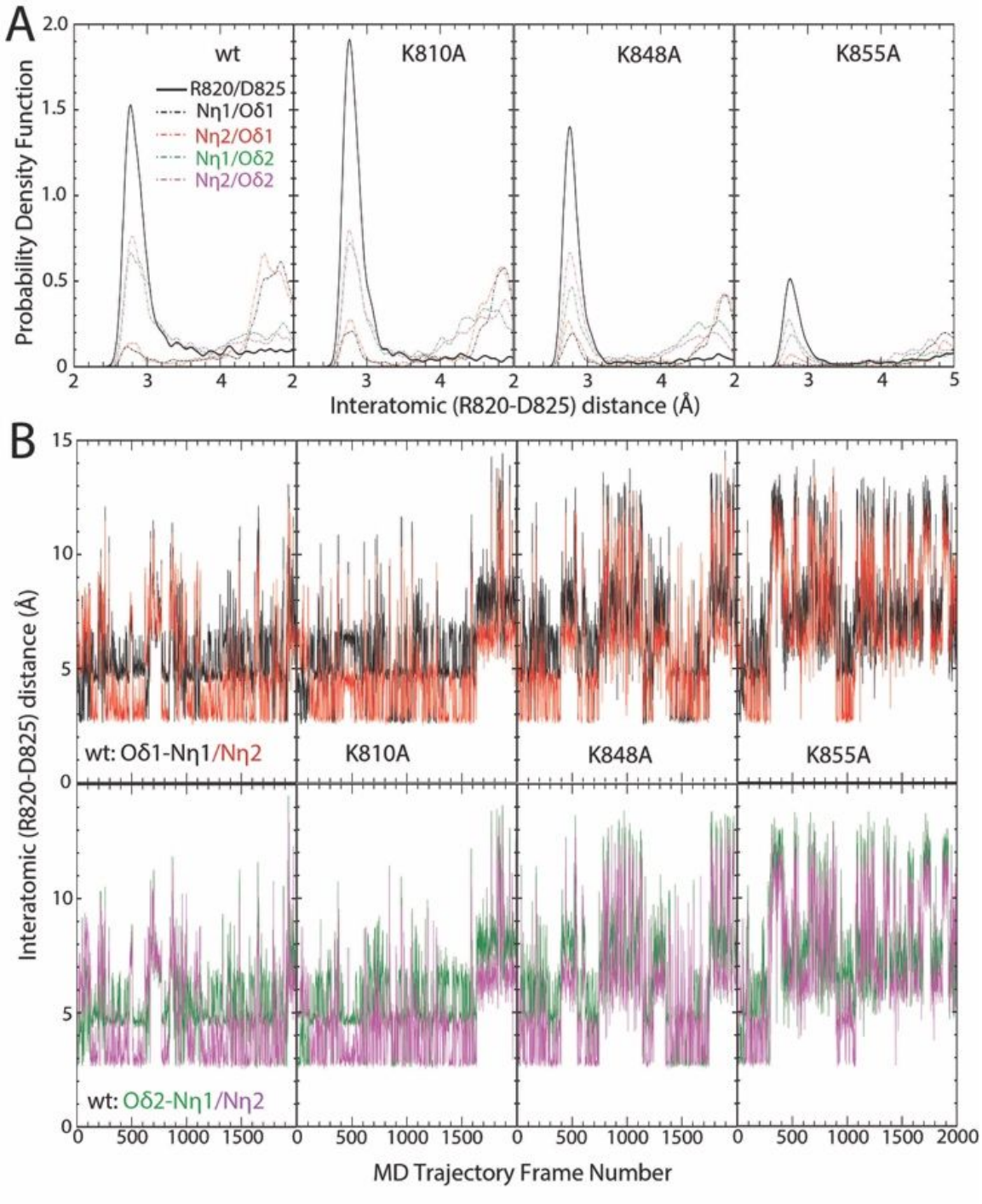


Figure S4. Generalized correlation matrix plots for wt, K810A, K848A, and K855A mutants


A
wt
K810A
K848A
K855A
R820/D825
Nη1/Oδ1
Nη2/Oδ1
Nη1/Oδ2
Nη2/Oδ2
Probability Density Function
Interatomic (R820-D825) distance (Å)
B
wt: Oδ1-Nη1/Nη2
wt: Oδ2-Nη1/Nη2
Interatomic (R820-D825) distance (Å)
MD Trajectory Frame Number

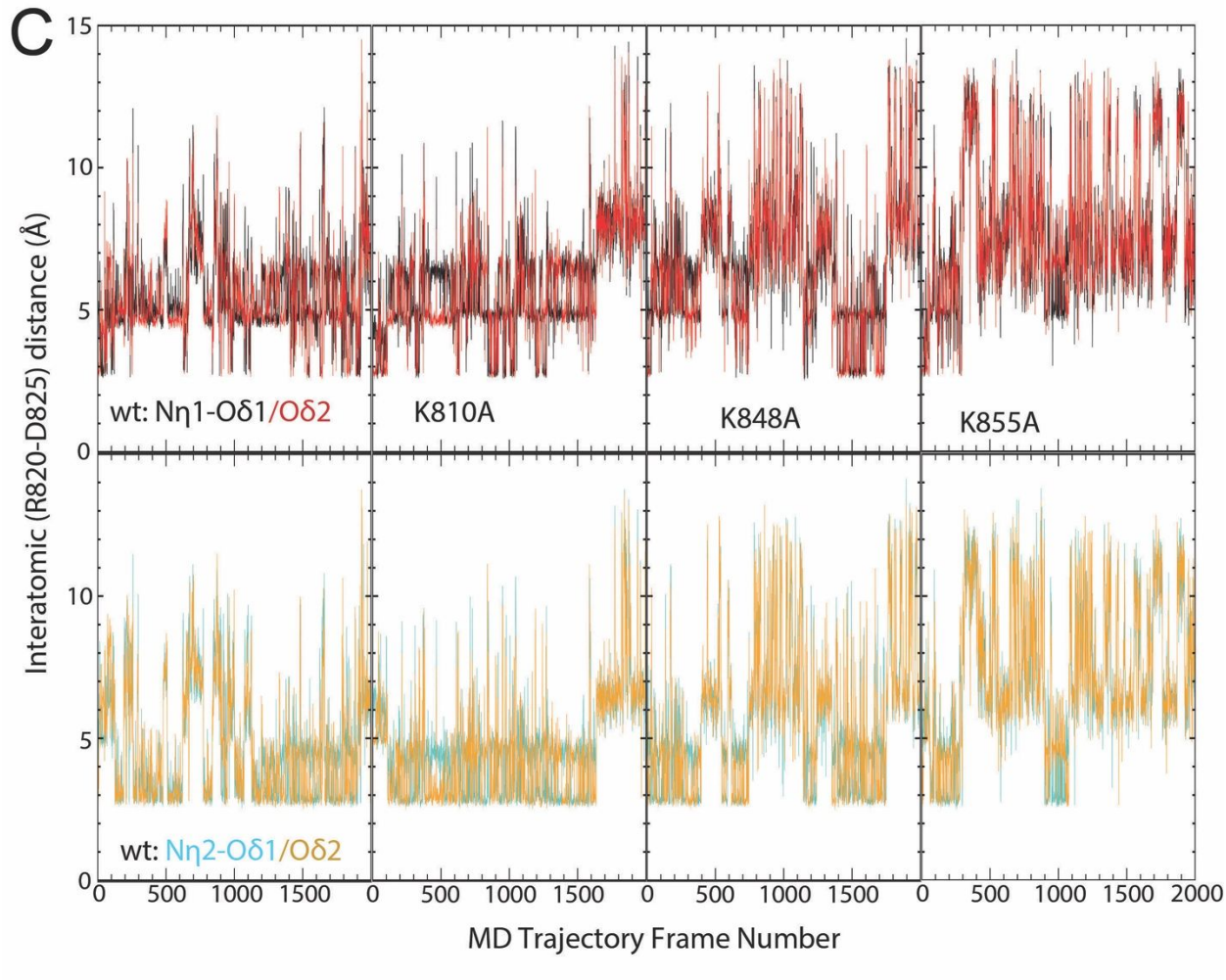


Figure S5. Equilibrated properties and kinetics for R820-D825 H-bond formation and breakage as measured by inter-atomic distances as a function of MD trajectory frame number. (A) Probability density functions of the R825-D825 distance for the wt, K810A, K848A, and K855A mutants (from left to right). The solid black curve is the sum of four possible H-bond pairs, R820Nη1-D825Oδ1(black dashed curves), Nη2-Oδ1 (red dashes), Nη1-Oδ2 (green dashes), and Nη2-Oδ2 (magenta dashes). (B) As a function of flipping R820 sidechain with a fixed Oδ1 (middle panel) or Oδ2 atom (bottom) of D825. See Figure S5 for the kinetics of flipping the D725 sidechain with fixed Nη1 or Nη2 atoms of R820. (C) Alternative interatomic R820-D825 distance as a function of trajectory frame number.