Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information: Protein flexibility reduces solvent-mediated friction barriers of ligand binding to a hydrophobic surface patch

Christopher Päslack,[†] Lars V. Schäfer,[†] and Matthias Heyden[*,‡]

[†]*Theoretical Chemistry, Ruhr University Bochum, D-44780 Bochum, Germany*

[‡]*School of Molecular Sciences, Arizona State University, Tempe, AZ 85287-1604, USA*

E-mail: mheyden1@asu.edu

Phone: +1 480 965-3980

Molecular dynamics simulations

The starting structure used in the simulations was the X-ray crystal structure of human erythrocytic ubiquitin refined at 1.8 Å resolution (PDB ID: 1UBQ).[S1] The model ligand was described as a particle with mass 39.948 a.m.u. and (12,6)-Lennard-Jones parameters $\sigma = 4$ Å and $\varepsilon = 0.025$ kJ/mol, yielding a moderate binding free energy to the hydrophobic patch. All simulations were performed using the GROMACS 4.6.1 software package.[S2] We used the CHARMM27[S3] force field for the protein and the TIP4P/2005 water model.[S4] The protein was placed in a very large 80 Å×80 Å×80 Å periodic cubic box and solvated with 16,545 water molecules. The ligand was positioned at 4 Å from the hydrophobic patch. The system was energy minimized and pre-equilibrated for 20 ps in the isobaric-isothermal (NPT) ensemble at 300 K and 1 bar with harmonic restraining potentials applied to all protein heavy atoms and the ligand (force constant 1000 kJ mol^{-1} nm^{-2}). We employed the SETTLE and LINCS algorithms[S5–S7] to constrain the internal degrees of freedom of water molecules and the bonds in the protein, respectively. The equations of motion were integrated using the leapfrog algorithm.[S8] Neighbor searching was performed every 10 steps. We used a cut-off of 0.9 nm for van der Waals and electrostatic interactions. Long-range Coulomb interactions were treated with the PME method[S9] with a grid spacing of 1.2 Å and cubic spline interpolation. Analytical dispersion corrections were applied for energy and pressure to compensate for the truncation of the Lennard-Jones interactions.[S10]

For the equilibrations with position restraints, a time step of 1 fs was used for integrating the equations of motion. A Berendsen thermostat[S11] was applied with a reference temperature of 300 K and a time constant of 1 ps, coupling the water separately from the protein and the ligand. For pressure coupling, a Berendsen barostat[S11] was used with a reference pressure of 1 bar and a time constant of 1 ps.

We generated the starting structures for the subsequent umbrella sampling simulations by performing a constant-force pulling simulation in the NPT ensemble (force constant of the harmonic spring 50 kJ/mol/nm^2). For each generated ligand position, we performed an

additional equilibration simulation of 1 ns with the biasing potential enabled to allow for a relaxation of water molecules for the given reference position of the ligand. From this step onward, we ran three sets of simulations of flexible, restrained and frozen proteins, respectively. In all simulations of the frozen protein, protein motion was inhibited using freeze groups (i.e., no dynamics). The equations of motion were integrated with a time step of 1 fs. For the flexible and restrained proteins, a time step of 2 fs was used.

Umbrella sampling was then performed in the canonical ensemble (NVT) using 15 windows with 1 Å-spacings of the ligand position on the reaction coordinate, and 3 additional windows at 3.5 Å, 6.5 Å and 7.5 Å. The connecting vector between the binding site and the ligand was defined as the reaction coordinate q, which was aligned along the x-axis (*i.e.*, roughly perpendicular to the hydrophobic patch, see Fig. 1B of the main text). The force constant of the harmonic biasing potential was 1000 $\mathrm{kJ \cdot mol^{-1} \cdot nm^{-2}}$ and the ligand was further position-restrained by harmonic potentials in orthogonal directions (y,z) with a force constant of $k = 1500$ $\mathrm{kJ \cdot mol^{-1} \cdot nm^{-2}}$. Each umbrella window was simulated for 2.1 ns (10 ns for the frozen protein). In addition, we also performed longer simulations (4.2 ns) for 3 selected windows (3.5 Å, 6.5 Å and 7.5 Å) to verify convergence of our data. We employed the Nosé-Hoover thermostat[S12,S13] using a reference temperature 300 K and coupling time constant 1 ps. We recorded positions, velocities and forces every 5 steps (10 fs). This high-frequency output was necessary to properly sample the fast random force fluctuations.

For reference calculations of bulk water properties, we performed simulations of a $40 \times 40 \times 40$ $\mathrm{Å^3}$ cubic box with TIP4P/2005 water. This water box was equilibrated subsequently in the NPT and NVT ensembles for 10 ps, respectively. Then a production simulation of 2.1 ns was performed as was done for the protein–ligand systems.

The first 100 ps of all production simulations were considered additional equilibration time and therefore excluded from the subsequent analyses. For the longer simulations of 4.2 ns we omitted the initial 200 ps from the analysis.

To verify that the restraining potentials on the 6 anchor C-alpha atoms did not affect

protein motions, we computed the root mean squared fluctuations (RMSF) of the protein atoms. To that end, we carried out one further simulation of 2.1 ns of a fully unrestrained protein, i.e. without any position restraints. Figure S1 compares the RMSF between the flexible and unrestrained proteins.

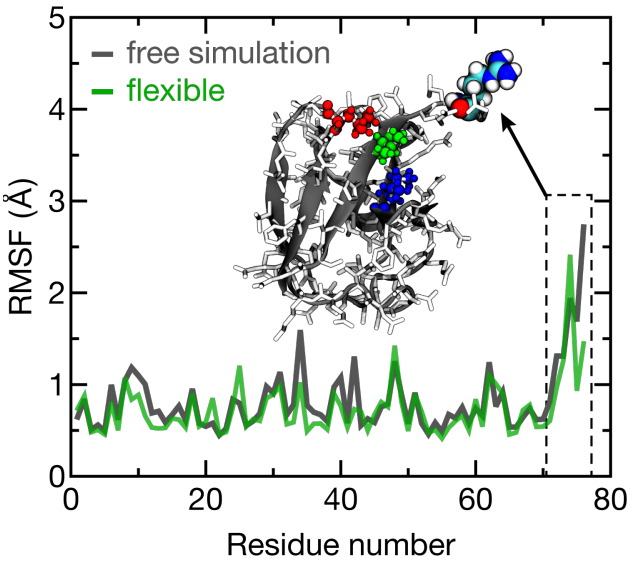

Figure S1: RMSF of protein heavy atoms computed for flexible (green) and fully unrestrained (gray) UBQ. The inset visualizes the location of the terminal atoms that form a highly flexible tail.

Potentials of mean force

Convergence of the free energy profiles was investigated by analyzing consecutively longer parts of the trajectories. Figure S2 shows that the PMFs were converged after 2 ns of sampling time (10 ns for the frozen protein). Generally, the error bars (Fig. S2) should not be larger than the difference between the final curve and the curve obtained for 50% of the data, which is the case here.

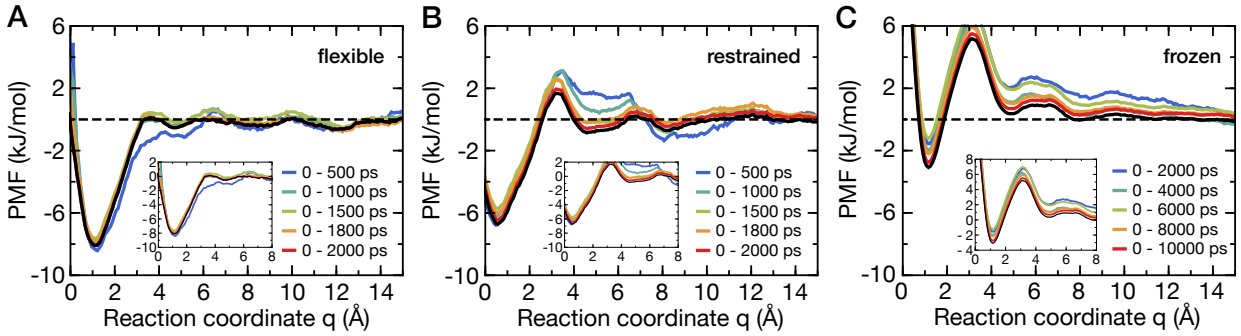

Figure S2: Convergence of the PMFs of protein–ligand interactions. Shown are the results of the analyses of successively longer parts of the trajectories for flexible (A), restrained (B), and frozen (C) proteins. Insets show a close-up view of the region around the minimum of the PMFs.

The histograms of ligand positions in Fig. S3 show that there was sufficient overlap between the distributions to properly sample the free energy profiles.

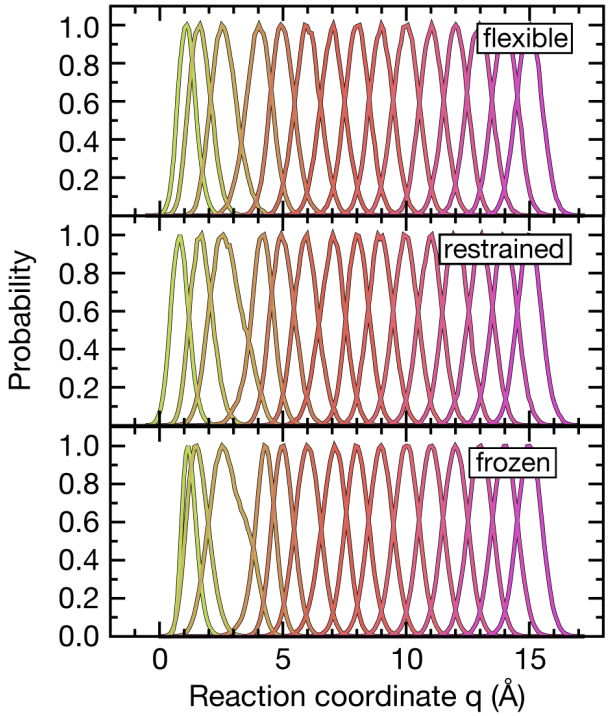

Figure S3: Histograms of ligand positions during the umbrella sampling simulations underlying the PMFs for the flexible, restrained and frozen proteins.

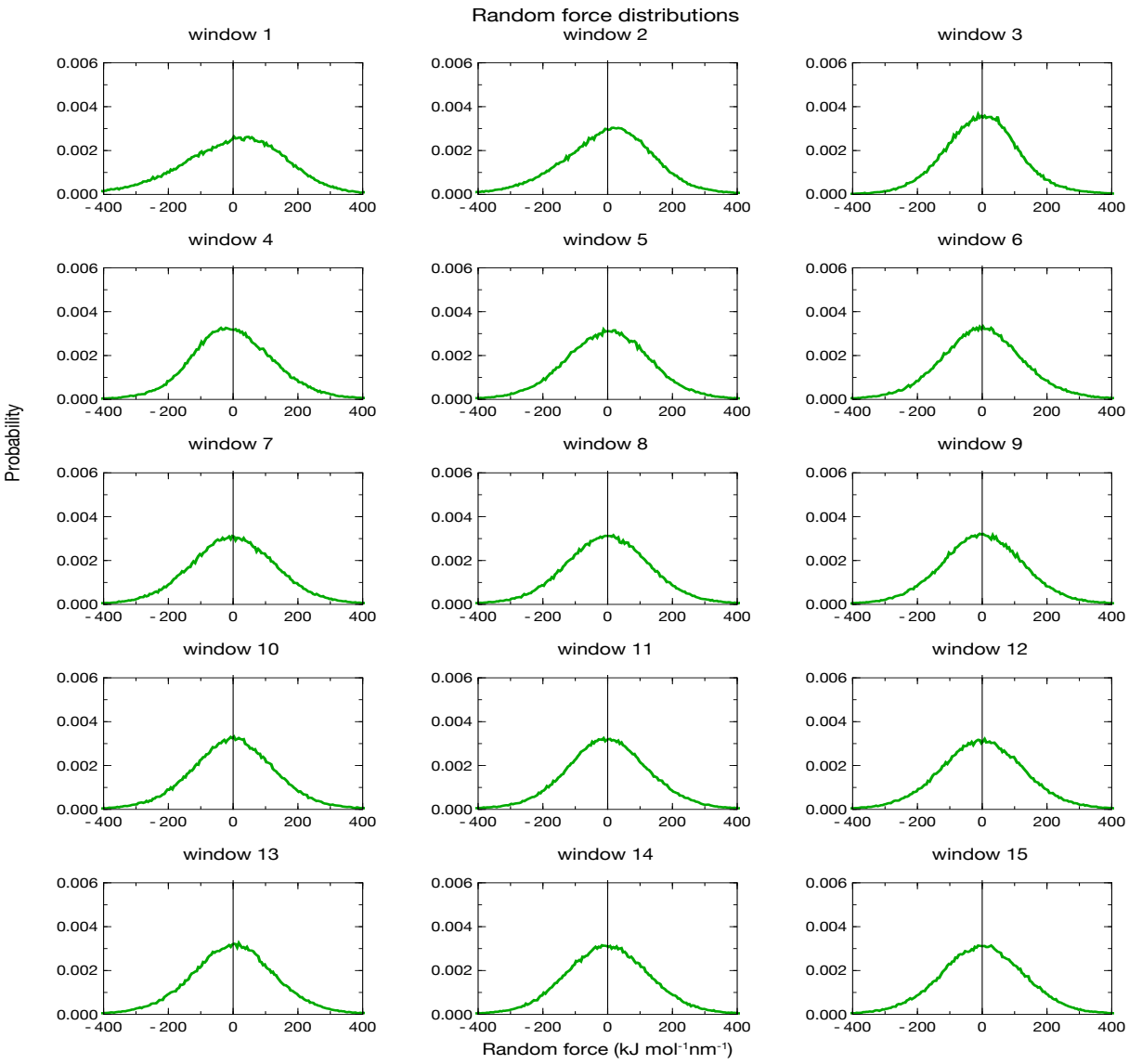

Figure S4: Random force distributions over 2 ns of production simulations for each umbrella window, shown for the flexible protein.

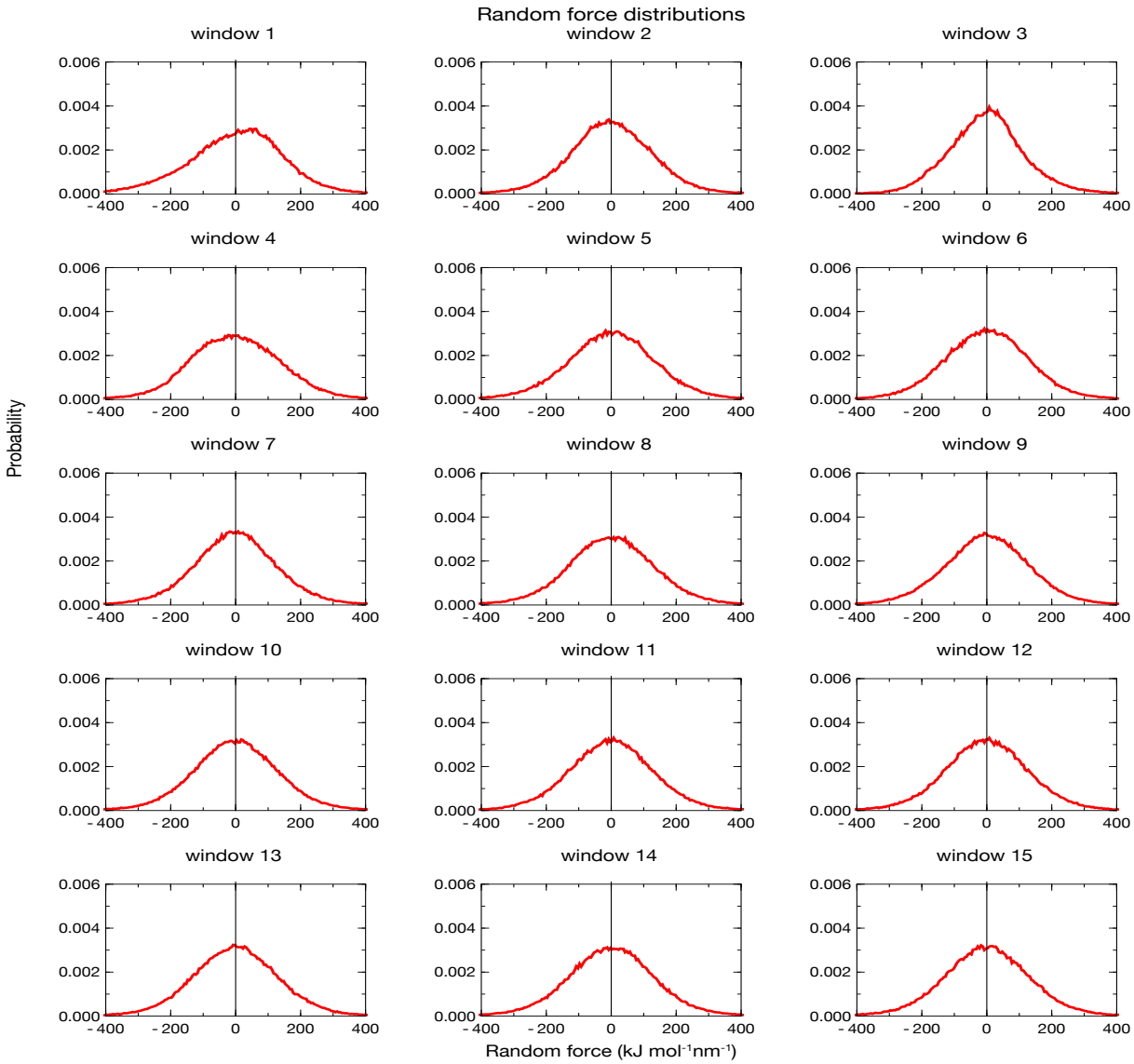

Figure S5: Random force distributions over 2 ns of production simulations for each umbrella window, shown for the restrained protein.

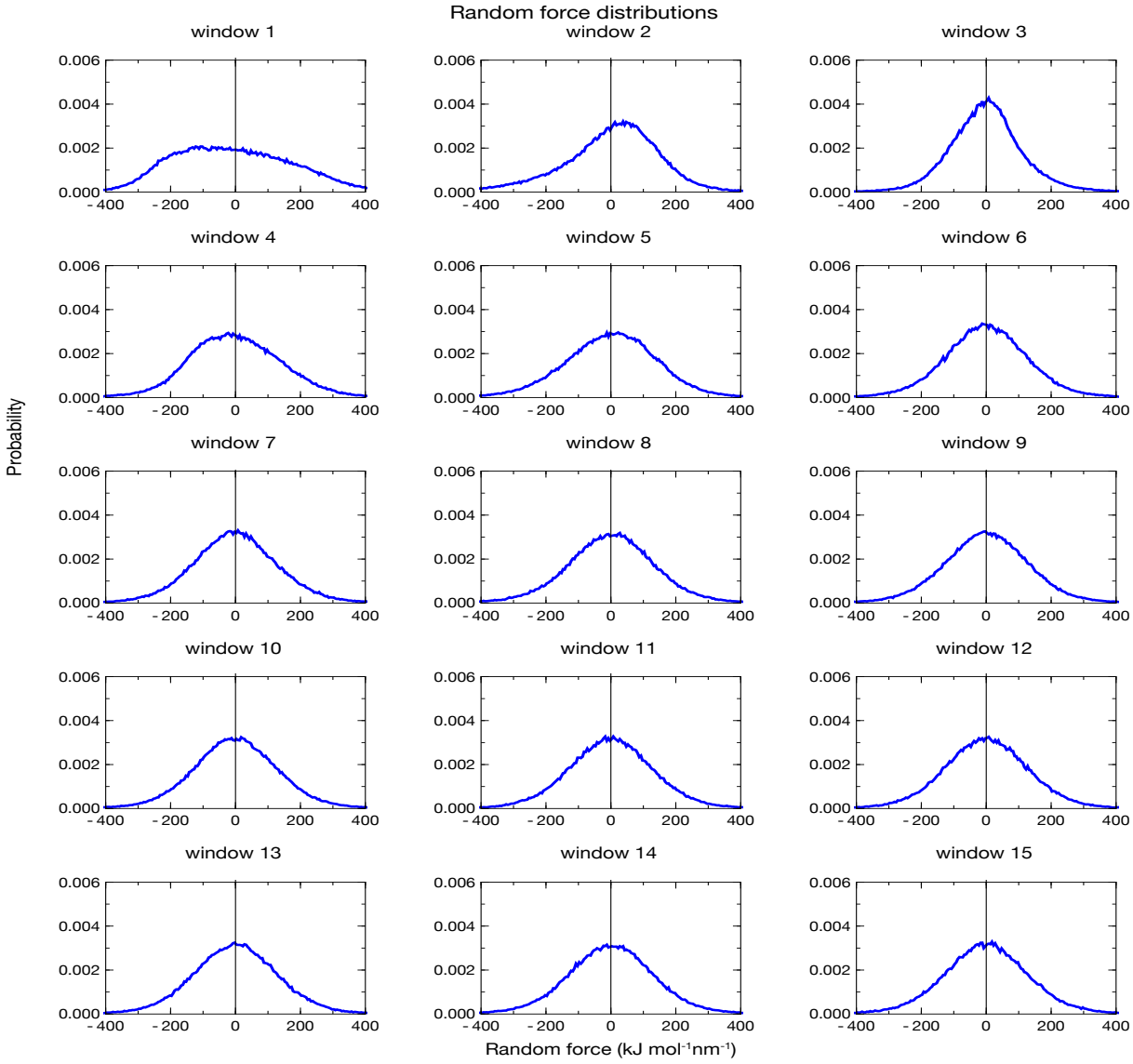

Figure S6: Random force distributions over 2 ns of production simulations for each umbrella window, shown for the frozen protein.

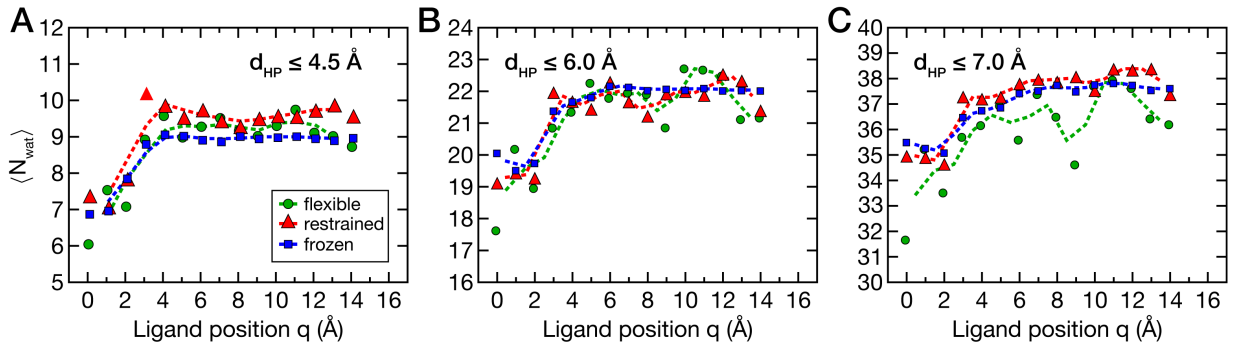

Figure S7: Average number of water molecules $\langle N_{wat} \rangle$ near the hydrophobic patch over 2 ns simulation time as a function of ligand position q. Shown are the results for flexible (green), restrained (red) and frozen (blue) proteins. Water molecules within (A) 4.5 Å of the HP surface, (B) 6.0 Å of the HP surface, and (C) 7.0 Å of the HP surface.

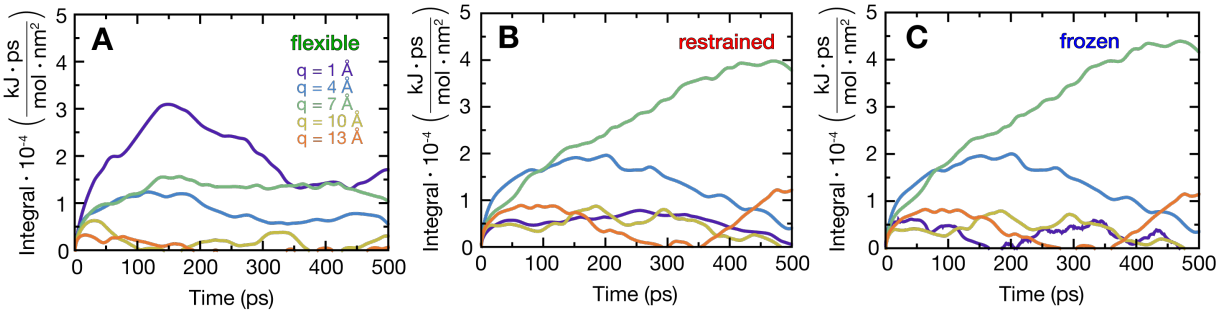

Figure S8: Integrals of the FACs. Shown are the results for the flexible (A), restrained (B) and frozen (C) proteins for selected ligand positions q along the reaction coordinate.

References

(S1) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **1987**, *194*, 531 – 544.

(S2) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(S3) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(S4) Abascal, J. L. F.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.

(S5) Miyamoto, S.; Kollman, P. A. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(S6) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(S7) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

(S8) Hockney, R. W.; Eastwood, J. W. *Computer simulation using particles*; crc Press, 1988.

(S9) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(S10) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. Accurate and Efficient Corrections for Missing Dispersion Interactions in Molecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 13052–13063.

(S11) 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.

(S12) Nosé, S. Constant-temperature molecular dynamics. *J. Phys. Condens. Matter* **1990**, *2*, SA115.

(S13) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97*, 2635–2643.