Supporting Information

Overcoming Convergence Issues in Free-Energy Calculations of Amide-to-Ester Mutations in the Pin1-WW Domain

Daniel Markthaler, Hamzeh Kraus, and Niels Hansen*

Institute of Thermodynamics and Thermal Process Engineering, University of Stuttgart, D-70569 Stuttgart, Germany

E-mail: hansen@itt.uni-stuttgart.de

Phone: +49 (0)711 685-66112. Fax: +49 (0)711 685-66140

EDS simulations

Enveloping distribution sampling (EDS) simulations[1,2] were conducted in the present work with the aim to remove a sampling issue related to the use of a dual topology in previous work,[3] in which two copies of the side chain are present that do not interact with each other but with the environment. In many cases, these two copies (one belonging to the native and one to the ester end state) sampled different conformations at a particular time step leading to strong solute-solvent overlap and strong modifications of the potential energy landscape through the need to use a low EDS smoothness parameter (see Table S1 in [3]). Through the application of distance restraints that keep the two copies of the side chain within the same conformation (see Table S1), this sampling issue can be removed allowing an EDS smoothness parameter of unity to be used for all cases. The sampling quality is improved considerably as demonstrated in Figures S1 to S3 for the case of tri- to heptapepdides (see previous work[3] for computational details and Figures S4 to S9 therein for an analysis of sampling quality). However, as is discussed in the main text of the present work, the use of distance restraints does not remedy the observed starting structure dependence and might even hamper the sampling of conformational transitions in the backbone. Therefore, we resorted to a single topology approach in the present work combined with an enhanced sampling method.

Table S1: Overview of applied distance restraints between corresponding side chain atoms of the two non-interacting side chain branches (dual topology approach) in the EDS simulations. A force constant of 500 kJ mol^{-1} nm^{-2} and a zero reference distance was applied for the harmonic restraining potential.

Mutation	Residue Type [a]	Restrained Atoms [a]
L7λ	LEU	-
W11ω	TRP	CD1, CE3, CZ2
E12ϵ	GLU	CB, CG, CD
K13κ	LYSH	CE
R14ρ	ARG	CD, NH1, NH2
M15μ	MET	CG, CE
S16σ	SER	-
R17ρ	ARG	CD, NH1, NH2
S19σ	SER	-
V22ϖ	VAL	-
Y23ψ	TYR	CG, CE1, CE2
Y24ψ	TYR	CG, CE1, CE2
F25ϕ	PHE	CG, CE1, CE2
N26ν	ASN	CG
H27η	HISA/HISB	CG, CD2, CE1
N30ν	ASN	CG
A31α	ALA	-
S32σ	SER	-
Q33θ	GLN	CB, CG, CD
W34ω	TRP	CD1, CE3, CZ2

[a] Naming convention of residue types (protonation states) and atoms according to the GROMOS force field.

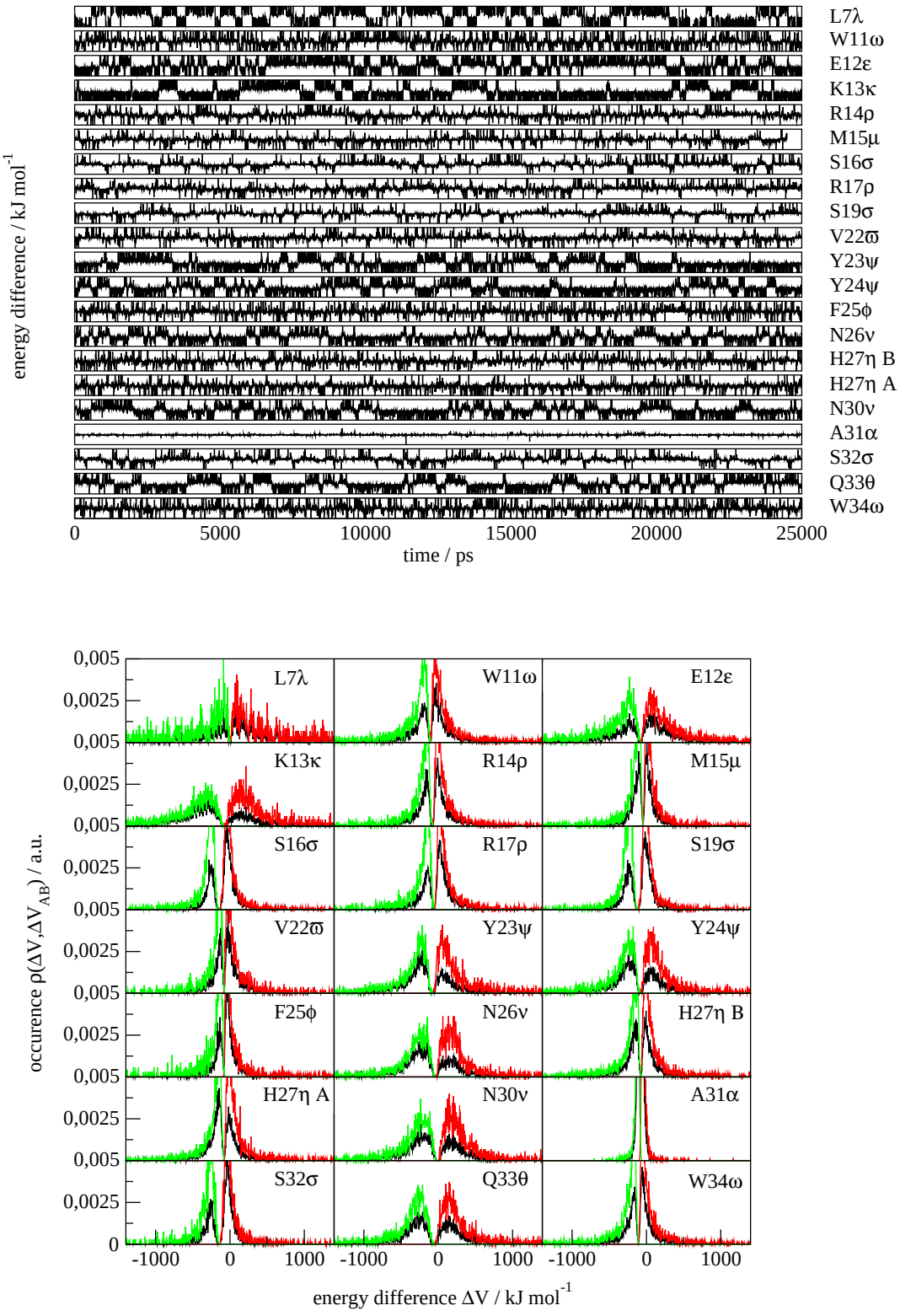


Figure S1: Upper panel: Time series of the potential energy differences, sampled from the EDS reference state simulations in case of tripeptides. The y-axes for all rows cover a range from -2000 to +2000 kJ mol^{-1}. For improved sampling, distance restraints between atoms of the two non-interacting copies of the side chains (dual topology approach) were applied, as specified in Tab. S1. Lower panel: Corresponding energy difference distributions for the reference state (black), the amide (green) and the ester state (red). Distributions of the two end states were obtained from reweighting of the reference state distribution.[3]

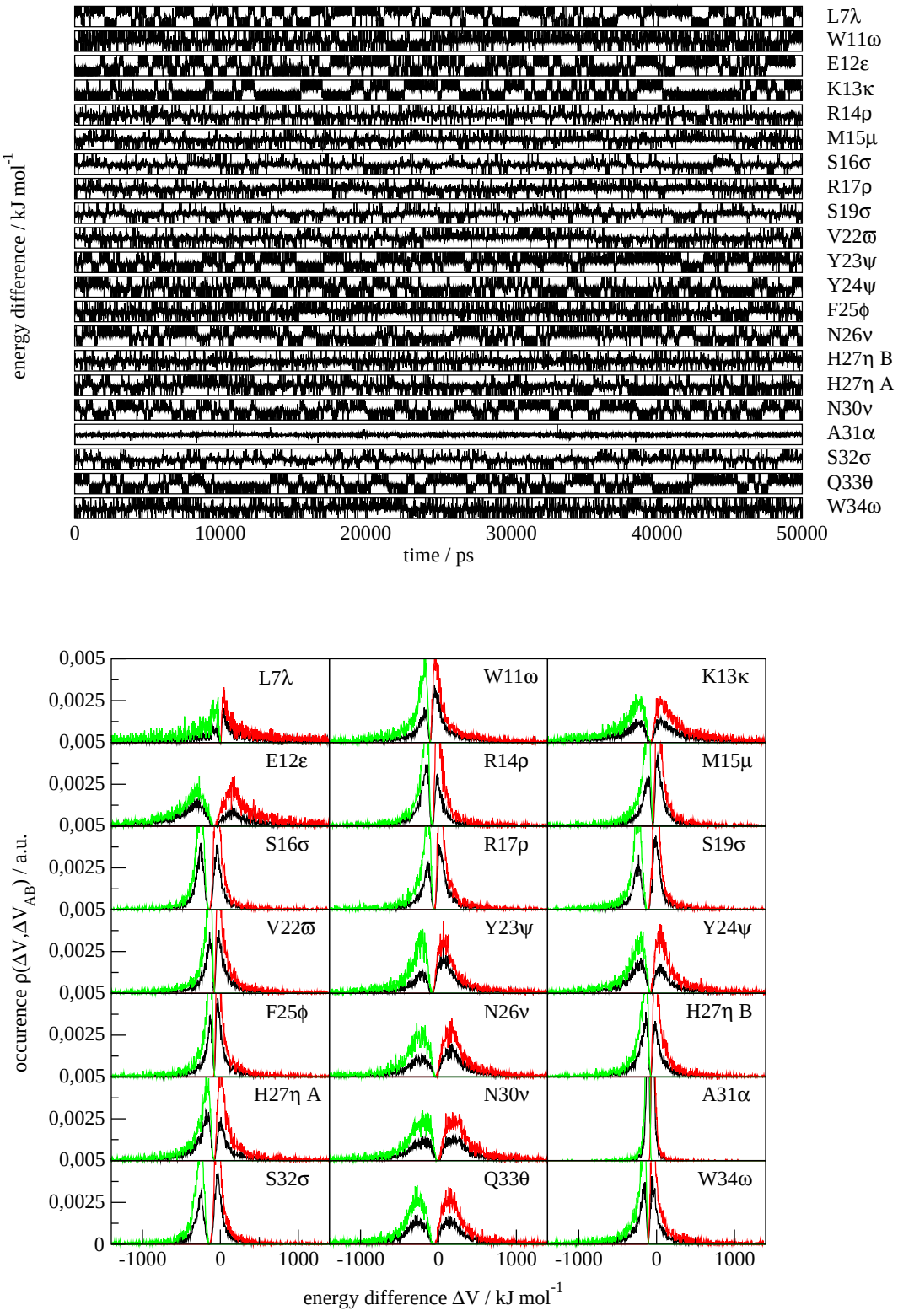


Figure S2: Upper panel: Time series of the potential energy differences, sampled from the EDS reference state simulations in case of pentapeptides. The y-axes for all rows cover a range from -2000 to +2000 kJ mol^{-1}. For improved sampling, distance restraints between atoms of the two non-interacting copies of the side chains (dual topology approach) were applied, as specified in Tab. S1. Lower panel: Corresponding energy difference distributions for the reference state (black), the amide (green) and the ester state (red). Distributions of the two end states were obtained from reweighting of the reference state distribution.[3]

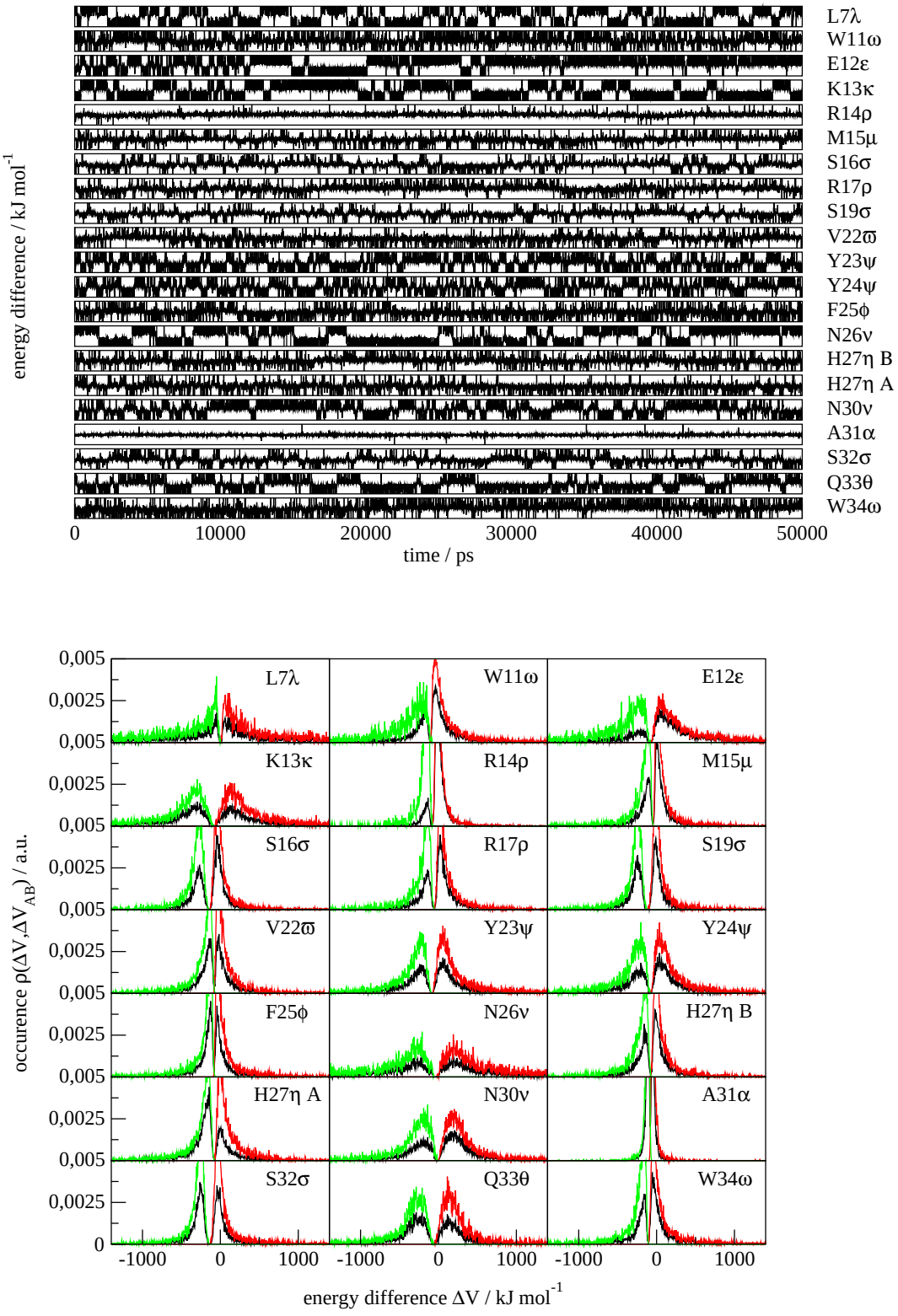


Figure S3: Upper panel: Time series of the potential energy differences, sampled from the EDS reference state simulations in case of heptapeptides. The y-axes for all rows cover a range from -2000 to +2000 kJ mol^{-1}. For improved sampling, distance restraints between atoms of the two non-interacting copies of the side chains (dual topology approach) were applied, as specified in Tab. S1. Lower panel: Corresponding energy difference distributions for the reference state (black), the amide (green) and the ester state (red). Distributions of the two end states were obtained from reweighting of the reference state distribution.[3]

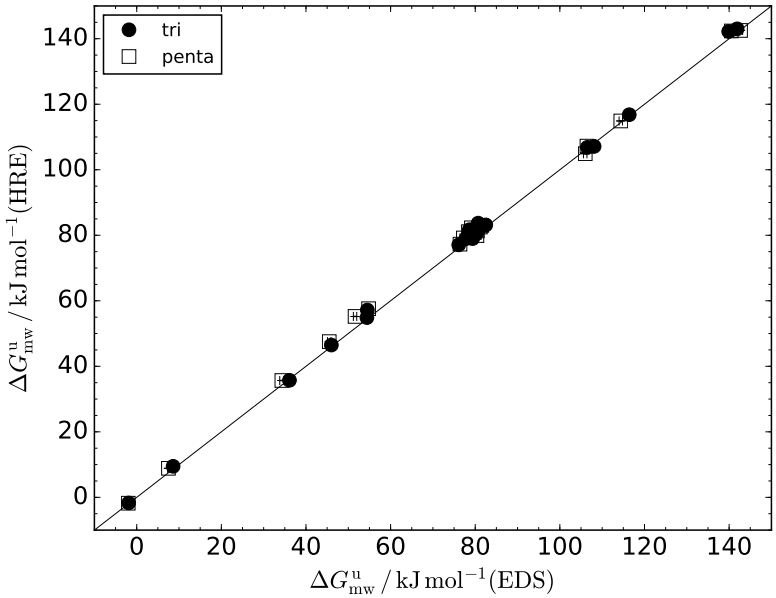


Figure S4: Agreement between alchemical free-energy differences for the unfolded states ($\Delta G^{\mathrm{u}}_{\mathrm{mw}}$), as obtained from EDS simulations and the combination of stratification and Hamiltonian replica exchange (HRE) simulations as used in the present study. The two data sets correspond to different peptide-lengths as used to approximate the unfolded state (tri-, pentapeptides). The solid line is intended as guide to the eye along $\Delta G^{\mathrm{u}}_{\mathrm{mw}}(\mathrm{EDS}) = \Delta G^{\mathrm{u}}_{\mathrm{mw}}(\mathrm{HRE})$.

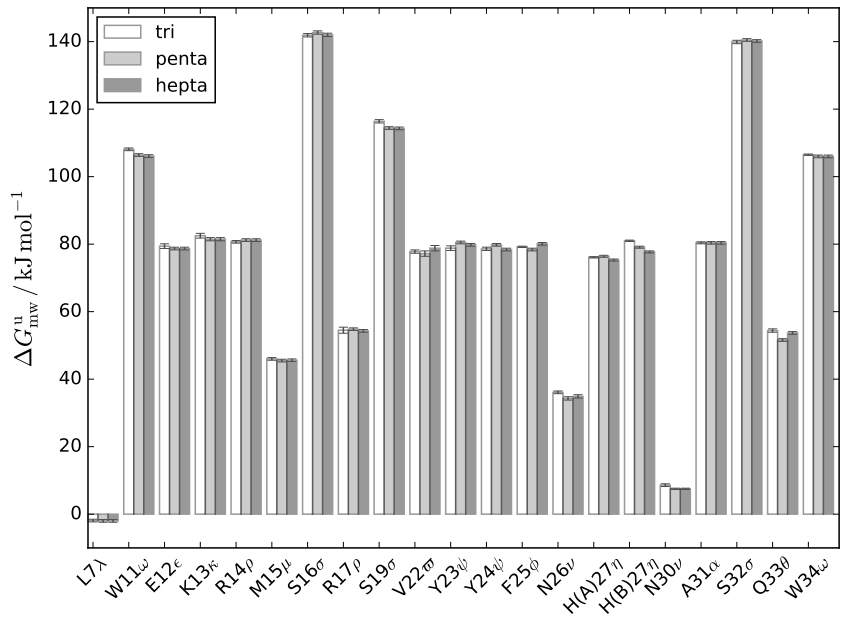


Figure S5: Dependence of the alchemical free-energy change ($\Delta G^{\mathrm{u}}_{\mathrm{mw}}$) obtained from EDS simulations on the type of the perturbed residue and on the length of the peptide as used to approximate the unfolded state (tri-, penta-, heptapeptides).

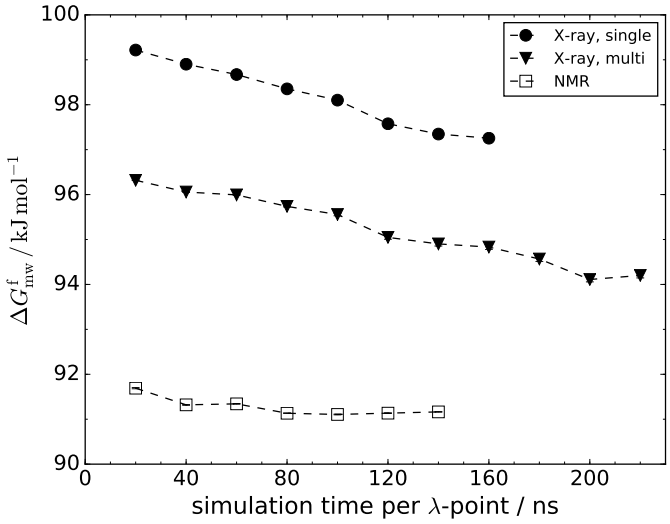


Figure S6: Alchemical free energy difference $\Delta G^{\mathrm{f}}_{\mathrm{mw}}$ of the mutation Y24ψ in the folded state, as function of the simulation time per λ-point. The different data sets correspond to different folded state starting structures (X-ray structure, NMR set). In case of the X-ray structure, further comparison is made between (i) simulations, where all λ-points are initiated from a single structure (single) and (ii) simulations, where each λ-point is initiated with a slightly perturbed structure from a synthetically generated conformational set (multi).

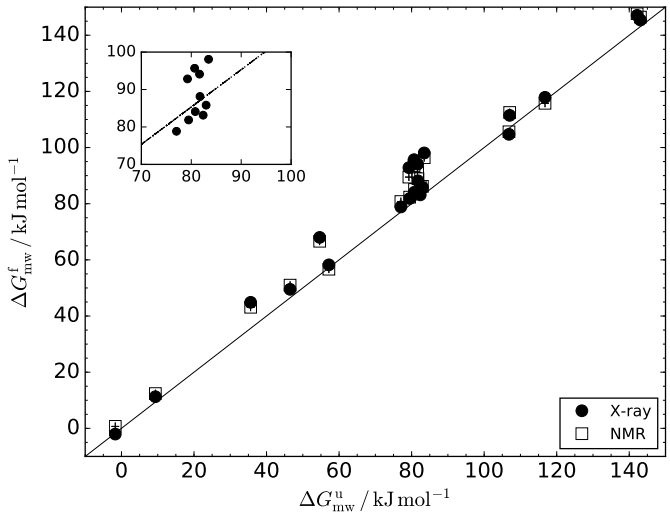


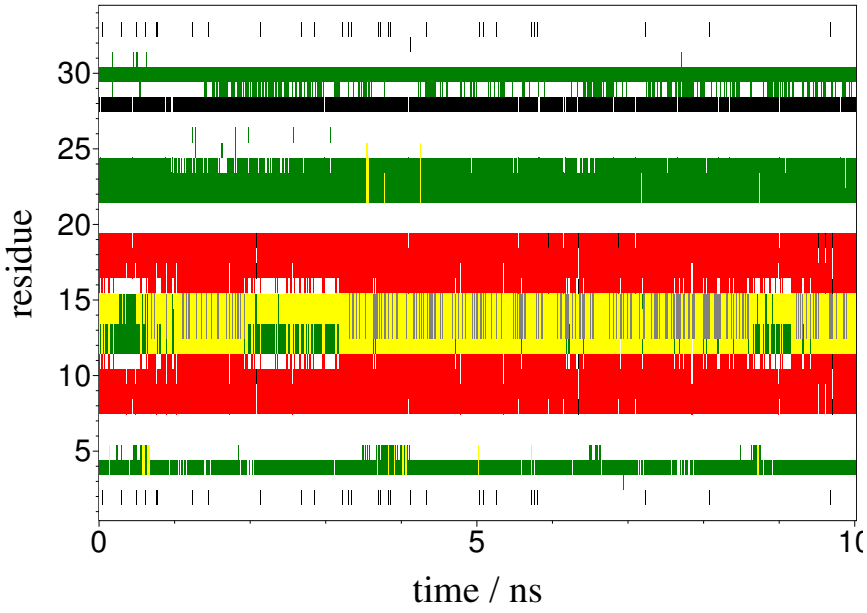
Figure S7: Correlation between alchemical free energy differences in the unfolded stated ($\Delta G^{\mathrm{u}}_{\mathrm{mw}}$ from tripeptide simulations) and the folded state ($\Delta G^{\mathrm{f}}_{\mathrm{mw}}$). The two data sets correspond to simulations based on different folded state starting structures (X-ray, NMR). The solid line is intended as guide to the eye along $\Delta G^{\mathrm{u}}_{\mathrm{mw}} = \Delta G^{\mathrm{f}}_{\mathrm{mw}}$. The inset shows a closer look at the highly populated medium-free energy region in case of the X-ray data set. The dashed line corresponds to a linear-least squares fit to the complete X-ray data set.

Table S2: Results from a thermodynamic analysis, conducted for tripeptides of a selected subset of A-to-E mutations, in terms of enthalpic ($\Delta H^{\mathrm{u}}_{\mathrm{mw}}$), entropic ($T_0\,\Delta S^{\mathrm{u}}_{\mathrm{mw}}$ with $T_0 = 278$ K) and free energy differences ($\Delta G^{\mathrm{u}}_{\mathrm{mw}}$).

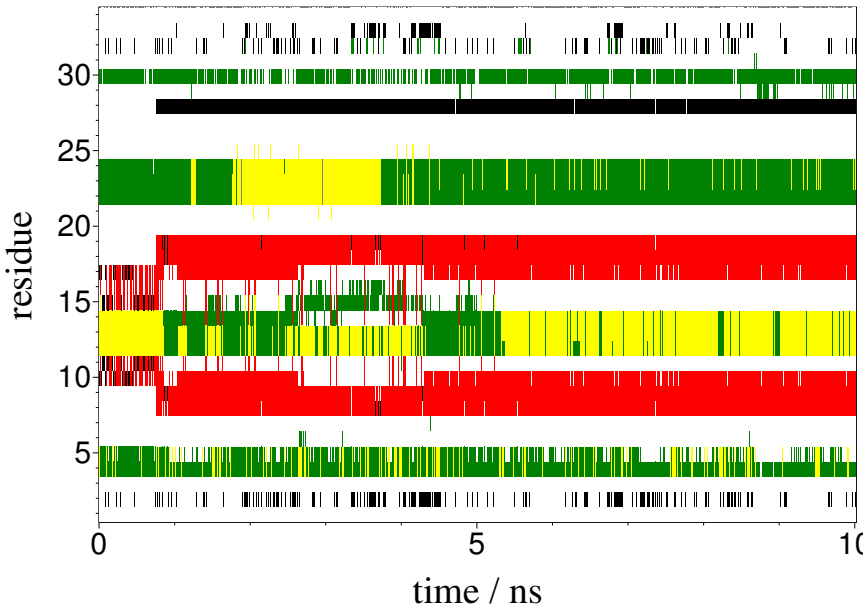
Mutation	$\Delta G^{\mathrm{u}}_{\mathrm{mw}}$ [kJ mol^{-1}]				$\Delta H^{\mathrm{u}}_{\mathrm{mw}}$	$T_0\Delta S^{\mathrm{u}}_{\mathrm{mw}}$
	278 K	298 K	318 K	338 K	[kJ mol^{-1}]	[kJ mol^{-1}]
W11ω	107.1	106.9	106.9	106.6	109.2	2.1
E12ϵ	79.0	79.5	78.9	78.6	81.0	2.0
N26ν	35.7	35.5	35.1	34.7	40.4	4.7
N30ν	9.5	9.4	9.0	9.0	12.0	2.5
Y23ψ	81.6	81.6	81.2	80.9	85.1	3.6
Y24ψ	81.7	80.7	80.6	80.3	87.7	6.0
S19σ	116.8	116.6	116.3	116.3	119.2	2.4
S32σ	142.2	141.8	142.0	141.8	143.7	1.4

Table S3: Occurrence (%) of the 11 hydrogen bonds (see Fig. 1 of the main text) for the protein wild-type within 10 ns MD simulations (isobaric, isothermal), initiated with the X-ray structure[4] and all 20 conformers (C1 to C20) of the NMR model set.[5]

H-bond No.	Pairing residues D→A	X-ray	NMR																			
			C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
1	W11 → P8	62.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	1.70	0.00	0.00	0.00	0.00	0.00
2	E12 → F25	97.80	95.35	94.40	95.95	96.40	94.70	93.05	93.65	90.05	95.30	95.85	75.96	97.20	90.50	0.00	0.00	91.85	32.98	95.85	73.01	93.15
3	F25 → E12	96.55	97.55	96.20	95.60	95.95	95.55	96.35	95.60	94.40	95.55	96.55	97.75	96.90	98.10	83.36	77.91	92.55	67.17	97.35	94.85	95.60
4	R14 → Y23	97.10	95.75	96.40	96.60	95.85	97.15	96.75	96.05	98.40	94.40	96.00	95.65	92.30	96.70	99.25	92.05	95.70	97.75	96.25	95.40	81.46
5	Y23 → R14	99.15	98.90	98.75	98.75	99.10	98.70	98.05	97.10	95.00	99.40	98.95	99.05	96.40	98.10	96.75	88.91	92.65	96.00	97.95	98.95	56.17
6	S16 → R21	89.56	83.81	72.01	82.01	90.90	89.31	75.41	74.41	26.84	83.81	82.91	77.41	70.06	87.36	89.91	68.27	63.27	88.91	91.10	91.20	26.59
7	S19 → S16	1.65	0.00	9.05	4.10	0.80	2.05	3.30	3.55	14.44	1.70	0.00	0.95	3.50	0.10	0.00	13.14	0.45	4.20	0.00	0.50	0.00
8	N30 → N26	0.10	1.75	7.90	12.99	3.90	0.00	0.00	0.05	0.00	0.00	4.00	0.00	0.00	0.30	0.05	18.74	0.00	0.00	2.45	0.00	2.35
9	N26 → A31	88.36	93.65	95.55	94.55	93.40	89.61	82.46	85.11	88.46	91.05	86.51	86.96	82.81	75.71	0.00	0.00	87.26	28.29	87.06	58.42	88.76
10	Q33 → Y24	96.70	97.55	96.00	93.50	96.80	95.40	97.75	94.55	95.65	93.95	93.35	87.41	97.55	95.40	76.16	76.86	97.40	39.33	95.65	92.65	97.95
11	Y24 → Q33	97.40	93.05	91.60	94.40	95.25	97.25	88.26	85.61	95.10	95.60	95.30	91.95	93.65	87.36	96.40	89.26	94.45	93.10	89.46	94.85	92.70



(a)



(b)

Figure S8: Time evolution of secondary structure elements, obtained from 10 ns MD simulations (isobaric, isothermal) of the protein wild-type, initiated from two conformers of the NMR model set ((a): conformer C14, (b): conformer C15). Every residue is assigned to a particular secondary structure element according to the DSSP algorithm:[6] coil (white), β-sheet (red), β-bridge (black), bend (green), turn (yellow).

References

(1) Christ, C.; van Gunsteren, W. Enveloping Distribution Sampling: A Method to Calculate Free Energy Differences from a Single Simulation. *J. Chem. Phys.* **2007**, *126*, 184110.

(2) Christ, C.; van Gunsteren, W. Multiple Free Energies from a Single Simulation: Extending Enveloping Distribution Sampling to Nonoverlapping Phase-Space Distributions. *J. Chem. Phys.* **2008**, *128*, 174112.

(3) Eichenberger, A.; van Gunsteren, W.; Riniker, S.; von Ziegler, L.; Hansen, N. The Key to Predicting the Stability of Protein Mutants Lies in an Accurate Description and Proper Configurational Sampling of the Folded and Denatured States. *Biochim. Biophys. Acta-General Subjects* **2015**, *1850*, 983–995.

(4) Ranganathan, R.; Lu, K.; Hunter, T.; Noel, J. Structural and Functional Analysis of the Mitotic Rotamase Pin1 Suggests Substrate Recognition is Phosphorylation Dependent. *Cell* **1997**, *89*, 875–886.

(5) Kowalski, J.; Liu, K.; Kelly, J. NMR Solution Structure of the Isolated Apo Pin1 WW Domain: Comparison to the X-Ray Crystal Structures of Pin1. *Biopolymers* **2002**, *63*, 111–121.

(6) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers: Original Research on Biomolecules* **1983**, *22*, 2577–2637.