Supporting Information

Controlling Protein Surface Orientation by Strategic Placement of Oligo-Histidine Tags

Dorothee Wasserberg, Jordi Cabanas-Danés, Jord Prangsma, Shane O'Mahony, Pierre-Andre Cazade, Eldrich Tromp, Christian Blum, Damien Thompson, Jurriaan Huskens, Vinod Subramaniam, Pascal Jonkheijm.

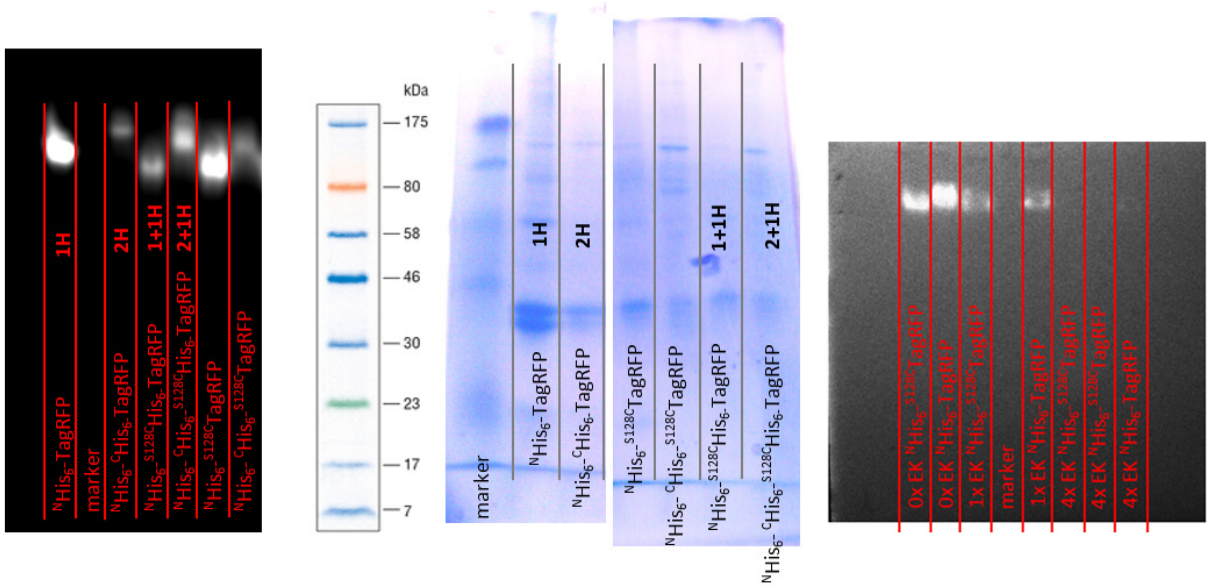


Figure S1. (Left) Fluorescence image (λ_{ex} = 535 nm, λ_{em} = 600 nm) of a native PAGE-gel loaded with relevant mutants and variants of TagRFP as indicated. (Middle) Color image of an SDS-PAGE-gel loaded with relevant mutants and variants of TagRFP as indicated, included are the bands of the prestained protein ladder used (marker). (Right) Bioluminescence (luminol) image of a Western blot from a PAGE-gel, loaded with digestion products of enterokinase (EK) digests of different TagRFP variants (as indicated), stained with primary mouse anti-His$_6$ antibody (Ab) and secondary HRP-conjugated anti-mouse Ab. 0xEK, 1xEK and 4xEK denote the equivalents of enterokinase used to cleave the proteins' N-terminal His$_6$-tag at the DDDDK enterokinase recognition site.

Table S1. Spectroscopic and mass spectrometric data for TagRFP variants: absorption and emission maxima in nm, fluorescence lifetimes in ns, calculated masses in g/mol, and masses determined by MALDI-TOF MS in m/z.

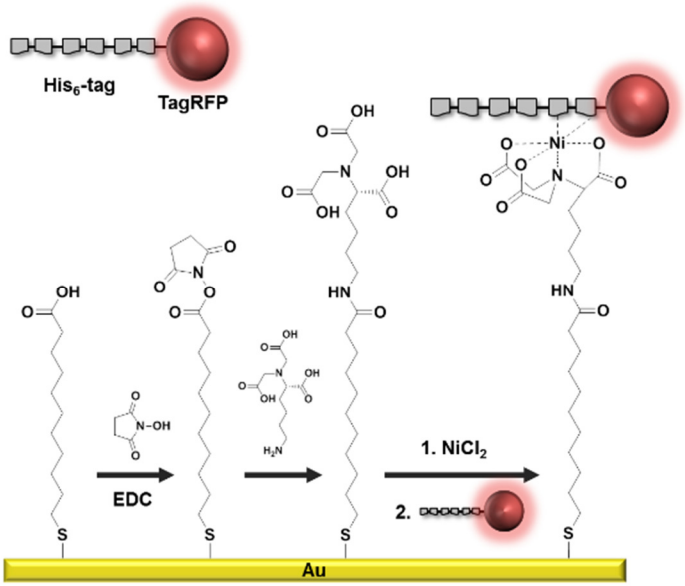
protein variant	absorption maximum (nm)	emission maximum (nm)	fluorescence lifetime (ns)	mass calculated (g/mol)	mass measured (m/z)
$^{N}His_6$-^{wt}TagRFP	555	582	2.4	30309	30350
$^{N}His_6$-TagRFP (**1H**)	555	579	2.4	30277	30000
$^{N}His_6$-$^{C}His_6$-TagRFP (**2H**)	555	582	2.4	31100	30400
$^{N}His_6$-$^{S128C}His_6$-TagRFP (**1+1H**)	555	585	2.4	31116	30800
$^{N}His_6$-$^{C}His_6$-$^{S128C}His_6$-TagRFP (**2+1H**)	-	-	-	32164	31700

Table S2. Fitted parameters of the mono-exponential fit to the anisotropy decay in time from integrated fluorescence polarization lifetime images: lifetime in ns and sum of the peak intensities of the parallel and perpendicular time traces in 10^4 cts.

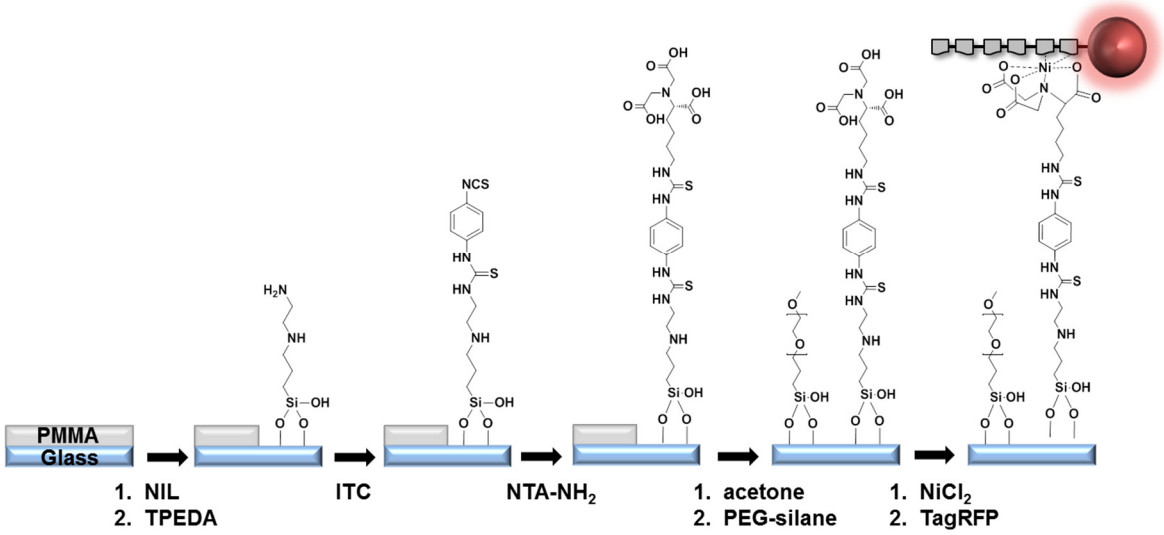
Data set	Lifetime (ns)	Peak Intensities (10^4 cts)
2H sample 1	0.84	12
2H sample 2	6.5	14
2H sample 3	1.1	9.3
2H sample 4	0.86	15
2+1H sample 5	25.5	1.34
2+1H sample 6	20.9	1.6
2+1H sample 7	17.9	1.7
2+1H sample 8	17.1	2.98
2+1H sample 9	43.4	3.09
2+1H sample 10	17.6	1.68
2+1H sample 11	34.9	1.58

Table S3. Dimensionless experimental and expected (calculated) ratios of relative abundances (averages of fits to three separate scans) from XPS of sulfur to carbon (S/C) and where present sulfur to nitrogen (S/N) in monolayers on gold covered glass substrates.

XPS samples		S/C	S/N
1. MUA	experimental	0.050	n.a.
	calculated	0.090	n.a.
2. NHS-activated MUA	experimental	0.067	1.62
	calculated	0.056	1.00
3. NTA-NH_2-functionalized MUA	experimental	0.051	2.70
	calculated	0.047	0.50
4. Ni^{2+}:NTA-functionalized MUA	experimental	0.040	10.7
	calculated	0.047	0.50



Scheme S1. Schematic representation of the functionalization of gold substrates with Ni^{2+}:NTA for protein immobilization via a 11-mercaptoundecanoic acid (MUA) self-assembled monolayer (SAM) on gold, activation with NHS/EDC, subsequent immobilization of N_α,N_α-bis(carboxymethyl)-L-lysine (NTA-NH_2), complexation of Ni^{2+} and immobilization of His_6-tagged proteins.



Scheme S2. Schematic representation of the patterning procedure using nanoimprint lithography (NIL) to generate Ni^{2+}:NTA/PEG-line-patterns on poly(methyl methacrylate) (PMMA)-spincoated glass substrates. After NIL and removal of the residual PMMA layer the exposed areas of the glass substrate are functionalized with N-[3-(trimethoxysilyl)propyl]ethylenediamine (TPEDA, step 1) and the resulting amine-terminated monolayer reacted with p-phenylene diisothiocyanate (ITC, step 2). The resulting isothiocyanate layer is then funtionalized with N_α,N_α-bis(carboxymethyl)-L-lysine (NTA-NH_2, step 3). Then, the remaining PMMA is stripped and the exposed areas of bare glass are backfilled with 2-[methoxy(polyethyleneoxy)propyl]trimethoxysilane (PEG-silane, step 4) to prevent non-specific protein interactions with the bare substrate. Finally, NTA-terminated areas are complexed with Ni^{2+} and His_6-tagged proteins are immobilized (step 5).

As an example, the following equilibrium (Equation S1) and mass balances (Equation S2, S3) were used for a doubly His_6-tagged TagRFP variant (**2H**):

$$\mathbf{2H} + \text{NiNTA} \underset{K_1}{\rightleftharpoons} \mathbf{2H}\cdot\text{NiNTA} + \text{NiNTA} \underset{K_2}{\rightleftharpoons} \mathbf{2H}\cdot(\text{NiNTA})_2 \qquad \text{(Equation S1)}$$

$$\theta_{\text{NiNTAtot}}/\theta_{\text{tot}} = \theta_{\text{NiNTAfree}}/\theta_{\text{tot}} + \theta_{\mathbf{2H}\cdot\text{NiNTA}}/\theta_{\text{tot}} + \theta_{\mathbf{2H}\cdot(\text{NiNTA})2}/\theta_{\text{tot}} \qquad \text{(Equation S2)}$$

$$[\mathbf{2H}]_{\text{tot}} = [\mathbf{2H}]_{\text{free}} + [\mathbf{2H}]_{\text{tot}}\,\theta_{\mathbf{2H}\cdot\text{NiNTA}}/\theta_{\text{tot}} + [\mathbf{2H}]_{\text{tot}}\,\theta_{\mathbf{2H}\cdot(\text{NiNTA})2}/\theta_{\text{tot}} \qquad \text{(Equation S3)}$$

where Θ is the absolute surface coverage in mol/unit area, while species in solution are given as concentrations in mol/L and with the corresponding equilibrium constants as follows:

$$K_1 = \frac{\theta_{\mathbf{2H}\cdot\text{NiNTA}}/\theta_{\text{tot}}}{[\mathbf{2H}]\cdot\theta_{\text{NiNTA}}/\theta_{\text{tot}}} = 2K_{i,His_6} \qquad \text{(Equation S4)}$$

$$K_2 = \frac{\theta_{\mathbf{2H}\cdot(\text{NiNTA})_2}/\theta_{\text{tot}}}{\theta_{\mathbf{2H}\cdot\text{NiNTA}}/\theta_{\text{tot}}\cdot\theta_{\text{NiNTA}}/\theta_{\text{tot}}} = \frac{1}{2}C_{eff}K_{i,His_6} \qquad \text{(Equation S5)}$$

Substituting Equations S4, S5 into Equations S2, S3 yields a set of numerically solvable equations with $\theta_{\text{Ni(II)}\cdot\text{NTAtot}}/\theta_{\text{tot}}$ as variable. Since the experiment was carried out under continuous flow conditions, [**2H**] is assumed to be constant and equal to $[\mathbf{2H}]_{\text{tot}}$. The equations were solved in a spreadsheet approach.[S4]

Initially, $K_{i,His6}$ and the maximum SPR resonance angle shift ($\Delta\alpha_{max}$) were iteratively optimized to fit the theoretical to the experimentally obtained values of SPR resonance shifts for **1H** binding to the Ni^{2+}:NTA-functionalized surface. This $K_{i,His6}$ value (obtained for **1H**) was used as a fixed parameter for all the other variants during optimization, as it is assumed intrinsic to all His_6-tags, while $\Delta\alpha_{max}$ and C_{eff} were iteratively optimized for the other variants.

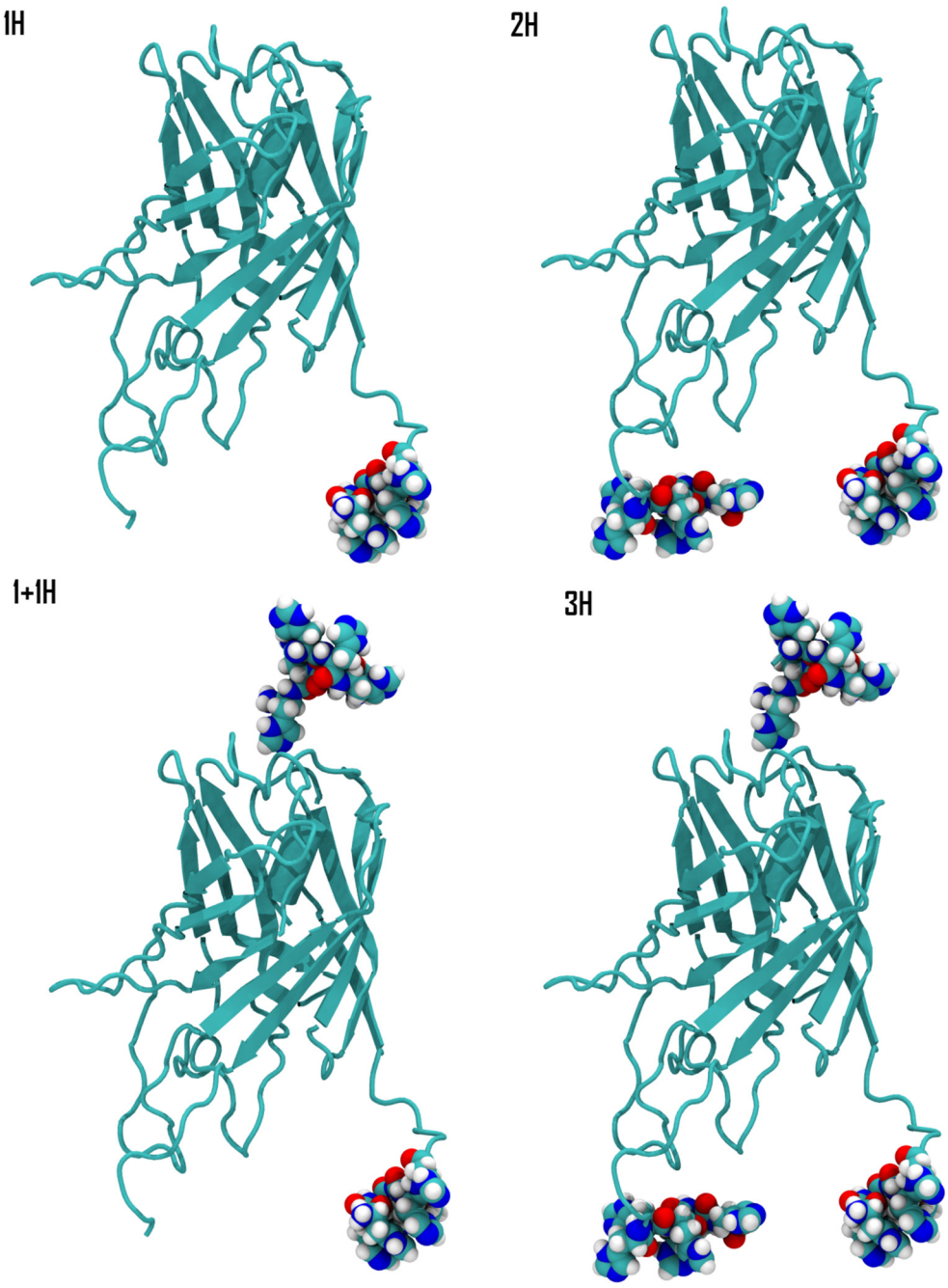


Figure S2. Schematic representation of TagRFP variants with the His_6-tags that adsorb in each model (**3H** is the same as **2+1H**).

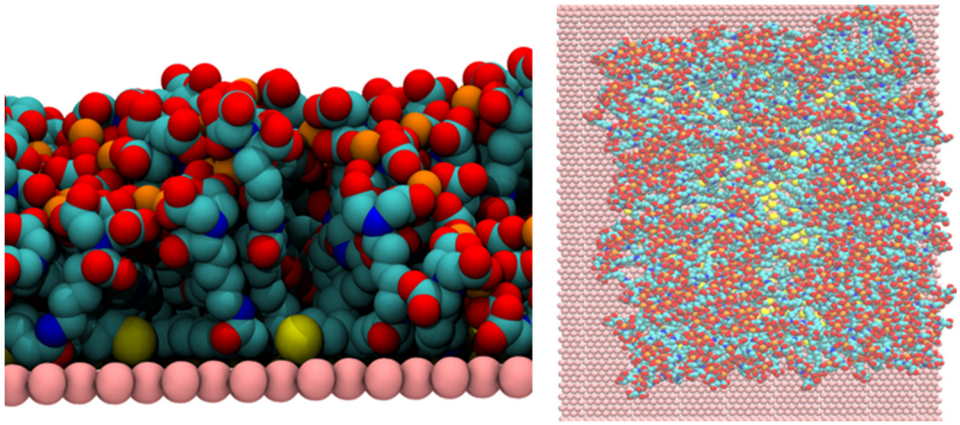


Figure S3. The computed Ni^{2+}:NTA-functionalized SAM viewed from the side (left) and top (right) with atoms shown as space-filling spheres. Ni^{2+} ions are colored in orange, oxygen atoms are red, nitrogen atoms are dark blue, carbon atoms are light blue, sulfur atoms are yellow and gold atoms are pink.

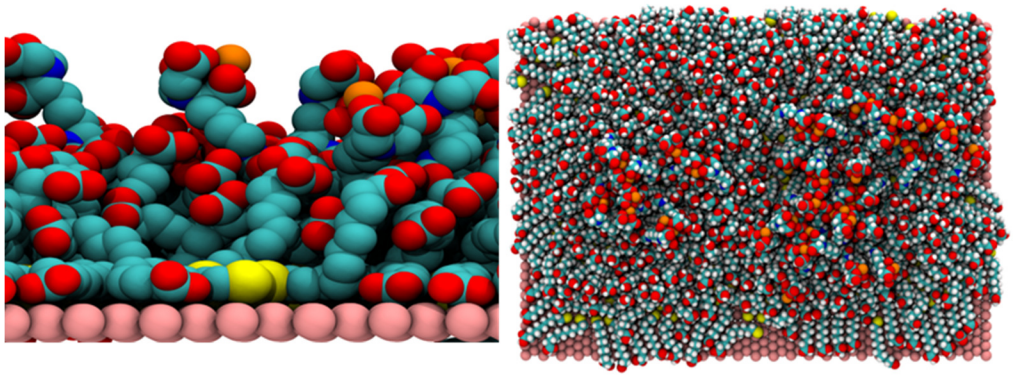


Figure S4. The computer mixed Ni^{2+}:NTA/acid-functionalized SAM viewed from the side (left) and top (right) with atoms shown as space-filling spheres. Ni^{2+} ions are colored in orange, oxygen atoms are red, nitrogen atoms are dark blue, carbon atoms are light blue, sulfur atoms are yellow and gold atoms are pink.

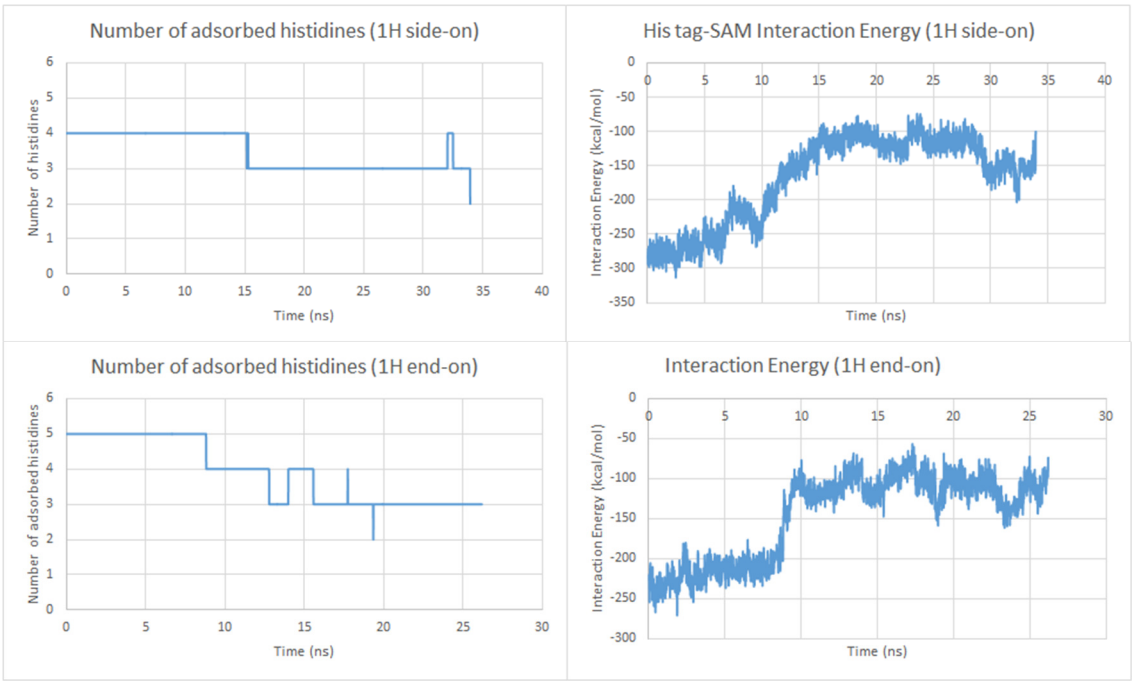


Figure S5. Computed timelines for number of complexed histidine residues and His_6-tag-SAM interaction energy, for **1H** adsorbed on the mixed SAM in side-on and end-on orientations.

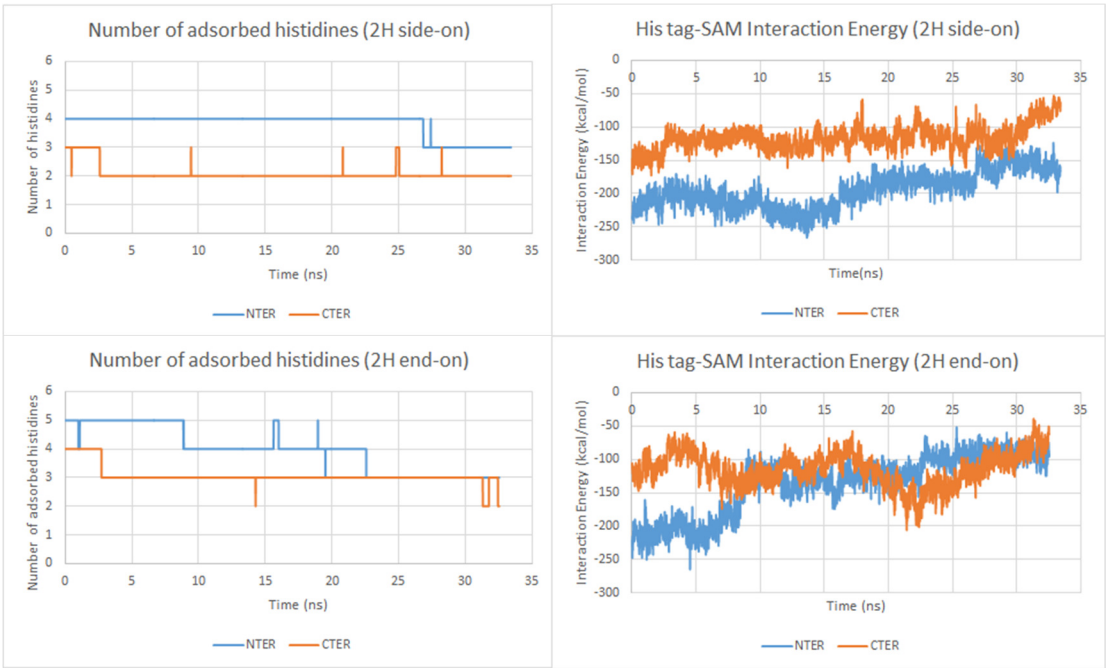


Figure S6. Computed timelines for number of complexed histidine residues and His_6-tag-SAM interaction energy, for **2H** adsorbed on the mixed SAM in side-on and end-on orientations.

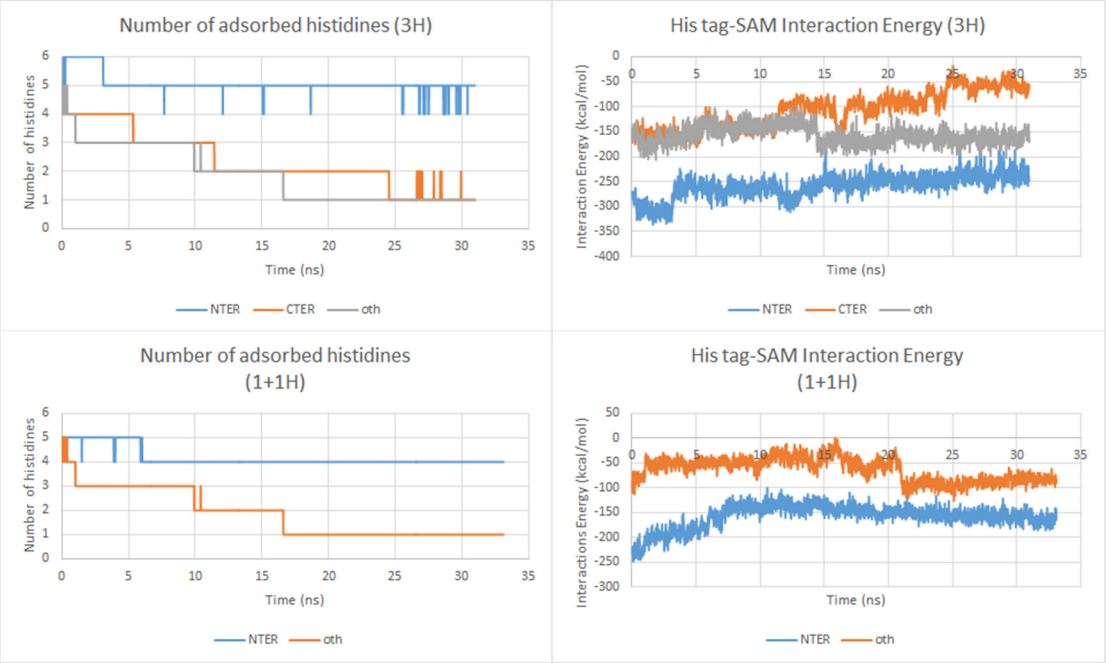


Figure S7. Computed timelines for number of complexed histidine residues and His_6-tag-SAM interaction energy, for **3H** (**2+1H**) and **1+1H** adsorbed on the mixed SAM. Both use His tags at opposite ends of the protein and can only bind in a side-on orientation.

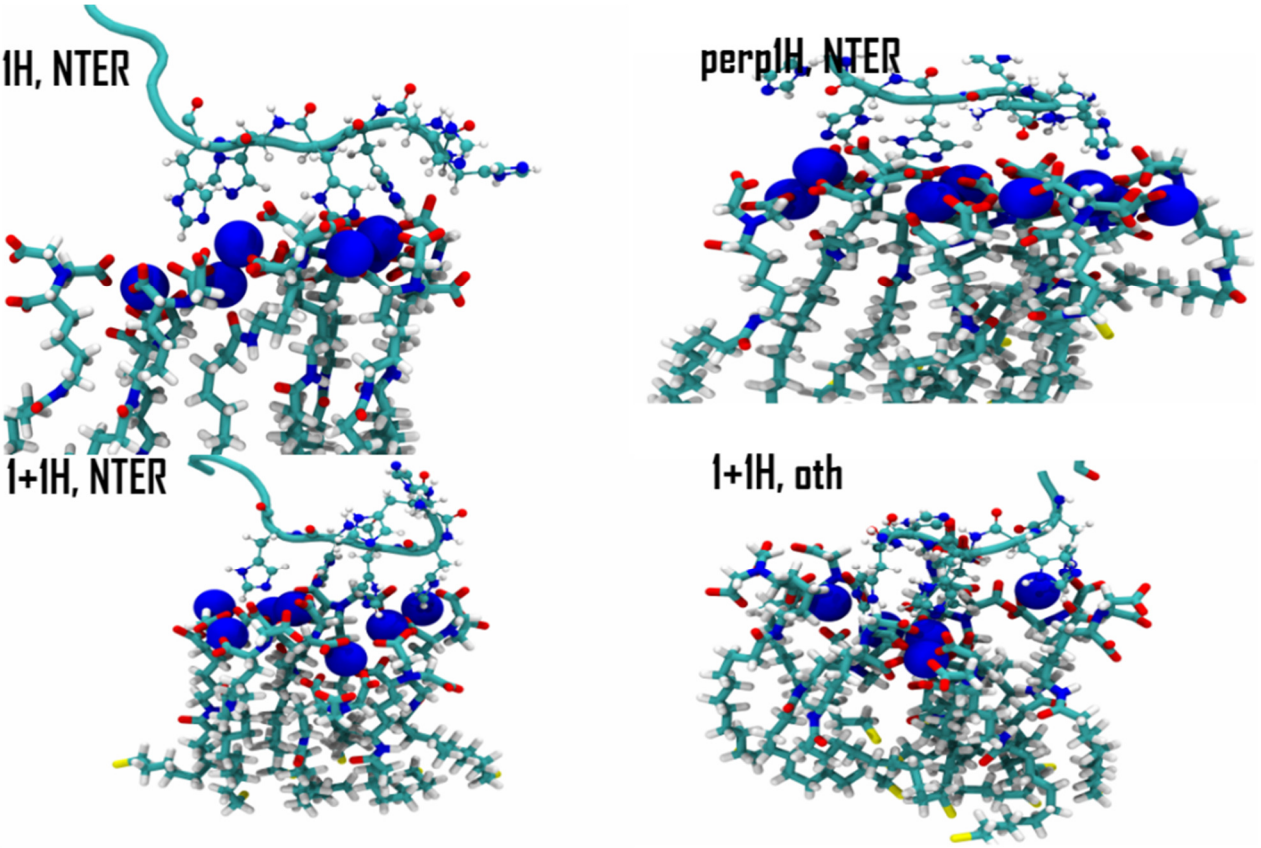


Figure S8. Final snapshots of the **1H** and **1+1H** His_6-tags adsorbing onto the uniform Ni^{2+}:NTA SAM. The SAM is in liquorice representation, the His_6-tag is in CPK representation, Ni^{2+} ions are in space-filling representation. End-on oriented 1H is named "perp1H" in this Figure.

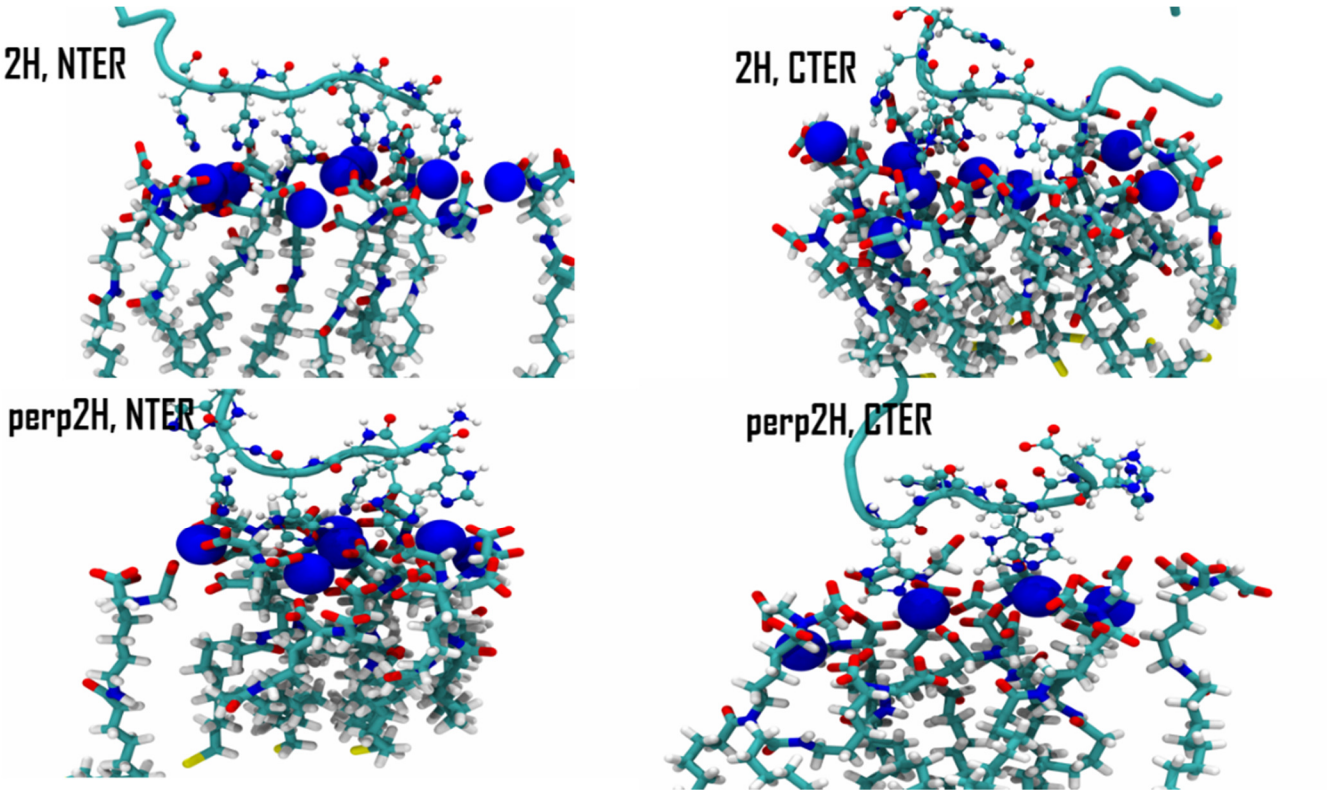


Figure S9. Final snapshots of the **2H** His_6-tags adsorbing onto the uniform Ni^{2+}:NTA SAM. The SAM is in liquorice representation, the His_6-tag is in CPK representation, Ni^{2+} ions are in space-filling representation. End-on oriented 2H is named "perp2H" in this Figure.

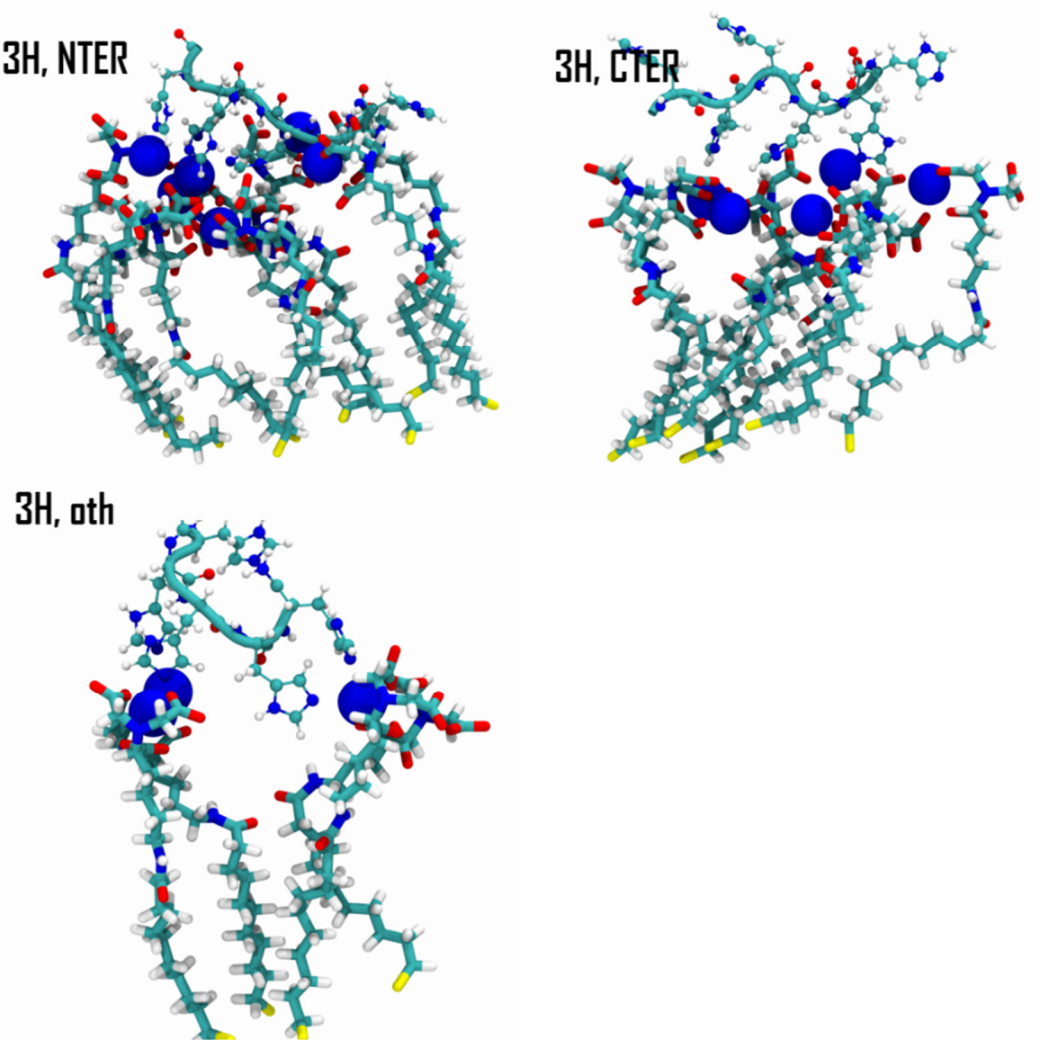


Figure S10. Final snapshots of the **3H** (**2+1H**) His_6-tags adsorbing onto the uniform Ni^{2+}:NTA SAM. The SAM is in liquorice representation, the His_6-tag is in CPK representation, Ni^{2+} ions are space-filling spheres.

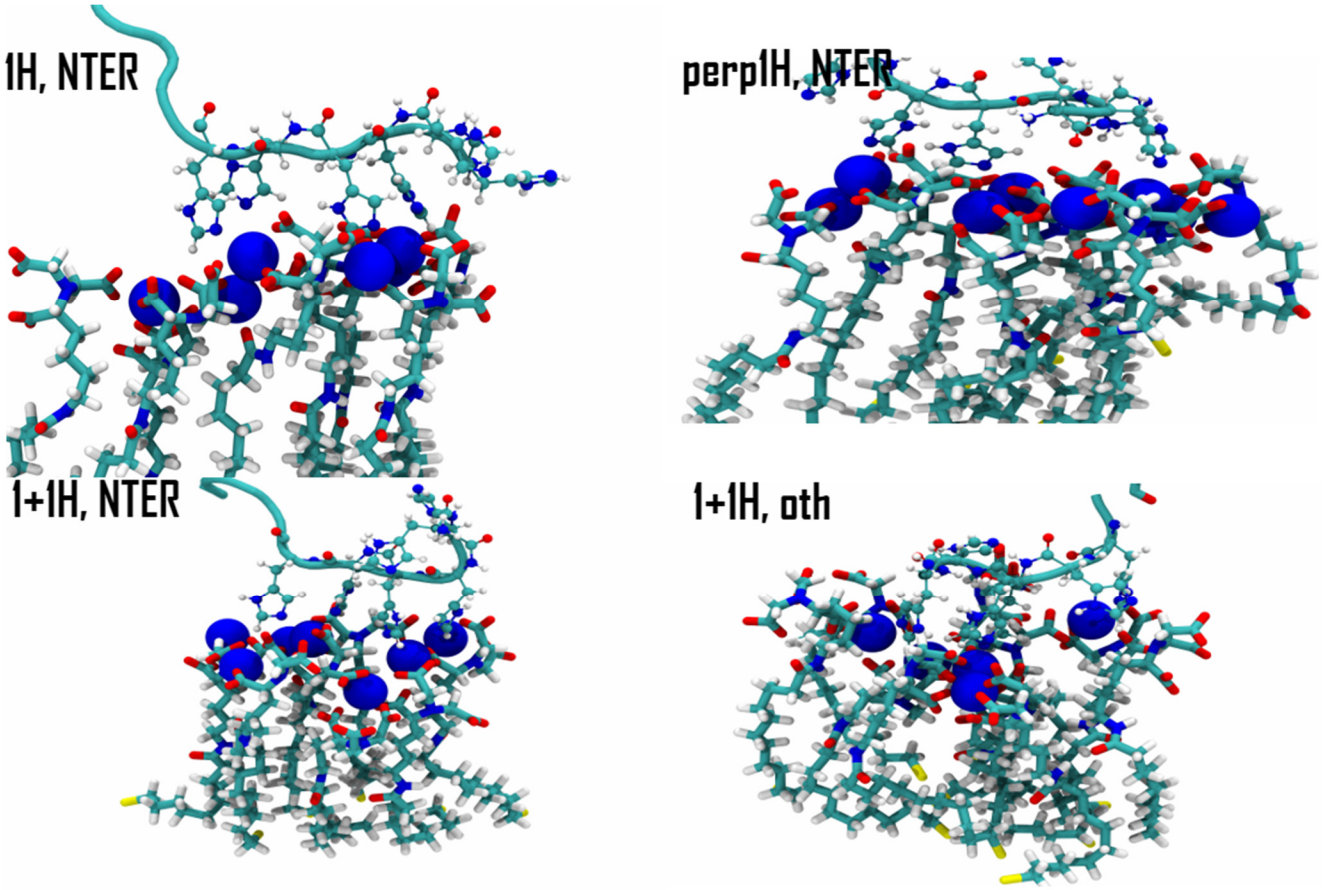


Figure S11. Final snapshots of the **1H** and **1+1H** His_6-tags adsorbing onto the mixed Ni^{2+}:NTA SAM. The SAM is in liquorice representation, the His_6-tag is in CPK representation, Ni^{2+} ions are in space-filling representation. End-on oriented 1H is named "perp1H" in this Figure.

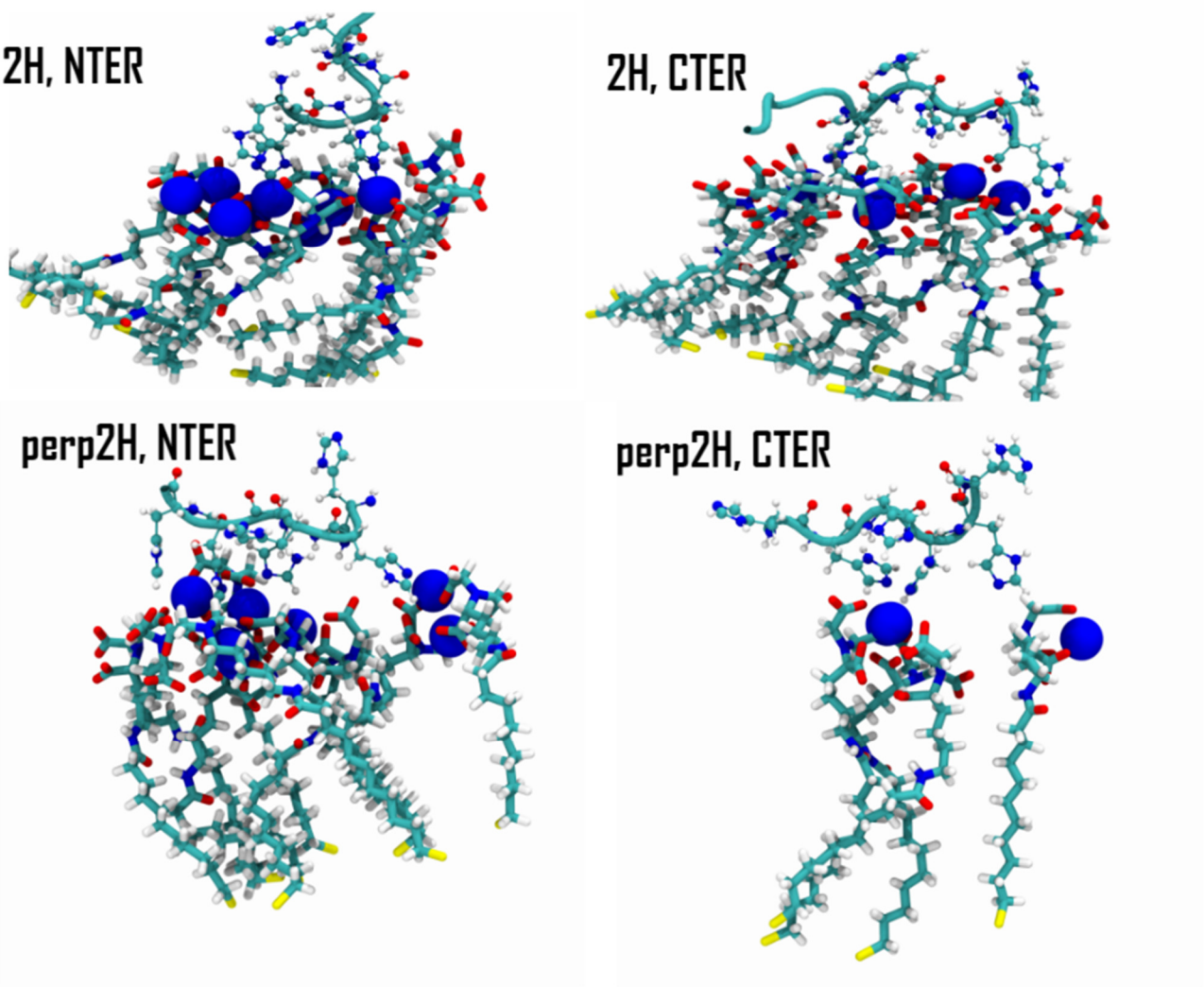


Figure S12. Final snapshots of the **2H** His_6-tags adsorbing onto the mixed Ni^{2+}:NTA SAM. The SAM is in liquorice representation, the His_6-tag is in CPK representation, Ni^{2+} ions are in space-filling representation. End-on oriented 2H is named "perp2H" in this Figure.

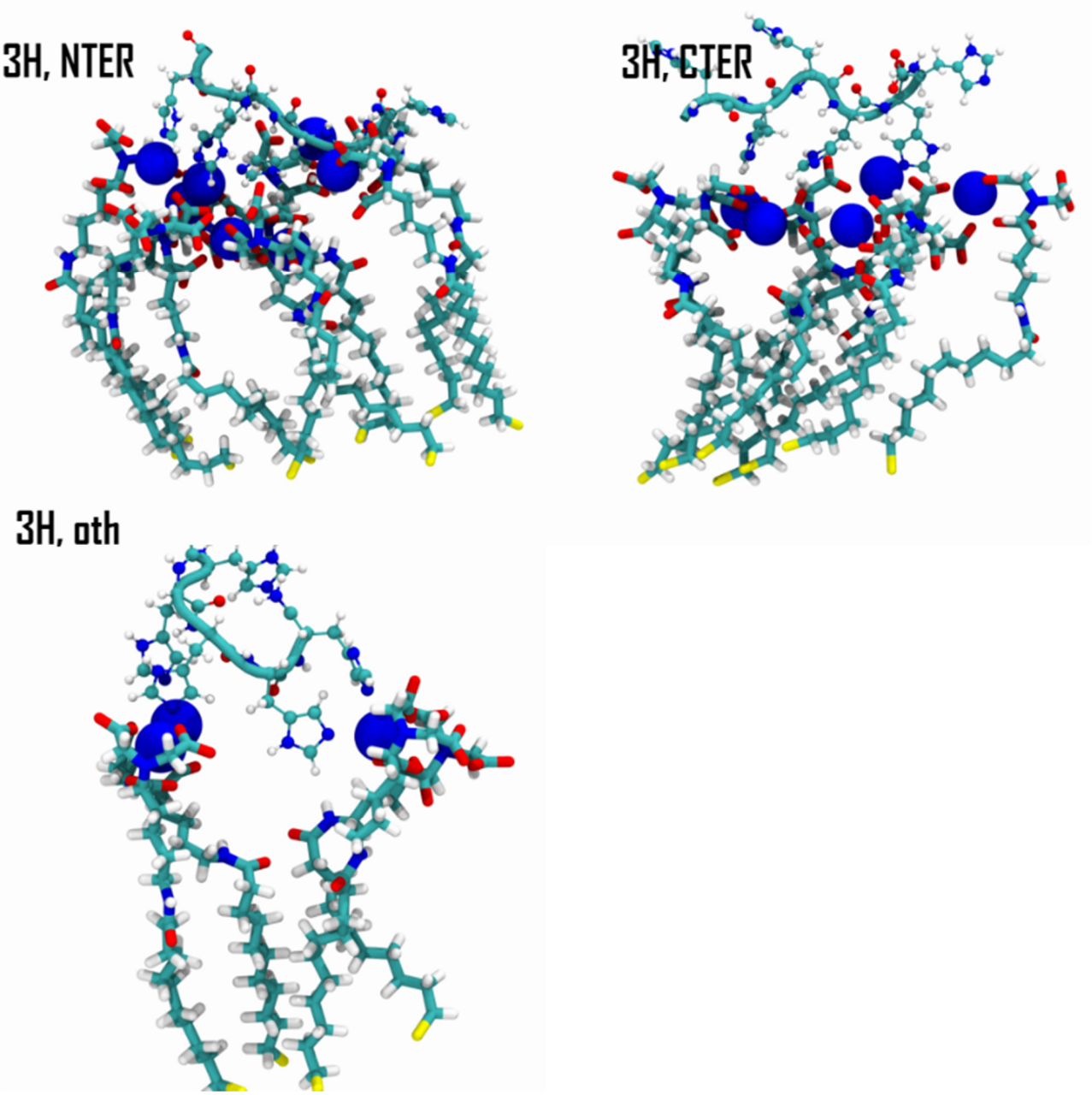


Figure S13. Final snapshots of the **3H** (**2+1H**) His_6-tags adsorbing onto the mixed Ni^{2+}:NTA SAM. The SAM is in liquorice representation, the His_6-tag is in CPK representation, Ni^{2+} ions are space-filling spheres.

Table S4. Calculated interaction energies for the uniform and mixed Ni^{2+}:NTA SAMs.

Variant (orientation)	Protein-SAM van der Waals (kcal/mol)		Protein-SAM Electrostatic (kcal/mol)		Total Protein-SAM (kcal/mol)		Protein-Water (kcal/mol)		Tag-SAM (kcal/mol)		Tag-Water (kcal/mol)		Total Interaction Energy (kcal/mol)	
	Uniform	*Mixed*	*Uniform*	*Mixed*	*Uniform*	*Mixed*	*Uniform*	*Mixed*	*Uniform*	*Mixed*	*Uniform*	*Mixed*	*Uniform*	*Mixed*
1H (side-on)	-44±7	-11±6	-875±48	-187±71	-919±47	-198±73	-6666±211	-6921±206	-201±35	-201±35	-186±28	-186±28	-7972±221	-7506±223
2H (side-on)	-59±8	-12±5	-1118±51	-149±76	-1177±52	-161±74	-6458±172	-7024±204	-402±49	-402±49	-372±40	-372±40	-8409±190	-7959±226
2+1H(side-on)	-62±8	-31±7	-1057±82	-394±42	-1119±80	-425±42	-6814±260	-6527±177	-603±57	-603±57	-570±48	-570±48	-9104±204	-8090±200
1+1H(side-on)	-40±8	-32±7	-673±56	-274±68	-713±57	-306±69	-6519±179	-6913±187	-402±49	-402±49	-372±40	-372±40	-8006±198	-7980±200
1H (end-on)	-0.02±0.01	-0.1±0.1	0.08±0.9	6±11	0.06±0.9	6±10	-7231±252	-6882±167	-201±35	-201±35	-186±28	-186±28	-7618±256	-7263±173
2H (end-on)	-0.3±0.2	-0.9±0.9	3±5	-17±19	3±5	-18±19	-7662±186	-7108±177	-402±49	-402±49	-372±40	-372±40	-8433±197	-7900±189
Ref. His_6 tag	-178±34	-178±34	-23±5	-23±5	-201±35	-201±35	-186±28	-186±28	-201±35	-201±35	-186±28	-186±28	-387±45	-387±45

Table S5. Number of water molecules within 10Å of the protein when bound to the mixed SAM.

Variant	**Solvated Protein**	**Side-on 1H**	**Side-on 2H**	**Side-on 2+1H**	**Side-on 1+1H**	**End-on 1H**	**End-on 2H**
No. of Water Molecules	3720	3350	3550	3400	3370	3570	3700

Table S6. Conformational energies and protein radii of gyration when bound to the uniform SAM (excluding the His_6-tags) and the corresponding values for reference unbound protein.

	Solvated Protein	**Side-on 1H**	**Side-on 2H**	**Side-on 2+1H**	**Side-on 1+1H**	**End-on 1H**	**End-on 2H**
Energy (kcal/mol)	4720±40	4770±50	4800±50	4750±50	4770±50	4750±50	4730±50
Radius of gyration (Å)	18.5±0.1	18.4±0.1	18.3±0.1	18.4±0.1	18.3±0.1	18.1±0.1	18.9±0.1

Table S7. Conformational energies and protein radii of gyration when bound to the mixed SAM (excluding the His_6-tags).

	Side-on 1H	**Side-on 2H**	**Side-on 2+1H**	**Side-on 1+1H**	**End-on 1H**	**End-on 2H**
Energy (kcal/mol)	4770±40	4700±40	4700±50	4690±40	4680±40	4690±40
Radius of gyration (Å)	18.1±0.1	18.3±0.1	18.4±0.0	18.4±0.1	18.2±0.1	18.4±0.1

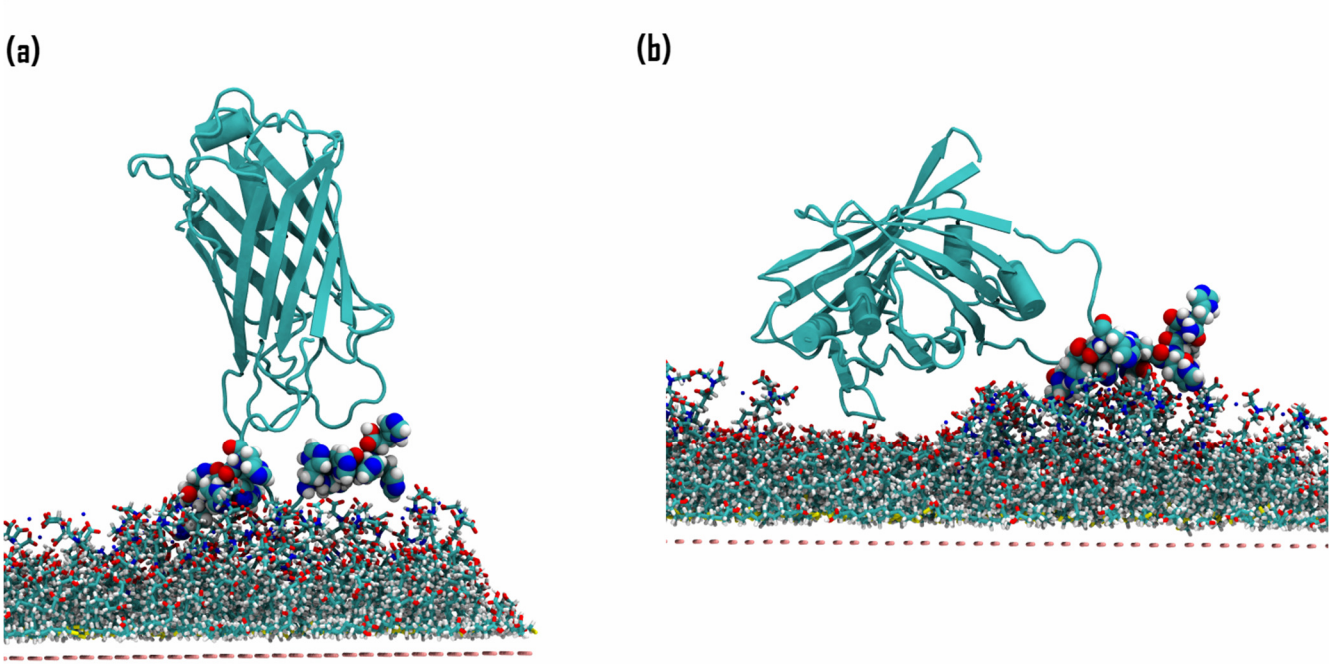
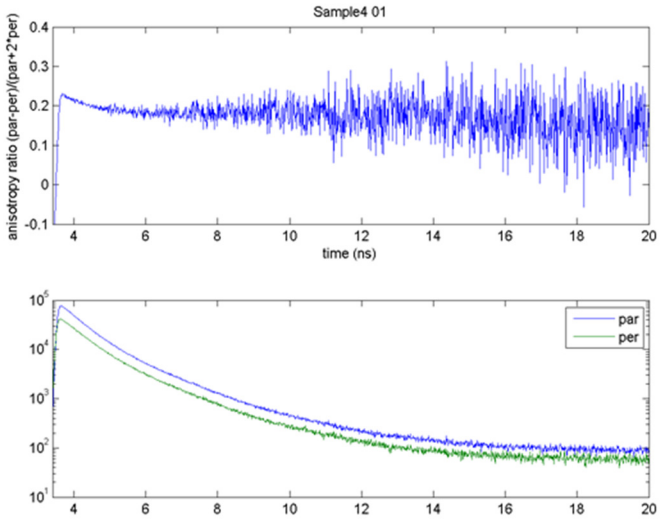


Figure S14. The starting bound configurations used for SMD simulations of **2H** unbinding from (a) end-on and (b) side-on orientations on a mixed SAM.

A) 2H



B) 2+1H

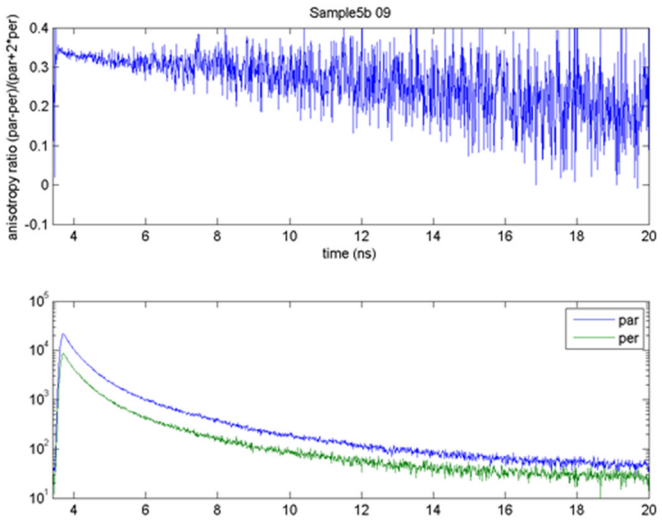


Figure S15. Anisotropy decay (top) and time traces for parallel and perpendicular polarization for immobilized protein **2H** (A) and **2+1H** (B).