

Supplementary Information

Simulating Biological Template to Nanoscale Architecture: Core Concept of Clamp-based Binding-Pocket-Favored N-terminal- Domain Assembly

Lunna Li^{1,2}, Angela M. Belcher^{1,*} and Desmond K. Loke^{2,3*}

*¹Department of Biological Engineering, David H. Koch Institute for Integrative
Cancer Research, Massachusetts Institute of Technology (MIT), Cambridge,
Massachusetts, 02139, United States*

*²Department of Science, Mathematics and Technology, Singapore University of
Technology and Design, Singapore 487372, Singapore*

³Office of Innovation, Changi General Hospital, Singapore, 529889, Singapore

*Authors to whom correspondence should be addressed. Electronic mail: belcher@mit.edu,

desmond_loke@sutd.edu.sg

Supplementary Note 1

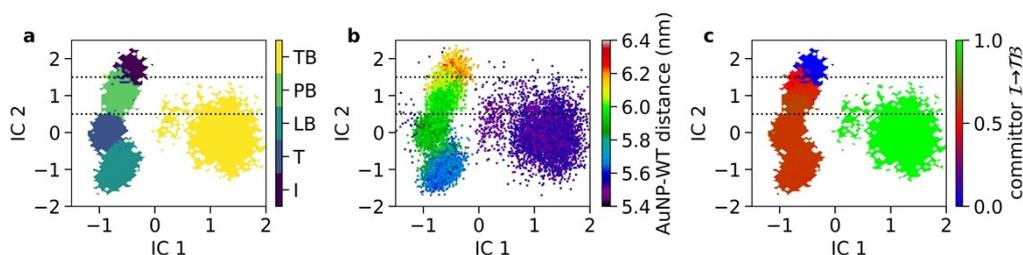


Figure S1 MSM analysis for backbone-torsion angles of BPs #1-3 for AuNP #3 in WT simulations, shown with the first two (most slowly changing) TICA collective variables IC 1 and IC 2, and plotted with **a**) state-space discretization of initial state (I), intermediate (T) state, pre-bound (PB) state, loosely-bound (LB) state and tightly-bound (TB) state; **b**) overall AuNP–M13 distance; and **c**) committor probabilities between the I state and TB state, which may identify the PB region as a possible transition state, characterized by committor values of approximately 0.5.

To probe the conformational dynamics of BPs #1-3, backbone-torsion angles of twenty-four amino acids were analyzed for an aggregate simulation length of approximately seventy nanoseconds using Markov-state-models¹ (MMS). Lag time of one hundred picoseconds was chosen to extract the subspace of twelve most-slowly-changing collective variables in forty-two-dimensional backbone-torsion-angle space. Feature selection, time-lagged independent component analysis² (TICA) dimension reduction and *k*-means clustering algorithm³ were performed using PyEMMA¹; and the estimate of macrostates was obtained using Robust Perron Cluster Analysis⁴ (PCCA++) and transition path theory⁵ (TPT) via PyEMMA. In order to relate BP conformations with observed AuNP–phage binding, we plot a colour map of perpendicular AuNP–M13 distance for WT phage, with space-state discretization at different stages of biotemplating (Figs. S1b). The evidence reveals a close relationship between the dynamics of disordered proteins and gold–M13 binding.

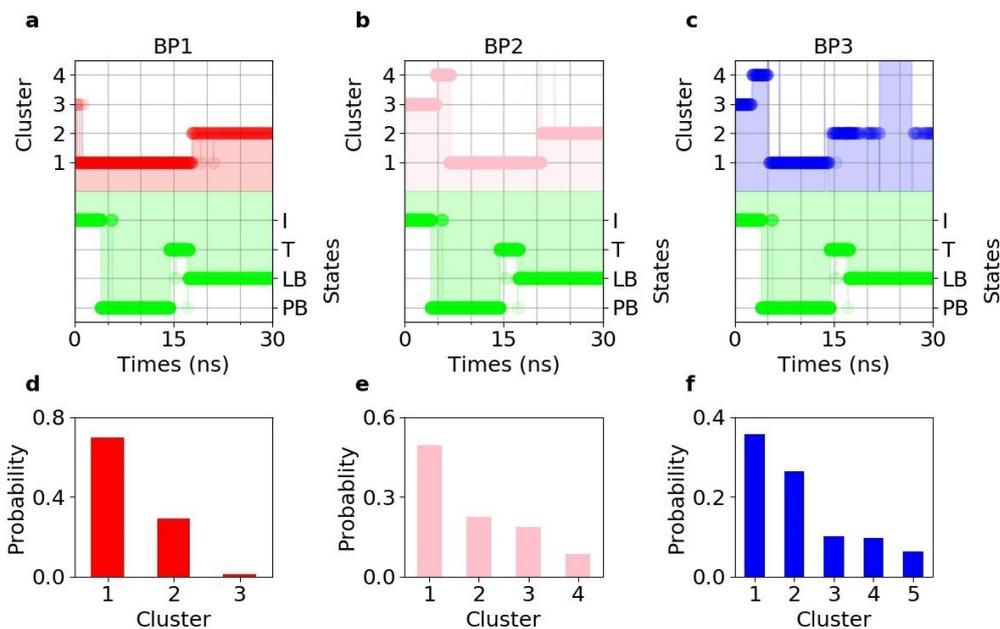


Figure S2 Daura clustering¹ analysis of BPs #1 (red region) **a**), 2 (pink region) **b**), and 3 (blue region) **c**) for a sample-unbiased-MD trajectory of WT simulations, in comparison with *k*-means clustering of the same BPs for the same trajectory (green regions). Cluster population probability of Daura analysis for BPs #1 (red bar) **d**), 2 (pink bar) **e**), and 3 (blue bar) **f**).

Daura clustering⁶ analysis was carried out on BPs #1-3 of AuNP #3 of WT phage for a sample-unbiased-MD trajectory. The conformations were analyzed via the Daura algorithm in GROMACS tools⁷. We used a root-mean-square-deviation (RMSD) cutoff of 0.18 nm for all atoms in BPs. Clustering analysis was performed every 10 frames (approximately one hundred picoseconds). Population probability of each cluster was calculated as ratio of the number of frames that were assigned to the specific cluster to the total number of frames in the trajectory (~500 frames).

Supplementary Note 2

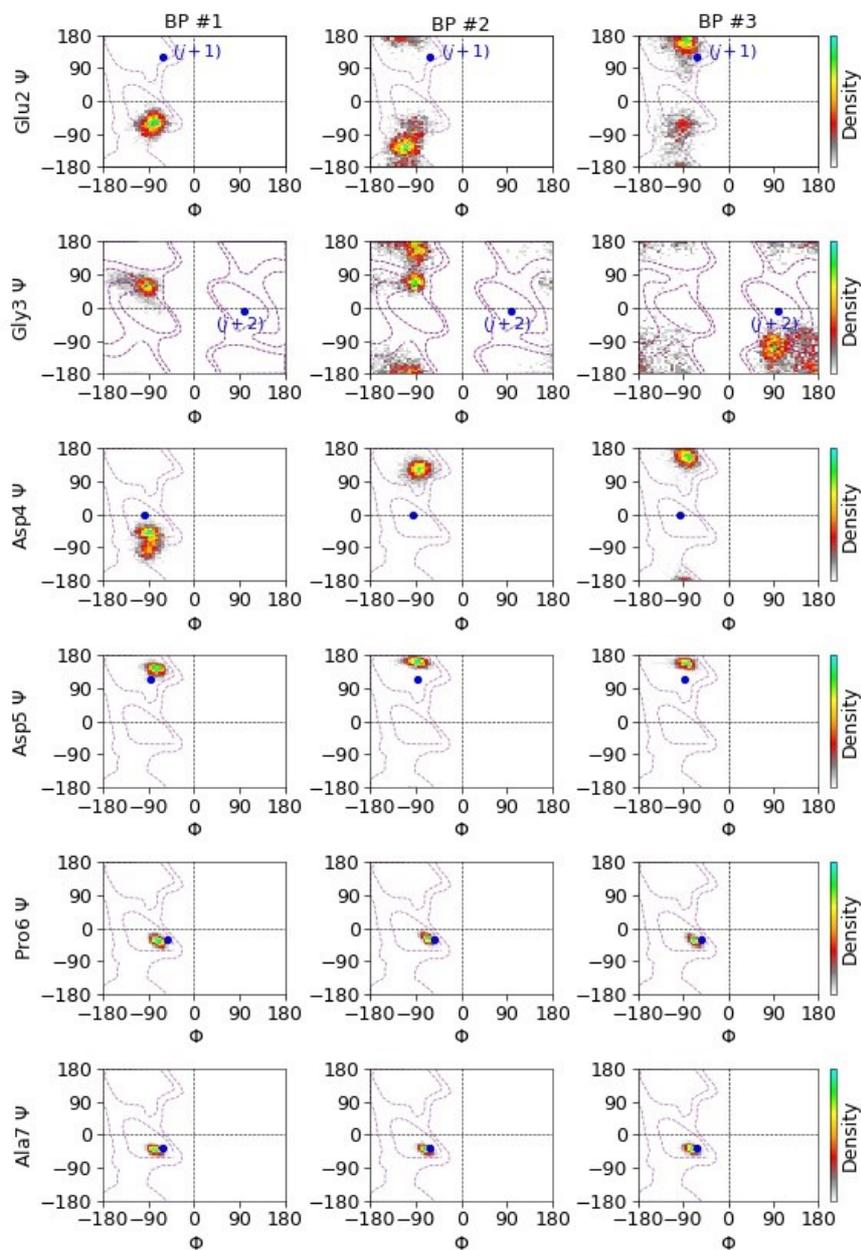


Figure S3 Ramachandran plots of Glu2-Gly3-Asp4-Asp5-Pro6-Ala7 for BPs #1-3 of LB state in WT simulations. Initial Φ and Ψ of all amino acids in the native configuration are highlighted as blue dots, and the residues $j + 1$ (Glu2) and $j + 2$ (Gly3) of type-II β -turn are labelled⁸. General α -helix and β -sheet regions¹⁰ are highlighted (dashed purple lines) for Glu2, Asp4, Asp5, Pro6 and Ala7. Gly dihedral distribution from the literature is shown (dashed purple lines) for Gly3⁹.

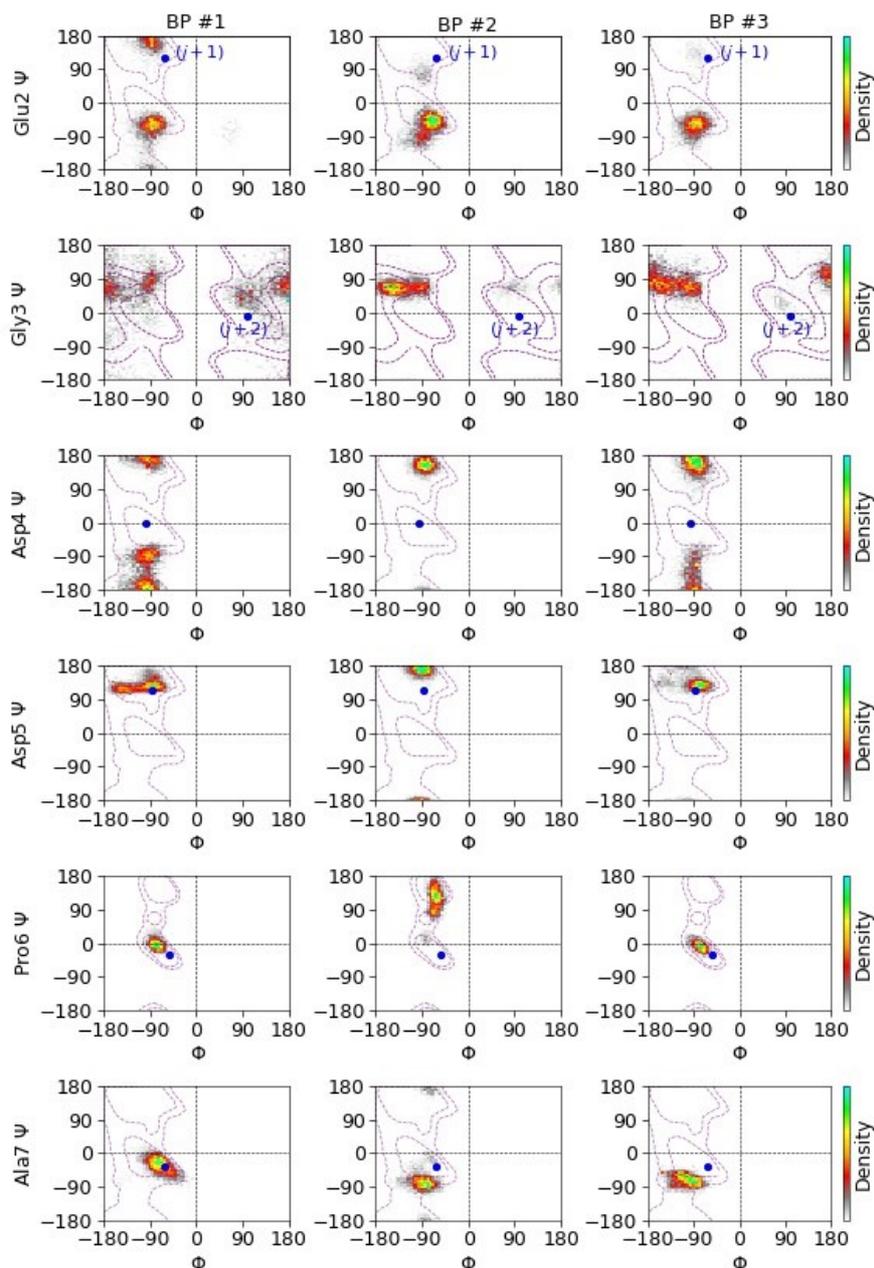


Figure S4 Ramachandran plots of Glu2-Gly3-Asp4-Asp5-Pro6-Ala7 for BPs #1-3 of TB state in WT simulations. Initial Φ and Ψ of all amino acids in the native configuration are highlighted as blue dots, and the residues $j + 1$ (Glu2) and $j + 2$ (Gly3) of type-II β -turn are labelled⁸. General α -helix and β -sheet regions¹⁰ are highlighted (dashed purple lines) for Glu2, Asp4, Asp5 and Ala7. Gly dihedral distribution from the literature is shown (dashed purple lines) for Gly3⁹. Pro dihedral distribution from the literature is described (dashed purple lines) for Pro6⁹.

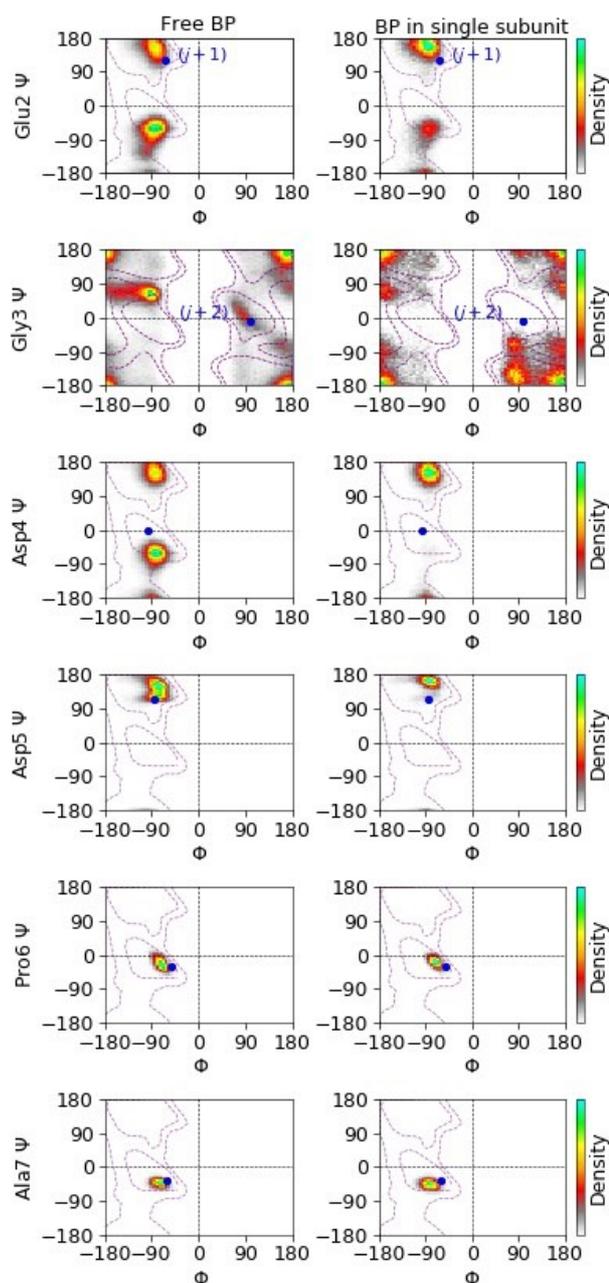


Figure S5 Ramachandran plots of Glu2-Gly3-Asp4-Asp5-Pro6-Ala7 of free BPs in WT simulations and of BP in a single WT subunit that was separately simulated. Initial Φ and Ψ of all amino acids in the native configuration are highlighted as blue dots, and the residues $j + 1$ (Glu2) and $j + 2$ (Gly3) of type-II β -turn are labelled⁸. General α -helix and β -sheet regions⁹ are highlighted (dashed purple lines) for Glu2, Asp4, Asp5, Pro6 and Ala7. Gly dihedral distribution from the literature is shown (dashed purple lines) for Gly3⁹. The conformations of Glu2, Gly3, Asp4 and Asp5 can be similar to/subsets of general random-coil regions in the literature¹⁰.

Table S1 STRIDE analysis of BPs #1-3 of the LB and TB conformations of AuNP #3-BP #1-3 complex in WT simulations. Alphabet coding of words: T, Turn (red); C, random coil (green); H, helix (blue).

WT	STRIDE secondary structure	STRIDE details
Initial native conformation	AEGDDPAK TTTTTHHH	Alpha Helix (P6 to A49) Turn II (A1 to D4) Turn IV (E2 to D5)
BP #1 (LB)	AEGDDPAK TTTTCHHH	Alpha Helix (P6 to A49) Turn VIII (A1 to D4)
BP #2 (LB)	AEGDDPAK TTTTCHHH	Alpha Helix (P6 to A49) Turn IV (A1 to D4)
BP #3 (LB)	AEGDDPAK CCCCCHHH	Alpha Helix (P6 to A49)
BP #1 (TB)	AEGDDPAK CCCCCCH	Alpha Helix (K8 to A49)
BP #2 (TB)	AEGDDPAK TTTTCCHH	Alpha Helix (A7 to A49) Turn IV (A1 to D4)
BP #3 (LB)	AEGDDPAK TTTTTTTH	Alpha Helix (K8 to A49) Turn IV (A1 to D4) Turn IV (D5 to K8)

The sequences of Ala1-Glu2-Gly3-Asp4 and Asp5-Pro6-Ala7-Lys8 can be denoted as flexible and rigid domains, respectively, based on the knowledge that Pro6-Ala7-Lys8 can generally be part of helical assemblies. Asp5 is not part of α -helix but it can display limited conformations in a simulation, so it is included in the rigid domain as a threshold residue. The differences between the free BPs and single BP subunit lie mainly in the flexible domain. For example, Fig. S5 shows that some type-II β -turn may still be observed for free BPs that are not involved in AuNP–M13 binding. It is possible that the differences can be associated with the assemblies of densely-packed coat-proteins, which is not present in the simulation of a single protein subunit.

Collectively, Figs. S3, S4 and Table S1 suggest that the distribution of random coils or type-IV β -turns at N-termini for TB state can generally be higher compared to that for LB state. Figure S4 and Table S1 also indicate that the α -helical part of BPs can be disrupted in TB state, which can be the result of enhanced sampling. On the whole, Figs. S3–S5 suggest that the conformations of the BPs in bound states can be subsets of the conformations of free BPs, and the trend is most significantly observed for Gly3.

Supplementary Note 3

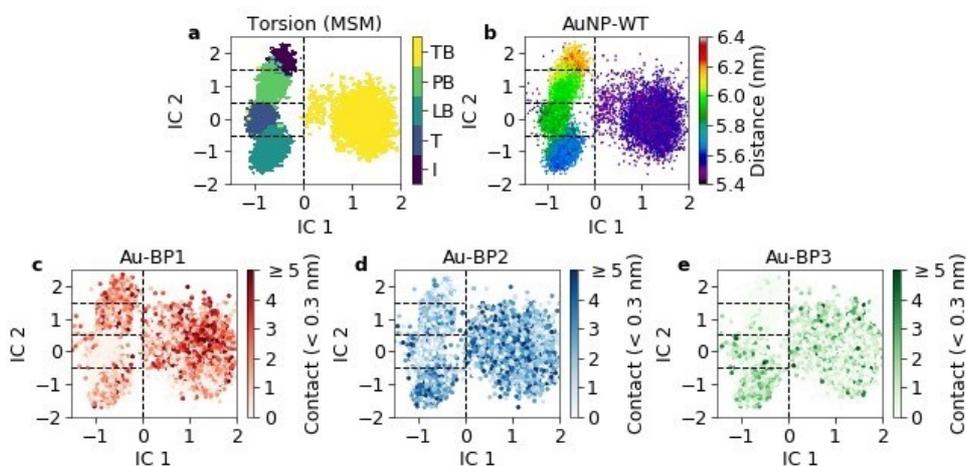


Figure S6 MSM analysis for backbone-torsion angles of BPs #1-3 for AuNP #3 in WT simulations, shown with the first two (most slowly changing) TICA collective variables IC 1 and IC 2, and plotted with the **a**) state-space discretization of initial state (I), intermediate (T) state, pre-bound (PB) state, loosely-bound (LB) state and tightly-bound (TB) state; **b**) overall AuNP-M13 distance; and total number of <0.3 nm contacts between the AuNP #3 (surface atoms) and all atoms of BPs #1 **c**), 2 **d**) and 3 **e**).

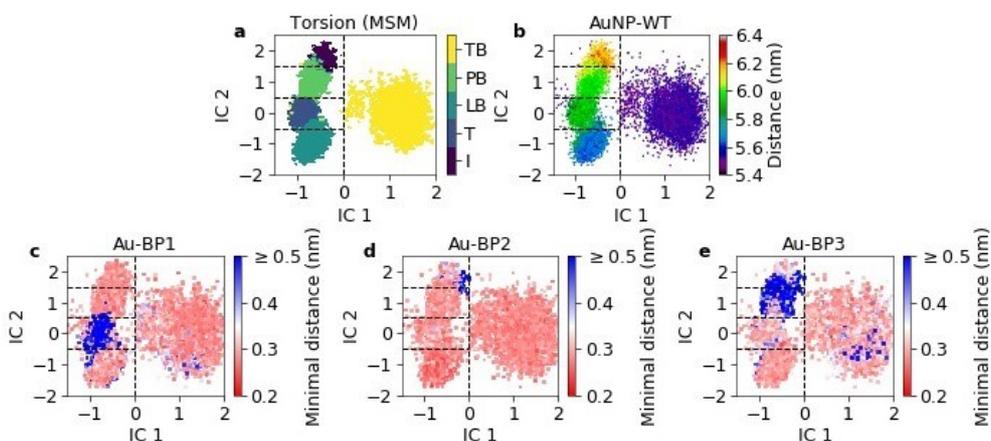


Figure S7 MSM analysis for backbone-torsion angles of BPs #1-3 for AuNP #3 in WT simulations, shown with the first two (most slowly changing) TICA collective variables IC 1 and IC 2, and plotted with the **a**) state-space discretization of initial state (I), intermediate (T) state, pre-bound (PB) state, loosely-bound (LB) state and tightly-bound (TB) state; **b**) overall AuNP-M13 distance; and minimal distance between AuNP #3 (surface atoms) and all atoms of BPs #1 **c**), 2 **d**) and 3 **e**). Note that Figs. S6a, b are the same as Figs S7a, b).

Supplementary Note 4

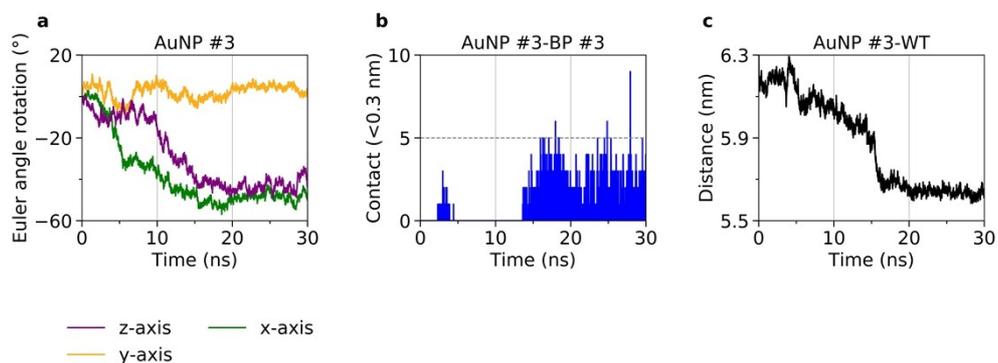


Figure S8 **a)** Time evolution of Euler angle rotation about three principle axes for a sample-unbiased-MD trajectory of AuNP #3-WT complex; **b)** number of <0.3 nm contacts between the AuNP #3 (surface atoms) and BP #3 (all atoms); and **c)** distance between the AuNP and WT for the sample trajectory.

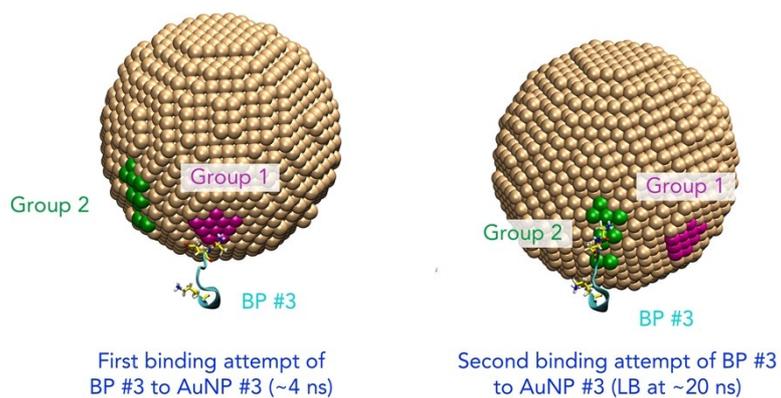


Figure S9 Snapshots of orientation of the AuNP #3 and surfaces atoms with minimal distances to BP #3 at two different simulation times shown in Fig. S8.

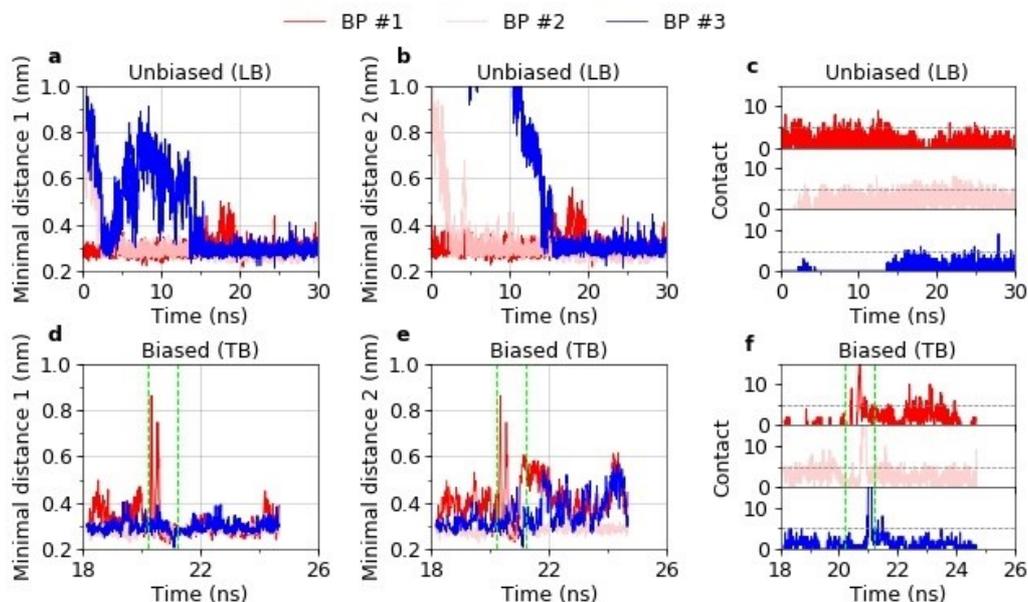


Figure S10 AuNP–BP binding characteristics for unbiased-and-biased sample-MD-trajectory of AuNP #3-WT complex. Time evolution of **a**) minimal distance between the AuNP #3 (surface atoms) and BPs #1-3 (all-atom) for a sample-unbiased-MD trajectory; **b**) minimal distance between the AuNP #3 (specific surface atoms) and BPs #1-3 (all-atom) – each specific group of Au surface atoms contains ten surface atoms that show the smallest distances to each BP in one final frame of LB state (around twenty nanoseconds); **c**) total number of <0.3 nm contacts between the AuNP #3 (surface atoms) and BPs #1-3 (all-atom). **d-f**) Similar set of plots to **a-c**) except under biased conditions, with the green dashed lines indicating the period of biasing based on the distance between the COM of AuNP #3 and COM of BPs #1, 2 and 3. Before the bias, the plots in **d-f**) represents a continued run of the unbiased ensemble in **a-c**).

The largest rotational displacements can be observed during the first five to fifteen nanoseconds of the sample trajectory (Fig. S8a), which approximately coincides with the gradual association of AuNP #3 to phage surface (Fig. S8c) before stable binding of BP #3 to the nanoparticle (Fig. S8b). The Euler angles are defined similar to those described in MATLAB¹¹. Figures S9 and S10a, b further show that during the rotational displacement, BP #3 is interacting with one group of AuNP #3 surface atoms (first binding attempt) that are different from the group of stably-bound AuNP #3 surface atoms (second binding attempt), indicating that nanoparticle rotation can frequently occur before formation of a stable AuNP–BPs #1-3 complex.

Supplementary Note 5

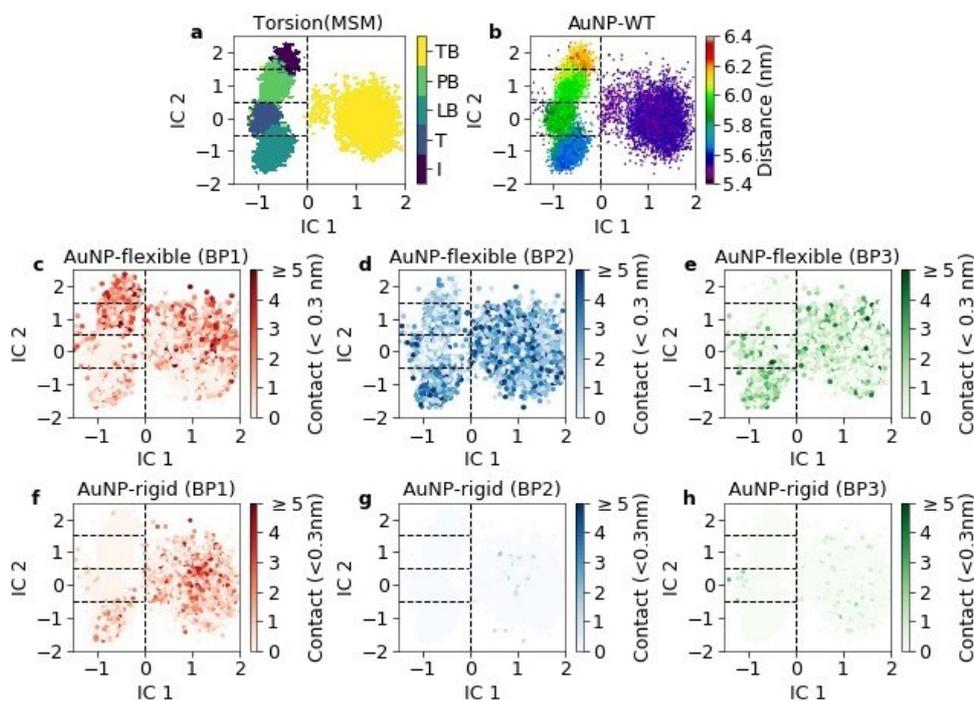


Figure S11 MSM analysis for backbone-torsion angles of BPs #1-3 for AuNP #3 in WT simulations, shown with the first two (most slowly changing) TICA collective variables IC 1 and IC 2, and plotted with the **a**) state-space discretization of initial state (I), intermediate (T) state, pre-bound (PB) state, loosely-bound (LB) state and tightly-bound (TB) state; **b**) overall AuNP–M13 distance; total number of <0.3 nm contacts between the AuNP #3 (surface atoms) and the flexible domain (Ala1-Glu2-Gly3-Asp4) of BPs #1 **c**), 2 **d**) and 3 **e**); and total number of <0.3 nm contacts between the AuNP #3 (surface atoms) and rigid domain (Asp5-Pro6-Ala7-Lys8) of BPs #1 **f**), 2 **g**) and 3 **h**).

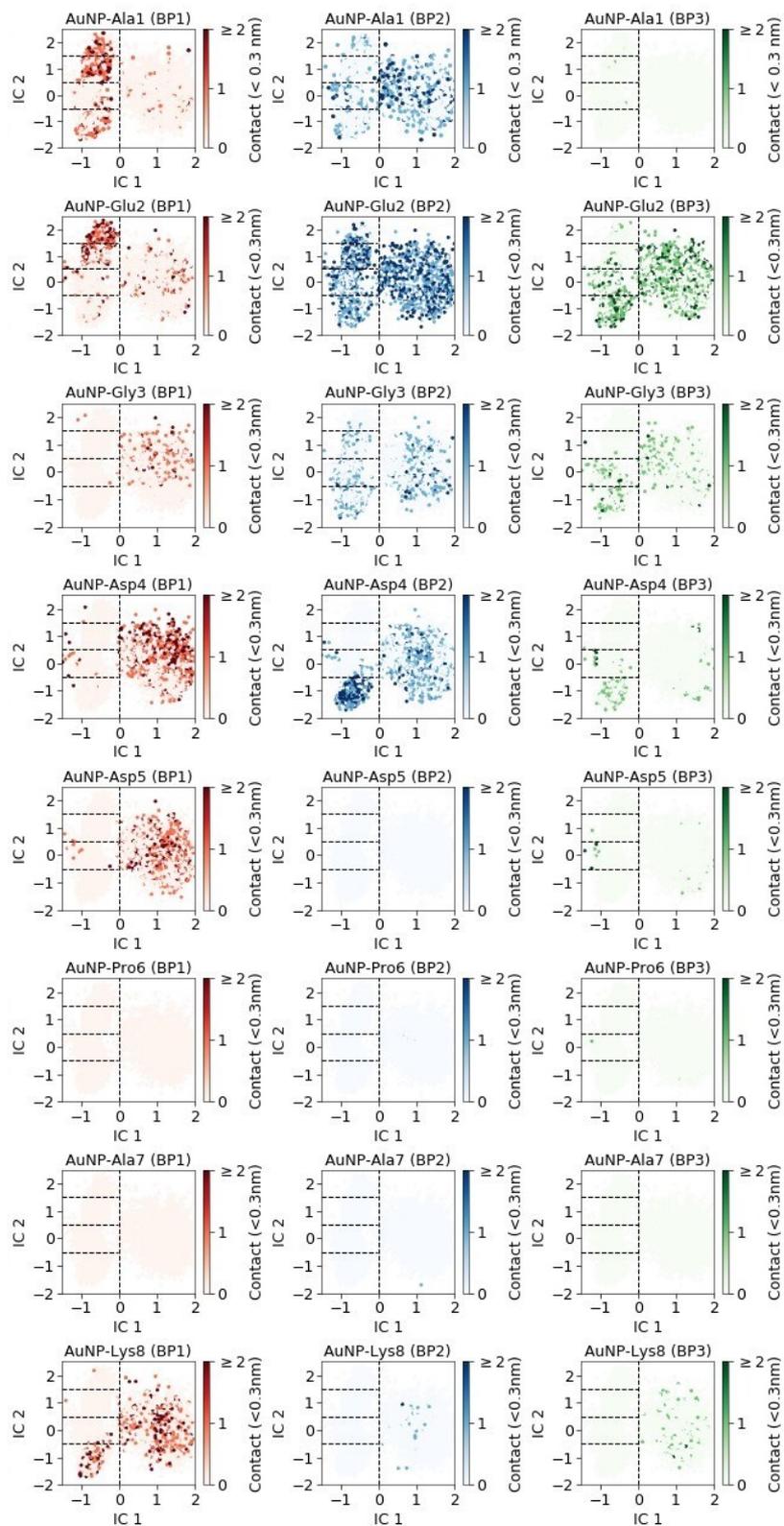


Figure S12 Total number of $<0.3\text{ nm}</math> contacts between the AuNP #3 (surface atoms) and individual amino acid for BPs #1-3, plotted with the first two TICA collective variables IC 1 and IC 2 in MSM analysis.$

Supplementary Note 6

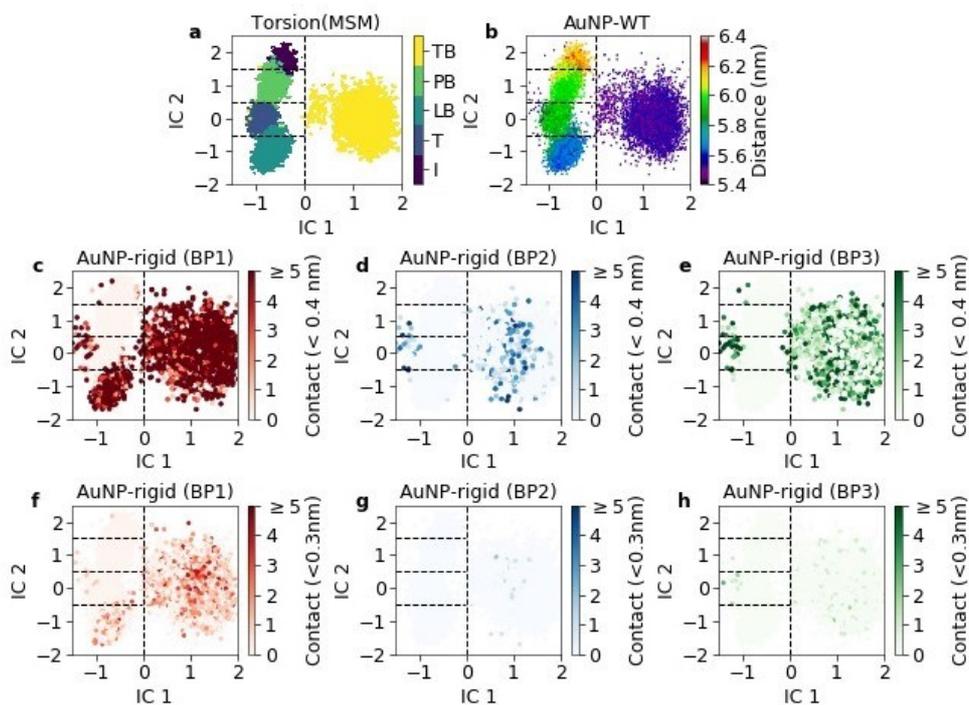


Figure S13 MSM analysis for backbone-torsion angles of BPs #1-3 for AuNP #3 in WT simulations, shown with the first two (most slowly changing) TICA collective variables IC 1 and IC 2, and plotted with the **a**) state-space discretization of initial state (I), intermediate (T) state, pre-bound (PB) state, loosely-bound (LB) state and tightly-bound (TB) state; **b**) overall AuNP–M13 distance; total number of <0.4 nm contacts between the AuNP #3 (surface atoms) and the rigid domain of BPs #1 **c**), 2 **d**) and 3 **e**); and total number of <0.3 nm contacts between the AuNP #3 (surface atoms) and the rigid domain of BPs #1 **f**), 2 **g**) and 3 **h**).

Supplementary Note 7

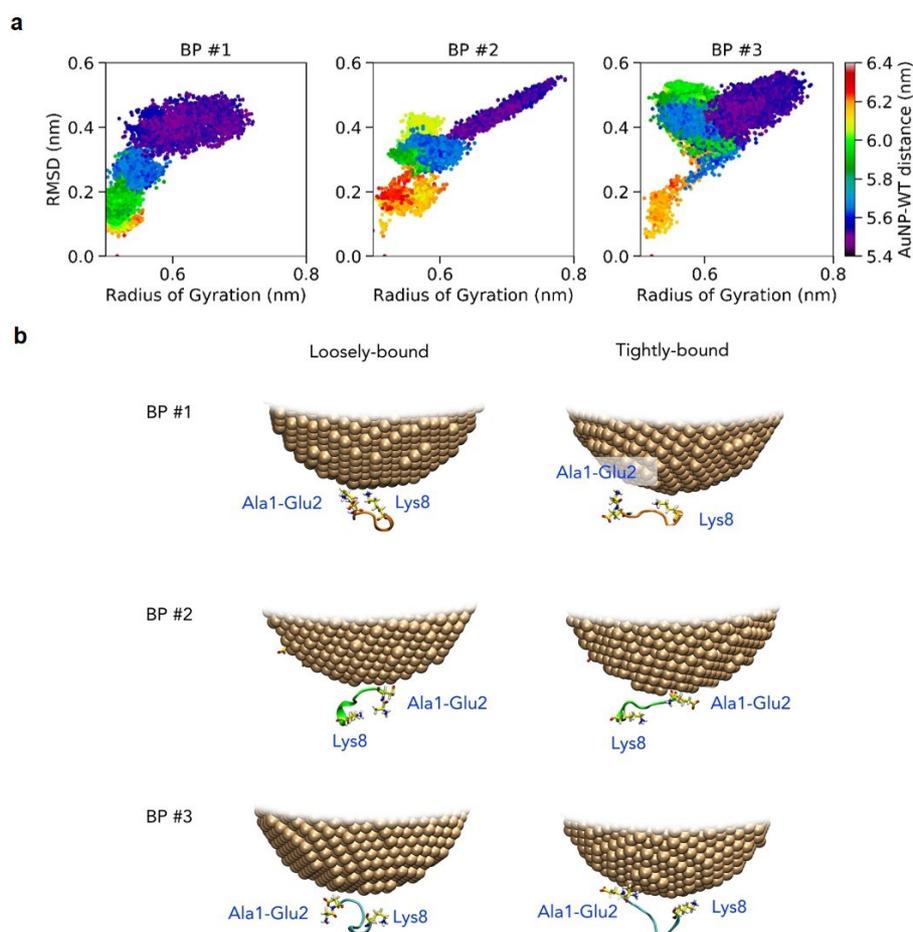


Figure S14 a) RMSD-RG colour map with overall gold–phage distance for BPs #1, 2 and 3; **b)** snapshots of binding-peptide conformations in LB state (blue region shown in **a**) and TB state (purple region shown in **a**).

In TB conformations, BPs can be more extended and aligned more parallel to AuNP surface (Fig. S14b); also, N-terminal Ala1-Glu2 can be more separated from Lys8 compared to those in LB conformations. Such conformations may allow the nanoparticle to be assembled closer to the phage surface, as observed in the AuNP–M13 complex in TB state.

Supplementary Note 8

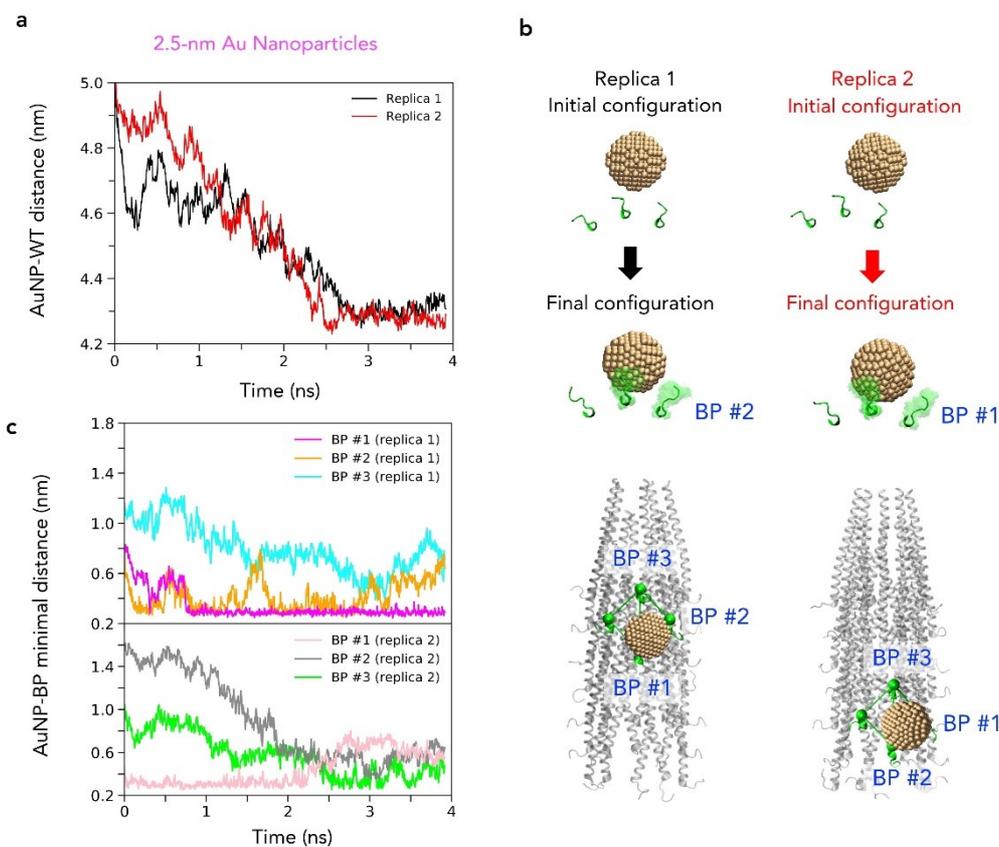


Figure S15 Simulation of AuNP–M13 biotemplating for the model with 2.5 nm-wide AuNPs and WT phage. **a)** Time evolution of absolute gold–phage distance (between the COM of AuNPs and phage axis); **b)** binding pattern of the AuNP–WT model in the initial and final configurations viewed parallel (top panel) and perpendicular (bottom panel) to the phage axis; green clouds show overlay of the frames captured towards end of the simulation (last one nanosecond); and **c)** minimal distance between the AuNPs of replicas 1 and 2 and their BPs #1-3.

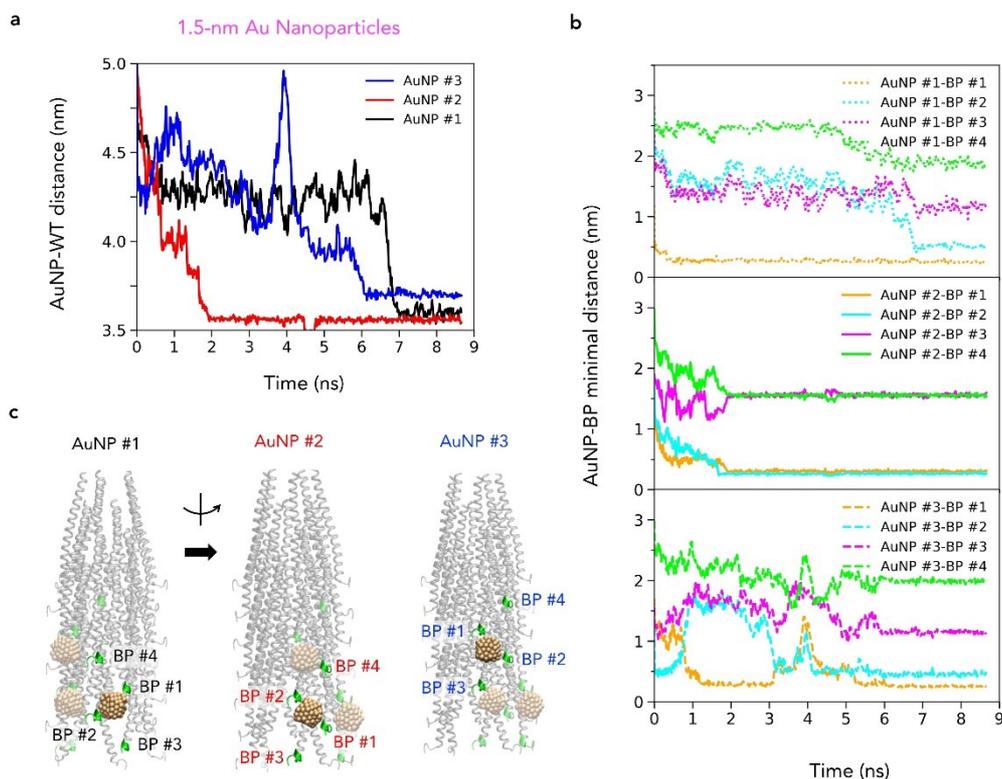


Figure S16 Simulation of AuNP-M13 biotemplating for the model with 1.5 nm-wide AuNPs using CHARMM-METAL force field and WT phage. **a)** Time evolution of absolute gold-phage distance (between the COM of AuNPs and phage axis); **b)** minimal distance between the AuNPs #1-3 and their BPs #1-4; and **c)** binding pattern of the AuNP-WT model in final configurations.

The 2.5 nm-wide nanoparticles may wobble slightly in the binding pocket after gold-phage assembling, as indicated by the evolution of minimal AuNP-BP distance for BP #1-3 (Fig. S15c), and the AuNPs may show a tendency to bind closer to the shorter edge of BPs #1-2 (Fig. S15b). One reason can be that the size of nanoparticle is small relative to the size of binding pocket (Fig. S15b). We note that such effect is not prominent for the 2.5 nm-wide nanoparticles. And, for the 1.5 nm-wide nanoparticles, the wobbling is less evident, potentially as a combined result of the much smaller size of nanoparticle and implemented inflexibility of N-termini (Fig. S16b). However, it is more evident that the nanoparticles can bind closer to the shorter edge

of BPs #1-2 (Fig. 16b, c). In this case, the total number of primary binding peptides for 1.5 nm-wide nanoparticles can typically be two rather than three. Nevertheless, such effect is not generally observed for the much larger 5 nm-wide nanoparticles (Fig. S10), as the AuNPs reside on top of the shallow binding pockets due to their large size. Experimentally available AuNPs can typically be equal to or greater than 5 nm in diameter, and it is likely that the binding pattern observed for the 5 nm-wide AuNP can be more typical for these cases.

Supplementary Note 9

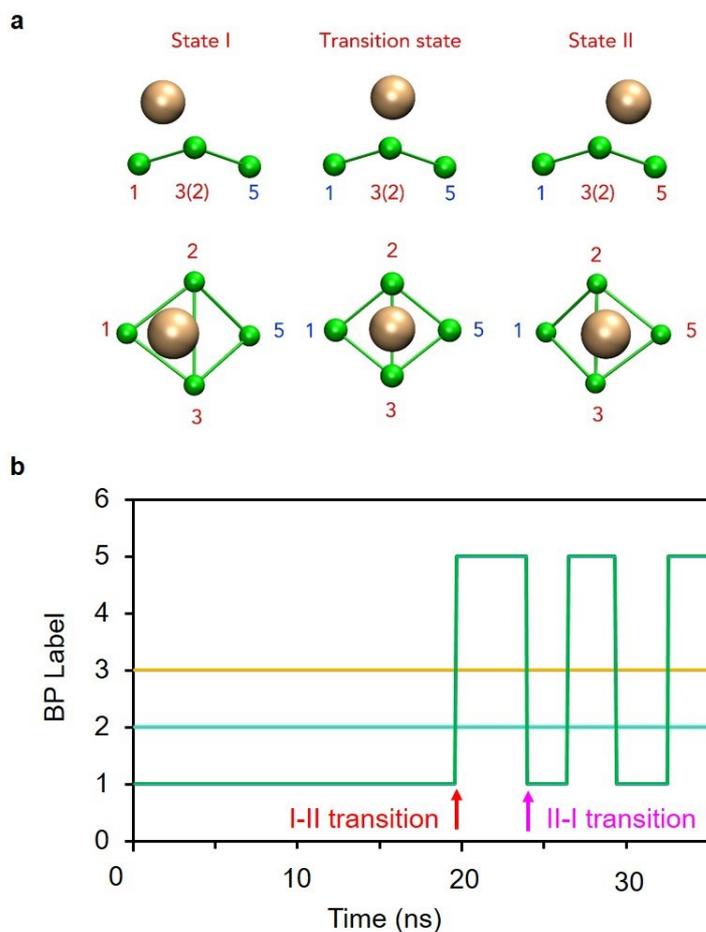


Figure S17 a) Molecular view of binding-pocket transition from state I to state II in CG simulations; and **b)** time evolution of the three closest CG BPs (based on the distance between the COM of AuNP and COM of surrounding BPs) during the first thirty-five nanoseconds. The red arrow marks onset of the binding-pocket transition depicted in **a)**, and the magenta arrow indicates the reverse binding-pocket transition at around twenty-five nanoseconds.

Supplementary Note 10

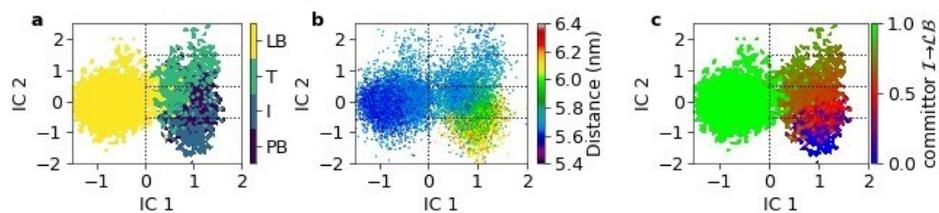


Figure S18 MSM analysis for backbone-torsion angles of BPs #1-3 for AuNP #3 in p8#9-like simulations, shown with the first two (most slowly changing) TICA collective variables IC 1 and IC 2, and plotted with the **a**) state-space discretization of initial state (I), intermediate (T) state, pre-bound (PB) state and loosely-bound (LB) state; **b**) overall AuNP-M13 distance; **c**) committor probabilities between the state I and LB state, which may identify the PB region as a possible transition state, characterized by committor values of approximately 0.5.

Supplementary Note 11

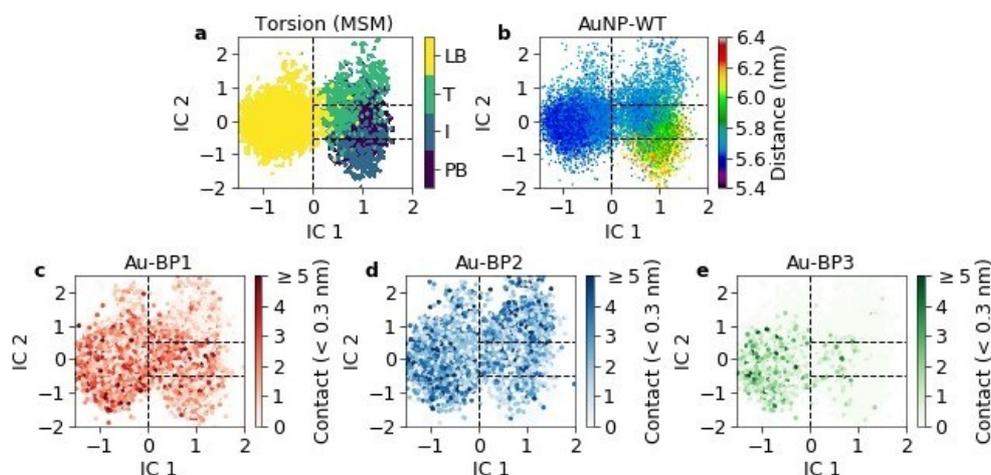


Figure S19 MSM analysis for backbone-torsion angles of BPs #1-3 for AuNP #3 in p8#9-like simulations, shown with the first two (most slowly changing) TICA collective variables IC 1 and IC 2, and plotted with **a**) state-space discretization of initial state (I), intermediate (T) state, pre-bound (PB) state and loosely-bound (LB) state; **b**) overall AuNP–M13 distance; and total number of <0.3 nm contacts between the AuNP #3 (surface atoms) and all atoms of BPs #1 **c**), 2 **d**) and 3 **e**).

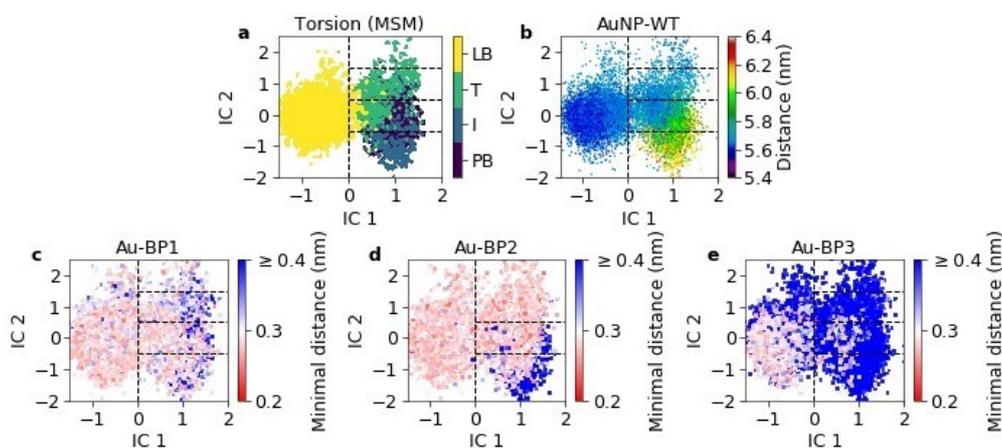


Figure S20 MSM analysis for backbone-torsion angles of BPs #1-3 for AuNP #3 in p8#9-like simulations, shown with the first two (most slowly changing) TICA collective variables IC 1 and IC 2, and plotted with **a**) state-space discretization of initial state (I), intermediate (T) state, pre-bound (PB) state and loosely-bound (LB) state; **b**) overall AuNP–M13 distance; and minimal distance between the AuNP #3 (surface atoms) and all atoms of BPs #1 **c**), 2 **d**) and 3 **e**). Note that Figs. S20a, b are the same as Figs. S19a, b.

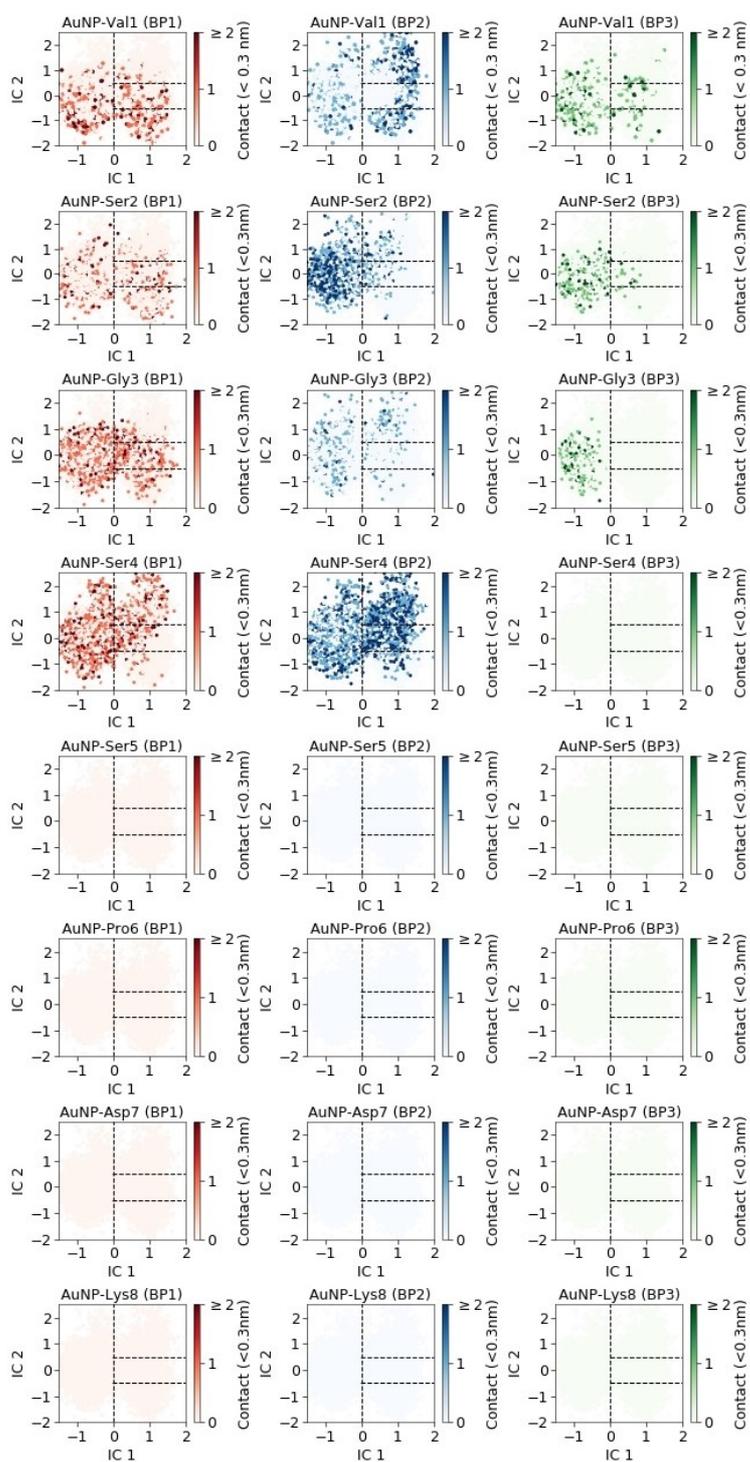


Figure S21 Total number of <0.3 nm contacts between the AuNP #3 (surface atoms) and individual amino acid for BPs #1-3 in p8#9-like simulations, plotted with the first two TICA collective variables IC 1 and IC 2 in MSM analysis.

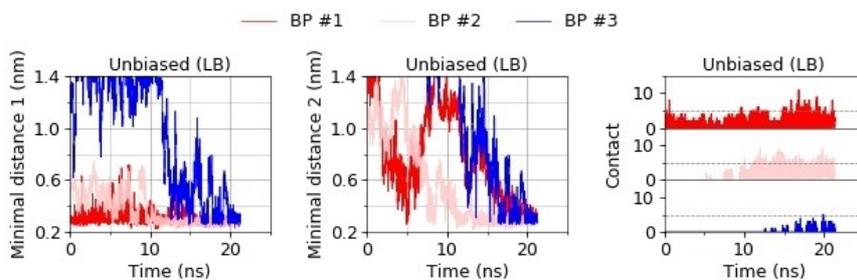


Figure S22 AuNP–BP binding characteristics of a sample trajectory of AuNP #3-p8#9-like complex. Time evolution of the **a**) minimal distance between the AuNP #3 (surface atoms) and BPs #1-3 (all-atom); **b**) minimal distance between the AuNP #3 (specific surface atoms) and BPs #1-3 (all-atom) – each specific group of Au surface atoms contains ten surface atoms that show the smallest distances to each BP in one final frame of LB state (around twenty nanoseconds); and **c**) total number of $<0.3\text{ nm}$ contacts between the AuNP #3 (surface atoms) and BPs #1-3 (all-atom).

Supplementary Note 12

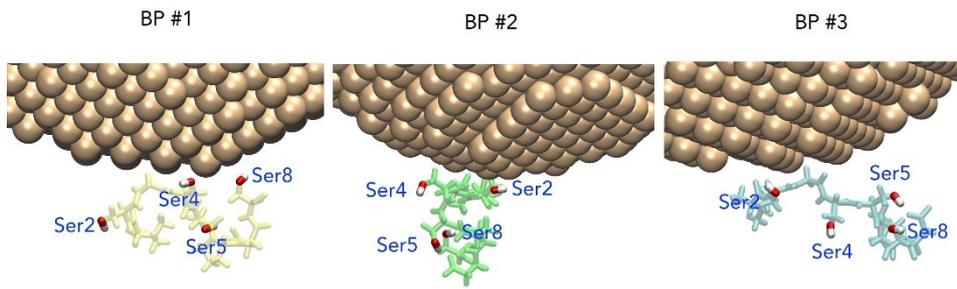


Figure S23 Representative snapshots of $-OH$ groups in gold-phage binding of p8#9-like mutant.

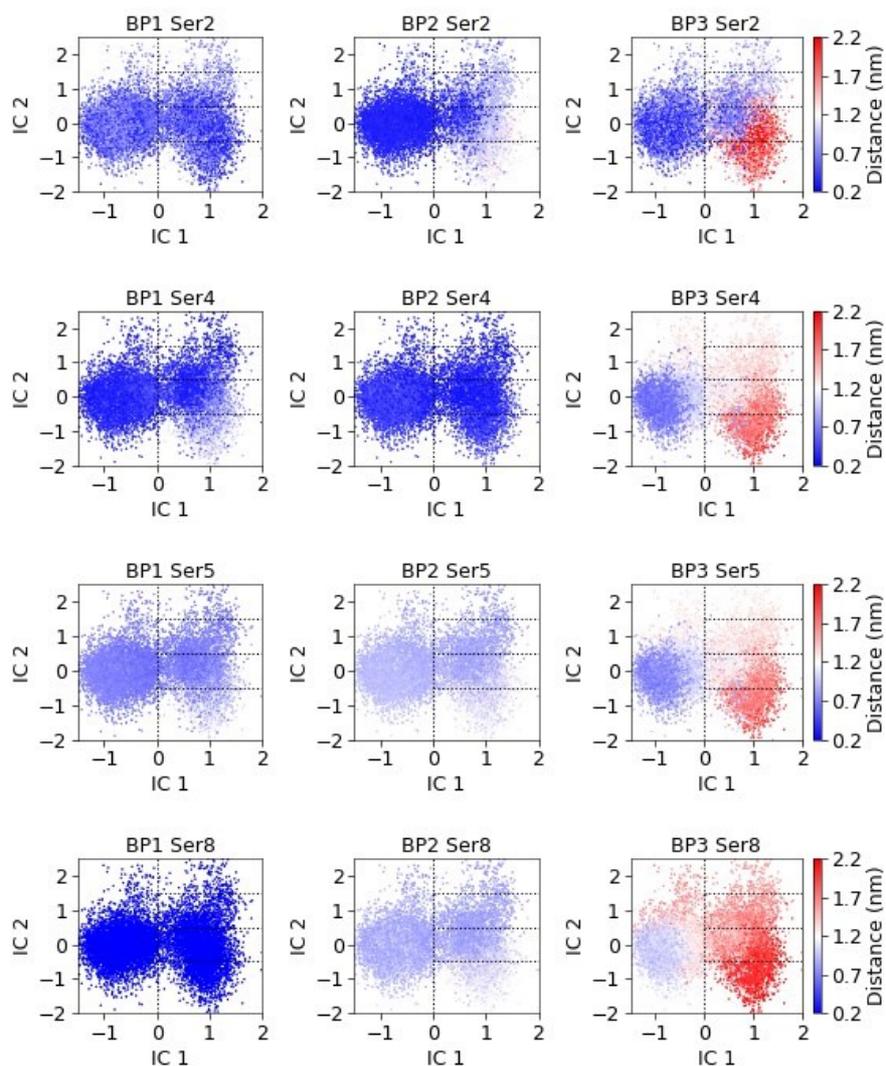


Figure S24 Minimal distance between the AuNP #3 surface atoms and -OH groups of Ser2, Ser4, Ser5 and Ser8 for BPs

#1-3 in p8#9-like simulations, plotted with the first two TICA collective variables IC 1 and IC 2 in MSM analysis (see Fig. S18).

Supplementary Note 13

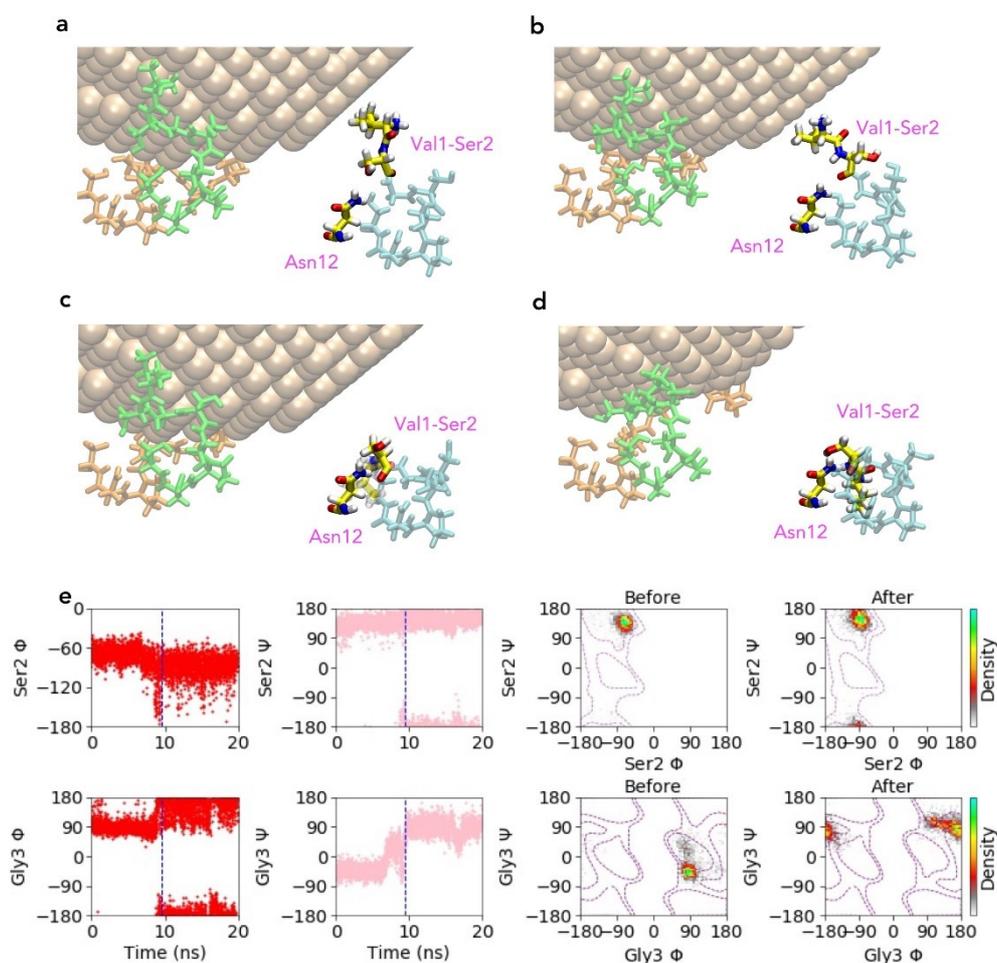


Figure S25 Conformational change around the Ser2-Gly3 peptide bond of pseudo-BP #3 for AuNP #4 with respect to the Asn12 carboxamide group, for a sample-MD trajectory of p8#9-like simulations. Snapshots of the structures at four different times, i.e., before (a–b) and after (c–d) the conformational change. e) Time evolution and distribution of Φ and Ψ for Ser2 and Gly3, and with the blue dashed vertical lines indicating the approximate time of conformational change.

In p8#9-like simulations, it can be that one conformational change in the proximity of Ser2-Gly3 bond occurs during a (unsuccessful) binding attempt of pseudo-BP #3 to AuNP #4. It is likely that due to the conformational change and subsequent hydrogen bond-like interaction¹² between the amide oxygen and carboxamide hydrogen, N-terminal domain of pseudo-BP #3 may experience a lower probability of orientating towards AuNP #4. Consequently, after a failed binding attempt, it is more

difficult for the pseudo-BP #3 to carry out another binding attempt to AuNP #4. The absence of binding of pseudo-BP #3 to AuNP #4 indicates that a stable gold–phage complex may not be formed. Subsequent observation may confirm the hypothesis, i.e., complete bindings of all three BPs are crucial to stabilize a gold–phage complex – AuNP #4 could deviate away from the phage surface without BP #3 (Fig. 6d). Note that AuNP #4 does not unbind from the phage surface as it remains bound with BPs #1 and 2 (see Fig. 6d).

Supplementary Note 14

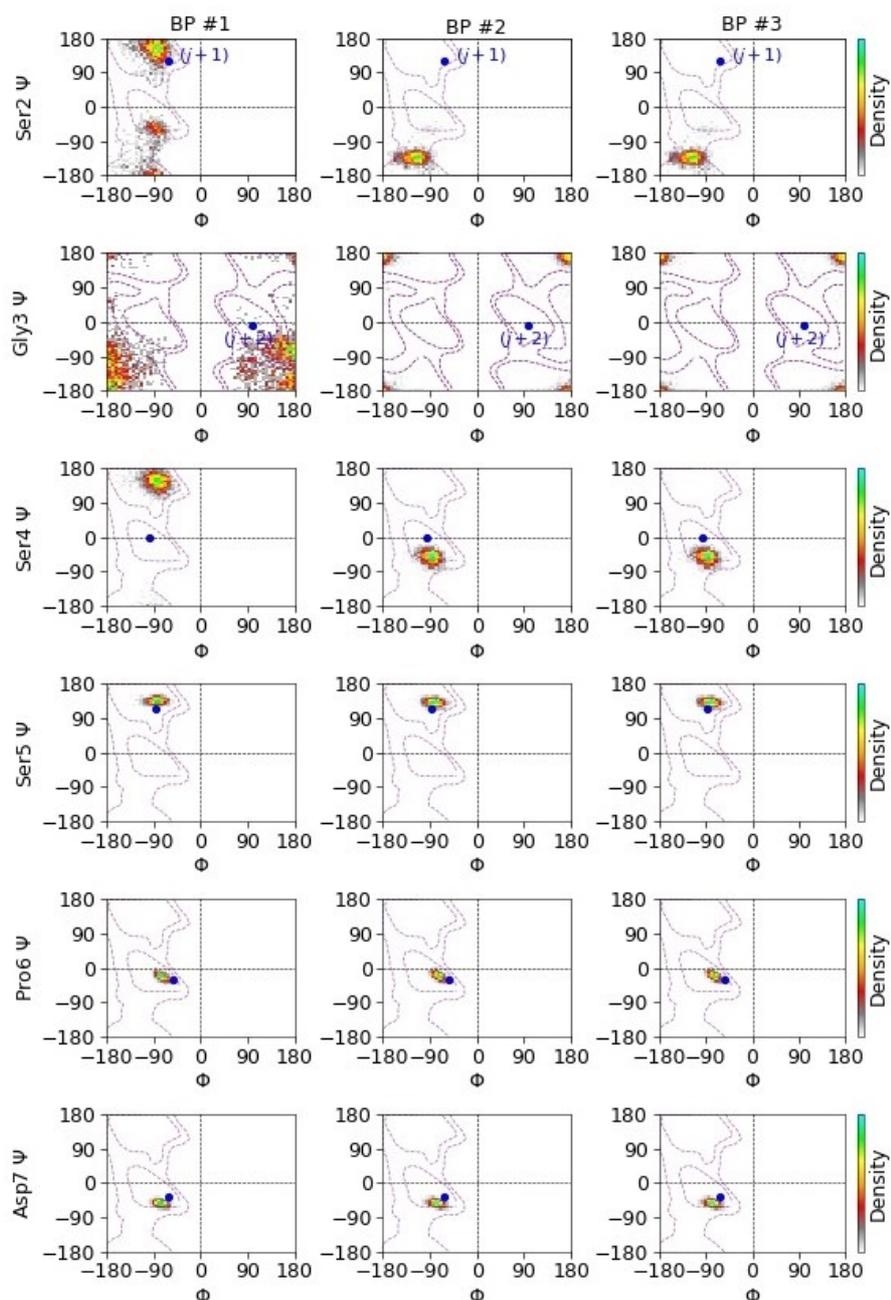


Figure S26 Ramachandran plots of Ser2-Gly3-Ser4-Ser 5-Pro6-Asp7 in binding peptides #1-3 for the loosely-bound state in p8#9-like simulation. Φ and Ψ of all amino acids in the initial native configuration are highlighted as blue dots, and the residues $j + 1$ and $j + 2$ of the initial type II β -turn (Val1-Ser4) are labelled⁸. General α -helix and β -sheet regions⁹ are highlighted (dashed purple lines) for Glu2, Ser4, Ser5, Pro6 and Asp7. Gly dihedral distribution from the literature is also highlighted (dashed purple lines) for Gly3⁹. Ramachandran plot of Serine can be similar to/subsets of the general random coil regions for Serine¹⁰.

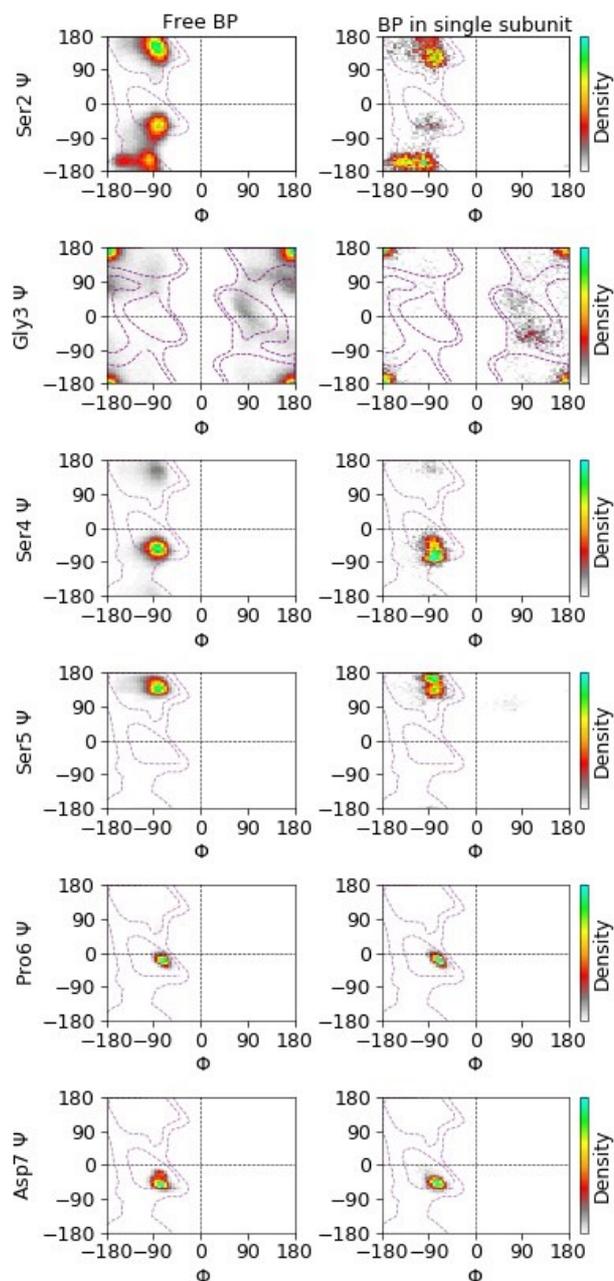


Figure S27 Ramachandran plots of Ser2-Gly3-Ser4-Ser5-Pro6-Asp7 of free BPs in p8#9-like simulations and for BP in a single p8#9-like subunit that was separately simulated. General α -helix and β -sheet regions⁹ are highlighted (dashed purple lines) for Glu2, Ser4, Ser5, Pro6 and Asp7. Gly dihedral distribution from the literature⁹ is also included (dashed purple lines) for Gly3. Ramachandran plot of Serine can be similar to/subsets of the general random coil regions for Serine¹⁰.

Table S2 STRIDE analysis of BPs #1-3 for the LB conformation of AuNP #3-BP #1-3 complex in p8#9-like simulations. Alphabet coding of words: T, Turn (red); C, random coil (green); H, helix (blue).

WT	STRIDE secondary structure	STRIDE details
Initial conformation	VSGSSPDS TTTTTHHH	Alpha Helix (P6 to A49) Turn II (V1 to S4) Turn IV (S2 to S5)
BP #1 (LB)	VSGSSPDS TTTTCHHH	Alpha Helix (P6 to A49) Turn IV (V1 to S4)
BP #2 (LB)	VSGSSPDS CCTTTHHH	Alpha Helix (P6 to A49) Turn IV (G3 to P6)
BP #3 (LB)	VSGSSPDS CCCCHHH	Alpha Helix (P6 to A49)

Supplementary Note 15

AuNP #4 (p8#9-like)

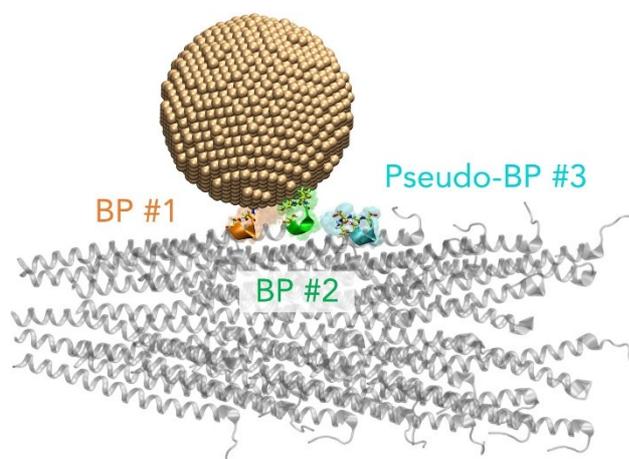


Figure S28 Representation of a transition-state-like structure for AuNP #4-p8#9-like complex.

References

1. M.K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, Ch. Wehmeyer, J. H. Prinz and F. Noé. PyEMMA 2: a Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.*, 2015, **11**, 5525–5542.
2. G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis and F. Noé. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.*, 2013, **139**, 015102–015113.
3. S. Lloyd. Least Squares Quantization in PCM. *IEEE T. Inform. Theory*, 1982, **28**, 129-137.
4. S. Röblitz and M. Weber. Fuzzy Spectral Clustering by PCCA+: Application to Markov State Models and Data Classification. *Adv. Data Anal. Classi.*, 2013, **7**, 147–179.
5. E. Weinan & E. Vanden-Eijnden. Towards a Theory of Transition Paths. *J Stat. Phys.*, 2006, **123**, 503–523.
6. X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Edit.*, 1999, **38**, 236–240.
7. M. J. Abraham, T. S. Murto, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl. GROMACS: High Performance Molecular Simulations through Multi-level Parallelism from Laptops to Supercomputers. *SoftwareX*, 2015, **1–2**, 19–25.
8. A. G. De Brevern. Extension of the Classical Classification of β -turns. *Sci. Rep.*, 2016, **6**, 33191.
9. S. C. Lovell, I. W. Davis, W. B. Arendall III, P. I. De Bakker, J. M. Word, M. G. Prisant, J. S. Richardson and D.C. Richardson. Structure Validation by $C\alpha$ Geometry: ϕ , ψ and $C\beta$ Deviation. *Proteins: Struct. Funct. Bioinf.*, 2003, **50**, 437–450.
10. S. Hovmöller, T. Zhou and T. Ohlson. Conformations of Amino Acids in Proteins. *Acta. Crystallogr. D*, 2002, **58**, 768–776.
11. MATLAB R2018b. Natick, Massachusetts: The MathWorks Inc.

12. L. Shimoni and J. P. Glusker. Hydrogen Bonding Motifs of Protein Side Chains: Descriptions of Binding of Arginine and Amide Groups. *Protein Sci.*, 2008, **4**, 65–74.