SI Appendix

Tunable Assembly of Amyloid-forming Peptides into Nanosheets as A Retrovirus Carrier

Bin Dai[a, 1], Dan Li[a, 1], Wenhui Xi[b], Fang Luo[a], Xiang Zhang[a], Man Zou[a], Mi Cao[c], Jun Hu[d], Wenyuan Wang[a], Guanghong Wei[b, 2], Yi Zhang[d, 2], Cong Liu[a, 2]

[a]Interdisciplinary Research Center on Biology and Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China; [b]State Key Laboratory of Surface Physics, Key Laboratory for Computational Physical Sciences (Ministry of Education), Department of Physics, Fudan University, Shanghai 200433, China; [c]National Center for Protein Science (Shanghai), Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201210, China; [d]Laboratory of Physical Biology, Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China

[1]These authors contributed equally to this work.
[2]To whom correspondence should be addressed. E-mails: liulab@sioc.ac.cn, zhangyi@sinap.ac.cn, ghwei@fudan.edu.cn

This PDF file includes:

Materials and Methods

Length estimation of KLVFFAK in β-conformation

The length of heptapeptide KLVFFAK in β-strand conformation was measured from its structural model. The structural model was generated using the crystal structure of hexapeptide KLVFFA (PDB ID: 3OW9) as template. We extended the terminus using ideal β-strand phi/psi angles, and then mutated the extended terminal residue to lysine side chains. Rotamers were selected to avoid steric clash. Modeling was performed using the program COOT (1).

Circular dichroism (CD) spectroscopy.

CD Spectra were acquired using a JASCO J-715 spectrometer equipped with a JASCO PTC-348 temperature controller. Measurements were made at 23°C. Far-UV spectra (260-190 nm) were collected in 0.1 cm path-length quartz cells with nanosheet pellets suspended in water. The concentration is equivalent to 0.5 mg/ml peptide monomer. Raw data were processed by smoothing and subtraction of the blank according to the manufacturer's protocol.

Analysis of the secondary structure was carried out using the software CDNN (2). The database consisting of 33 reference proteins was used in the deconvolution analysis.

Molecular dynamics (MD) simulations details

1. Initial states for all peptide systems

The initial β-sheet structures were built using AmberTools as follows: eight peptides were aligned with well-formed anti-parallel hydrogen bonds (HB) and the side chain pointed to the same orientation. Two layers (2×8) (here 2 refers to the number of β-sheet and 8 is the number of peptide chains) were parallel aligned with a distance of 1.2 nm between the center of mass of the Cα atoms in the two layers and the minimum distance between atoms in the different layers was no less than 0.4 nm. A larger bilayer β-sheet of KLVFFAK and VQIVYK (2×16) were built in the same way. For explicit solvent MD simulations, the constructed β-sheet structure was placed in a box of 7.5 nm x 7.5 nm x 8.5 nm. Counter ions were added to neutralize the simulated system. The box size is large enough to avoid atomic interactions with its images. It is noted that there was no β-sheet twisting in the initial states.

2. MD simulations

MD simulations were mainly performed with implicit solvent model. The initial structures of each peptide underwent 5000 steps energy minimization, followed by 1-ns MD simulation with main chain restrained. Starting from this

MD-generated structure, a 30-ns MD simulation was performed for each system. In the first 5 ns, the backbone HBs in β-sheets were restrained loosely (the distance between donor and acceptor is less than 0.7 nm). Meanwhile the distance between the two layers was also restrained loosely (less than 1.5 nm) during this period of simulation.

Implicit solvent MD simulations were performed using the generalized Born (GB) model developed by Onufriev et al. (3, 4). The whole systems were set in elastic sphere boundary at a constant temperature of 300 K. Explicit solvent MD simulations were conducted in the isothermal-isobaric (NPT) ensemble at 300 K. Particle mesh Ewald (PME) methods was employed to calculate the electrostatic interactions (5). Hydrogen atoms were constraint by SHAKE algorithm in all simulations, which allows an integration time step of 2 fs (6). The cutoff was set to 3.0 nm in implicit-solvent MD and 0.8 nm in explicit-solvent MD.

The last 15 ns trajectories were used for analyses. Due to the finite size effect, the peptide chains in the two ends of a β-sheet layer (first and eighth) sometimes dissociate from the β-sheet. To examine influence the simulation time on the stability of preformed β-sheet bilayer, one of the implicit solvent MD simulation was extended to 80 ns for the KLVFFAK (2x8) system with A to B packing interface. RMSD analysis shows that the β-sheet bilayer structure does not change with simulation time.

To examine the reliability of the simulation results obtained from implicit solvent MD simulations, we performed additional MD simulations with explicit solvent of TIP3P water model for all 2x8 β-sheet bilayer (Table S3). In good agreement with the simulation results using implicit solvent model, we found that the constructed KLVFGAK/VQIVYK β-sheet bilayer with A to B and B to B packing interfaces dissociate within a 30-ns MD simulation, while the β-sheet bilayer with A to A packing interface maintains.

3. Binding energy calculation with MM/GBSA and MM/PBSA methods

MM/GBSA method is a frequently used method in estimating binding free energy for biomolecule association (7-10). The binding free energy reported here is the relative binding free energy, which was calculated using the formula given below

$$\begin{aligned}\Delta G_{bind} &= G_{complex} - G_{Layer1} - G_{Layer2} \\ &= \Delta E_{vdw} + \Delta E_{electrostatic} + \Delta E_{PB/GB} + \Delta E_{np} - T\Delta S\end{aligned}$$

Where ΔE_{elec} and ΔE_{vdw} are respectively the electrostatic (ΔE_{elec}) and van der Waals (ΔE_{vdw}) terms in gas phase. $\Delta E_{PB/GB}$ is polar solvation energy and ΔE_{np} is non-polar solvation energy component. The calculation of the entropic contribution to the binding free energy remains a great challenge (11). Approaches used to estimate the entropy of a molecule are very time-consuming

and the magnitude of standard error is high in comparison with the other energetic terms (8, 12). We have tried to get the entropy contribution with quasi-harmonic methods but obtained over-estimated values (about -150 ~ -200 kcal/mol). In addition, the net contribution of entropy ($T\Delta S$) is usually small and several studies reported that including corrections for changes in conformational entropy of the system lead to only a small improvement in the correlation with experiment (8, 13). Thus, in this study, the contribution of conformational entropy of peptides was ignored in accordance with a number of previous computational studies (14-18).

To examine the accuracy and reliability of MM/GBSA method, we also calculated the binding free energy using MM/PBSA method for all simulated systems. The results are given in Table S4. In spite of the differences in the absolute value of the solvation energy, the relative magnitude of binding energy for β-sheets with different packing interfaces obtained using GB method is consistent with those calculated using PB method. The consistence is also seen for the systems simulated using implicit model and explicit solvent model. These results demonstrate that the relative binding free energy based on implicit MD simulation in combination with MM/GBSA method is qualitatively consistent with the one based on explicit MD simulation in combination with MM/GB(PB)SA method, and it is also consistent with the relative stabilities of β-sheets with different packing interfaces. For brevity, the binding energies reported in the main text were based on the implicit-solvent MD simulations with MM/GBSA analysis.

References

1. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**:2126-2132.
2. Böhm G, Jaenicke R (1992) Correlation functions as a tool for protein modeling and structure analysis. *Protein Sci* **1**(10): 1269– 1278.
3. Onufriev A, Bashford D, Case DA (2000) Modification of the generalized born model suitable for macromolecules. *J Phys Chem B* **104**(15): 3712-3720.
4. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55**(2): 383-394.
5. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. *J Chem Phys* **103**(19): 8577-8593.
6. Ryckaert J-P, Ciccotti G, Berendsen HJ (1977) Numerial integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J Comput Phys* **23**(3): 327-341.
7. Kollman PA, et al. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* **33**(12): 889-897.
8. Homeyer N, Gohlke H (2012) Free energy calculations by the molecular mechanics poisson−boltzmann surface area method. *Mol Inf* **31**(2): 114-122.
9. Reyes CM, Kollman PA (2000) Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *J Mol Biol* **297**(5): 1145-1158.
10. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE (2012) *MMPBSA.py*: an efficient program for end-state free energy calculations. *J Chem Theory Comput*, **8**(9): 3314-3321.
11. Polyansky AA, Kuzmanic A, Hlevnjak M, Zagrovic B (2012) On the contribution of linear correlations to quasi-harmonic conformational entropy in proteins. *J Chem Theory Comput* **8**(10): 3820-3829.
12. Yang T, et al. (2011) Virtual screening using molecular simulations. *Proteins* **79**(6): 1940-1951.
13. Rastelli G, Del Rio A, Degliesposti G, Sgobba M (2010) Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J Comput Chem* **31**(4): 797-810.
14. Wang J, Morin P, Wang W, Kollman PA (2001) Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J Am Chem Soc* **123**(22): 5221-5230.
15. Wu C, Bowers MT, Shea JE (2010) Molecular structures of quiescently grown and brain-derived polymorphic fibrils of the Alzheimer amyloid abeta9-40 peptide: a comparison to agitated fibrils. *PLoS Comput Biol* **6**(3): e1000693.

16. Xie L, et al. (2014) The molecular mechanism of fullerene-inhibited aggregation of Alzheimer's beta-amyloid peptide fragment. *Nanoscale* **6**(16): 9752-9762.
17. Berhanu WM, Hansmann UH (2013) The stability of cylindrin beta-barrel amyloid oligomer models-a molecular dynamics study. *Proteins* **81**(9): 1542-1555.
18. Kumari R, Kumar R, Open Source Drug Discovery C, Lynn A (2014) g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model* **54**(7): 1951-1962.

Table S1. Comparison of secondary structure contents between KLVFFAK and its variants.*

	KLVFFA**K**	**K**LVFFA**R**	**R**LVFFA**K**	**K**LVFFA**D**
Helix	7.5 %	5.7 %	6.1 %	3.8 %
Antiparallel	49.6 %	40.1 %	46.0 %	43.3 %
Parallel	4.0 %	3.2 %	3.5 %	3.3 %
Beta-Turn	15.1 %	18.6 %	17.0 %	11.1 %
Rndm. Coil	23.7 %	32.2 %	27.4 %	38.5 %
Total Sum	100.0%	100.0 %	100.0 %	100.0 %

* The secondary structure contents are analyzed from CD spectra using the software CDNN (2). For each peptide, the percentage of secondary structure content is normalized to a total sum of 100.0 %.

Table S2. Surface zeta potential of four peptide nanosheets.

Sequence	Isoelectric point (pI)	Zeta potential
KLVFFAK	10.6	+7.8 ± 0.3 mv
RLVFFAK	11.4	+17.1 ± 1.2 mv
KLVFFAR	11.4	+14.1 ± 2.1 mv
KLVFFAD	6.2	-5.8 ± 0.2 mv

Table S3. Details of all simulated systems, including packing interfaces between two β-sheets, size of β-sheet bilayers (the first number refers to the number of β-sheet and the second number refers to the peptide chains in one sheet), number of MD runs, and the name of solvent models.

Peptides	Packing interface	Size of β-sheet bilayer	Number of MD runs (30 ns/MD run)	
			Implicit solvent GB model	Explicit solvent TIP3P water
KLVFFAK	A to A B to B A to B	2x8	2	2
		2x16	2	
KLVFGAK	A to A B to B A to B	2x8	2	1
VQIVYK	A to A B to B A to B	2x8	2	1
	A to A	2x16	2	

Table S4. Binding energy (in kcal/mol) in all MD simulations with different solvent models. Im. and Ex. denote respectively the implicit and explicit solvent models.

MM/GBSA							
Peptide	**Interface**	**MD run**	**ΔE_{vdw}**	**ΔE_{ele}**	**ΔE_{GB}**	**ΔE_{np}**	**ΔG_{bind}**
KLVFFAK	A to A (V-F-A)	Im. MD1	-105.23	3340.28	-3290.90	-13.55	-69.40 ±5.19
		Im. MD2	-111.97	3340.18	-3288.33	-14.48	-74.59 ±4.56
		Ex. MD1	-125.77	3406.59	-3344.04	-16.14	-79.36 ±5.80
		Ex. MD2	-112.78	3418.63	-3358.39	-14.62	-67.16 ±4.57
	B to B (K-V-F-K)	Im. MD1	-81.94	3415.45	-3362.33	-10.74	-39.57 ±9.04
		Im. MD2	-80.37	3391.80	-3339.78	-10.64	-38.99 ±7.15
		Ex. MD1	-114.91	3468.34	-3401.50	-14.90	-62.97 ±9.38
		Ex. MD2	-116.22	3484.48	-3414.25	-15.33	-61.32 ±7.12
	A to B	Im. MD1	-108.45	3367.54	-3308.90	-13.90	-63.71 ±8.56
		Im. MD2	-115.74	3403.51	-3341.83	-15.08	-69.13 ±9.89
		Ex. MD1	-119.26	3359.00	-3297.14	-15.41	-72.81 ±9.72
		Ex. MD2	-122.96	3492.50	-3423.48	-15.85	-69.78 ±6.01
KLVF GAK	A to A (L-F-A)	Im. MD1	-88.68	2770.76	-2725.90	-11.60	-55.42 ±8.86
		Im. MD2	-89.09	2776.16	-2731.17	-11.65	-55.75 ±8.75
		Ex. MD1	-116.39	3268.71	-3208.88	-15.32	-71.88 ±7.35
VQIVYK	A to A (V-I-Y)	Im. MD1	-84.34	674.33	-631.59	-11.47	-53.07 ±10.65
		Im. MD2	-94.96	691.83	-645.74	-12.95	-61.83 ±9.07
		Ex. MD1	-145.51	697.68	-635.46	-19.90	-103.19 ±8.10
MM/PBSA							
Peptide	**Interface**	**MD run**	**ΔE_{vdw}**	**ΔE_{ele}**	**ΔE_{PB}**	**ΔE_{np}**	**ΔG_{bind}**
KLVFFAK	A to A (V-F-A)	Im. MD1	-105.23	3340.28	-3259.45	-11.73	-36.13 ±5.47
		Im. MD2	-111.97	3340.18	-3252.01	-12.06	-35.86 ±5.74
		Ex. MD1	-125.77	3406.59	-3314.24	-12.31	-45.74 ±7.10
		Ex. MD2	-112.78	3418.63	-3330.89	-11.74	-36.78 ±3.24
	B to B (K-V-F-K)	Im. MD1	-81.94	3415.45	-3322.30	-10.83	0.37 ±8.06
		Im. MD2	-80.37	3391.80	-3297.76	-10.55	3.12 ±7.42
		Ex. MD1	-114.91	3468.34	-3364.09	-12.65	-23.31 ±7.80
		Ex. MD2	-116.22	3484.48	-3377.10	-13.71	-22.55 ±8.80
	A to B	Im. MD1	-108.45	3367.54	-3270.23	-13.02	-24.16 ±8.60
		Im. MD2	-115.74	3403.51	-3300.23	-13.84	-26.29 ±9.15
		Ex. MD1	-119.26	3359.00	-3270.28	-13.01	-43.56 ±9.01
		Ex. MD2	-122.96	3492.50	-3393.18	-13.30	-36.94 ±7.43
KLVFG AK	A to A (L-F-A)	Im. MD1	-88.68	2770.76	-2696.69	-11.04	-25.66 ±6.37
		Im. MD2	-89.09	2776.16	-2701.82	-11.08	-25.83 ±6.37
		Ex. MD1	-116.39	3268.71	-3187.57	-11.95	-47.20 ±6.42
VQIVYK	A to A (V-I-Y)	Im. MD1	-84.34	674.33	-623.33	-11.05	-44.38 ±9.72
		Im. MD2	-94.96	691.83	-640.13	-12.36	-55.63 ±10.29
		Ex. MD1	-145.51	697.68	-642.28	-14.95	-105.06 ±8.88

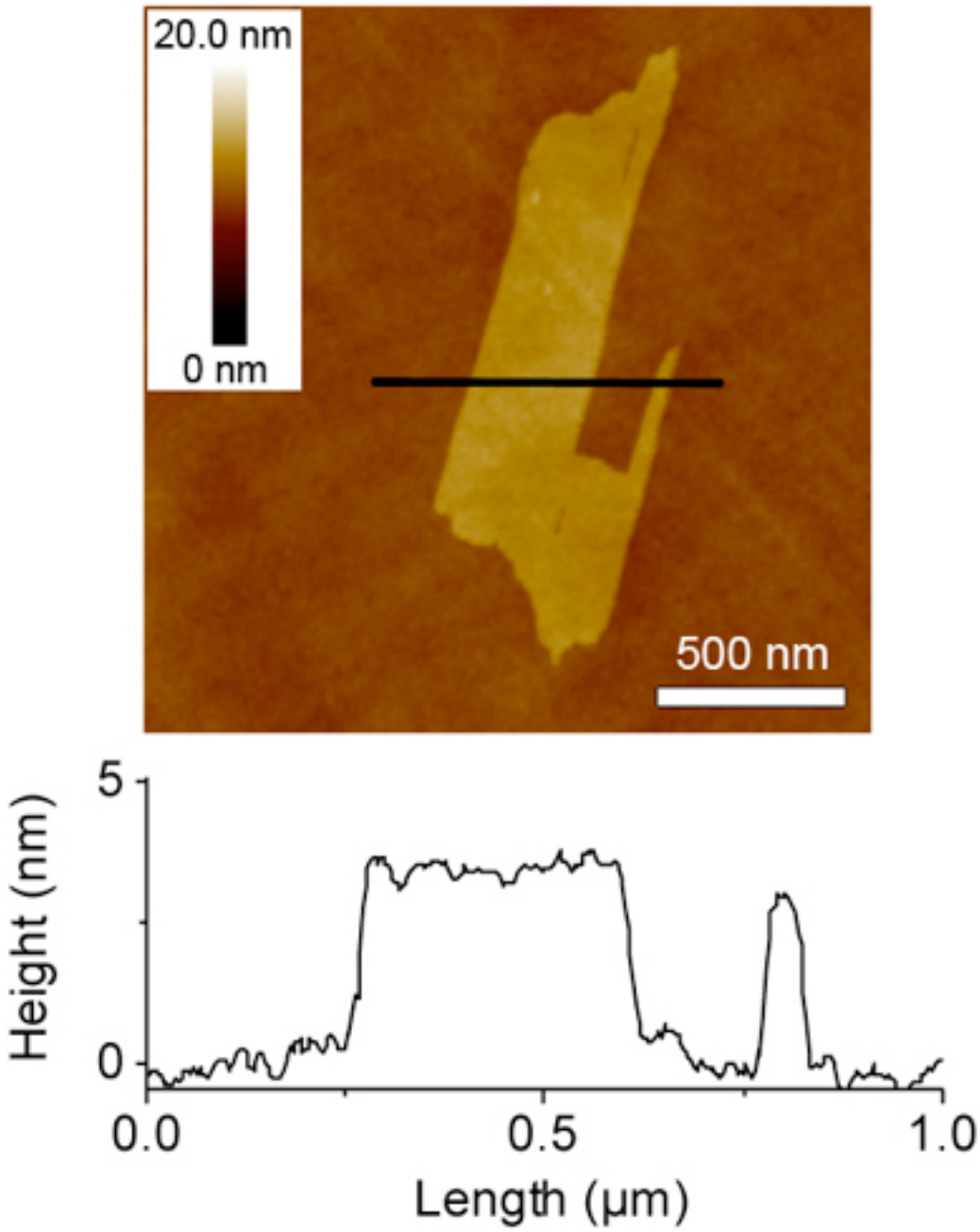


Fig. S1. AFM image of KLVFFAK nanosheet on Si wafer. The lower panel is the linear section profile of the nanosheet.

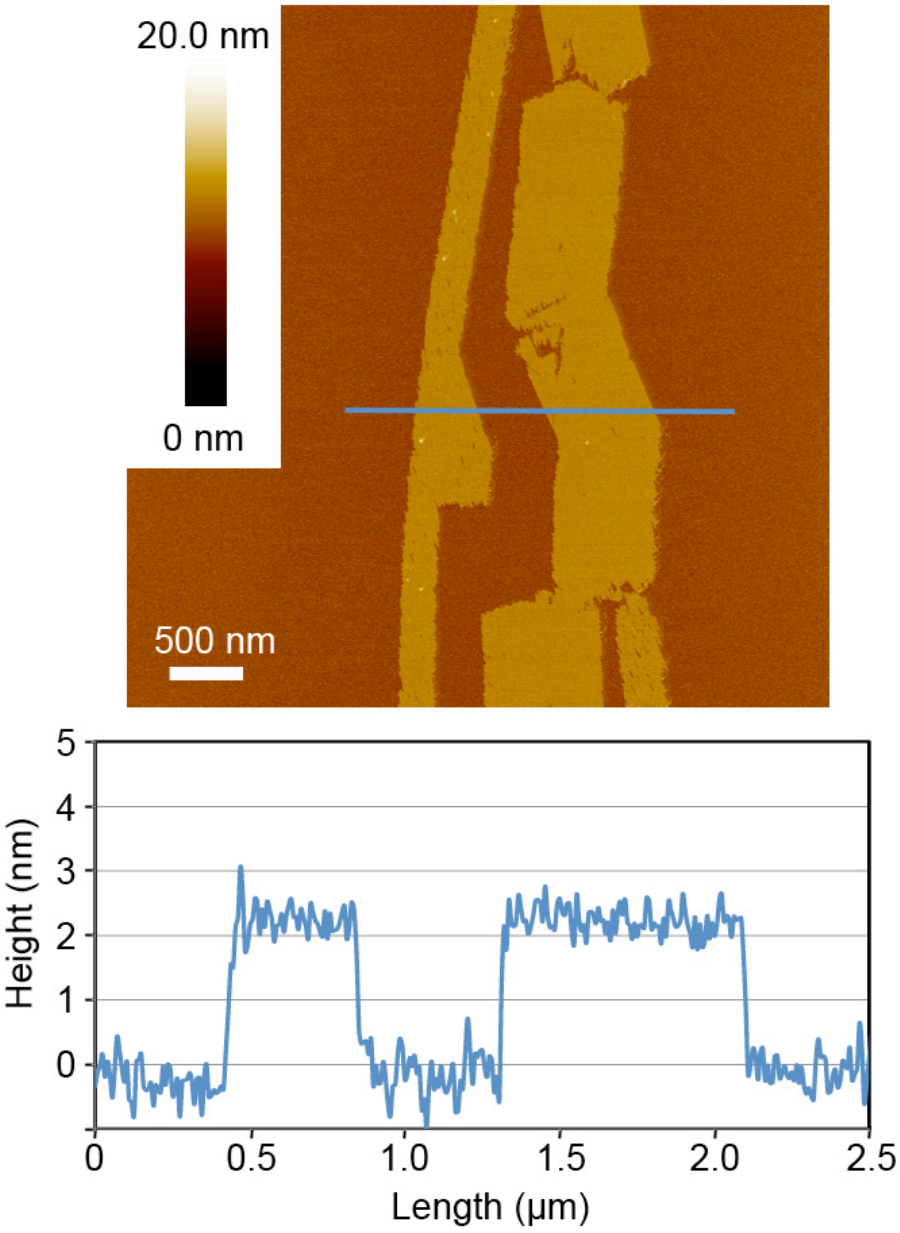
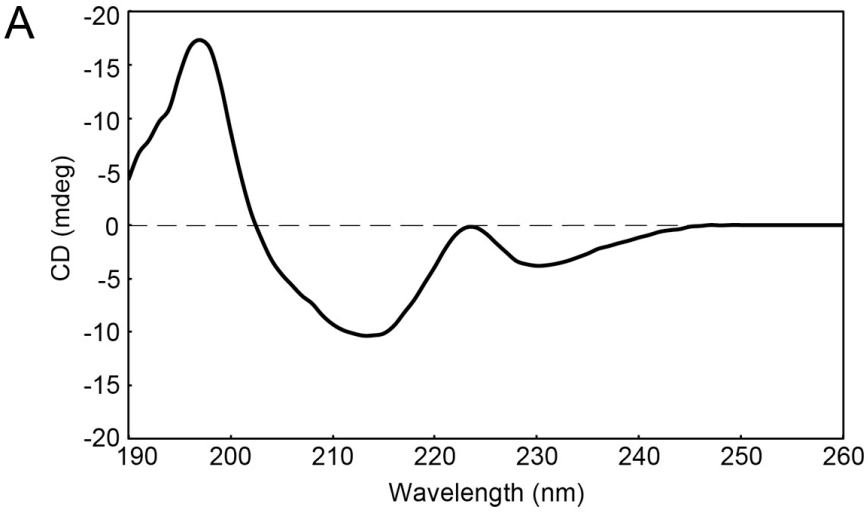


Fig. S2. AFM image of KLVFFAK nanosheet in liquid on mica. The lower panel is the linear section profile of the nanosheet.

A



B

	190-260 nm	195-260 nm	200-260 nm	205-260 nm	210-260 nm
Helix	8.10%	9.20%	7.90%	4.40%	4.80%
Antiparallel	53.50%	41.60%	39.60%	40.70%	42.10%
Parallel	4.30%	5.90%	5.80%	5.20%	5.30%
Beta-Turn	16.30%	17.50%	18.40%	19.50%	19.30%
Rndm. Coil	25.60%	31%	35.50%	34.80%	35.30%
Total Sum	107.80%	105.20%	107.20%	104.60%	106.80%

Fig. S3. Secondary structure content of KLVFFAK nanosheet. (*A*) Far-UV CD spectrum of KLVFFAK nanosheet. (*B*) Analysis of the secondary structure of KLVFFAK naosheet using CDNN (2). The major secondary structure element is antiparallel β-sheet. According to the CDNN instruction, most reliable is the lowest-wavelength prediction (190-260 nm). The other predictions, which do not differ too far from each other, indicate the reliability of the predictions. Also, the sum of the secondary structure elements deviate no more than 10 % from 100%, indicating the reliability of the prediction.

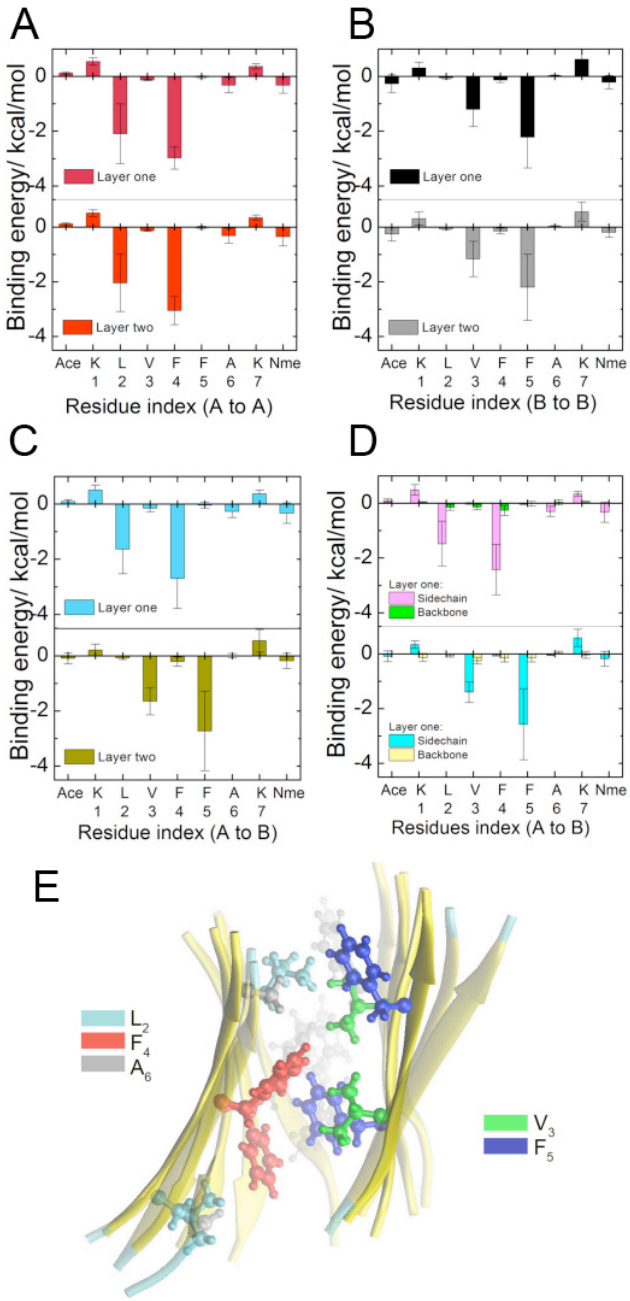


Fig. S4. Binding energy of each residue and the side-chain steric packing of KLVFFAK nanosheet. (*A*-*C*) Binding energy of each residue in one β-sheet layer with the residues in the other layer calculated for a 2x8 KLVFFAK β-sheet with three different packing interfaces (*A*) "A to A", (*B*) "B to B", and (*C*) "A to B". (*D*) Comparison of main chain and side chain binding energy of each residue in a 2x8 KLVFFAK β-sheet with "A to B" interface. The per residue binding energy is the average binding energy of the residue in the 2^{nd}∼7^{th} peptide chains in one layer. (*E*) A snapshot of a 2x8 KLVFFAK β-sheet with the "A to B" interface. The side chains of the five residues LVFFA are shown in ball-stick representations with different colors.

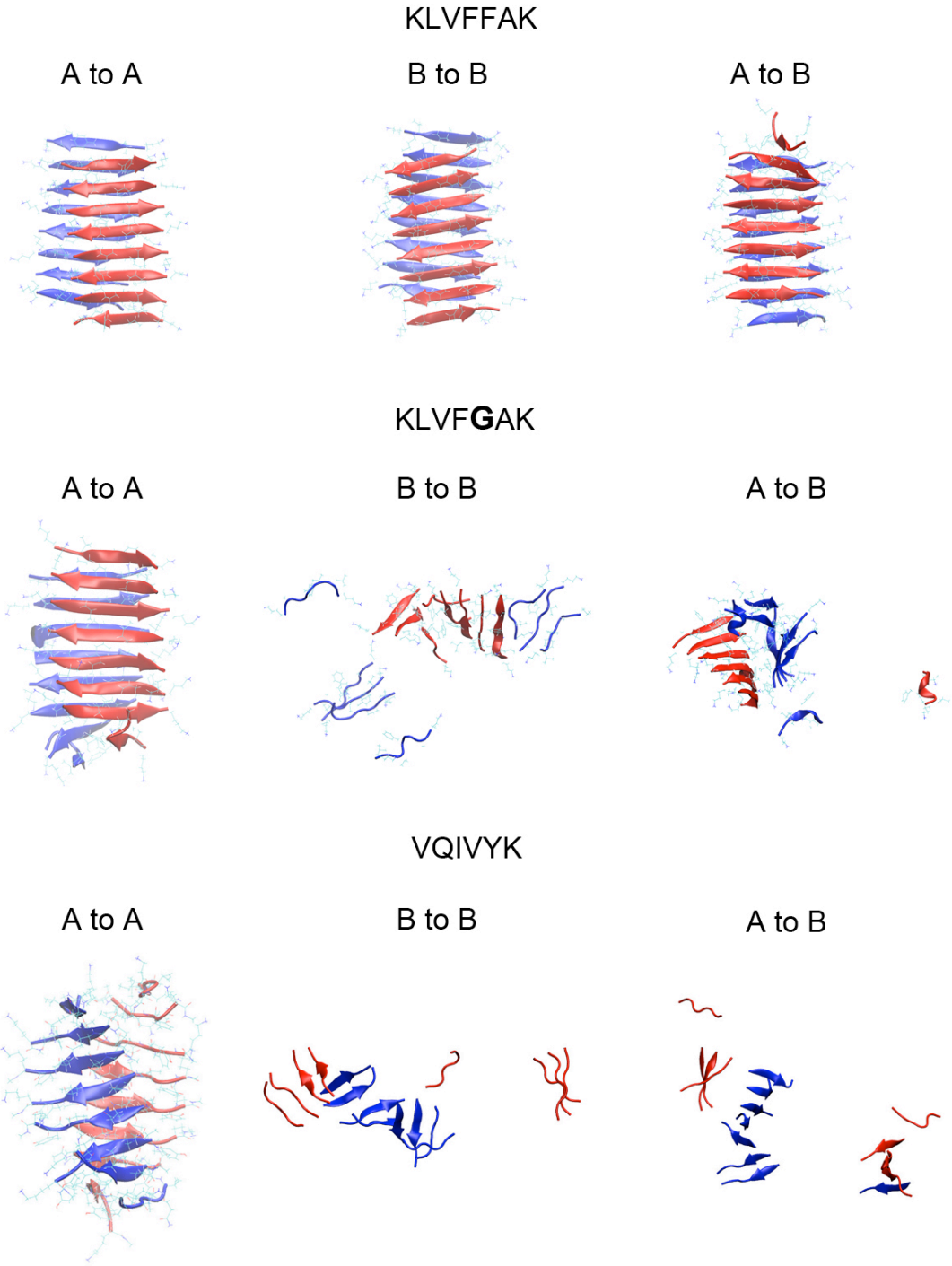


Fig. S5. MD simulations snapshots of the three zipper spine geometries: A to A, B to B, and A to B. Simulations started from 2 (two sheet layers) × 8 (eight strands per layer) β-sheet pairs. The three geometries are all stable for KLVFFAK, whereas only "A to A" is stable for KLVFGAK and VQIVYK.

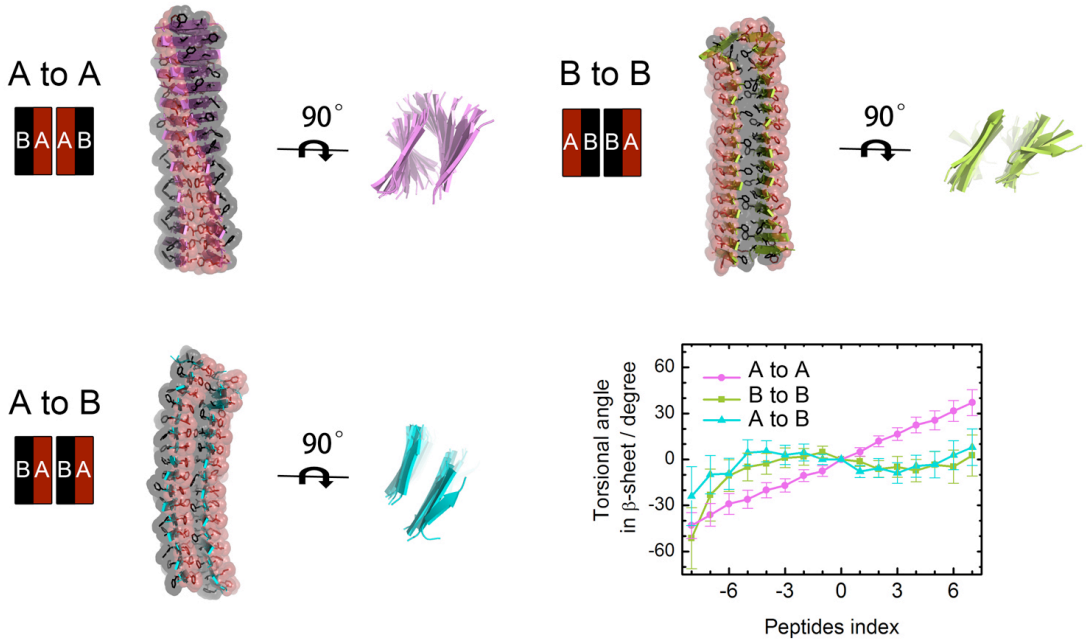


Fig. S6. Twist of KLVFFAK zipper spine. The twist of three possible zipper spines was predicted by MD simulations of a 2 × 16 β-sheet pair. The zipper spine with "A to A" interface is significantly twisted, whereas, the other two interfaces do not cause a twist of the zipper spine. The chart shows the torsional angles.

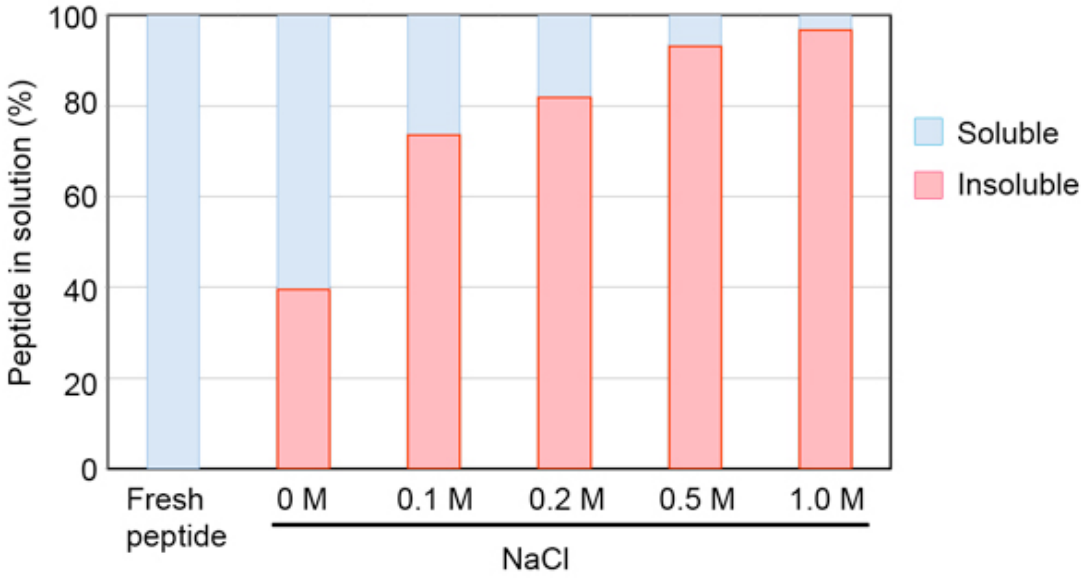


Fig. S7. Increasing KLVFFAK nanosheet yield with the addition of NaCl. The nanosheet yield was quantified by measuring the soluble peptide concentration in the supernatant at 205 nm using NANODROP 2000C (Thermo Scientific).

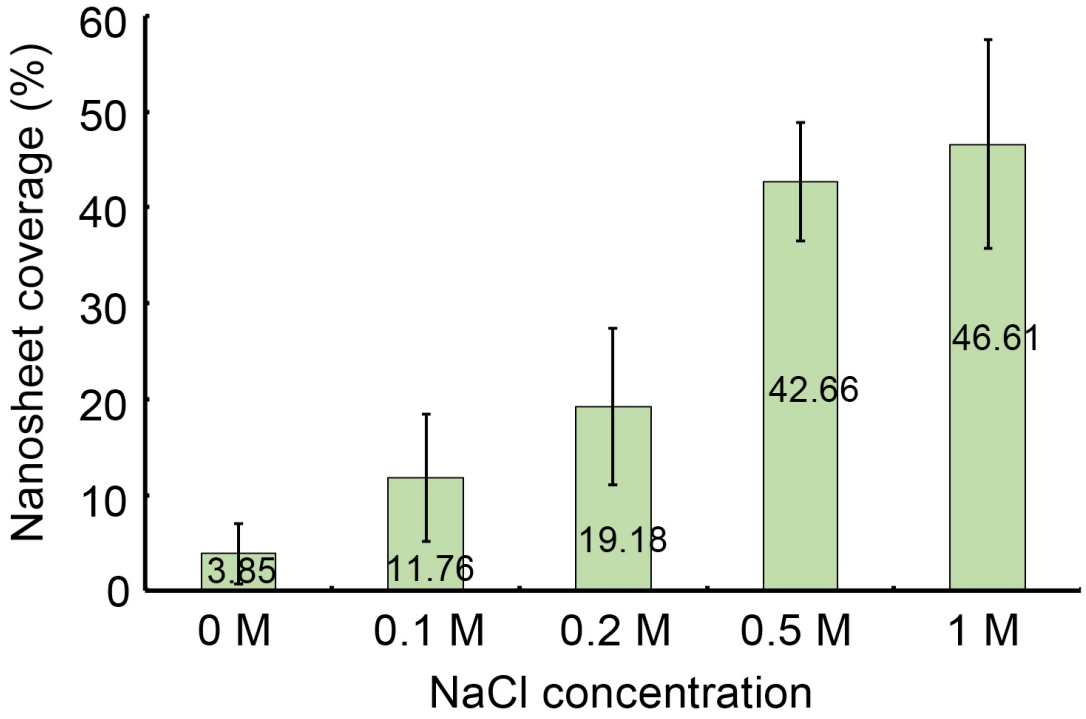


Fig. S8. Systematic quantification of nanosheet yield and width increase. For each NaCl concentration, 18 AFM images (10 μm × 10μm) were randomly sampled. In each image, the coverage of nanosheet is calculated. The average coverage of the 18 images for each NaCl concentration is labeled. The result showed that the coverage of nanosheet in the AFM images increases as the salt concentration increases.

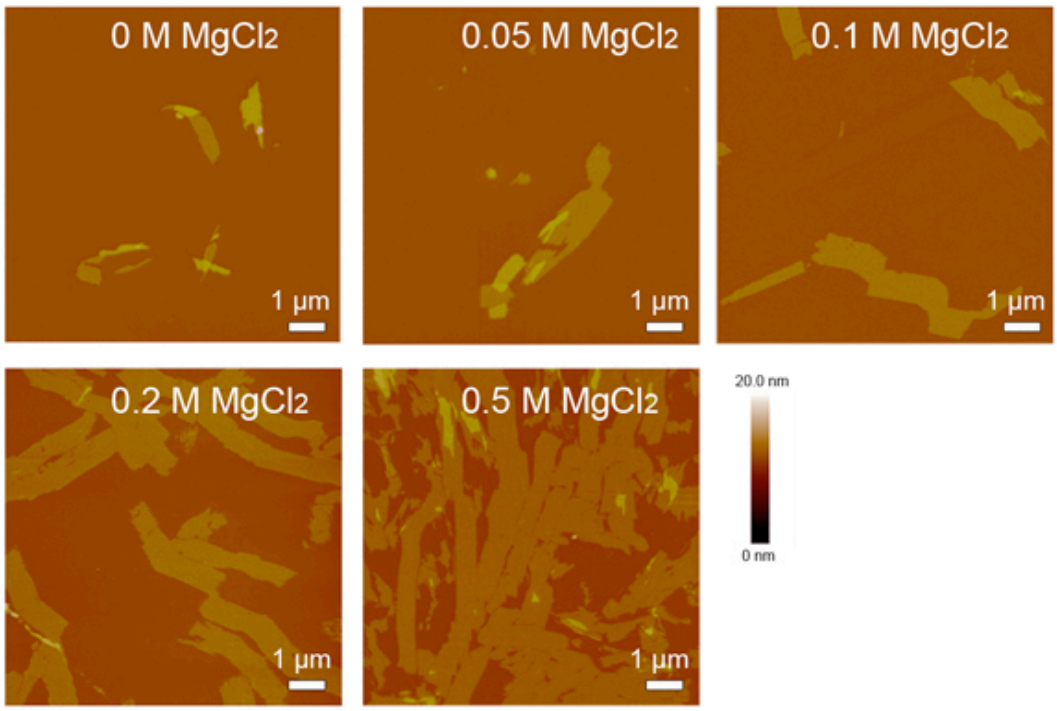


Fig. S9. AFM images of the KLVFFAK nanosheet formed under a gradient of $MgCl_2$ concentration. The nanosheet yield and size are growing as the $MgCl_2$ concentration increases.

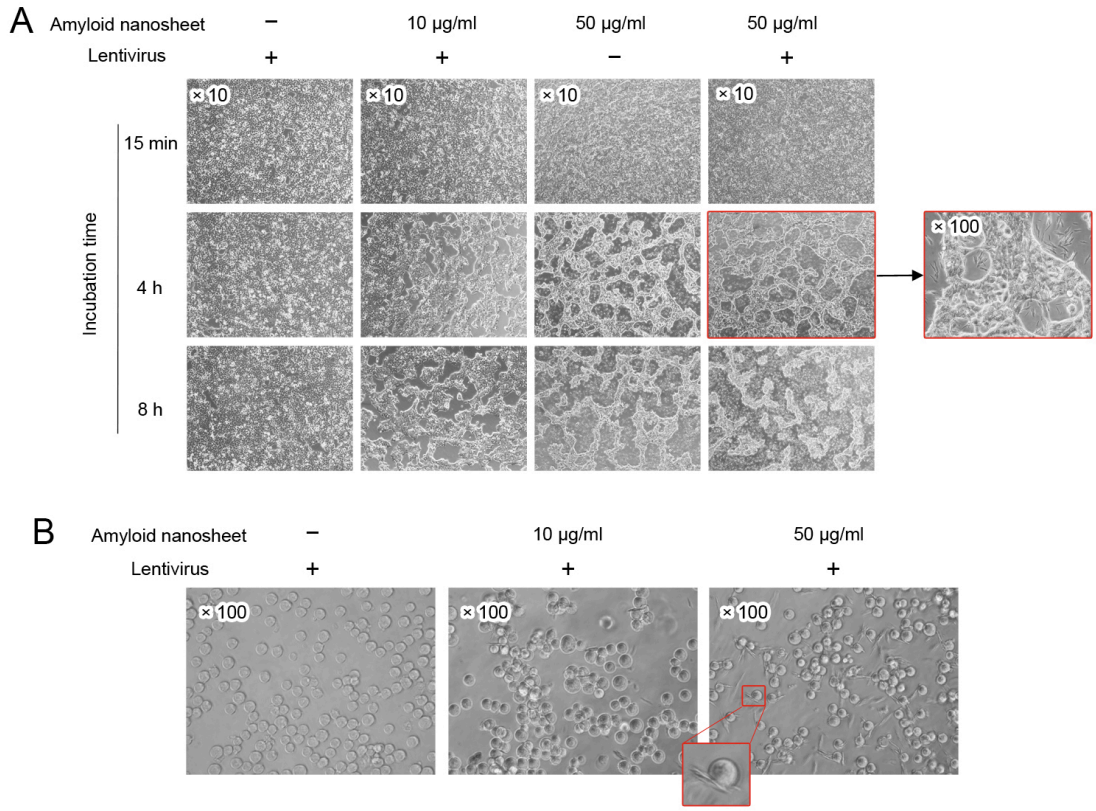
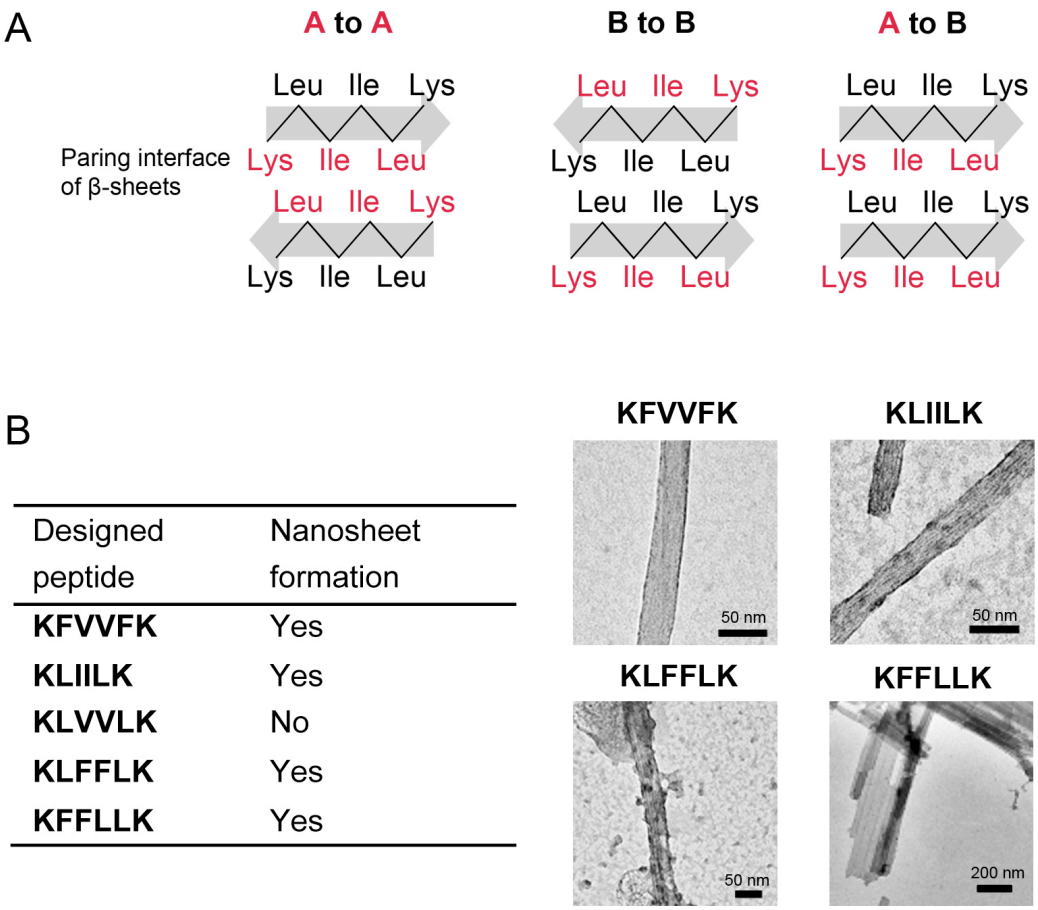


Fig. S10. Amyloid-like nanosheets guide the distribution of HEK293T cells. (*A*) On well surfaces, the nanosheets cause a grid-like distribution of cells rather than an even distribution without the nanosheet. (*B*) When suspended in culture after trypsin treatment, it is clear that the cells and the nanosheets adhere to each other. The addition of lentivirus does not interfere with the cell distribution.



Designed peptide	Nanosheet formation
KFVVFK	Yes
KLIILK	Yes
KLVVLK	No
KLFFLK	Yes
KFFLLK	Yes

Fig. S11. *De novo* design of nanosheet-forming peptides. (*A*) Illustration of design strategy using the designed nanosheet-forming peptide KLIILK as an example. An energetic requirement for nanosheet formation is either "A to A" and "B to B" interfaces are equally preferred, or "A to B" interface is preferred. The requirement was met in a simple way, that is, the peptides were designed to have the same residues to compose sides A and B of the β-strand. As shown in (*A*), both sides of KLIILK are composed of residues Lys, Ile, and Leu, thus the three interfaces ("A to A", "B to B", and "A to B") can be equivalent. (*B*) Five peptides were designed. Four of them form nanosheet-like structures observed by TEM.