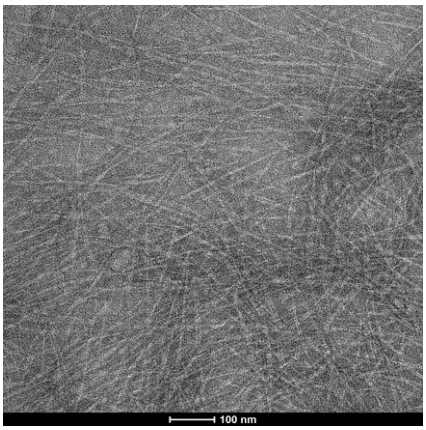
# Supplementary Information for

## Real-time Visualization of Collagen Assembly Uncovers Metastable Properties in Hierarchical Organization

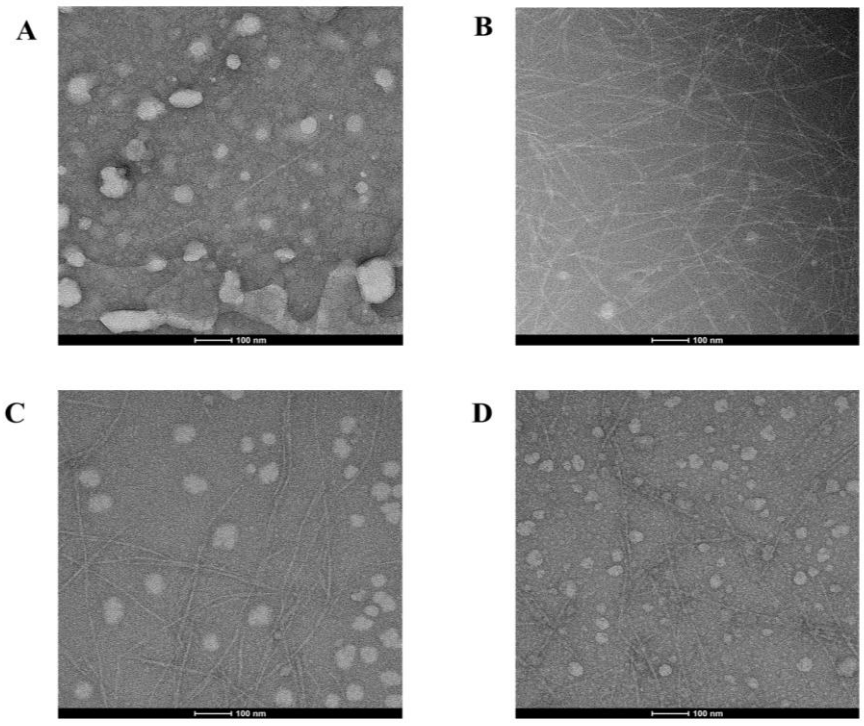
Bohuan Fang *et al.*

Corresponding authors: Daidi Fan, fandaidi@nwu.edu.cn，
Lujia Zhang, ljzhang@chem.ecnu.edu.cn

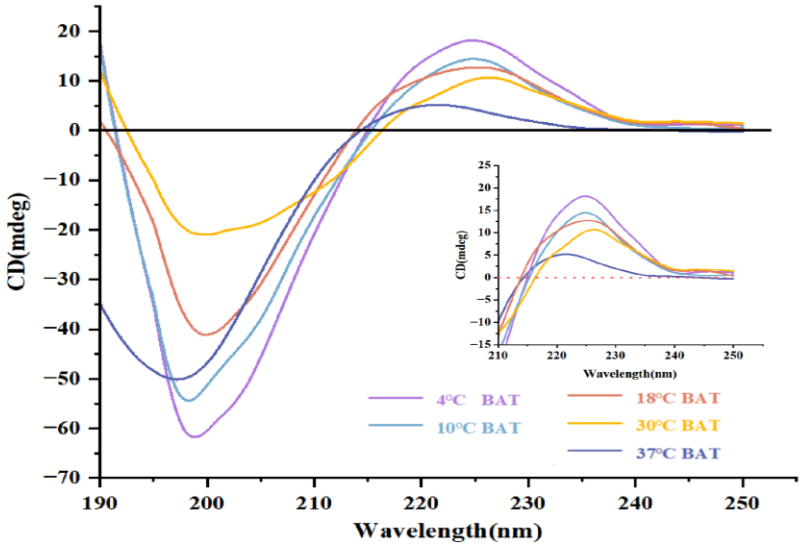
The PDF file includes:
Supplementary Fig. 1 to Fig. 17
Supplementary Table 1 to Table 3
Python scripts for model translation



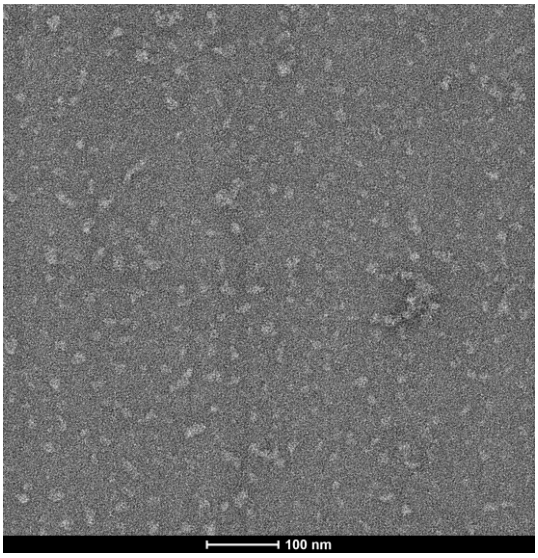

**Supplementary Fig. 1 CF-1552 assembly pre-experiment**
For pre-experiment, the sample was prepared at 10 mg/mL in 10 mM phosphate buffer with 150 mM NaCl (pH=7) and incubated at 18 °C for 24 h. The morphological observation indicated that CF-1552 exhibited strong assembly capability, with the assemblies displaying a well-defined fibrous structure. (Related to Results Model Protein-Recombinant Collagen CF-1552)



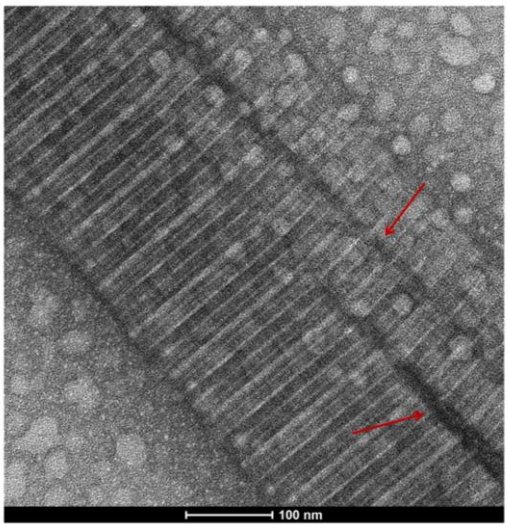

**Supplementary Fig. 2 Regulation of CF-1552 self-assembly**
(A)and(B) TEM images of the CF-1552 assemblies at 4 ℃
(C) TEM images of the CF-1552 assemblies at pH 8
(D) TEM images of the CF-1552 assemblies at pH 9
(Related to Fig. 2)



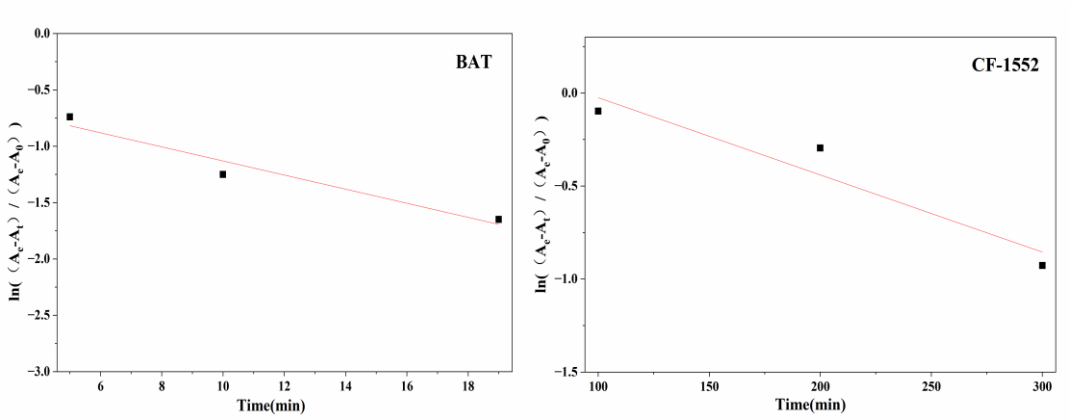

**Supplementary Fig. 3 Circular dichroism (CD) spectra of BAT assemblies at various temperatures**
(Related to Fig. 3A)



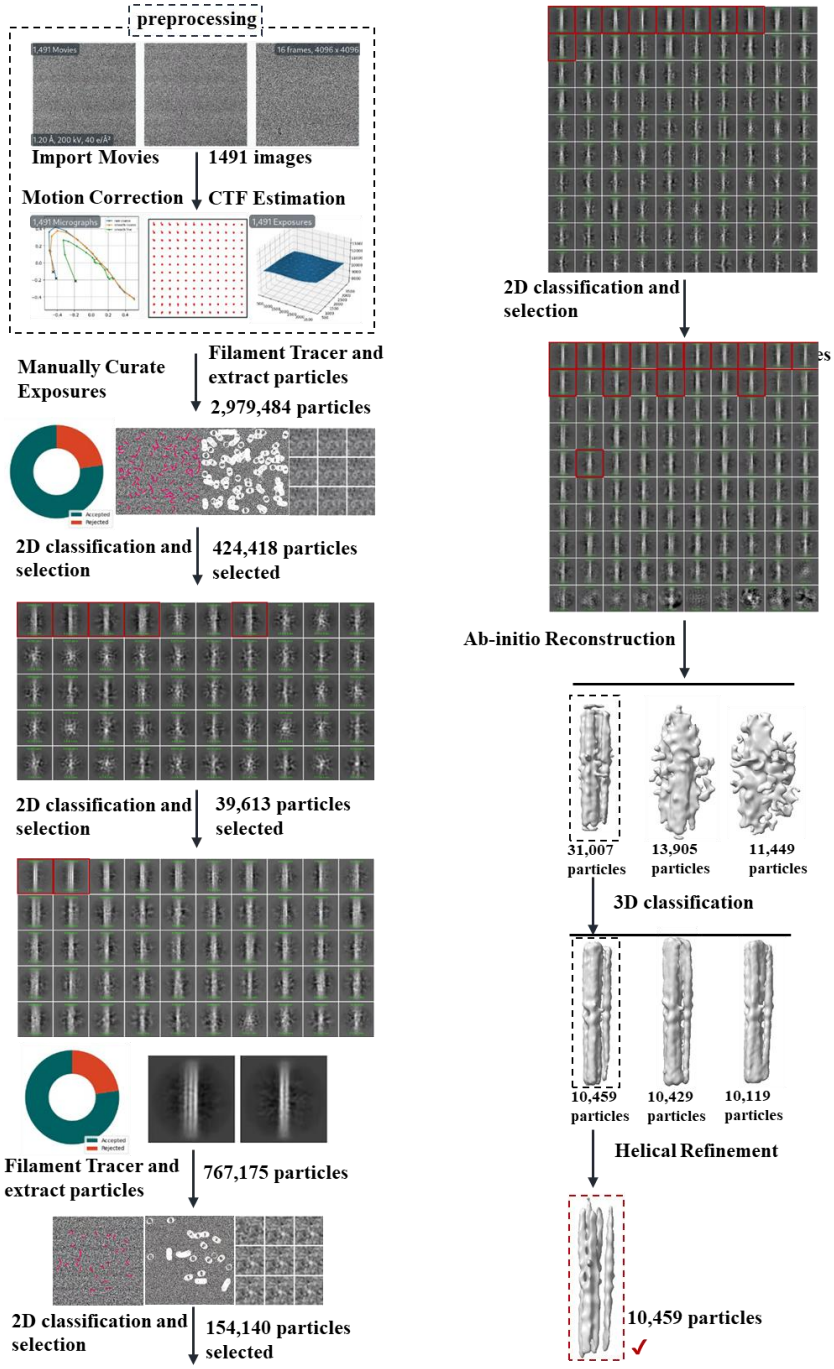

**Supplementary Fig. 4 CF-1552 assemblies at the initial time point (0 min)**
(Related to Fig. 3B)



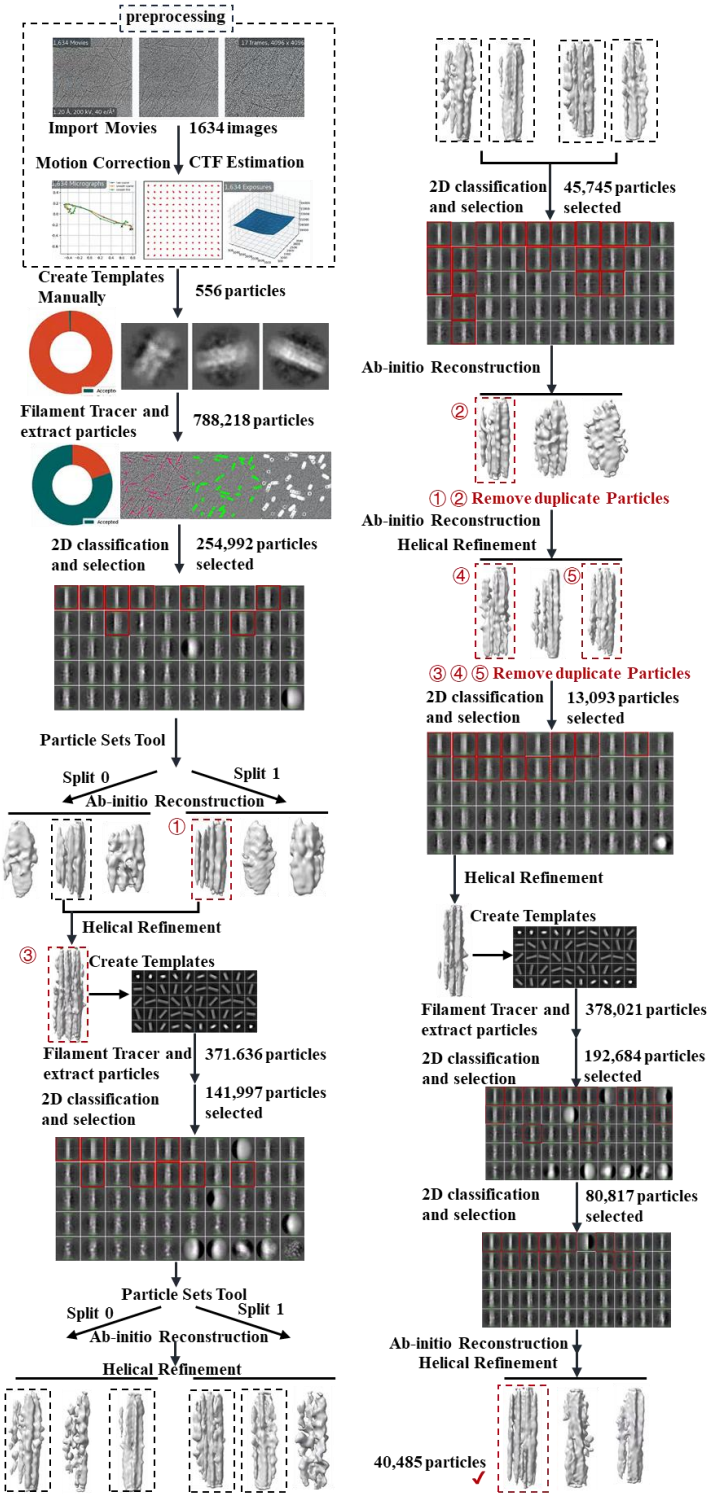

**Supplementary Fig. 5 The lateral growth of fibirls in the growth phase**
Small fibrils underwent lateral interactions, merging to form larger D-banded structures (Related to Fig. 4B 0.5 h)




**Supplementary Fig. 6 Assembly kinetics of BAT and CF-1552**
**(**Related to Fig. 5**)**



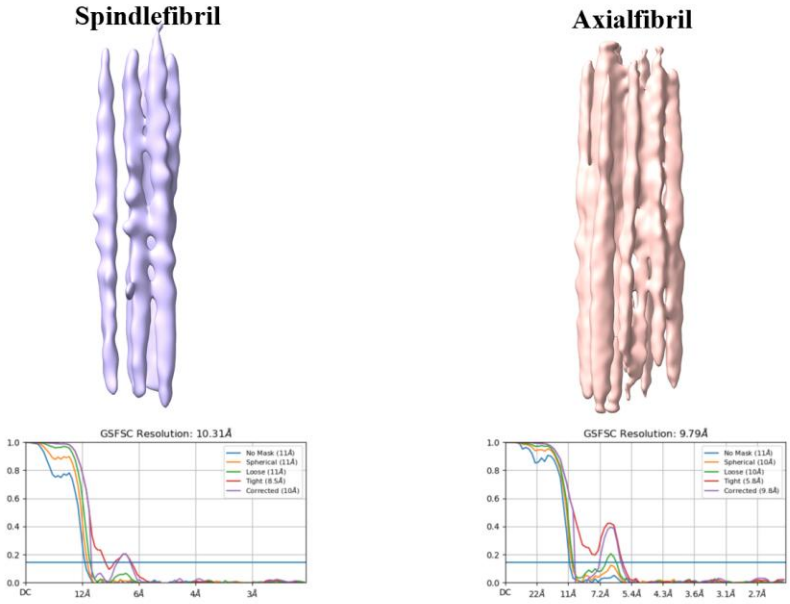

**Supplementary Fig. 7 Cryo-EM image processing workflow of spindlefibril**
After image preprocessing, the Filament Tracer was first used for template-free particle selection. Taking two rounds of 2D Classification to obtain suitable 2D templates for auto-picking. The selected particles underwent two rounds of 2D Classification before being used for 3D reconstruction. Finally, 3D classification and Helix Refinement were processed to obtain the final map
(Related to Fig. 5B)




**Supplementary Fig. 8 Cryo-EM image processing workflow of axialfibirl**

After image preprocessing, a small subset ofthe micrographs was chosen to generate reliable 2D templates for particleauto-picking, Then the selected particles were deviled into two splits for Ab-initio Reconstruction and HelixRefimement. A fter several rounds of 2 Classification-Ab-initio Reconstruction-Helix Refimement, the final map of axialfibril was obtained

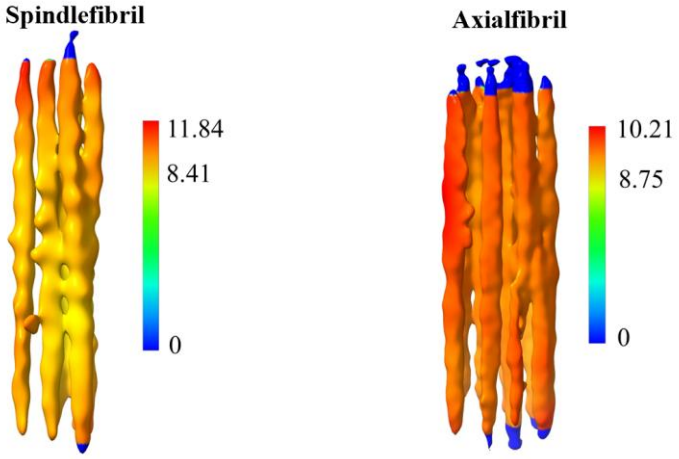
(Related to Fig. 5B)




**Supplementary Fig. 9 Fourier shell correlation (FSC) curves of Spindlefibril and Axialfibril**

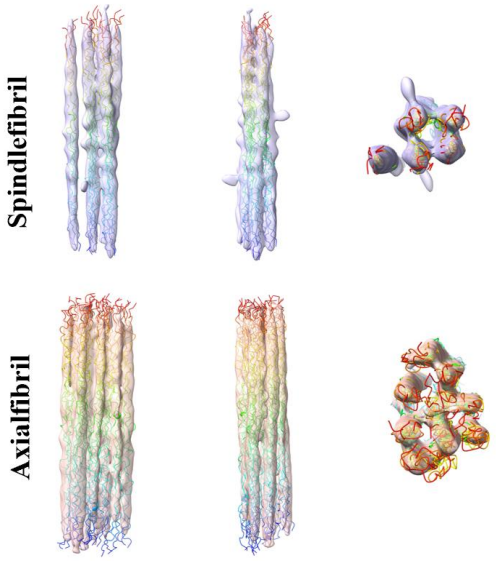
The resolutions were estimated using the 0.143 FSC criteria.

(Related to Fig. 5B)

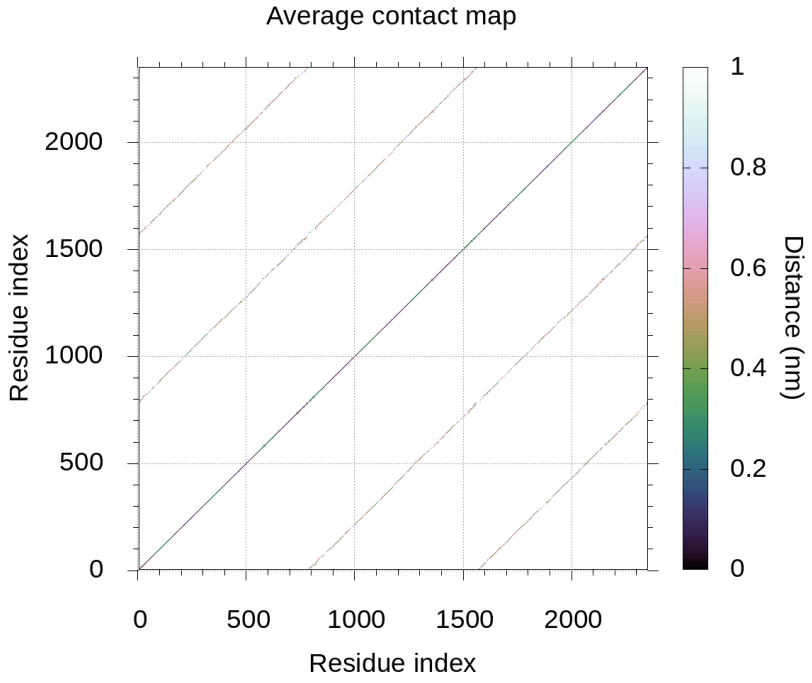



**Supplementary Fig. 10 The maps colored to local resolution of Spindlefibril and Axialfibril from UCSF ChimeraX**

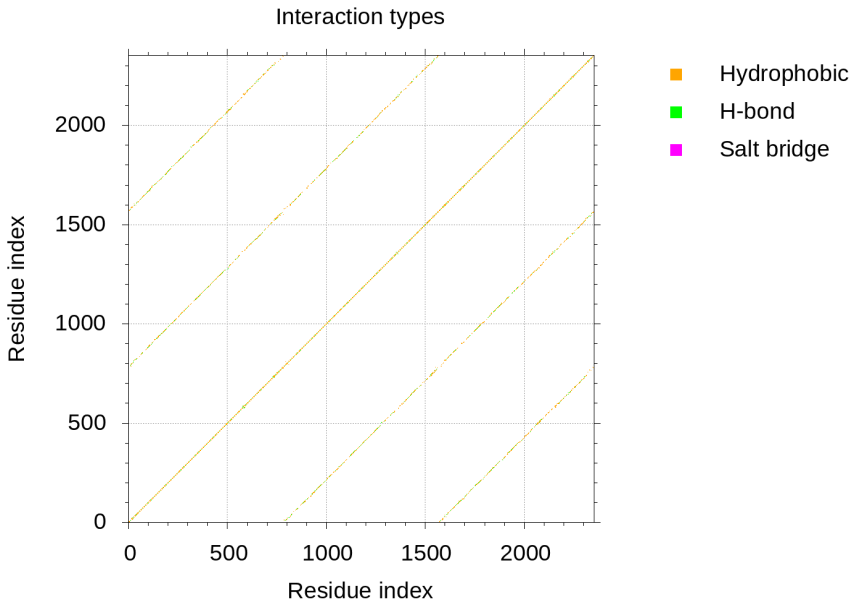
(Related to Fig. 5B)




**Supplementary Fig. 11 Different views of the model-to-map fit**
(Related to Fig. 5B)

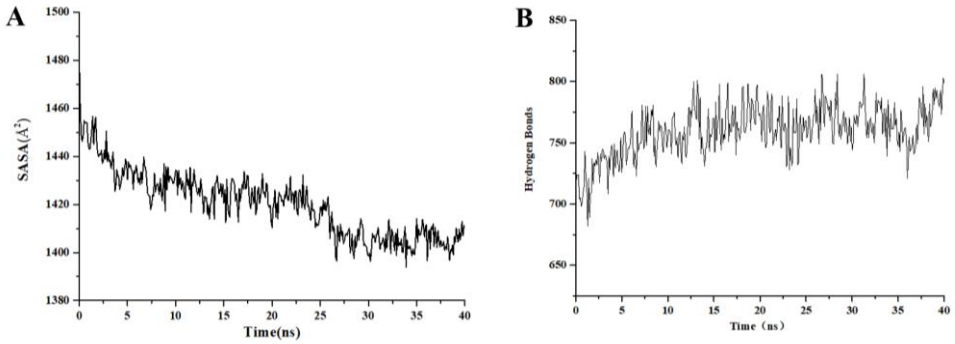



**Supplementary Fig. 12 A full-size version of the average interchain contact map derived from MD simulations**
(Related to Fig. 6D)




**Supplementary Fig. 13 A full-size version of the interaction type map of the three chains**

(Related to Fig. 6E)

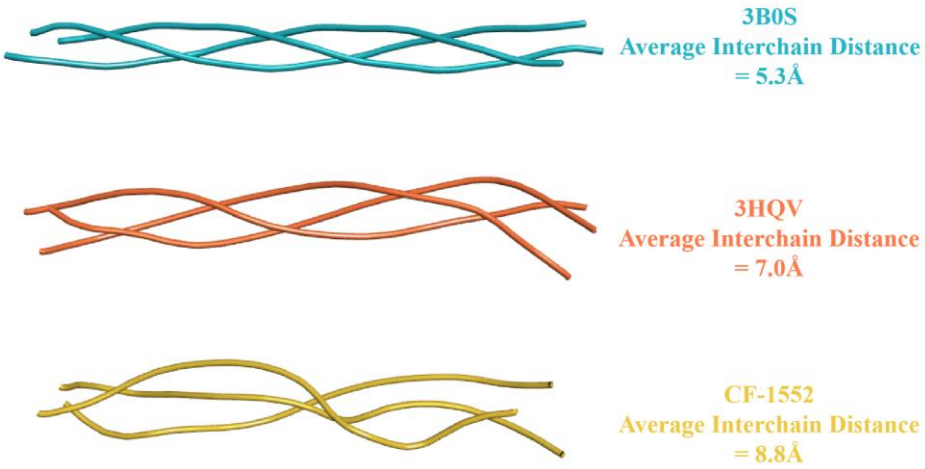



**Supplementary Fig. 14 Molecular dynamic simulation of triple helix**

(A) Variation of the solvent accessible surface area (SASA) of triple-helix along MD simulations,

(B) Hydrogen bond numbers of triple-helix along MD simulations

(Related to Fig. 6)



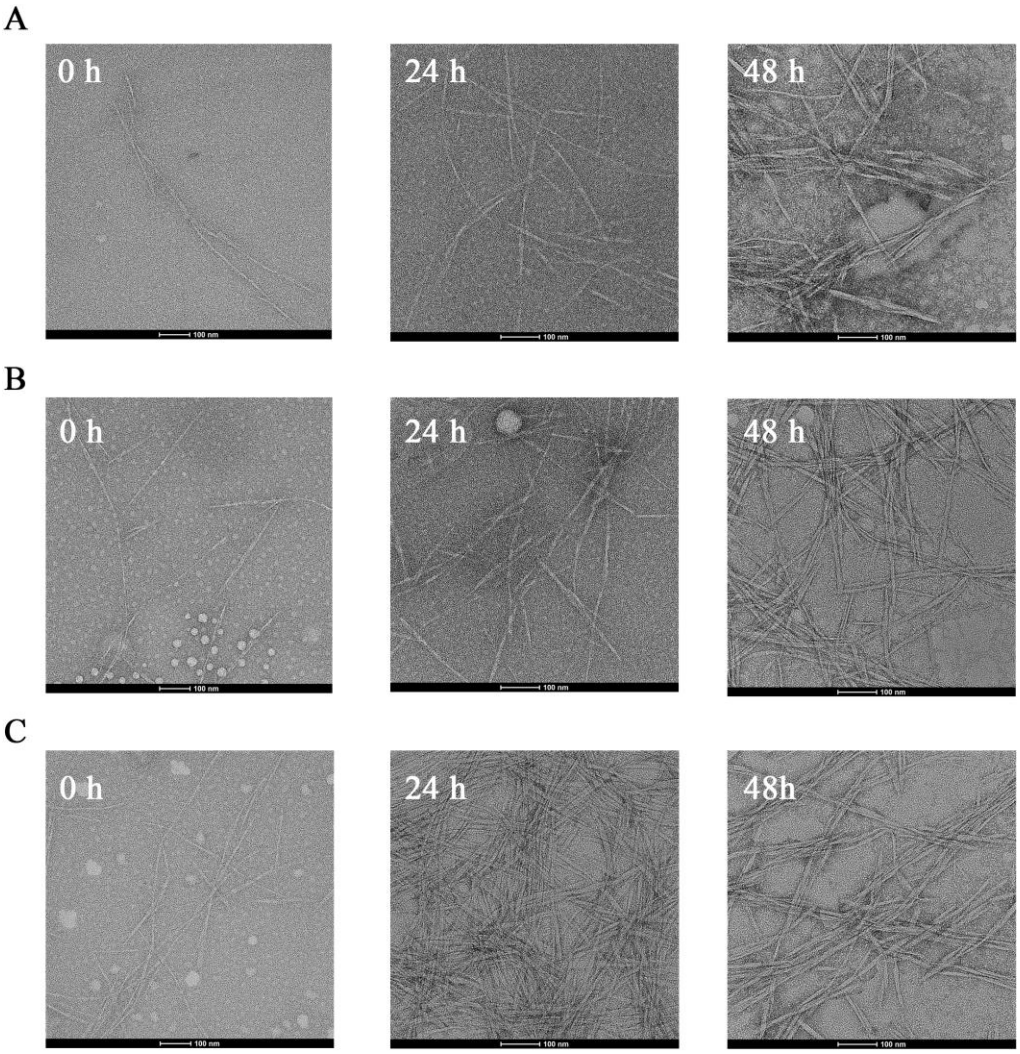

**Supplementary Fig. 15 Triple helix structure alignment**

**(A)** Natural collagen（PDB:3HQV）

**(B)** $(GPO)_9$ (PDB:3B0S)

**(C)** CF-1552

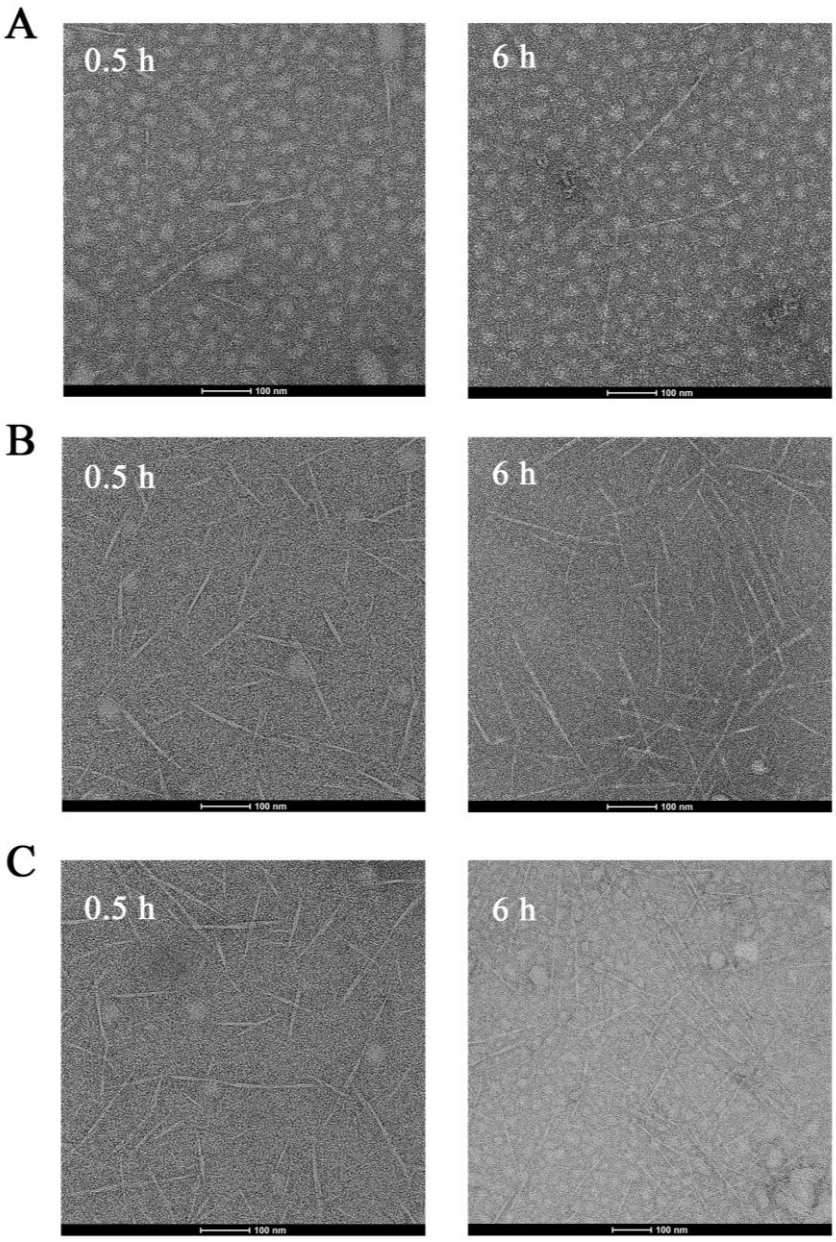
(Related to Fig. 6)




**Supplementary Fig. 16 The stability of spindlefibrils at different concentrations during the cell culture process**

(A) 0.25 mg/mL, (B)1 mg/mL, (C) 4 mg/mL

(Related to Fig. 7)




**Supplementary Fig. 17 The effect of concentration on CF-1552 assembly**
(A) 1 mg/mL，(B) 3 mg/mL, (C) 4 mg/mL
(Related to Method: In Vitro Regulation of CF-1552 Assembly)

**Supplementary Table 1** Secondary structure assignment and content of CF-1552 and BAT assemblies

| name | Peak position(cm-1) | Band area (%) | Assignment |
|---|---|---|---|
| CF-1552 assemblies | 1605 | 5.9 | β-fold |
| | 1614 | 10.35 | β-fold |
| | 1621 | 2.94 | β-fold |
| | 1631 | 21.37 | β-fold |
| | 1643 | 7.82 | β-fold |
| | 1650 | 6.9 | PP II |
| | 1660 | 16.48 | PP II |
| | 1670 | 8.08 | β-turn |
| | 1681 | 13.07 | β-turn |
| | 1694 | 7.09 | β-turn |
| BAT assemblies | 1617 | 25.03 | β-fold |
| | 1634 | 23.1 | β-fold |
| | 1649 | 13.87 | PP II |
| | 1660 | 20.34 | PP II |
| | 1674 | 10.04 | β-turn |
| | 1691 | 7.27 | β-turn |

**Supplementary Table 2** Contribution of each component of interaction for triple helix using MMPBSA calculation.

| | Van-der Waal energy(kJ/mol) | Electrostatic energy(kJ/mol) | Polar solvation energy (kJ/mol) | SASA energy (kJ/mol) | ΔH (kJ/mol) | ΔG (kJ/mol) |
|---|---|---|---|---|---|---|
| Triple helix | -1.397±0.25 | -5.592±1.01 | 2.516±0.75 | -1.157±0.24 | -5.630±1.24 | -4.745±1.24 |

**Supplementary Table 3** Cryo-EM data collection, refinement and validation statistics

| | #1 spindlefibirl (EMD-60543) (PDB 8ZXL) | #2 axialfibirl (EMD-60913) (PDB 9IUS) |
|---|---|---|
| **Data collection and processing** | | |
| Magnification | 120k | 120k |
| Voltage (kV) | 200 | 200 |
| Electron exposure (e–/Å$^2$) | 40 | 40 |
| Defocus range (μm) | -1.5,-1.8,-2.0,-2.3,-2.5,-2.8 | -1.5,-1.8,-2.0,-2.3,-2.5,-2.8 |
| Pixel size (Å) | 1.2 | 1.2 |
| Symmetry imposed | C1 | C1 |
| Initial particle images (no.) | 2979484 | 788218 |
| Final particle images (no.) | 10459 | 40485 |
| Map resolution (Å) | 8.6 | 9.8 |
| FSC threshold | | |
| Map resolution range (Å) | 8.5-11 | 5.8-11 |
| | | |
| **Refinement** | | |
| Model composition | | |
| Non-hydrogen atoms | 6178 | 12812 |
| Protein residues | 1052 | 2183 |
| Ligands | 0 | 0 |
| *B* factors (Å$^2$) | | |
| Protein | 149.45/1077.60/619.71 | 207.03/999.99/715.44 |
| Ligand | --- | --- |
| R.m.s. deviations | | |
| Bond lengths (Å) | 0.18 | 0.19 |
| Bond angles (°) | 0.43 | 0.47 |
| Validation | | |
| MolProbity score | 3.74 | 3.84 |
| Clashscore | 249 | 289.69 |
| Poor rotamers (%) | 0.00 | 0.01 |
| Ramachandran plot | | |
| Favored (%) | 64.29 | 59.30 |
| Allowed (%) | 34.15 | 39.24 |
| Disallowed (%) | 1.57 | 1.46 |

**Python scripts for model translation**

```
from Bio import PDB

def decrease_coordinate(pdb_file_path, column_to_change, decrement,
output_file_path):
    parser = PDB.PDBParser(QUIET=True)
    structure = parser.get_structure('protein', pdb_file_path)
    model = structure[0]
    atoms = PDB.Selection.unfold_entities(model, 'A')
    for atom in atoms:
        current_coord = atom.get_coord()
        current_coord[column_to_change] -= decrement
        atom.set_coord(current_coord)
    io = PDB.PDBIO()
    io.set_structure(structure)
    io.save(output_file_path)
    print(f"Coordinates successfully decreased and saved to {output_file_path}")
if __name__ == "__main__":
    input_pdb_file = "2.pdb"
    column_to_change = 1
    decrement = 50
    output_pdb_file = "output2.pdb"
    decrease_coordinate(input_pdb_file, column_to_change, decrement,
output_pdb_file)
```