Supporting information for:

A high-throughput system combining microfluidic hydrogel droplets with deep learning for screening the antisolvent-crystallization conditions of active pharmaceutical ingredient

Zhenning Su [a], Jinxu He[a], Peipei Zhou[a], Lu Huang[a], and Jianhua Zhou [a, b*]

[a] Key Laboratory of Sensing Technology and Biomedical Instruments of Guangdong Province, School of Biomedical Engineering, Sun Yat-sen University, Guangzhou 510275, China

[b] Division of Engineering in Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Cambridge, MA 02139, USA

*Corresponding author:

Tel.: +86 20 39387890; Fax: +86 20 39387890.

Email: zhoujh33@mail.sysu.edu.cn (J. H. Zhou);

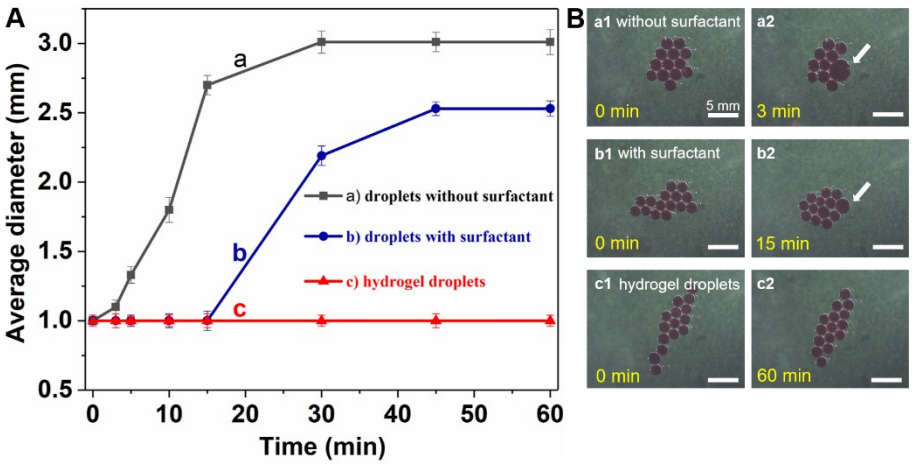

Fig. S1 Stability of three types of droplets. (A) Time-dependent average sizes of different types of droplets. (B) Optical images of droplets over time, (a1-a2) The water-in-oil droplets without surfactant began to fuse at 3 min. (b1-b2) The water-in-oil droplets with Span80 as surfactant began to fuse at 15 min. (c1-c2) The hydrogel droplets remained stable for 60 min. The scale bars in the images are all 5 mm.

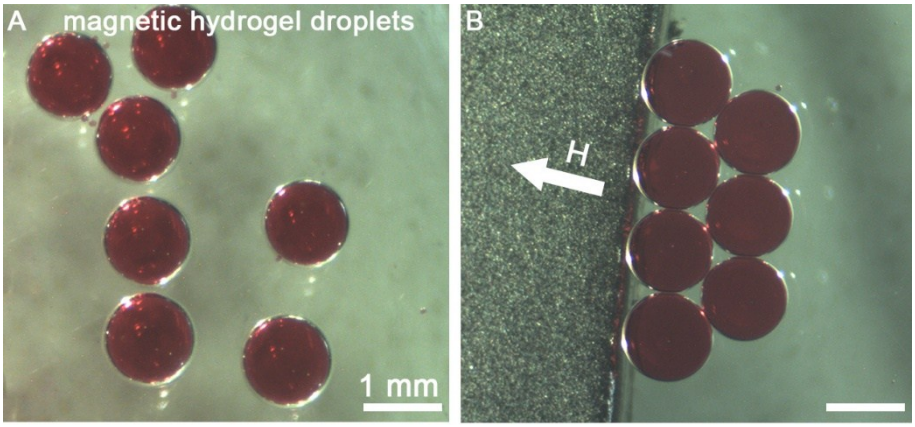

Fig. S2 Control movement of magnetic hydrogel droplets under a magnetic field. (A) Hydrogel droplets before exposure to a magnet. (B) Magnetic hydrogel droplets under a magnetic field (as indicated by “H”). The scale bars in the images are 1 mm.

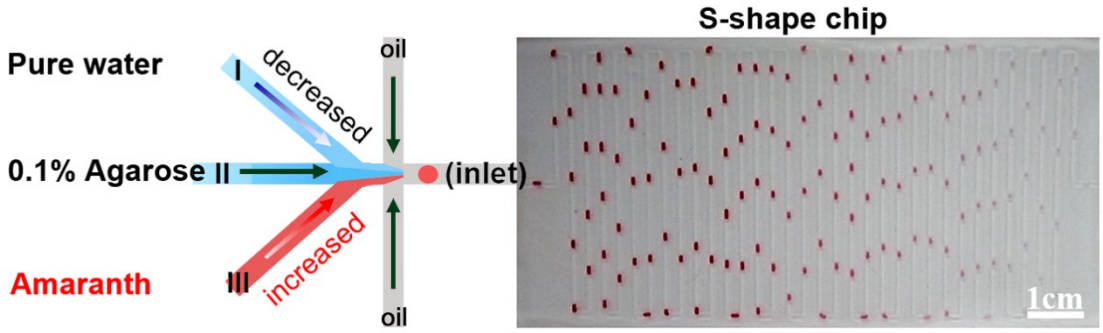

Fig. S3 Demonstrating the control of the concentration of amaranth in hydrogel droplets. The darkness of the color indicates that the concentration of amaranth (as red reagent) in hydrogel droplets can be controlled by altering the flow rates of dispersed phase. We injected pure water, 0.1% agarose solution, and 0.1% amaranth solution into channel I, channel II and channel III, respectively. According to the parameters of flow rates in the previous step for preparing hydrogel droplets, we set the flow rate of the different channels on the syringe pumps. Then, a continuous phase of oil was used to shear the dispersed phase into droplets on the microfluidic chip. After the temperature of the droplets dropped to the room temperature, the agarose solution was transformed into the state of semi-solid, and hydrogel droplets with different red darknesses were formed.

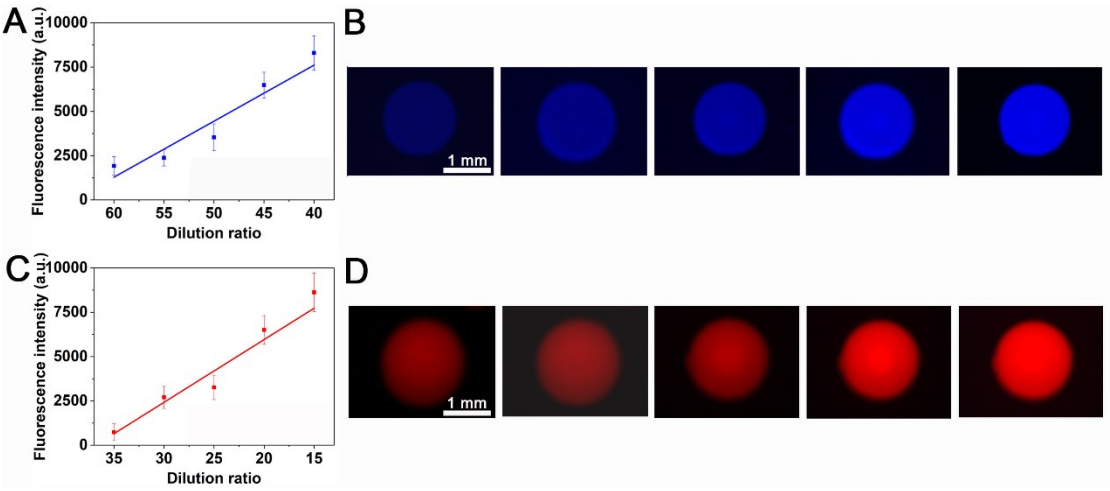

Fig. S4 (A) Standard curve showing the relationship between the concentration and fluorescence intensity of blue fluorescent dye (*i.e.* Hoechst 33342), y=20259.5-316.24x, R^2=0.8829. (B) Fluorescence images of hydrogel droplets containing blue dye with different concentrations. (C) Standard curve showing the relationship between the concentration

and the fluorescence intensity of red fluorescent dyes (*i.e.* propidium iodide), y=13037.1-353.89x, R^2=0.9257. (D) Fluorescence images of hydrogel droplets containing red dye with different concentrations.

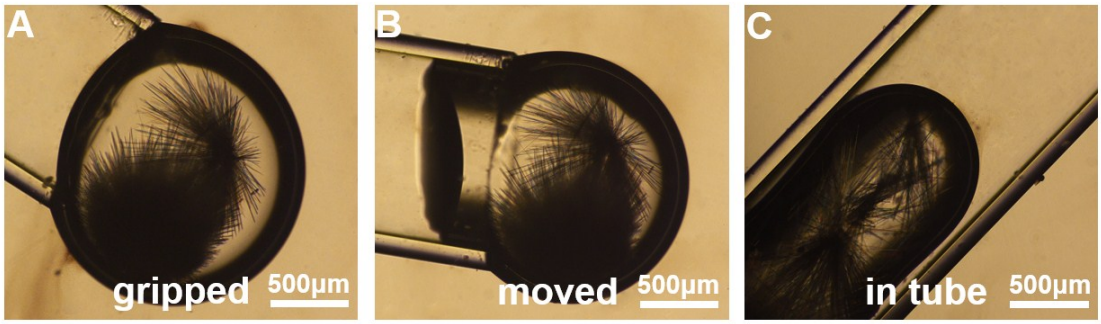

Fig. S5 Hydrogel droplets which contained indomethacin crystals could be moved and manipulated by a thin glass tube.

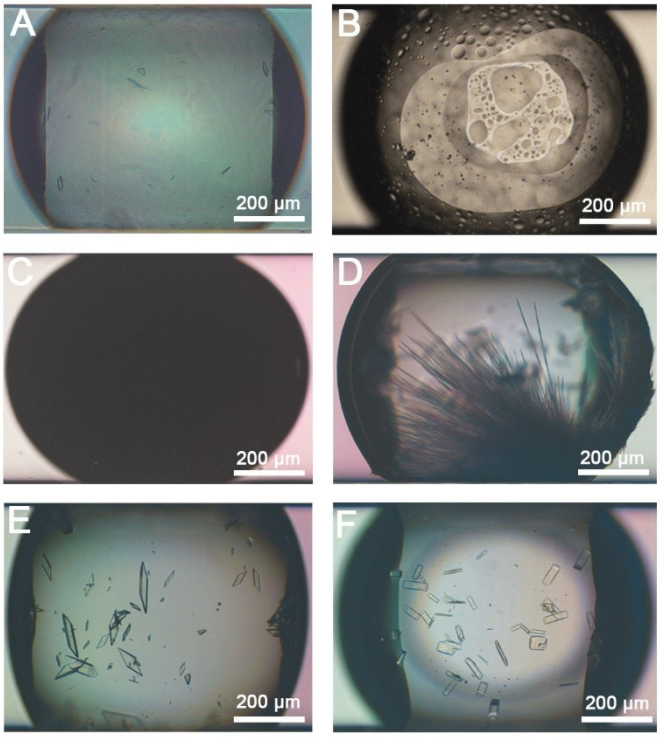

Fig.S6 Six representative images of droplets in the "S"-shape collection chip. (A) The image of an empty hydrogel droplet. (B) The image of a hydrogel droplet which contained a small amount of air bubbles. (C) The image of a hydrogel droplet which contained Jelly-

like phase. (D-F) The images of hydrogel droplets that contained wire, sheet and rod crystals of API, respectively.

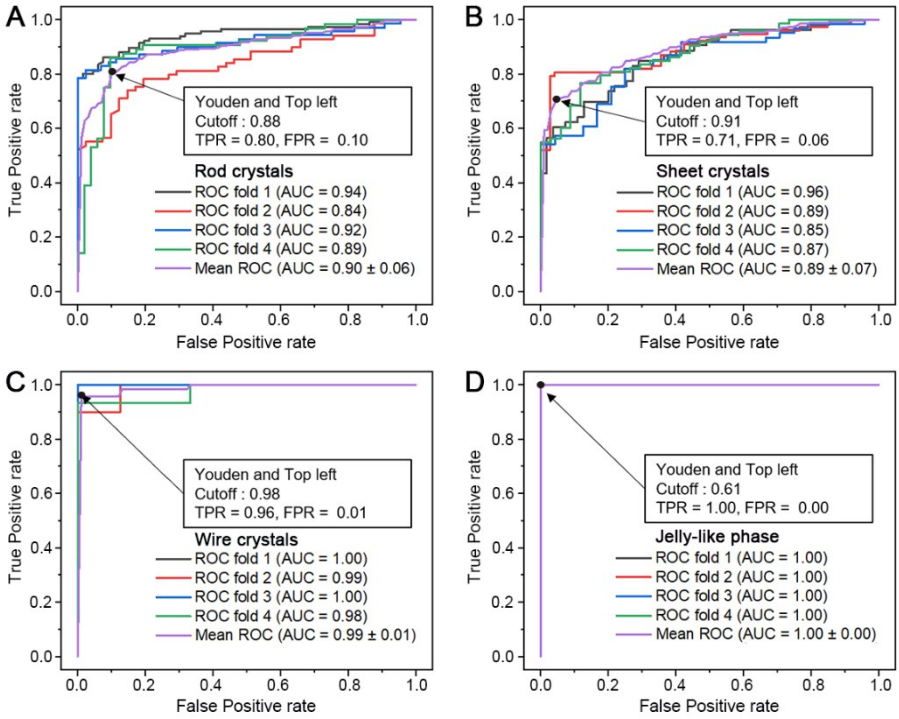

Fig. S7 ROC curves for each crystal morphology. (A) ROC curves of rod crystals. (B) ROC curves of sheet crystals. (C) ROC curves of wire crystals. (D) ROC curves of Jelly-like phase. In A, the AUC values of ROC fold 1, 2, 3 and 4 are 0.94, 0.84, 0.92 and 0.89, respectively, for rod crystal, and the AUC value of the mean ROC is 0.90±0.06, indicating that the neural network used in the experiment can predict the rod crystals well; the inset shows that, at the maximum Youden value, the difference between the sensitivity (*i.e.* True Positive rate) and False Positive rate is the largest in the curve, which means that the prediction sensitivity of neural network for rod crystals is 80% and the false positive rate is 10%. In this case, the corresponding threshold value (*i.e.* cutoff) is 0.88.

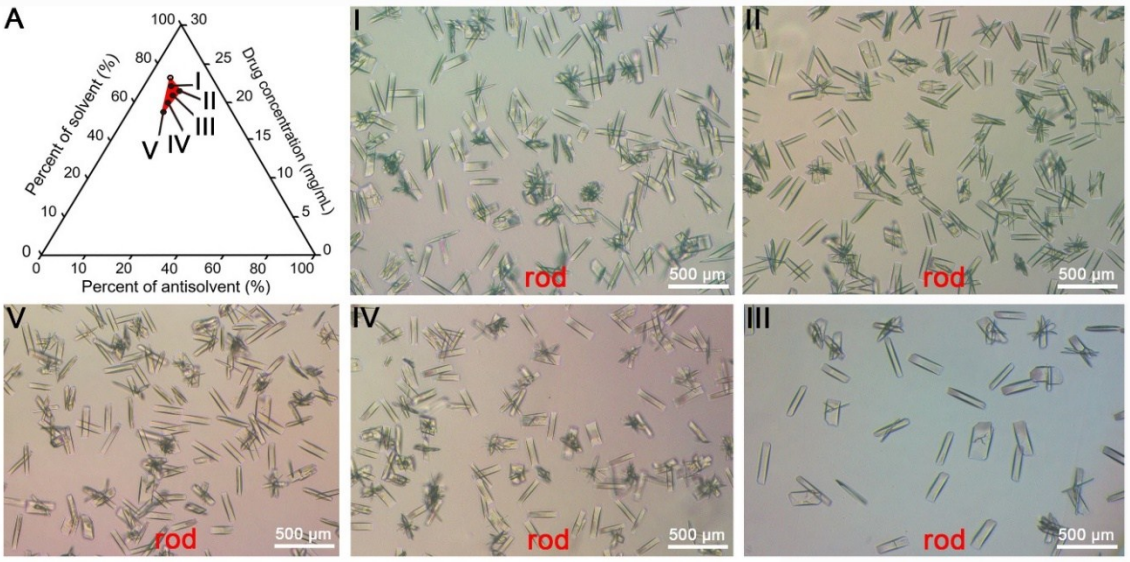

Fig. S8. (A) The scale-up preparations of indomethacin rod crystals according to the five locations of red region indicated in the ternary phase diagram. The location I, III and IV were randomly selected by using the RANDBETWEEN function of Excel on the inside of the phase diagram, where the volume ratios of solvent to antisolvent were 7.8: 2.6, 7.5: 3.3 and 6.9: 2.2, respectively, and the corresponding drug concentrations were 27.6 mg/mL, 26.3 mg/mL, and 25.7 mg/mL, respectively. The location II and V were selected by us on the edge of the phase diagram, where the volume ratios of solvent to antisolvent were 8: 4 and 6: 2, respectively, and the corresponding drug concentrations both were 25 mg/mL. (I-V) The rod crystals of indomethacin that were obtained from the scale-up experiments at the five antisolvent-crystallization conditions.

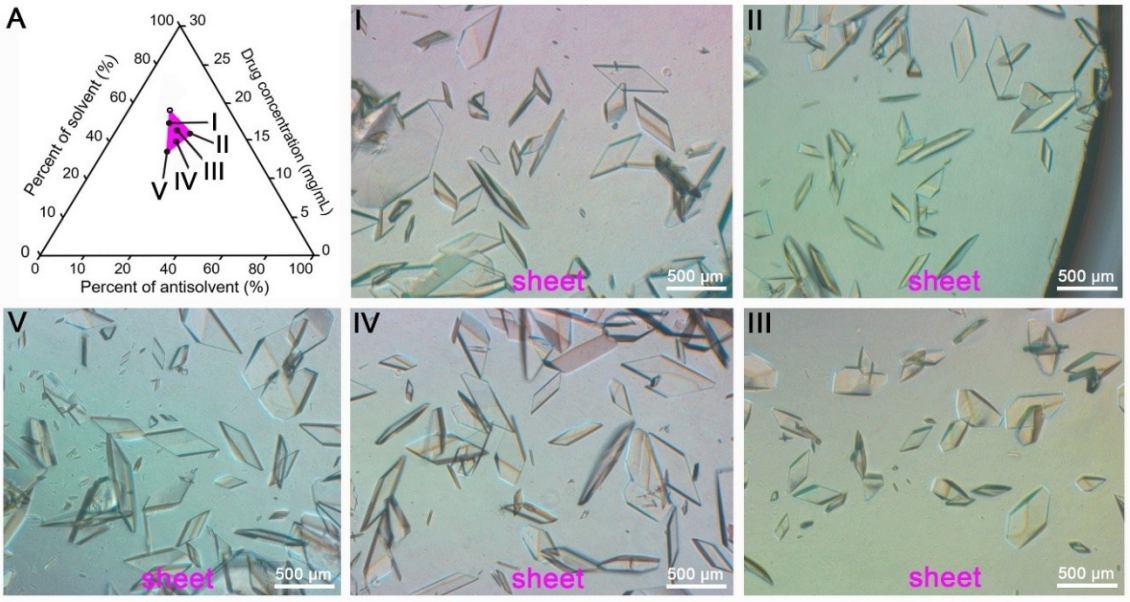

Fig. S9. (A) The scale-up preparations of indomethacin sheet crystals according to the five locations of purple region indicated in the ternary phase diagram. The location I, III and IV were randomly selected by using the RANDBETWEEN function of Excel on the inside of the phase diagram, where the volume ratios of solvent to antisolvent were 6.5: 3.2, 6.3: 4.1 and 5.9: 3.8, respectively, and the corresponding drug concentrations were 23.2 mg/mL, 22.5 mg/mL, and 20.9 mg/mL, respectively. The location II and V were selected by us on the edge of the phase diagram, where the volume ratios of solvent to antisolvent were 7: 5 and 5: 3, respectively, and the corresponding drug concentrations both were 20 mg/mL. (I-V) The sheet crystals of indomethacin that were obtained from the scale-up experiments at the five antisolvent-crystallization conditions.

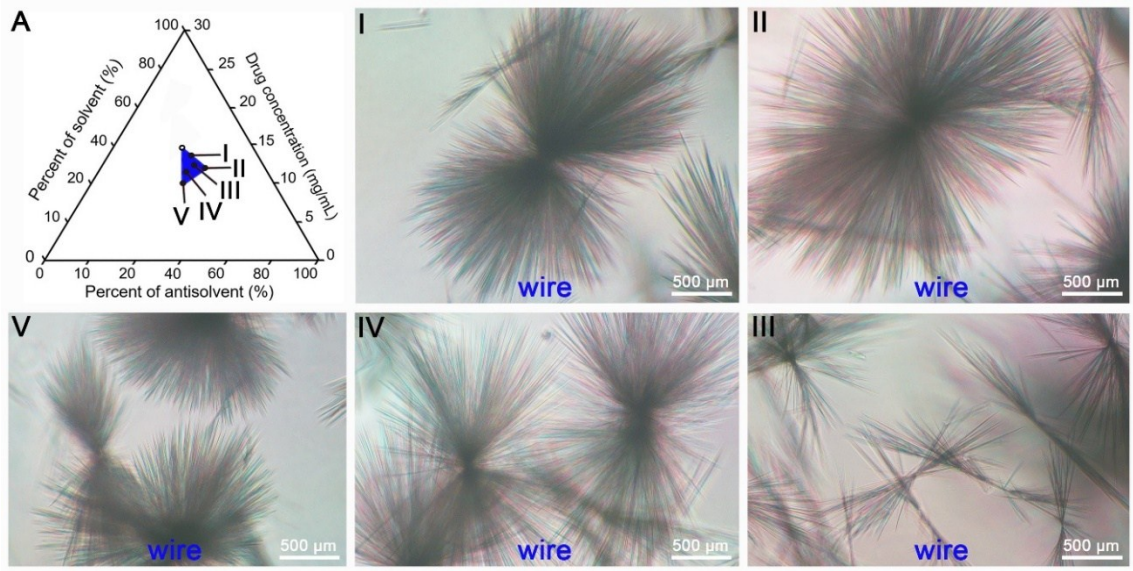

Fig. S10. (A) The scale-up preparations of indomethacin wire crystals according to the five locations of purple region indicated in the ternary phase diagram. The location I, III and IV were randomly selected by using the RANDBETWEEN function of Excel on the inside of the phase diagram, where the volume ratios of solvent to antisolvent were 5.9: 4.7, 5.4: 5.1 and 4.7: 4.3, respectively, and the corresponding drug concentrations were 18.1 mg/mL, 17.3 mg/mL, and 16.5 mg/mL, respectively. The location II and V were selected by us on the edge of the phase diagram, where the volume ratios of solvent to antisolvent were 6: 6 and 4: 4, respectively, and the corresponding drug concentrations both were 15 mg/mL. (I-V) The sheet crystals of indomethacin that were obtained from the scale-up experiments at the five antisolvent-crystallization conditions.

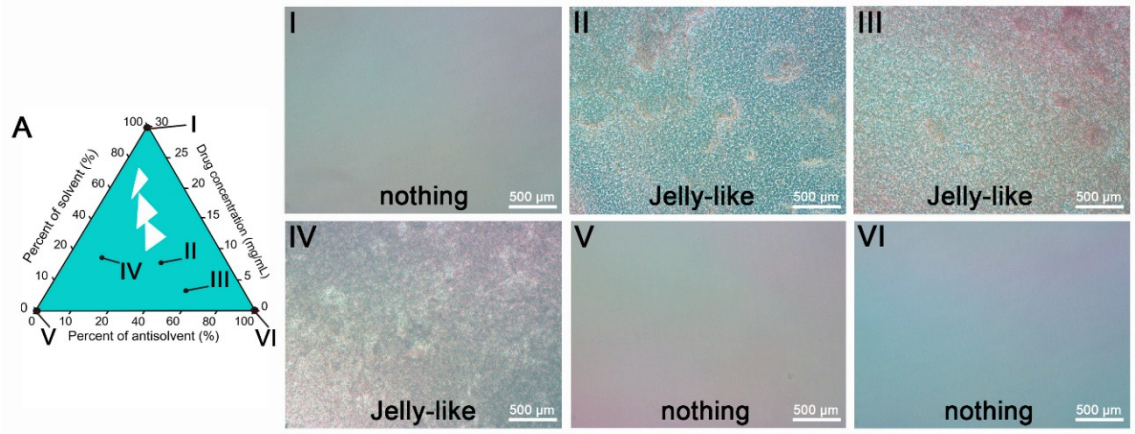

Fig. S11. (A) The scale-up generation of Jelly-like phase of indomethacin according to the six locations of green region indicated in the ternary phase diagram. The location II, III and IV were randomly selected by using the RANDBETWEEN function of Excel on the inside of the phase diagram, where the volume ratios of solvent to antisolvent were 3.5: 5.2, 1.7: 6.5 and 2.1: 1.5, respectively, and the corresponding drug concentrations were 11.6 mg/mL, 4.9 mg/mL, and 20.1 mg/mL, respectively. The location I, V and VI were selected by us on the edge of the phase diagram. (II-IV) The Jelly-like phase of indomethacin were obtained from the scale-up experiments at the three antisolvent-crystallization conditions. (I, V-IV) There are no jelly-like phase or other crystals from the scale-up experiments at the corresponding antisolvent-crystallization conditions.

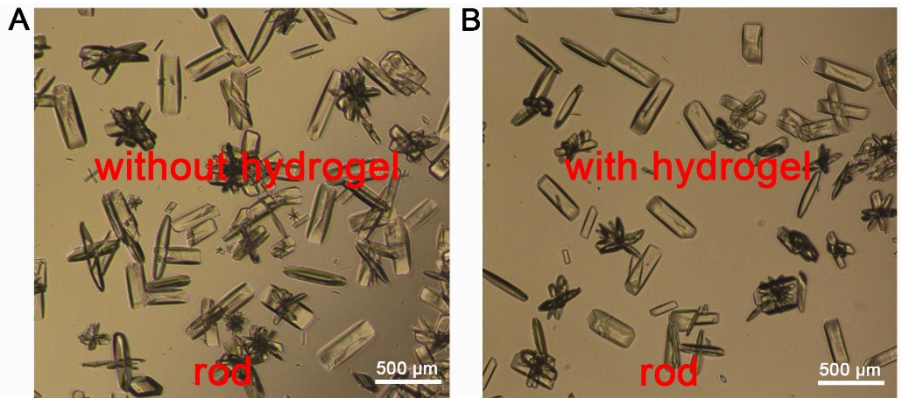

Figure S12. The scale-up preparations of the rod crystals (A) without and (B) with hydrogel.

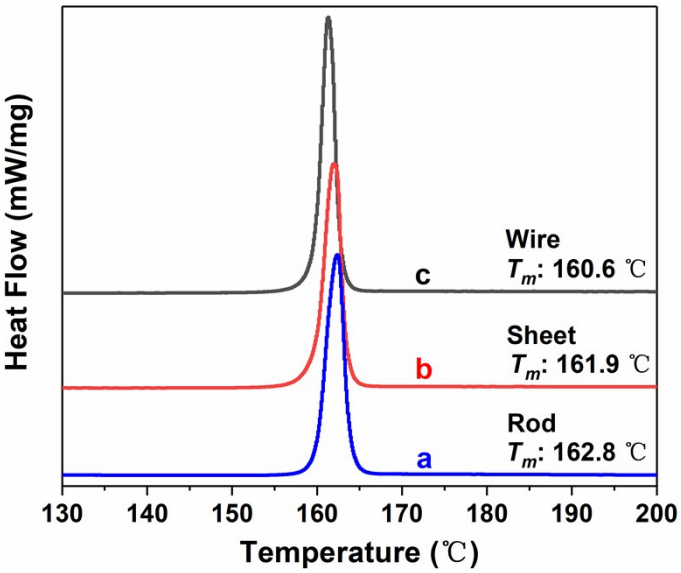

Fig. S13 Differential scanning calorimetry (DSC) curves of the indomethacin crystals of a) rod, b) sheet and c) wire.

Table S1. The number of indomethacin crystal images in the training set and test set, and the number of images of different crystal morphologies.

Total image number	Crystal morphologies	Image number
498*	Sheet	129
	Rod	151
	Wire	103
	Jelly-like phase	115

* The total images were randomly divided into training dataset and testing dataset at a ratio of 3:1.