Electronic Supplementary Material (ESI) for CrystEngComm.

Supplementary Information

Construction and application of a qualitative and quantitative analysis system of three boscalid polymorphs based on solid-state analytical methods and chemometric tools

Zeren Shang[a], Mingdi Liu[a,b], Weiguo Hu[a,c], Tong Deng[a], Xin Su[a], Baohong Hou[a], Jingkang Wang[a], Junbo Gong[a,*]

[a] State Key Laboratory of Chemical Engineering, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, People's Republic of China

[b] College of Chemistry and Chemical Engineering, Qinghai Nationalities University, Xining 810007, People's Republic of China

[c] North China Pharmaceutical Co., Ltd., Shijiazhuang 050015, People's Republic of China

AUTHOR INFORMATION

Corresponding Author

[*]E-mail: junbo_gong@tju.edu.cn (J. Gong).

Tel: 86-22-27405754, Fax: 86-22-27374971.

➢ The detailed introduction to the use of Unscrambler 10.4 software

(1) The corresponding average spectra (XRPD or ATR-FTIR) of samples were imported into the Unscrambler 10.4 software.

(2) Utilize the Transform module of the software to preprocess these spectra. The preprocessing algorithms (such as normalization, MSC and etc.) used in the establishment of qualitative and quantitative systems can be found in this module.

(3) Use the PCA, PCR or PLSR algorithms in Analyze module to deal with preprocessed spectra of model training/calibration samples to build the corresponding qualitative and quantitative models. The different spectrum ranges and leave-one-out cross-validation could be selected in the displayed dialog box.

(4) Use the Predict module to verify the predictive ability of the corresponding models. The models and spectra to be validated can be selected in the displayed dialog box.

(5) The corresponding results to evaluate the models could be found in the dialog box that appears automatically after the calculation is complete.

Table S1. The error analysis of the established quantitative analysis system[a]

Source of error	Evaluation approach
Instrument reproducibility	Four consecutive measurements were performed on the sample
Intra-day reproducibility	The sample was measured four times a day
Inter-day reproducibility	A single measurement was performed on the sample each day over four days
Sample mixing	The sample was divided into four sub-samples and each of them was measured once
Overall method error	Four independent mixtures of the same composition were prepared and analyzed by performing three measurements on each sample over one day

[a] The corresponding mass ratio (form I: form II: form H1) of the sample used in the error analysis was (1/3:1/3:1/3).

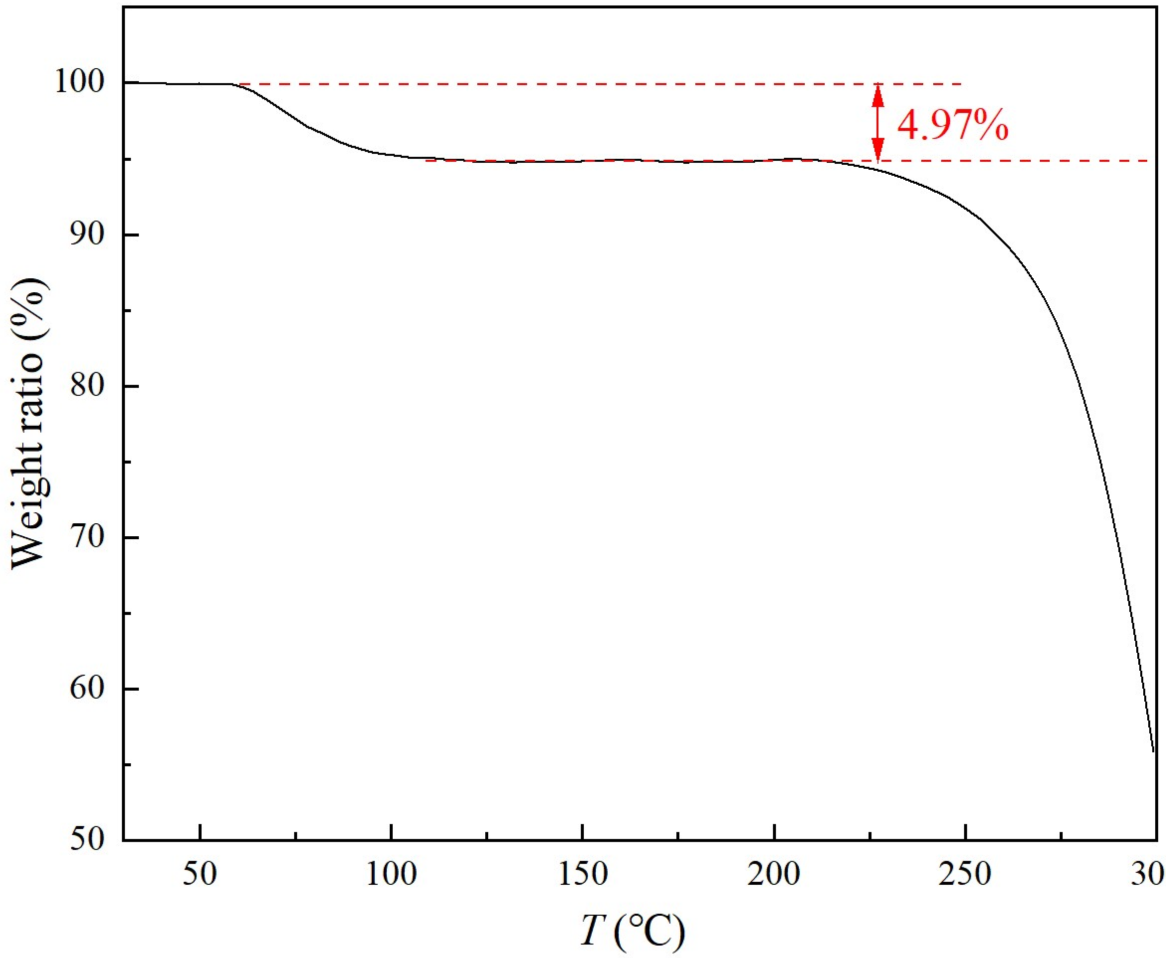


Fig. S1 TGA result of form H1.

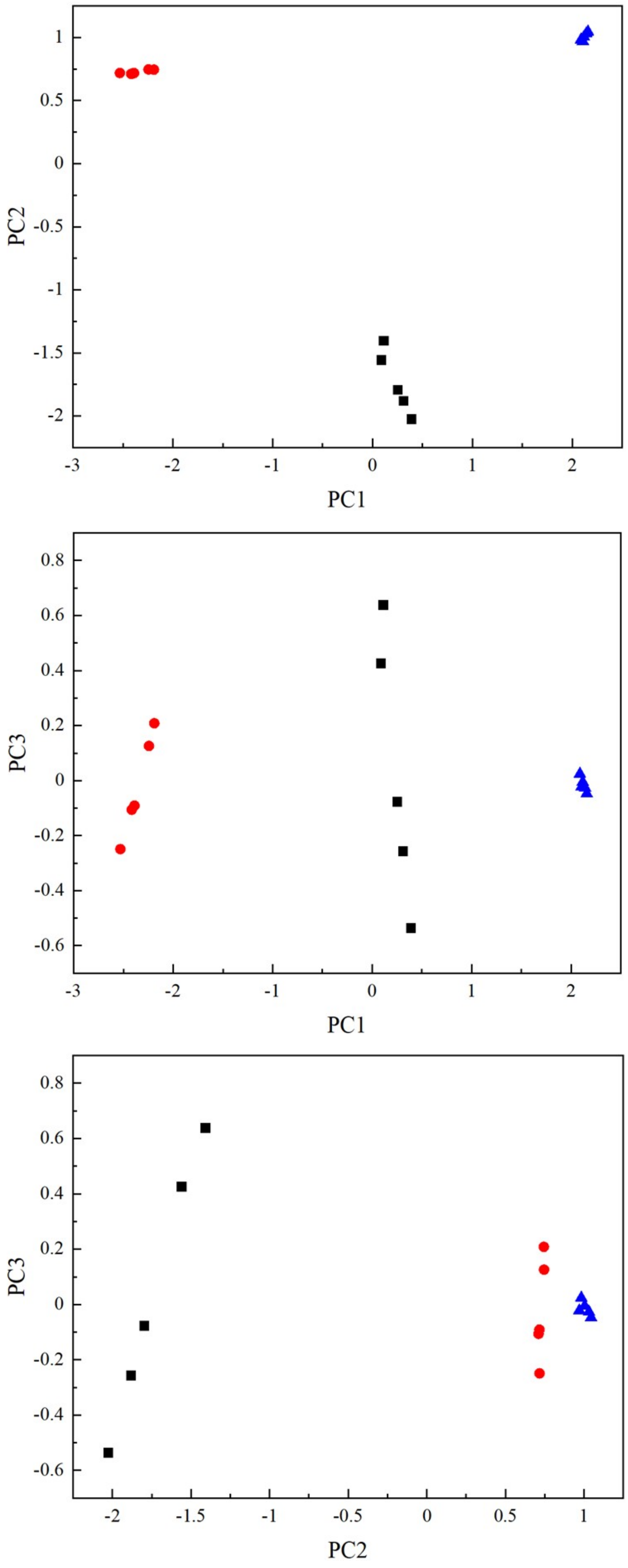


Fig. S2 Screening to determine the best combination of two-PCs to distinguish the three polymorphs: form I (■), form II (●) and form H1 (▲) of boscalid into distinct classes on the basis of ATR-FTIR spectra.

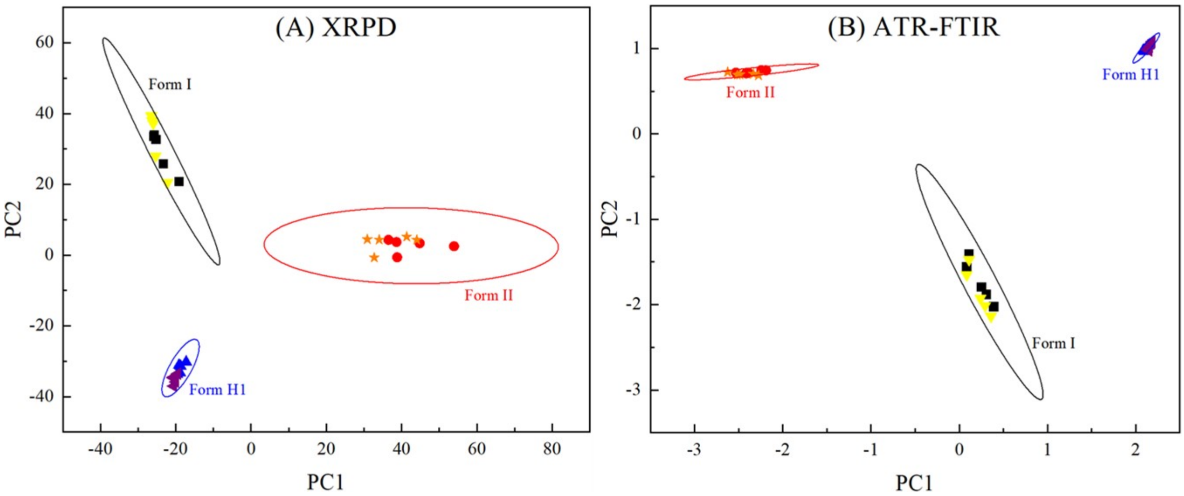
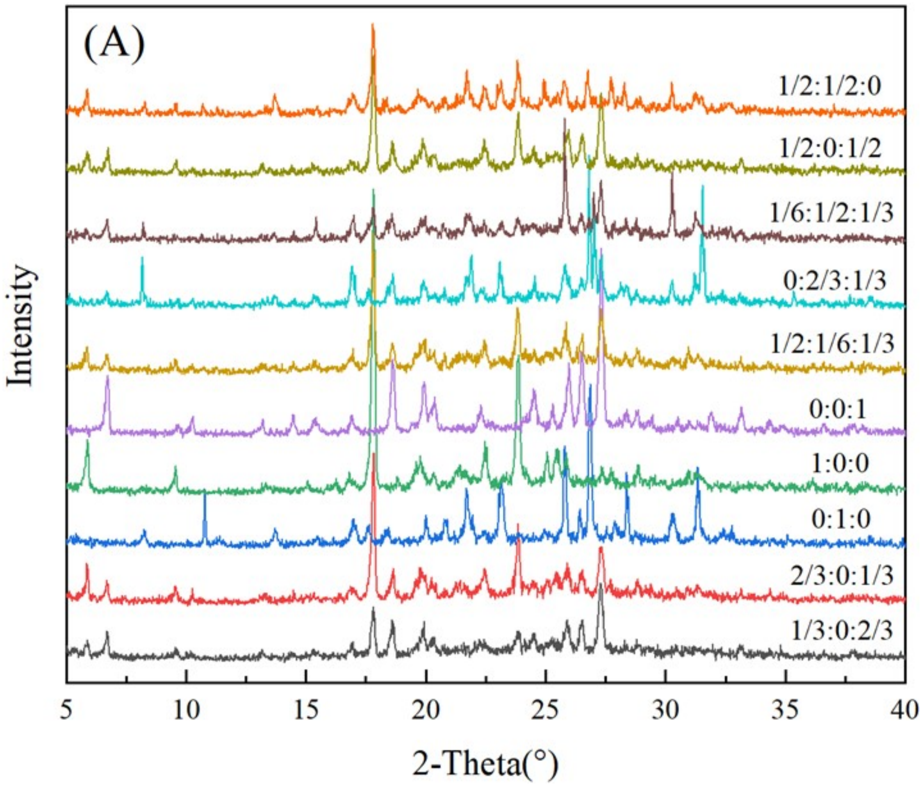


Fig. S3 PCA score plot (PC1 vs. PC2) based class separation of model training samples and cross-validation samples: model training sample of form I (■), model training sample of form II (●), model training sample of form H1 (▲), cross-validation sample of form I (▼), cross-validation sample of form II (★) and cross-validation sample of form H1 (◀). The corresponding 95% confidence ellipses are shown as a guide to the eye.



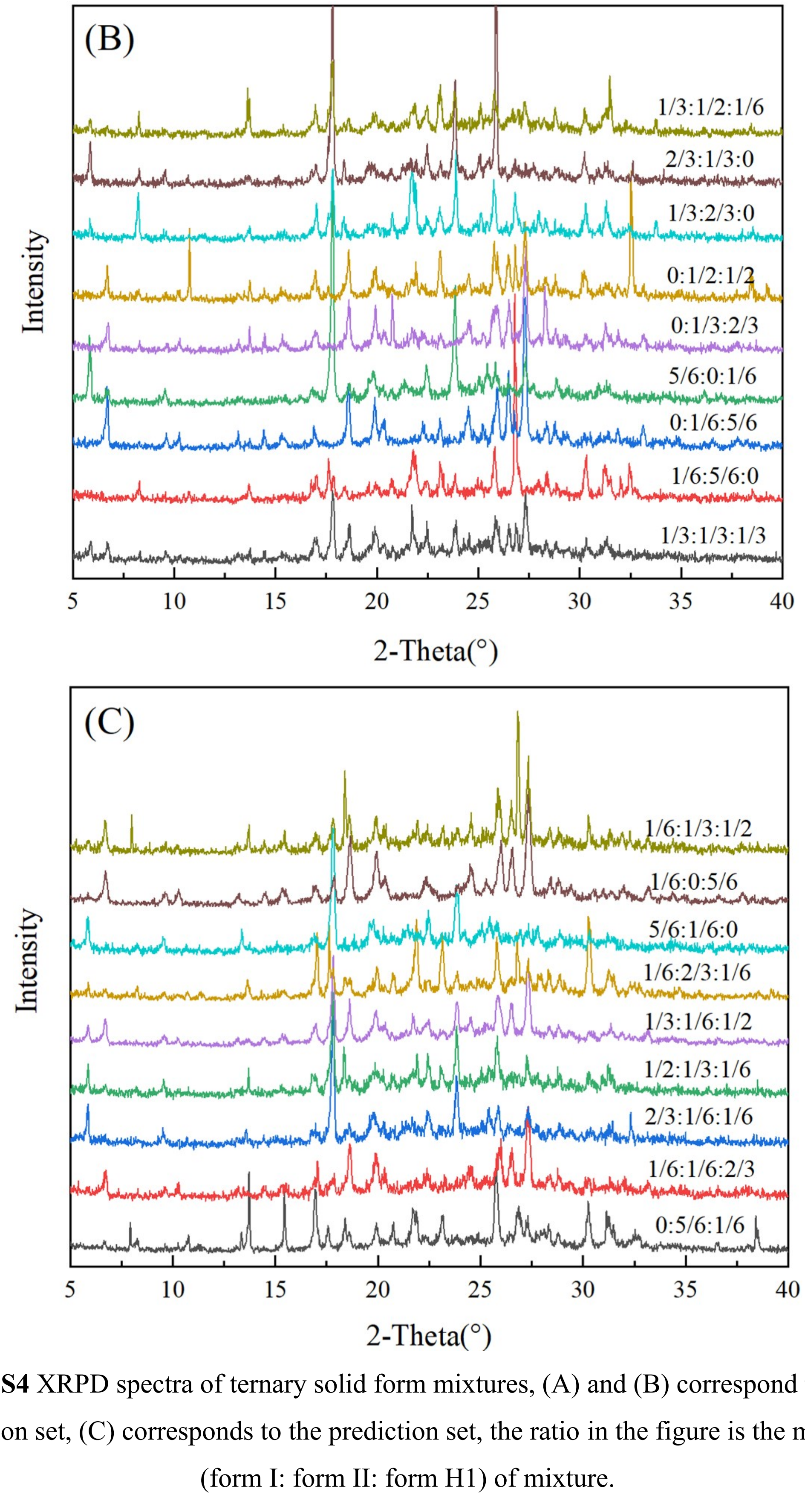
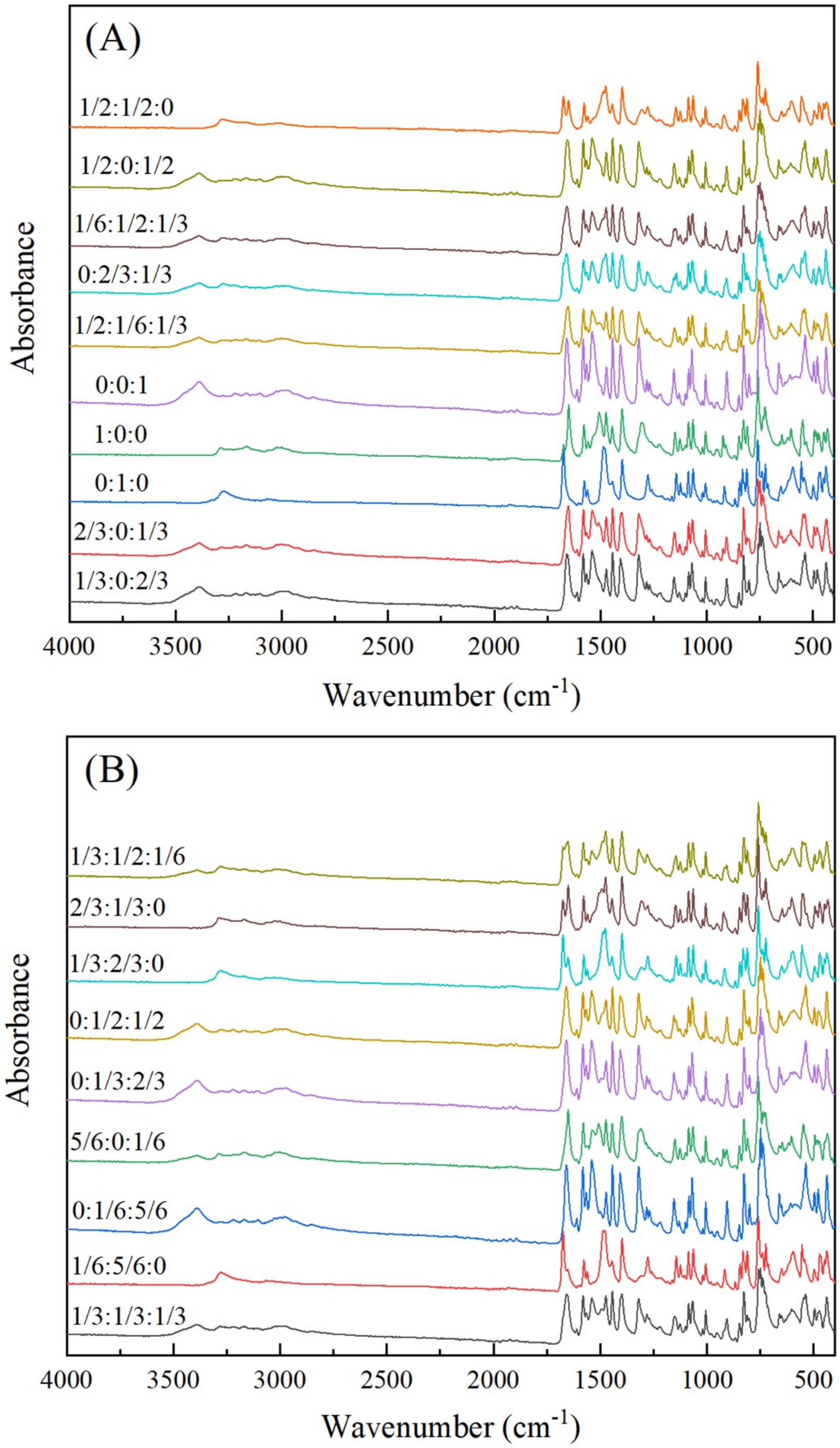


Fig. S4 XRPD spectra of ternary solid form mixtures, (A) and (B) correspond to the calibration set, (C) corresponds to the prediction set, the ratio in the figure is the mass ratio (form I: form II: form H1) of mixture.


(A)
1/2:1/2:0
1/2:0:1/2
1/6:1/2:1/3
0:2/3:1/3
1/2:1/6:1/3
0:0:1
1:0:0
0:1:0
2/3:0:1/3
1/3:0:2/3
Absorbance
4000
3500
3000
2500
2000
1500
1000
500
Wavenumber (cm-1)
(B)
1/3:1/2:1/6
2/3:1/3:0
1/3:2/3:0
0:1/2:1/2
0:1/3:2/3
5/6:0:1/6
0:1/6:5/6
1/6:5/6:0
1/3:1/3:1/3
Absorbance
4000
3500
3000
2500
2000
1500
1000
500
Wavenumber (cm-1)

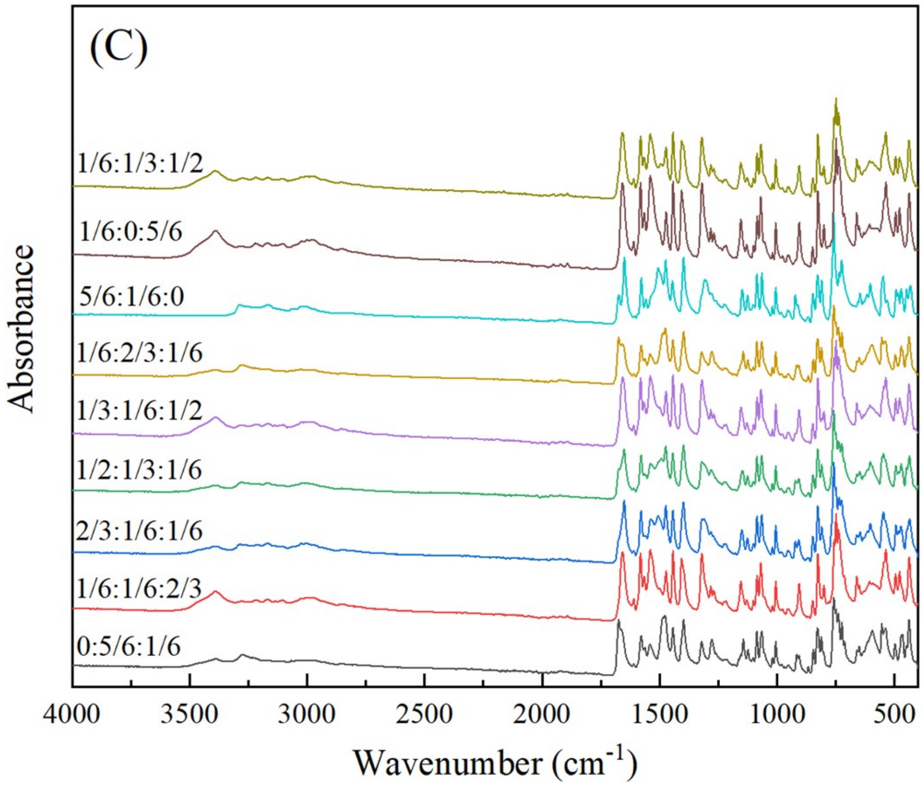


Fig. S5 ATR-FTIR spectra of ternary solid form mixtures, (A) and (B) correspond to the calibration set, (C) corresponds to the prediction set, the ratio in the figure is the mass ratio (form I: form II: form H1) of mixture.

Table S2. Performance of PCR quantification results for XRPD spectra using different preprocessing algorithms and different spectrum regions[a]

Preprocessing	5-40°			5-34°		
	Form I PC RMSEC (%) RMSECV (%) RMSEP (%)	Form II PC RMSEC (%) RMSECV (%) RMSEP (%)	Form H1 PC RMSEC (%) RMSECV (%) RMSEP (%)	Form I PC RMSEC (%) RMSECV (%) RMSEP (%)	Form II PC RMSEC (%) RMSECV (%) RMSEP (%)	Form H1 PC RMSEC (%) RMSECV (%) RMSEP (%)
Without	2	2	2	2	2	2
	4.119	7.235	4.373	4.118	7.257	4.389
	5.007	8.564	6.728	4.982	8.557	6.722
	9.968	10.177	8.699	9.966	10.190	8.686
SNV	2	2	2	2	2	2
	5.888	6.958	4.847	5.926	6.969	4.842
	7.342	8.719	6.817	7.366	8.686	6.774
	6.311	8.090	5.399	6.353	8.111	5.416
MSC	6	6	6	**6**	**6**	**6**
	3.137	4.258	3.254	**3.077**	**4.228**	**3.229**
	4.515	6.940	5.266	**4.525**	**6.948**	**5.237**
	3.909	4.907	4.854	**3.865**	**4.806**	**4.826**

[a] The preprocessing algorithm (without) means the spectra were only transformed by mean normalization, the bold fields represent the optimal results.

Table S3. Performance of PLSR quantification results for XRPD spectra using different preprocessing algorithms and different spectrum regions[a]

Preprocessing	5-40°			5-34°		
	Form I PLS factors RMSEC (%) RMSECV (%) RMSEP (%)	Form II PLS factors RMSEC (%) RMSECV (%) RMSEP (%)	Form H1 PLS factors RMSEC (%) RMSECV (%) RMSEP (%)	Form I PLS factors RMSEC (%) RMSECV (%) RMSEP (%)	Form II PLS factors RMSEC (%) RMSECV (%) RMSEP (%)	Form H1 PLS factors RMSEC (%) RMSECV (%) RMSEP (%)
Without	2	2	2	**2**	2	2
	1.188	2.786	4.358	**1.114**	2.850	4.302
	3.088	5.351	5.414	**2.796**	4.997	5.551
	5.741	5.022	6.762	**5.354**	5.133	6.387
SNV	4	4	4	4	**4**	4
	3.877	2.178	3.036	3.374	**2.072**	3.154
	5.265	4.560	4.314	4.924	**4.282**	4.027
	7.100	4.922	4.167	7.662	**4.063**	4.204
MSC	3	3	3	3	3	**3**
	2.378	2.944	2.310	2.165	2.970	**2.227**
	3.956	5.779	3.723	3.587	5.637	**3.871**
	6.429	4.492	4.541	5.976	4.273	**4.153**

[a] The preprocessing algorithm (without) means the spectra were only transformed by mean normalization, the bold fields represent the optimal results.

Table S4. Performance of PCR quantification results for ATR-FTIR spectra using different preprocessing algorithms and different spectrum regions[a]

	4000-400 cm^{-1}			1700-400 cm^{-1}			3600-2800 cm^{-1}			1700-400 cm^{-1} and 3600-2800 cm^{-1}		
	Form I	Form II	Form H1	Form I	Form II	Form H1	Form I	Form II	Form H1	Form I	Form II	Form H1
Preprocessing	PC factors RMSEC (%) RMSECV (%) RMSEP (%)	PC factors RMSEC (%) RMSECV (%) RMSEP (%)	PC factors RMSEC (%) RMSECV (%) RMSEP (%)	PC factors RMSEC (%) RMSECV (%) RMSEP (%)	PC factors RMSEC (%) RMSECV (%) RMSEP (%)	PC factors RMSEC (%) RMSECV (%) RMSEP (%)	PC factors RMSEC (%) RMSECV (%) RMSEP (%)	PC factors RMSEC (%) RMSECV (%) RMSEP (%)	PC factors RMSEC (%) RMSECV (%) RMSEP (%)	PC factors RMSEC (%) RMSECV (%) RMSEP (%)	PC factors RMSEC (%) RMSECV (%) RMSEP (%)	PC factors RMSEC (%) RMSECV (%) RMSEP (%)
Without	3	3	3	3	3	3	3	3	3	3	3	3
	3.942	6.158	4.485	4.252	6.224	4.246	5.253	7.338	5.655	4.209	6.236	4.278
	5.015	8.044	5.955	5.482	8.071	5.587	6.586	9.408	7.675	5.462	8.126	5.675
	3.041	5.326	3.318	3.185	5.140	4.122	3.099	5.282	3.161	3.038	5.205	3.847
SNV	3	3	3	3	3	3	2	2	2	3	3	3
	7.350	5.385	5.566	7.330	5.432	5.489	8.381	9.159	10.592	7.337	5.450	5.549
	9.485	7.048	7.966	9.428	7.258	7.710	10.056	10.404	12.365	9.459	7.258	7.853
	4.390	3.820	3.968	4.448	3.780	3.957	4.577	8.231	9.228	4.440	3.699	3.964
MSC	6	6	6	6	6	6	2	2	2	6	6	6
	3.524	2.725	4.149	3.429	2.316	3.922	9.240	10.077	10.634	3.329	2.539	3.932
	6.388	5.604	6.853	5.624	4.604	6.350	11.007	11.463	12.506	5.932	5.134	6.589
	2.918	2.686	3.492	3.334	2.838	3.419	4.740	9.541	9.895	3.032	2.611	3.521
1st Der	4	4	4	4	4	**4**	3	3	3	4	4	4
	3.608	2.934	3.498	3.667	2.961	**3.449**	7.245	6.510	6.846	3.665	2.962	3.461
	5.248	4.383	5.019	5.312	4.327	**4.959**	9.531	7.914	9.219	5.309	4.336	4.974
	2.846	3.797	2.567	2.894	3.860	**2.528**	4.953	5.595	6.802	2.885	3.859	2.532
2nd Der	4	4	4	4	4	4	3	3	3	4	4	4
	4.007	3.462	3.515	4.078	3.556	3.491	7.835	7.790	9.633	4.073	3.549	3.496
	5.730	4.866	4.986	5.790	4.897	4.971	9.609	10.288	11.471	5.788	4.902	4.975
	2.858	3.487	3.438	2.905	3.511	3.454	6.174	7.323	9.346	2.893	3.499	3.451
SNV+1st Der	5	5	5	**6**	6	6	2	2	2	6	6	6
	3.760	3.210	3.651	**3.062**	2.941	3.662	8.799	9.317	10.485	3.063	2.076	3.661
	5.814	6.345	5.655	**4.656**	4.325	5.738	10.189	10.668	12.090	4.669	4.368	5.743
	2.738	3.437	4.248	**2.559**	3.981	4.222	7.112	7.592	9.556	2.582	3.931	4.225
SNV+2nd Der	5	5	5	5	5	5	2	2	2	5	5	5
	4.254	3.604	3.679	4.132	3.636	3.697	9.740	10.067	10.793	4.137	3.632	3.699
	6.602	5.869	5.560	6.375	5.943	5.622	11.228	11.607	12.269	6.392	5.933	5.624
	2.910	2.967	3.757	2.882	2.979	3.834	6.536	8.286	10.288	2.886	2.975	3.820
MSC+1st Der	5	5	5	5	**5**	5	2	2	2	5	5	5
	3.710	3.123	3.608	3.516	**2.076**	3.606	9.607	10.089	10.330	3.552	2.977	3.608
	5.706	5.319	5.430	4.342	**4.023**	4.494	11.303	11.686	11.945	5.407	5.088	5.491
	2.777	1.795	3.262	2.844	**1.784**	3.324	6.089	8.707	10.066	2.859	1.822	3.309
MSC+2nd Der	5	5	5	5	5	5	2	2	2	5	5	5
	4.651	3.954	3.509	4.323	3.697	3.487	10.579	10.473	10.612	4.351	3.722	3.491
	7.165	6.084	5.284	6.611	5.627	5.345	12.084	11.907	12.050	6.664	5.673	5.343
	3.438	2.123	2.599	4.351	1.820	2.573	5.931	9.220	10.742	4.321	1.872	2.586

[a] The preprocessing algorithm (without) means the spectra were only transformed by maximum normalization, the bold fields represent the optimal results.

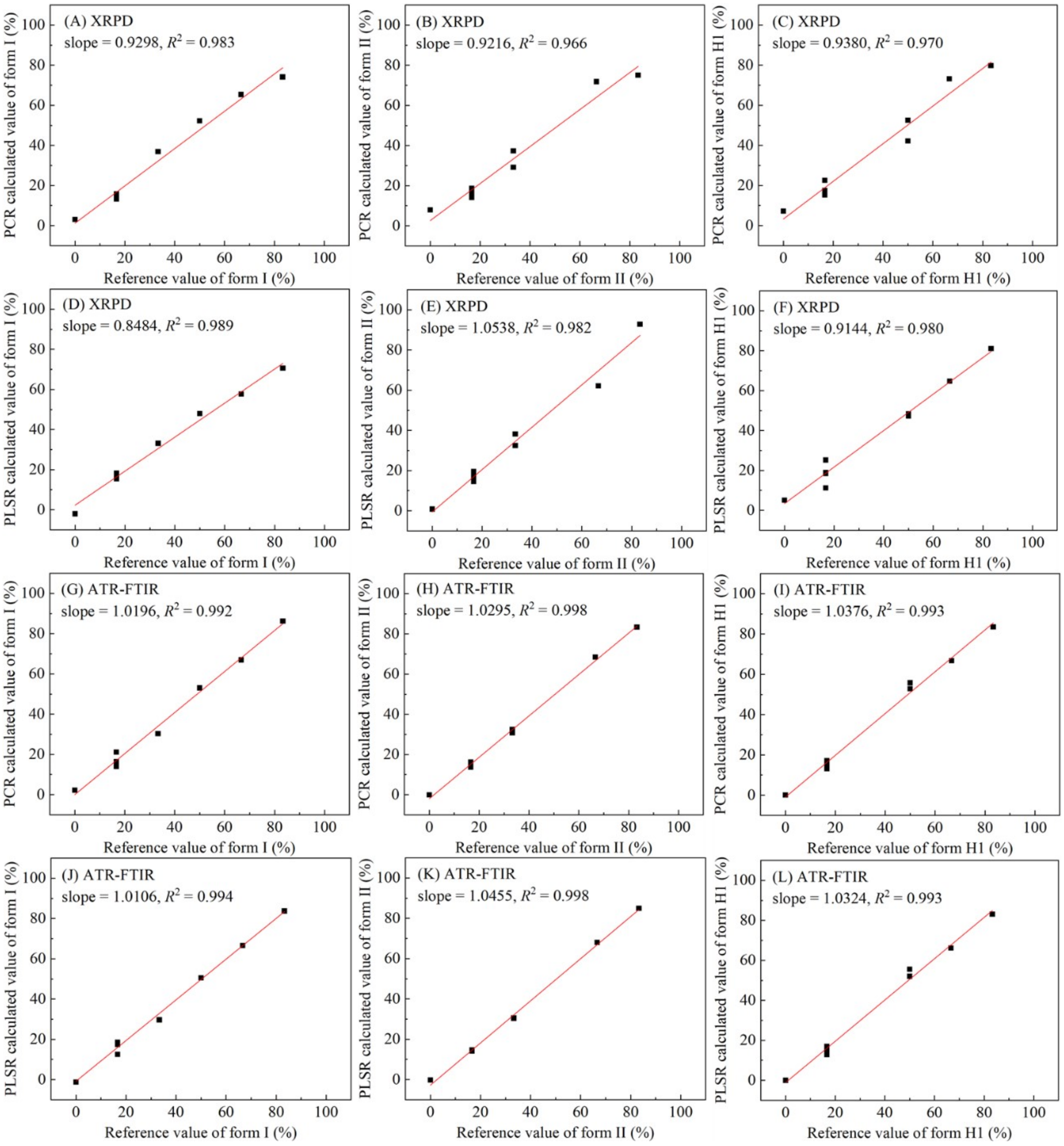


Fig. S6 Calculated vs. reference value (mass ratio, %) of three polymorphs in the prediction set from the model based on (A)-(C) XRPD and PCR, (D)-(F) XRPD and PLSR, (G)-(I) ATR-FTIR and PCR, (J)-(L) ATR-FTIR and PLSR.