Supplementary Information

MCC-GCN: An Interpretable Graph Learning Framework for Multicomponent Crystal Classification and Discovery

Yuehua Deng[1,2], Fanyu Zhao[3], Xinliang Zhou[4], Minqi Fu[1], Stephanie Chow[1], Zhi Wei[5], Qingsong Wen[6]*, Shing Fung Chow[1,2]*

[1] Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China.

[2] Advanced Biomedical Instrumentation Centre, Organization, Hong Kong Science Park, Shatin, Hong Kong SAR, China.

[3] College of Computer Science and Artificial Intelligence, Fudan University, Shanghai, China.

[4] College of Computing and Data Science, Nanyang Technological University, Singapore.

[5] Department of Computer Science, Ying Wu College of Computing, New Jersey Institute of Technology, Newark, NJ, USA.

[6] Squirrel Ai Learning, Bellevue, WA, USA.

* *Corresponding author*

Contents

1. Data Curation and Experimental Benchmarking

1.1. Data sources and repository characteristics

1.1.1 The Cambridge Structural Database (CSD)

The primary source of positive multicomponent crystal (MCC) data is the Cambridge Structural Database (CSD, v2024.1) [1], maintained by the Cambridge Crystallographic Data Centre (CCDC). As the world's repository for small-molecule organic and metal-organic crystal structures, the CSD provides high-fidelity atomic coordinates determined via X-ray and neutron diffraction. For this study, the CSD Python API was employed to programmatically access crystal entries and component information [2], enabling the construction of a large-scale dataset of 33,569 MCC entries comprising 15,878 salts, 11,118 cocrystals, and 6,573 solvates. We note that minor count variations may occur when using different CSD releases or when applying additional downstream quality-control steps.

1.1.2 CSD object model and terminology

To make the curation procedure transparent across disciplines, we briefly summarize the CSD object model that our pipeline operates on [3]. Each structure is indexed by a *refcode* (identifier) and can be accessed programmatically via the CSD Python API. The same refcode can be queried at multiple levels, including an entry-level view (deposition and bibliographic metadata), a crystal-level view (unit-cell parameters and symmetry), and a molecule-level view (chemical structure with 3D coordinates). Importantly, the molecule-level view provides an explicit decomposition into chemical components (such as distinct molecular/ionic species and optional lattice solvent/water). Our MCC dataset construction is built around this component decomposition: we use it to (i) enforce a consistent MCC setting (two primary components with an optional lattice solvent/water), (ii) assign crystal-form labels based on component count and charge state [4], and (iii) standardize each entry into a two-component molecular-pair representation for machine learning.

1.1.3 Rationale for using CSD and known limitations

The CSD offers two practical advantages for MCC learning. First, it provides experimentally resolved 3D structures, enabling consistent component decomposition and high-fidelity atomistic representations. Second, it aggregates structures across a wide chemical space, supporting large-scale model training and benchmarking. A known limitation is that crystallographic repositories inherently over-represent *successful* crystallizations (structures that were solved and deposited) and do not systematically capture negative experiments. This motivates the incorporation of literature-curated negative samples (CCG-Net) when training and evaluating formation vs. negative tasks [5].

1.1.4 Software and database resources

To support reproducibility and provide authoritative background material, we summarize the scope and typical use of the main resources employed in this study and provide direct links to their official documentation pages.

Cambridge Crystallographic Data Centre (CCDC). CCDC is the organization responsible for curating and distributing crystallographic structural resources and related software solutions, with a strong focus on small-molecule crystallography. In the context of this study, CCDC serves as the primary steward of the Cambridge Structural Database (CSD) and the associated tooling that enables standardized access, validation, and downstream analysis of deposited structures. We used the CCDC ecosystem specifically because it provides a stable object model for accessing refcodes, crystal metadata, and component-resolved molecular structures, which is essential for our entry-level filtering and component-based MCC labeling.
Official link: https://www.ccdc.cam.ac.uk/

Cambridge Structural Database (CSD). The CSD is a curated repository of experimentally determined crystal structures, designed to support broad scientific use across chemistry, materials science, and life-science adjacent applications. Conceptually, it provides a high-quality mapping from a published crystal structure to standardized identifiers, chemical components, and crystallographic descriptors. For MCC learning, this is particularly valuable

because the component decomposition provides a natural representation of multicomponent systems, enabling systematic construction of large-scale positive datasets and consistent enforcement of MCC definitions (two primary coformers with optional lattice solvent/water). **Official link:** https://www.ccdc.cam.ac.uk/solutions/csd-system/

CSD Python API. The CSD Python API provides a programmatic interface to the CSD portfolio, enabling scripted retrieval and analysis of entries, crystals, and molecules. This API is critical for reproducible data curation because it allows the same query logic and quality filters to be executed deterministically across large collections of refcodes, while exposing the relevant crystallographic and chemical attributes needed for automated processing. In our workflow, the API enables refcode-based access to component lists, charge information, and structure-quality flags that are subsequently combined with cheminformatics normalization to construct ML-ready molecular pairs. **Official link:** https://www.ccdc.cam.ac.uk/solutions/csd-core/components/csd-python-api/

RDKit. RDKit is an open-source cheminformatics toolkit that provides core molecular representations and operations commonly required for data cleaning, standardization, and feature construction. It supports conversion between chemical encodings (such as SMILES and InChI), molecule sanitization routines, reaction/SMARTS-based transformations, and canonicalization utilities that ensure deterministic string identifiers for equivalent chemical graphs. In this study, RDKit underpins the component-level normalization steps (InChI-based reconstruction, sanitization with error handling, charge parentization, targeted structural normalization, and canonical SMILES generation), which collectively reduce representation noise and stabilize downstream graph construction. **Official link:** https://www.rdkit.org/docs/

PyTorch. PyTorch is a widely used open-source deep learning framework providing tensor computation, automatic differentiation, and a flexible programming model for defining neural networks. Its design supports both rapid prototyping and scalable training workloads, including

GPU acceleration and modular model composition. In our project, PyTorch provides the foundational training runtime for neural models and optimization, and supports our implementation choices around batching, gradient-based interpretability[6], and reproducible inference. **Official link:** https://pytorch.org/docs/stable/

PyTorch Geometric (PyG). PyTorch Geometric is a graph learning library built on top of PyTorch that provides graph data structures, sampling/loading utilities, and a collection of graph neural network operators. It is designed to support efficient message-passing computations on sparse graphs, making it well-suited for molecular graph learning where atoms and bonds form natural graph primitives. In this work, the availability of standard GNN layers (for example, GCN-style convolutions) and pooling/readout utilities aligns directly with our MCC-GCN implementation, and motivates our emphasis on deterministic molecular graph construction during curation. **Official link:** https://pytorch-geometric.readthedocs.io/en/latest/

scikit-learn. scikit-learn[7] is a general-purpose machine learning library that provides a consistent interface for classical ML models, preprocessing, and evaluation utilities. It is commonly used to implement descriptor-based baselines and to compute standard performance metrics in a reproducible way. In our benchmarking, scikit-learn is used for baseline models (for example, RF/SVM/MLP) trained on descriptor representations, enabling direct comparison to the graph-based MCC-GCN under the same evaluation protocol. **Official link:** https://scikit-learn.org/stable/

pandas and NumPy. pandas provides high-level data structures and IO utilities for tabular data, while NumPy provides foundational numerical arrays and vectorized operations. Together they form the core data processing stack used to assemble, validate, and summarize large curation outputs. In this study, they support intermediate dataset tables, merging of positive and negative samples, and sanity checks on label distributions and schema consistency across pipeline stages. **Official links:** https://pandas.pydata.org/docs/; https://numpy.org/doc/stable/

tqdm. tqdm is a lightweight progress-bar utility commonly used to monitor long-running loops and data processing pipelines. While not affecting scientific results directly, it improves transparency during large-scale dataset processing and helps detect performance bottlenecks or unexpectedly slow failure modes when iterating through large entry collections. We use it throughout our scripts and notebooks to provide consistent progress reporting during curation and analysis. **Official link:** https://github.com/tqdm/tqdm

1.1.5 Negative samples from literature (CCG-Net)

Since crystallographic databases inherently lack "failed" experiments, negative samples (pairs that do not form MCCs) were sourced from the CCG-Net dataset[5]. These 1,052 entries represent experimentally verified negative cases manually curated from literature. The integration of these "real-world failures" is critical for training the binary classification head to distinguish between MCC formation and negative.

1.2. Data processing and cleaning pipeline

The transformation of raw CSD entries into machine-learning-ready molecular graphs followed a multi-stage algorithmic pipeline.

1.2.1 Stage-wise pipeline overview

For clarity and reproducibility, the dataset curation workflow can be summarized as a two-stage procedure: (Stage 1) *entry-level filtering and labeling* using CSD metadata and component patterns; and (Stage 2) *pair-level chemical standardization* to convert each retained entry into a canonical two-component representation for graph learning. Negative samples were subsequently integrated at the SMILES-pair level. This separation is practically important: Stage 1 ensures structural quality and a well-defined MCC setting, while Stage 2 ensures that the same underlying chemical entities are represented consistently even when different charge states, residues, or notational variants occur across CSD entries.

Algorithm (high-level): two-stage MCC dataset curation.

- **Input:** CSD entries (v2024.1), predefined list of common laboratory solvents, and literature-derived negative pairs.
- **Stage 1 — entry screening and labeling:** retain entries that satisfy (i) structural quality requirements (resolved 3D structure and no disorder), (ii) a valid MCC compo nent pattern (two primary components with an optional lattice solvent/water), and (iii) basic chemical constraints (organic-focused element set and component molecular weight cutoff). Assign crystal-form labels deterministically using component count and charge state.
- **Stage 2 — pair standardization:** convert each retained entry into a canonical two component molecular pair by applying chemistry-aware standardization (normalization of charge/protonation variants, removal of solvent components, and canonicalization of chemical identifiers). Discard entries that cannot be reduced unambiguously to two non-solvent parent components.
- **Negative integration:** standardize the negative pairs under the same SMILES-level conventions and merge them with the positive MCC pairs to form the final ML-ready dataset.

1.2.2 Rigorous filtering criteria

The raw CSD entries were subjected to a rigorous filtering protocol to ensure consistency and structural quality. Specifically, an entry was retained only if it simultaneously satisfied the following criteria:

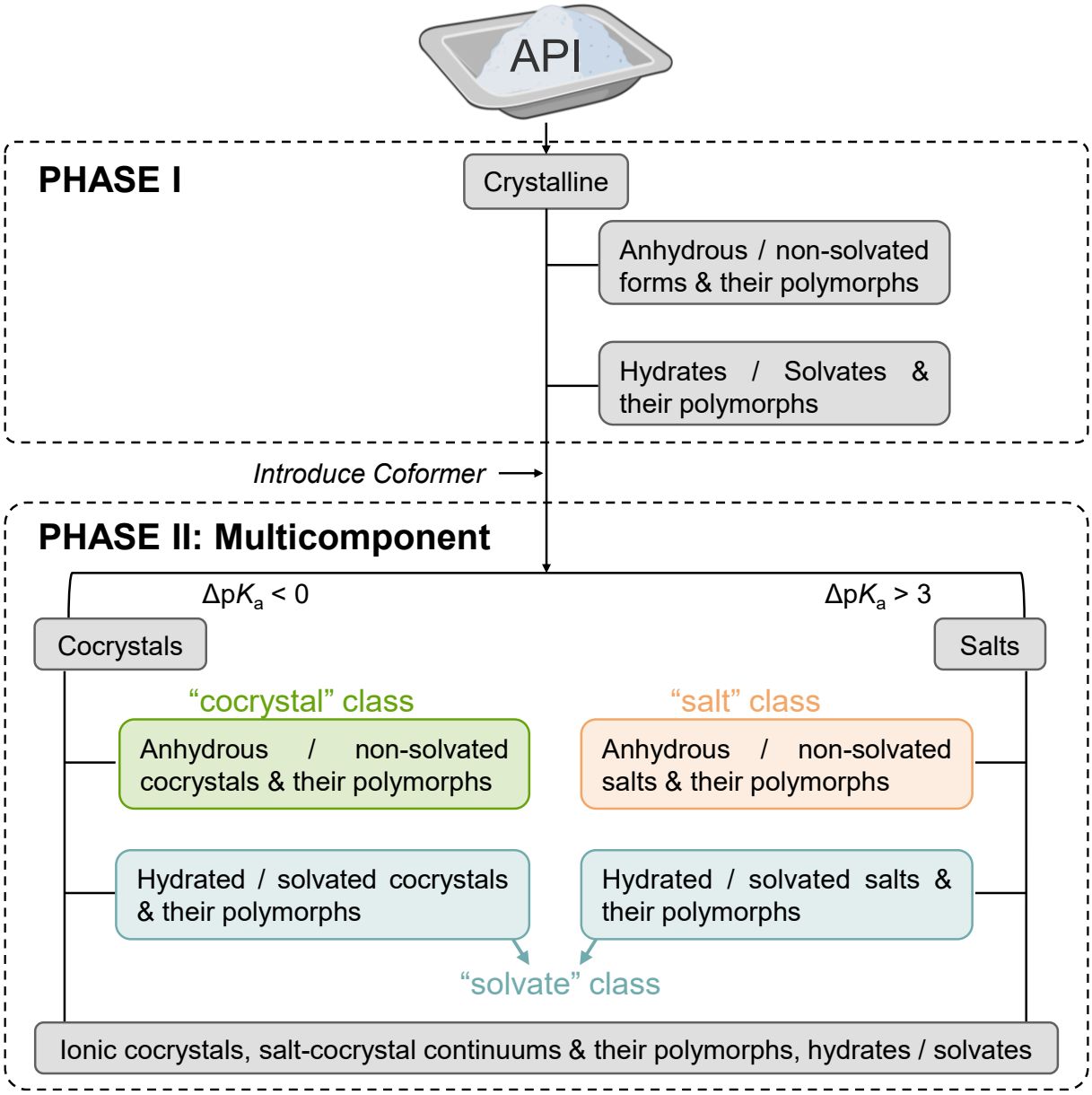
- **Polymorph removal (identifier-prefix de-duplication):** To reduce redundancy arising from polymorphism and repeated depositions, we de-duplicated entries by the identifier prefix (alphabetic part of the CSD refcode). Specifically, refcodes differing only by a numerical suffix were treated as describing the same underlying system at the refcode-prefix level, and only one representative entry was retained for each prefix. This rule is defined at the level of refcode prefixes and does not depend on entry ordering.
- **3D structure and disorder control:** Only entries with a resolved 3D structure and without disorder were retained.
- **Component pattern constraint:** The entry must contain either (i) exactly two chemically distinct components without common solvents, or (ii) exactly three components where one component is a recognized common solvent (see Supplementary

Table 3). This rule enforces an MCC setting composed of two main coformers, with an optional solvent/water.

- **Element whitelist:** Entries were restricted to contain only {C, H, O, N, P, S, F, Cl, Br, I, F and Si}, excluding metals and other elements outside typical pharmaceutical organic chemical space.
- **Molecular-weight cutoff:** The molecular weight of each component was restricted to < 700.
- **Structural ambiguity removal**: Pairs capable of existing in multiple crystal forms were excluded.
- **Dataset purity control**: Ionic cocrystals, salt-cocrystal continuums, and their polymorphs/solvates were excluded.

Rationale for the Classification of MCCs. The classification framework adopted in this study classifies MCCs into three principal types: *cocrystals*, *salts*, and *solvates*, based on the chemical identity of the components, the degree of ionization, and the type of intermolecular interactions governing the crystalline architecture[8]. This approach aligns with the widely accepted nomenclature proposed by Aitipamula *et al.* (2012)[9], which defines cocrystals as "solids that are crystalline single-phase materials composed of two or more different molecular and/or ionic compounds generally in a stoichiometric ratio which are neither solvates nor simple salts." This consensus definition, established during the Indo–U.S. Bilateral Meeting Forum titled *The Evolving Role of Solid State Chemistry in Pharmaceutical Science*, further refined by the systematic classification system introduced by Grothe *et al.* (2016)[8], remains a cornerstone in crystal engineering and pharmaceutical solid-state research.

In this study, focusing on the phase II of pharmaceutical solid form screening where coformer is introduced (Supplementary Figure 1) [10-12], MCCs are operationally defined as crystalline, single-phase materials composed of two distinct chemical components in a definite stoichiometric ratio, with or without the inclusion of an additional neutral solvent molecule.



Supplementary Figure 1. Pharmaceutical crystalline solid form screening strategy.

The three subcategories are delineated as follows:

- Cocrystals. Formed between two neutral molecules without neutral solvent participation.
- Salts. Composed of ionic species generated through proton transfer between an acid and a base, and likewise free of solvent molecules.
- Solvates. Unlike the two-component solvates, the three-component solvates here represent crystalline materials in which a third component (solvent) participates directly in lattice formation. In these systems, solvent molecules (mostly water) occupy well-defined positions in the crystal structure and engage in specific interactions (typically hydrogen bonding) with the host components. From a structural standpoint, these

cocrystal solvates and salt solvates lie at the intersections of the salt and solvate domains and the cocrystal and solvate domains. However, in this study, they are classified collectively as a distinct category (solvate) because the presence of the solvent (mostly water) as a third, structurally integral participant fundamentally differentiates them from binary systems. The solvent acts as a lattice-supporting component, stabilizing the crystal packing and occasionally dictating the overall supramolecular arrangement.

The proposed framework thereby captures the overlapping yet distinct nature of these MCCs and provides a scientifically coherent basis for comparing their formation, stability, and physicochemical properties. It also remains fully compatible with the consensus definition of cocrystals, which continues to provide the most robust and widely adopted standard in the field.

1.2.3 Standardization and representation

After entry-level filtering, we further standardized each retained MCC entry into a clean *pairwise* representation suitable for graph learning. In brief, we (i) retrieved the CSD entry by identifier and enumerated its chemical components; (ii) reconstructed each component in RDKit via InChI to reduce ambiguity[13]; (iii) performed molecule sanitization with robust error handling; (iv) neutralized charged components to obtain parent (charge-stripped) structures and applied a small set of chemically motivated normalization rules to improve consistency across recurrent edge cases; and (v) collapsed different residues/charge states of the same parent scaffold using an InChIKey connectivity identifier (first 14 characters). Common solvents were excluded using the same solvent list as in the entry-level filter. An entry was kept only if exactly two distinct non-solvent parent scaffolds remained after this procedure. Finally, canonical SMILES[14] were generated (RDKit canonicalization) and cached structure strings were stored to enable deterministic molecular graph construction.

1.2.4 De-duplication and ambiguity handling

In addition to the identifier-prefix heuristic for polymorph de-duplication (Stage 1), we applied several “ambiguity-reduction” rules motivated by MCC screening practice and by the needs of supervised learning:

- **Component-count enforcement:** Entries that could not be reduced to exactly two distinct *non-solvent* parent scaffolds after charge normalization and solvent removal were discarded in Stage 2. This includes cases where multiple residues map to more than two parent scaffolds or where the entry contains additional small molecules beyond the allowed solvent/water component.
- **Charge-state harmonization:** To avoid treating the same chemical entity as different molecules solely due to protonation/ionization state, we standardized components to parent forms and clustered residues by an InChIKey connectivity identifier (first 14 characters). This reduces redundant samples arising from different salt forms/resonance representations and stabilizes downstream graph feature extraction.
- **Special-case normalization:** A small set of targeted structure-normalization rules was applied to address recurring edge cases observed in pharmaceutical-like functional groups (for example, specific sulfonamide/azole patterns). These rules were designed to reduce systematic fragmentation of chemically equivalent scaffolds during standardization.

These design choices prioritize label clarity and representation consistency over maximal coverage. In practice, this trade-off reduces contradictory supervision and improves model robustness when predicting for APIs absent from the CSD.

1.2.5 Dataset summary tables

Supplementary Table 1. Datasets used in this study and their roles in model development and evaluation.

Dataset	Size	Label space	Primary role
CSD-derived MCC positives	33,569	Salt (15,878) Cocrystal (11,118) Solvate (6,573)	Model training and in-distribution evaluation (positive MCC classes)
Literature-derived negatives (CCG-Net)	1,052	Negative pairs (negative class)	Negative supervision for formation prediction and negative class in four-class setting
KPXKPR-64 experimental benchmark	64	four-class experimental outcomes (salt / cocrystal / solvate / negative)	Out-of-distribution evaluation on APIs absent from the CSD

Supplementary Table 2. Conceptual objectives of the two-stage curation workflow.

Stage	Primary goal	Key constraints	Output
Stage 1	Define a clean MCC from CSD entries	3D resolved; no disorder; Valid component pattern; Element/MW constraints	Entry-level labels (salt/cocrystal/ solvate)
Stage 2	Construct ML-ready molecular pairs	Charge and representation normalization; Solvent removal; Exactly two non-solvent parents	Canonical two-component pairs for deterministic graph construction

1.2.6 Class labeling and prioritization

Class labels were assigned at the entry level following a deterministic priority rule. Briefly, the number of distinct components in an entry was first evaluated. Three-component entries were treated as solvated systems: if solvent was present, the entry was labeled as a *solvate*. Two-component entries were then labeled by charge state: if any component had a non-zero formal charge, the entry was labeled as a *salt*; otherwise it was labeled as a neutral *cocrystal*. This priority ordering ensures that lattice-solvent inclusion is handled explicitly and consistently before distinguishing salts and cocrystals.

1.2.7 Negative sample integration and harmonization

To provide experimentally verified negative supervision, 1,052 negative pairs were imported from the literature-curated CCG-Net dataset (reported as SMILES pairs)[8]. These negatives were concatenated with the positive CSD-derived pairs after standardization. Because negative samples originate from a different data modality (literature-reported failures rather than resolved crystal structures), we applied the same SMILES-level conventions (canonicalization and consistent ordering) during dataset assembly to minimize spurious distribution cues unrelated to chemistry.

1.2.8 Implementation details and reproducibility

All entry-level filtering and labeling were performed using the CSD Python API with CSD release v2024.1. Pair-level standardization used RDKit-based chemical normalization (InChI reconstruction, sanitization, charge normalization, and canonical SMILES generation) to ensure deterministic and reproducible molecular identifiers. Common solvent recognition used a fixed curated list of solvent SMILES, applied consistently at both the entry-filtering and pair-standardization stages. Processing exceptions during large-scale conversion and sanitization

were recorded for auditability, and samples that could not be standardized unambiguously were excluded according to the rules described above.

1.2.9 Software stack and roles of key libraries

Although the overall workflow is conceptually straightforward, robust large-scale curation requires multiple software components, each responsible for a specific class of operations:

- **CSD Python API (CCDC):** Used for programmatic access to CSD refcodes and their associated entry-, crystal-, and molecule-level information, including component-level properties such as charge state, molecular weight, and structure-quality flags.
- **RDKit:** Used for chemistry-aware standardization and canonicalization, including conversion between chemical encodings, molecule sanitization, charge normalization to parent forms, targeted normalization for edge cases, and generation of canonical SMILES identifiers.
- **Data handling utilities (pandas/numpy) and workflow tooling (tqdm/logging):** Used to stream through large numbers of entries, assemble intermediate tables, cache repeated computations, and record processing failures for auditability.
- **Graph learning ecosystem (PyTorch / PyTorch Geometric):** Used downstream (see Supplementary Note 2) to construct molecular graphs and train GCN models on the standardized molecular-pair inputs. We highlight this here because the curation output format (canonical SMILES and cached MolBlock/SDF strings) was chosen specifically to provide deterministic graph construction inputs.

1.2.10 Identification of solvent components

To accurately distinguish solvates from salts or cocrystals, we implemented a solvent-recognition sub-routine. Each component in a multi-component entry was cross-referenced against a curated list of common laboratory solvent SMILES. This list is used consistently in both entry-level filtering and downstream pairwise standardization to (i) permit at most one solvent component in a three-component entry and (ii) exclude solvent molecules when forming the final two-coformer molecular pair. Isotopically labelled solvent variants (e.g., deuterated

solvents) were treated as their corresponding parent solvents for recognition purposes and are therefore not listed separately.

Supplementary Table 3. Curated list of recognized solvents for dataset labeling.

Solvent Name	SMILES	Category
Toluene	CC1=CC=CC=C1	Solvent
4-Chlorotoluene	CC1=CC=C(C=C1)Cl	Solvent
diglyme	COCCOCCOC	Solvent
1,3,5-trichlorobenzene	C1=C(C=C(C=C1Cl)Cl)Cl	Solvent
iodobenzene	C1=CC=C(C=C1)I	Solvent
gamma-Butyrolactone	C1CC(=O)OC1	Solvent
1,1,2-trichloroethane	C(C(Cl)Cl)Cl	Solvent
ethoxyethane	CCOCC	Solvent
DL-sec-Butyl acetate	CCC(C)OC(=O)C	Solvent
formic acid	C(=O)O	Solvent
methylamine	CN	Solvent
iodomethane	CI	Solvent
Dimethyl sulfoxide (DMSO)	CS(=O)C	Solvent
p-Xylene	CC1=CC=C(C=C1)C	Solvent
methanamide	C(=O)N	Solvent
3-methyl-1-butanol	CC(C)CCO	Solvent
1-butanol	CCCCO	Solvent
Tetrahydrofuran	C1CCOC1	Solvent
bromobenzene	C1=CC=C(C=C1)Br	Solvent
cyclohexanone	C1CCC(=O)CC1	Solvent
chlorobenzene	C1=CC=C(C=C1)Cl	Solvent
dimethoxymethane	COCOC	Solvent
1H-pyrrole	C1=CNC=C1	Solvent
Ethyl formate	CCOC=O	Solvent
2-butanone	CCC(=O)C	Solvent
2-butanol	CCC(C)O	Solvent
isobutanol	CC(C)CO	Solvent
N-Ethylmorpholine	CCN1CCOCC1	Solvent
1,1,2,2-tetrachloroethane	C(C(Cl)Cl)(Cl)Cl	Solvent
N, N, N', N'-Tetramethylethylenediamine	CN(C)CCN(C)C	Solvent
propan-2-ol	CC(C)O	Solvent
1,4-dioxane	C1COCCO1	Solvent
Ethanol	CCO	Solvent
2-methyl-2-propanol	CC(C)(C)O	Solvent
2-methylpyridine	CC1=CC=CC=N1	Solvent
3-methylpyridine	CC1=CN=CC=C1	Solvent
1-methyl-2-pyrrolidone	CN1CCCC1=O	Solvent
N, N-dimethylacetamide	CC(=O)N(C)C	Solvent
2,2'-Dichlorodiethyl ether	C(CCl)OCCCl	Solvent
2-butoxyethanol	CCCCOCCO	Solvent
diethylenetriamine	C(CNCCN)N	Solvent
2-methoxyethanol	COCCO	Solvent
dibromomethane	C(Br)Br	Solvent
Methyl acetate	CC(=O)OC	Solvent
cyclopentane	C1CCCC1	Solvent
benzyl alcohol	C1=CC=C(C=C1)CO	Solvent
benzene	C1=CC=CC=C1	Solvent
hexadecane	CCCCCCCCCCCCCCCC	Solvent
nitromethane	C[N+](=O)[O-]	Solvent
hexamethyldisiloxane	C[Si](C)(C)O[Si](C)(C)C	Solvent
Hexane	CCCCCC	Solvent
1-Chloro-2-Methylpropane	CC(C)CCl	Solvent

Solvent Name	SMILES	Category
acetic anhydride	CC(=O)OC(=O)C	Solvent
propanenitrile	CCC#N	Solvent
acetamide	CC(=O)N	Solvent
acetic acid	CC(=O)O	Solvent
Ethylene glycol	C(CO)O	Solvent
Diethylene glycol	C(COCCO)O	Solvent
Isopropyl acetate	CC(C)OC(=O)C	Solvent
Isopropyl ether	CC(C)OC(C)C	Solvent
tetrachloromethane	C(Cl)(Cl)(Cl)Cl	Solvent
acetone	CC(=O)C	Solvent
acetophenone	CC(=O)C1=CC=CC=C1	Solvent
nitrobenzene	C1=CC=C(C=C1)[N+](=O)[O-]	Solvent
propionic acid	CCC(=O)O	Solvent
1,2-Propanediol	CC(CO)O	Solvent
pentane	CCCCC	Solvent
1,1-Dichloroethane	CC(Cl)Cl	Solvent
butane-1,4-diol	C(CCO)CO	Solvent
1,3-dimethylbenzene	CC1=CC(=CC=C1)C	Solvent
1,2-dihydrostilbene	C1=CC=C(C=C1)CCC2=CC=CC=C2	Solvent
N, N-diethylethanamine	CCN(CC)CC	Solvent
tribromomethane	C(Br)(Br)Br	Solvent
2-propoxyethanol	CCCOCCO	Solvent
1,2-Dichloroethane	C(CCl)Cl	Solvent
1-propanol	CCCO	Solvent
water	O	Solvent
phenylamine	C1=CC=C(C=C1)N	Solvent
heptane	CCCCCCC	Solvent
trichloromethane	C(Cl)(Cl)Cl	Solvent
pyridine	C1=CC=NC=C1	Solvent
cyclohexene	C1CCC=CC1	Solvent
cyclohexane	C1CCCCC1	Solvent
Methanol	CO	Solvent
1,2-dimethoxyethane	COCCOC	Solvent
3-pentanone	CCC(=O)CC	Solvent
fluorobenzene	C1=CC=C(C=C1)F	Solvent
epichlorohydrin	C1C(O1)CCl	Solvent
acetonitrile	CC#N	Solvent
dichloromethane	C(Cl)Cl	Solvent
methanedithione	C(=S)=S	Solvent
1-Octanol	CCCCCCCCO	Solvent
butanedioic acid	C(CC(=O)O)C(=O)O	Solvent
N, N-dimethylformamide	CN(C)C=O	Solvent
1,2-ethanediamine	C(CN)N	Solvent
2,4-pentanedione	CC(=O)CC(=O)C	Solvent
o-Xylene	CC1=CC=CC=C1C	Solvent
Propylene glycol monomethyl ether acetate	CC(COC)OC(=O)C	Solvent
1,3,5-trimethylbenzene	CC1=CC(=CC(=C1)C)C	Solvent
2-phenylacetonitrile	C1=CC=C(C=C1)CC#N	Solvent
2-Chlorotoluene	CC1=CC=CC=C1Cl	Solvent
1,2-dichlorobenzene	C1=CC=C(C(=C1)Cl)Cl	Solvent
isophorone	CC1=CC(=O)CC(C1)(C)C	Solvent
morpholine	C1COCCN1	Solvent
nitric acid	[N+](=O)(O)[O-]	Solvent
quinoline	C1=CC=C2C(=C1)C=CC=N2	Solvent
benzonitrile	C1=CC=C(C=C1)C#N	Solvent
ethyl acetate	CCOC(=O)C	Solvent

1.3. Experimental benchmarking dataset (kpxkpr-64)

The KPXKPR-64 dataset serves as an independent, out-of-distribution (OOD) validation set. It consists of 64 API-coformer combinations synthesized and characterized in-house.

1.3.1 Coformer selection and diversity

The 32 coformers were selected to span a wide range of chemical spaces, including varying pK_a values, hydrogen-bonding capacities, and molecular shapes, ensuring a comprehensive assessment of the model's predictive power.

1.3.2 Experimental labeling protocol

Experimental outcomes were determined by comprehensive solid-state characterizations. Pairs were categorized into four classes: (0) Negative, (1) Salt, (2) Cocrystal, and (3) Solvate. This experimentally determined label set supports an independent evaluation of model generalizability on target molecules unseen in the CSD.

2. MCC-GCN Model Specification and Training Protocol

2.1 Graph representation of MCC pairs

Each MCC candidate comprises an ordered molecular pair ($\mathcal{M}_A$,$\mathcal{M}_B$) A. In MCC-GCN, the pair is represented as a single undirected graph $G = (V, E)$ by taking the disjoint union of the two molecular graphs and indexing all atoms in a shared node set. Let N_A and N_B denote the number of atoms in $\mathcal{M}_A$ and $\mathcal{M}_B$, respectively, and $N = N_A + N_B$. The node feature matrix is $X \in \mathbb{R}^{N\times F}$, where $F = 34$ and the atom- and bond-level feature definitions follow the main manuscript. Here we focus on the pairwise graph construction, the training protocol, and transfer learning settings.

2.1.1 Notation and label convention

Throughout this note, we denote by C the number of classes (C = 4 for crystal-form prediction) and by $y \in \{1,\ldots,C\}$ the class label for a sample. The corresponding one-hot representation is $y \in \{0,1\}^C$. Predicted logits are $\hat{y} \in \mathbb{R}^C$ and probabilities are $p = \mathrm{softmax}(\hat{y})$. We use $G_{A,B}$ and $G_{B,A}$ to denote the two graphs obtained by swapping the two coformers, and $\hat{p}$ to denote the permutation-invariant probability vector obtained by symmetrization.

2.1.2 Preprocessing assumptions

Graph construction assumes standardized molecular inputs such that (i) chemically equivalent representations map to consistent graphs, (ii) each sample corresponds to two primary components as defined in the dataset curation protocol, and (iii) bond perception yields a valid covalent bond graph for each component. These assumptions align with the dataset construction procedure described in Supplementary Note 1 and ensure that graph topology differences reflect chemistry rather than representation artefacts.

2.1.3 Bond-channel adjacency tensor

Covalent connectivity is encoded as a bond-type-resolved adjacency tensor $A \in \{0,1\}^{N\times R\times N}$ with $R = 4$ channels corresponding to single, double, triple, and aromatic bonds. In downstream message passing, the model uses the unweighted connectivity induced by A (i.e., edges are defined wherever any channel is active), while the channelized representation is retained for

reproducible graph construction and optional analyses. In practice, A provides a compact and lossless encoding of covalent topology and allows the same sample to be re-materialized into an edge list representation without ambiguity. This is particularly useful for cross-validation of preprocessing steps and for controlled ablations that alter the edge representation while keeping node features fixed.

2.1.4 Optional intermolecular interaction channels

To support mechanism-oriented analyses, we also implemented an augmented adjacency tensor in which additional channels encode heuristic intermolecular interaction candidates between $\mathcal{M}_A$ and $\mathcal{M}_B$, including donor–acceptor complementarity (hydrogen-bond candidates), aromatic–aromatic pairing (putative π–π stacking), and CH···acceptor contact candidates. Unless otherwise stated, the reported MCC-GCN models use covalent-bond connectivity only. These optional channels are designed to provide an explicit inductive bias toward supramolecular interaction motifs commonly implicated in solid-form outcomes. Importantly, they are constructed from chemically interpretable rules and therefore can be inspected directly, enabling qualitative consistency checks against crystallographic interaction patterns when such structures are available.

2.1.5 Permutation invariance

Because MCC identity is invariant to the ordering of the two coformers, we employ a permutation-invariant prediction rule by symmetrizing over the two input orders. Denoting the model logits by $f(G) \in \mathbb{R}^C$ with $C = 4$, the final predicted class probabilities are computed as

$$\hat{p} = \frac{1}{2}\left(\mathrm{softmax}(f(G_{A,B})) + \mathrm{softmax}(f(G_{B,A}))\right). \tag{S1}$$

In addition, order-swapping can be used as a simple data augmentation strategy to encourage permutation robustness during training. This symmetrization has two practical benefits. First, it reduces variance attributable to an arbitrary coformer ordering and thus yields more stable probabilities for downstream decision-making. Second, it provides a conservative estimate of prediction confidence: cases for which the two orders produce materially different probabilities can be flagged as sensitive samples that merit closer inspection.

2.1.6 Graph construction procedure (algorithmic form)

Algorithm 1 Construction of a pair graph for MCC-GCN

Require: Two standardized molecules $\mathcal{M}_A$ and $\mathcal{M}_B$

Extract atom sets V_A and V_B and covalent bond sets E_A and E_B

Define global node set $V \leftarrow V_A \cup (V_B + N_A)$ by shifting indices of $\mathcal{M}_B$ by N_A

Initialize node feature matrix $X \in \mathbb{R}^{(N_A + N_B) \times 34}$

for all $v \in V$ **do**

  Encode x_v using the 34-dimensional atom descriptor scheme specified in the main manuscript

  Assign $X[v,\cdot] \leftarrow x_v$

end for

Initialize bond-channel tensor $A \in \{0,1\}^{N \times 4 \times N}$

for all $(u,v) \in E_A \cup (E_B + N_A)$ **do**

  Set the corresponding bond-type channel in $A[u,\cdot,v]$ and $A[v,\cdot,u]$ to 1

end for

return $G = (V,E)$ with X and A

2.2 MCC-GCN architecture

MCC-GCN follows a three-layer graph convolutional network (GCN) backbone[15] with batch normalization and ReLU activations, followed by global mean pooling and an MLP classifier. Let $H^{(0)} = X$ denote input node features. Each layer performs

$$H^{(\ell+1)} = \mathrm{ReLU}\left(\mathrm{BN}_\ell\left(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(\ell)}W^{(\ell)}\right)\right), \tag{S2}$$

where $\widetilde{A}$ denotes the (binary) adjacency with self-loops, $\widetilde{D}$ is the corresponding degree matrix, and $W^{(\ell)}$ are trainable weights. Graph-level embeddings are obtained by global mean pooling,

$$z = \frac{1}{|V|}\sum_{v \in V} h_v^{(3)}, \tag{S3}$$

and the classifier is a two-layer MLP with dropout and batch normalization:

$$z_1 = \mathrm{Dropout}(\mathrm{ReLU}(\mathrm{BN}_4(W_4 z + b_4))), \tag{S4}$$

$$z_2 = \mathrm{Dropout}(\mathrm{ReLU}(\mathrm{BN}_5(W_5 z_1 + b_5))), \tag{S5}$$

$$\hat{y} = W_{\mathrm{out}} z_2 + b_{\mathrm{out}}. \tag{S6}$$

2.2.1 Regularization and stability considerations

Batch normalization is applied after each graph convolution and dense layer to reduce internal covariate shift and improve optimization stability across heterogeneous graphs with varying sizes. Dropout is applied in the dense classifier to reduce overfitting, which is particularly important for the fine-tuning stage where the number of labeled samples is limited. Weight decay is used as an additional regularizer and is increased during fine-tuning to constrain updates when only the classifier is trainable.

2.2.2 Graph size variability and pooling

MCC graphs vary substantially in atom count across chemical space. Global mean pooling provides a size-invariant readout that aggregates node-level representations into a fixed-length graph embedding. This choice mitigates biases toward larger graphs that can arise with sum pooling, while retaining sensitivity to distributed interaction patterns encoded in node embeddings. In addition, mean pooling yields embeddings whose scale is comparatively stable across samples, which empirically improves the behavior of subsequent batch normalization in the classifier.

2.3 Two-stage optimization (pretraining and fine-tuning)

We employ a two-stage procedure: (i) pretraining on the large curated MCC dataset to learn general structure–form associations, and (ii) fine-tuning on a compact domain-specific dataset to adapt the decision boundary under distribution shift.

2.3.1 Data partitioning and model selection

For pretraining, a held-out validation subset is used for model selection and learning-rate scheduling. The validation split is performed with a fixed random seed to facilitate reproducibility. For fine-tuning, the domain-specific set is treated as a small-sample adaptation problem; the pretrained representation is retained and only the classifier is updated. Model selection follows the same principle: the model state that yields the best validation balanced accuracy (BAcc) is retained.

2.3.2 Learning-rate scheduling

We adopt ReduceLROnPlateau scheduling to adapt the learning rate based on validation loss. This strategy is well suited to heterogeneous graph datasets where the loss landscape can exhibit plateaus and where a fixed step schedule may be sensitive to the particular split. By monitoring validation loss, the schedule reduces the learning rate when optimization progress slows, thereby improving stability without requiring manual tuning of decay milestones.

2.3.3 Training workflow overview (boxed schematic)

Workflow (high level): MCC-GCN training and deployment.

Input: molecular pairs ($\mathcal{M}_A$,$\mathcal{M}_B$) and crystal-form labels $y \in \{0, 1, 2, 3\}$.

Step 1: Graph construction. Build $G_{\mathrm{A,B}}$ by disjoint union of two molecular graphs; construct X and bond channel tensor A.

Step 2: Pretraining. Optimize MCC-GCN on the large MCC dataset with a held-out validation split for model selection.

Step 3: Fine-tuning. Initialize from the pretrained model and update only the final dense layers on the domain specific dataset.

Step 4: Permutation-invariant inference. Average predictions over (A, B) and (B, A) to obtain $\hat{p}$

2.3.4 Algorithmic description

Algorithm 2 Permutation-invariant inference for MCC-GCN

Require: Molecular pair ($\mathcal{M}_A$,$\mathcal{M}_B$), trained model $f(\cdot)$

Construct graphs $G_{A,B}$ and $G_{B,A}$

$p_1 \leftarrow \mathrm{softmax}\,(f(G_{A,B}))$

$p_2 \leftarrow \mathrm{softmax}\,(f(G_{B,A}))$

return $\hat{p} \leftarrow (p_1 + p_2)/2$

Algorithm 3 Two-stage optimization with transfer learning

Require: Pretraining dataset $\mathcal{D}_{\mathrm{pre}}$, fine-tuning dataset $\mathcal{D}_{\mathrm{ft}}$

Require: Model f_θ with parameters $\theta = \{\theta_{\mathrm{gcn}}, \theta_{\mathrm{mlp}}\}$

Stage I (pretraining):

Initialize θ and optimize $\min_\theta \mathbb{E}_{(G,y)\sim\mathcal{D}_{\mathrm{pre}}} \mathrm{L}_{\mathrm{CE}}(f_\theta(G), y)$

Select $\theta^\star$ based on validation balanced accuracy and/or validation loss

4: **Stage II (fine-tuning):**

5: Initialize $\theta \leftarrow \theta^{\star}$

6: Freeze θ_{gcn} and optimize $\min_{\theta_{\text{mlp}}} \mathbb{E}_{(G,y)\sim\mathcal{D}_{\text{ft}}} \mathrm{L}_{\text{WCE}}(f_{\theta}(G),y)$

7: **return** Fine-tuned parameters $\tilde{\theta}$

2.3.5 Pretraining hyperparameters

The pretraining configuration follows the protocol used in this study. A fixed random seed was used for reproducibility, and a validation split was reserved for model selection. Training employed a standard cross-entropy objective (implemented in a class-weighted form with all weights set to one), Adam optimization with weight decay, and a ReduceLROnPlateau scheduler (mode = min).

Supplementary Table 4. Pretraining configuration for MCC-GCN.

Setting	Value
Model	
Output classes C	4
GCN hidden widths	256, 256, 128
Readout	Global mean pooling
MLP widths	128, 64
Dropout rate	0.208
Optimization	
Optimizer	Adam
Learning rate	3×10^{-4}
Weight decay	1×10^{-4}
Batch size	64
Epochs	400
Loss	Cross-entropy (class weights: [1,1,1,1])
LR scheduler	ReduceLROnPlateau (factor=0.5, patience=5, min LR = 10^{-7})
Validation split	0.1
Random seed	42

2.3.6 Fine-tuning protocol and hyperparameters

Fine-tuning updates only the final dense layers of the pretrained network while freezing earlier layers. Specifically, the last three dense layers (including their batch-normalization parameters) are unfrozen, while graph convolutional layers remain fixed. This design aims to adapt class boundaries with limited domain-specific supervision while preserving pretrained graph representations. In addition to parameter-efficient adaptation, this strategy has an interpretability advantage: because the graph feature extractor remains fixed, changes in

predictions after fine-tuning can be attributed primarily to a reweighting of pretrained features, which can be examined via attribution analyses and error breakdowns.

Supplementary Table 5. Fine-tuning dataset comprising 20 samples of minoxidil and 14 samples of KPX and KPR. Identifiers in CSD or CAS numbers of coformers are shown.

API	Cocrystal	Salt	Solvate	negative
Minoxidil	WIYXEL	NUYRIJ	BOFWAY	52-90-4
	BOFWEC	WIYWUA	CAYJOF	57-10-3
	XAVTOG	WIYXIP	NUYROP	56-84-8
	NUYSIK	NUYREF	UDABOU	56-86-0
	BOFVUR	BOFWIG	CAYJUL	56-40-6
KPX	51-36-5	50-84-0	100-21-0	99-96-7
		88-99-3	150-13-0	121-91-5
KPR	303-07-1	10312-55-7	65-49-6	491-11-2
	51-44-5	98-73-7		
		156-38-7		

Rationale for Classifying Ionic Cocrystals as Cocrystals for Experimental Results in This Study

Ionic cocrystals typically comprise three distinct components, a cation (B^+), an anion (A^-), and an additional neutral molecule (N), and can be generally represented by the formula A^-B^+N[16-18]. According to widely accepted guidelines for salt–cocrystal screening, salt formation is favored when the ΔpK_a value ($\Delta pK_a = pK_a(\text{base}) - pK_a(\text{acid})$) exceeds 3, whereas cocrystal formation is more likely when ΔpK_a is less than 0[19,20]. Within the intermediate or "unpredictable region" ($0 < \Delta pK_a < 3$), cocrystals, salts, and ionic cocrystals may all occur[21]. In our screen of 64 combinations involving two APIs (KPX, KPR) and 32 cofomers where nearly all ΔpK_a values fall within this unpredictable range (Supplementary Table 6), four ionic cocrystals were identified: KPR-1-hydroxy-2-naphthalic acid ionic cocrystal, KPR-2,6-dihydroxybenzoic acid ionic cocrystal, KPR-2,6-pyridinedicarboxylic acid ionic cocrystal, KPR-L-tartaric acid ionic cocrystal. These ionic cocrystals, which occupy the conceptual intersection between cocrystals and salts, extend beyond the four-class framework examined in this study. For the purpose of our classification model, ionic cocrystals were assigned to the cocrystal category because they exhibit incomplete proton transfer between the API and the coformer. Among them, KPR-2,6-pyridinedicarboxylic acid ionic cocrystal was selected for fine-tuning and detailed analysis.

Supplementary Table 6. ΔpK_a values of 64 molecule pairs studied in this work.

CAS	Coformer	ΔpK_a (KPX)	ΔpK_a (KPR)
81-04-9	1,5-naphthalenedisulfonic acid	5.642	5.718
86-48-6	1-hydroxy2-naphthoic acid	2.019	2.095
50-84-0	2,4-dichlorbenzoic acid	2.36	2.436
89-86-1	2,4-dihydroxybenzoic acid	1.718	1.794
303-07-1	2,6-dihydroxybenzoic acid	3.743	3.819
499-83-2	2,6-pyridinedicarboxylic acid	2.339	2.415
10312-55-7	2-aminoterephthalic acid	1.091	1.167
120-18-3	2-naphthalenesulfonic acid	4.773	4.849
527-72-0	2-thiophenecarboxylic acid	1.533	1.609
51-44-5	3,4-dichlorobenzoic acid	1.441	1.517
99-50-3	3,4-dihydroxybenzoic acid	0.591	0.667
51-36-5	3,5-dichlorobenzoic acid	1.581	1.657
491-11-2	3-chloro-4-nitrophenol	-1.278	-1.202
150-13-0	4-aminobenzoic acid	0.181	0.257
65-49-6	4-aminosalicylic acid	1.466	1.542
99-96-7	4-hydroxybenzoic acid	0.471	0.547
156-38-7	4-hydroxyphenylacetic acid	0.54	0.616
50-85-1	4-methylsalicylic acid	1.873	1.949
98-73-7	4-tert-butylbenzoic acid	0.641	0.717
618-83-7	5-hydroxyisophthalic acid	1.632	1.708
124-04-9	adipic acid	0.652	0.728
65-85-0	benzoic acid	0.841	0.917
144060-53-7	febuxostat	1.623	1.699
110-17-8	fumaric acid	1.891	1.967
121-91-5	isophthalic acid	1.512	1.588
87-69-4	L-tartaric acid	1.972	2.048
24280-93-1	mycophenolic acid	0.331	0.407
118-90-1	o-toluic acid	1.096	1.172
88-99-3	phthalic acid	2.092	2.168
69-72-7	salicylic acid	2.037	2.113
110-15-6	succinic acid	0.798	0.874
100-21-0	terephthalic acid	1.552	1.628

The pK_a values were retrieved from the SciFinder database.

Supplementary Table 7. Fine-tuning configuration for MCC-GCN (domain adaptation).

Setting	Value
Batch size	16
Epochs	50
Optimizer	Adam
Learning rate	3×10^{-4}
Weight decay	3×10^{-1}
Weight decay scope	Applied only to trainable parameters during classifier-only fine-tuning
Loss	Cross-entropy (class weights: [1,1,1,2])
LR scheduler	ReduceLROnPlateau (factor=0.5, patience=3, min LR=10^{-5})
Trainable layers	Final dense layers (3-layer unfreezing)
Random seed	12

2.4 Computational environment

Unless otherwise stated, training was performed with GPU acceleration on NVIDIA GeForce RTX 4090 (24GB VRAM). The software stack comprises PyTorch, PyTorch Geometric, and CUDA (versions as reported in the main manuscript).

2.5 Inference protocol and output handling

2.5.1 Probability vectors and decision rules

For each query pair, the model outputs a probability vector $\hat{p} \in [0,1]^C$ with $\sum_c \hat{p}_c = 1$. The predicted class is $\mathrm{argmax}_c \hat{p}_c$. For binary tasks derived from the multiclass setting, the probability of MCC formation can be obtained by aggregating the relevant classes (e.g., summing the probabilities of salt, cocrystal, and solvate) and applying an application-specific threshold. In this work, classification performance is reported using standard thresholds implied by argmax for multiclass prediction.

2.5.2 Permutation sensitivity as a diagnostic

Although symmetrized probabilities are used for final predictions, the discrepancy between softmax ($f(G_{A,B})$) and softmax($f(G_{B,A})$) can be used as a diagnostic signal. Large discrepancies indicate that the model is sensitive to input ordering, which may arise from borderline samples, preprocessing edge cases, or distributional shift. Such cases are suitable candidates for manual review or for targeted fine-tuning set expansion.

2.6 Objective functions, class imbalance, and evaluation metrics

2.6.1 Cross-entropy objective (optionally class-weighted)

For the C-class classification task ($C = 4$ in this study), MCC-GCN outputs logits $\hat{y} \in \mathbb{R}^C$ for each input graph G. Let $p = \mathrm{softmax}(\hat{y})$ denote the predicted class probabilities and $y \in \{0,1\}^C$ denote the one-hot label vector. We optimize a weighted cross-entropy loss

$$L_{WCE}(p,y) = -\sum_{c=1}^{C} w_c \, y_c \log p_c, \tag{S7}$$

where $w_c > 0$ are class weights. In pretraining, uniform weights were used (w_c=1). In fine-tuning, class weights were adjusted to upweight the solvate class (Supplementary Table 7).

2.6.2 Permutation augmentation

To reduce sensitivity to the arbitrary ordering of coformers, we use order-swapping as a label-preserving augmentation. Given a labeled pair $(\mathcal{M}_A, \mathcal{M}_B, y)$, we generate an additional training instance $(\mathcal{M}_B, \mathcal{M}_A, y)$. This augmentation is complementary to the permutation-invariant inference rule and encourages the model to learn symmetric interaction patterns.

2.6.3 Balanced accuracy

Balanced accuracy (BAcc) is reported throughout as the primary selection metric under class imbalance. For the multiclass task, we define

$$\mathrm{BAcc} = \frac{1}{C}\sum_{c=1}^{C} \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FN}_c}, \tag{S8}$$

where TP_c and FN_c denote true positives and false negatives for class c. For binary tasks, this reduces to BAcc = (TPR + TNR)/2.

2.6.4 Macro-averaged precision/recall/F1

For completeness, macro-averaged precision and recall are computed as the unweighted mean across classes:

$$\mathrm{MacroP} = \frac{1}{C}\sum_{c=1}^{C} \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c}, \quad \mathrm{MacroR} = \frac{1}{C}\sum_{c=1}^{C} \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FN}_c}, \tag{S9}$$

and the macro-averaged F1 score is

$$\mathrm{MacroF1} = \frac{1}{C}\sum_{c=1}^{C} \frac{2\,\mathrm{P}_c\,\mathrm{R}_c}{\mathrm{P}_c + \mathrm{R}_c}, \tag{S10}$$

where P_c and R_c are the class-specific precision and recall.

2.6.5 Training loop (algorithmic form)

Algorithm 4 Training loop with model selection by validation BAcc

Require: Training set $\mathcal{D}_{\text{tr}}$, validation set $\mathcal{D}_{\text{val}}$, model f_θ
Require: Optimizer and learning-rate scheduler; class weights $\{w_c\}$
$\theta_{\text{best}} \leftarrow \theta$; $\text{BAcc}_{\text{best}} \leftarrow -\infty$
for $t = 1$ to T **do**
  Update θ by minimizing $\mathbb{E}_{(G,y)\sim\mathcal{D}_{\text{tr}}}\mathcal{L}_{\text{WCE}}(\text{softmax}(f_\theta(G)),y)$
  Compute validation loss and BAcc on $\mathcal{D}_{\text{val}}$
  Update scheduler using validation loss
  if BAcc > $\text{BAcc}_{\text{best}}$ **then**
    $\theta_{\text{best}} \leftarrow \theta$; $\text{BAcc}_{\text{best}} \leftarrow$ BAcc
  end if
end for
return θ_{best}

2.7 Practical considerations and limitations

2.7.1 Scope of the graph abstraction

MCC-GCN operates on static molecular graphs and therefore captures structural information encoded in covalent topology and local chemical environments. It does not explicitly model crystallization conditions such as solvent identity, temperature, stoichiometry, or process parameters. These factors are known to influence experimentally observed outcomes and can introduce label noise or apparent inconsistencies when comparing across studies. The present abstraction is intended as a chemistry-first baseline that enables high-throughput screening from molecular structure alone.

2.7.2 Label noise, de-duplication, and robustness considerations

Crystallographic repositories are not designed as controlled benchmark datasets and can contain heterogeneous reporting conventions, repeated depositions, and cases where multiple solid forms may be chemically plausible for the same molecular pair under different conditions. Consequently, residual label noise is difficult to eliminate completely even after conservative curation. In this study, potential sources of ambiguity were handled by exclusion during dataset construction (Supplementary Note 1), prioritizing label clarity over maximal coverage. From a modeling perspective, three design choices further mitigate sensitivity to such noise: (i) permutation augmentation and permutation-invariant inference reduce spurious dependence on input ordering; (ii) balanced accuracy is used as a primary selection metric to avoid dominance

by frequent classes; and (iii) classifier-only fine-tuning adapts decision boundaries in target chemical space without re-optimizing the full representation, reducing the risk of overfitting to small, potentially noisy adaptation sets.

2.7.3 Distribution shift and the role of fine-tuning

Performance degradation under distribution shift is expected when target APIs and their associated MCCs are absent from the pretraining corpus. Fine-tuning provides a practical mechanism to adjust decision boundaries using a small number of representative samples in the target domain. In particular, fine-tuning is most effective when the adaptation set covers all crystal-form classes and includes both positive and negative outcomes, thereby constraining class-conditional distributions in the target chemical space.

3. Baseline Method Implementation Details

3.1 Classical computational methods

Five widely used rule-based computational screening methods (MC[22], HSP[23], MEP[24], HBP[25], and COSMO-RS[26]) were evaluated as classical baselines in the binary formation prediction task. Because these workflows rely on external commercial software and collaborator-side execution, we do not reproduce step-by-step operational details in this Supplementary Note. Instead, we refer readers to the main manuscript *Methods* section for the exact definitions, equations, and decision thresholds used, and we report results exactly as summarized in the main text.

3.2 Machine learning baselines

This section provides an expanded, textbook-style description of the three descriptor-based machine learning (ML) baselines used in this study: Support Vector Machine (SVM)[27], Random Forest (RF)[28], and Multi-Layer Perceptron (MLP)[29]. These baselines are widely used in cheminformatics and pharmaceutical informatics because they offer strong performance with limited data and provide interpretable decision behaviors relative to deep graph models.

3.2.1 Problem formulation (binary and four-class)

Each data instance corresponds to an ordered molecular pair $(\mathcal{M}_A,\mathcal{M}_B)$ represented by a fixed-length descriptor vector $x \in \mathbb{R}^d$, with label y.

- **Binary formation prediction.** $y \in \{0,1\}$ indicates whether the pair forms any MCC (positive) or is a negative case (negative), consistent with the main manuscript.
- **Crystal-form prediction (four-class).** $y \in \{0,1,2,3\}$ denotes one of four classes (salt / cocrystal / solvate / negative) following the label space reported in the main manuscript.

All baseline results reported in the paper follow the same evaluation protocol and metrics as the main manuscript (balanced accuracy for binary; accuracy and macro-averaged metrics for four-class).

Why include these baselines?

From a pharmaceutical screening perspective, SVM/RF/MLP provide three complementary "classical" inductive biases for learning on physicochemical descriptors: (i) a *margin-based* decision boundary (SVM), (ii) a *rule-ensemble* partitioning of descriptor space (RF), and (iii) a *continuous nonlinear function approximator* (MLP). Together they constitute a strong and widely accepted reference point for evaluating whether a graph model yields benefit beyond standard tabular ML.

3.2.2 Descriptor input to baseline models

Descriptor-based baselines operate on fixed-length vectors rather than molecular graphs. In this work, the descriptor set follows the 16 molecular-level features described in the main manuscript (e.g., shape, polarity, hydrogen-bonding capacity, flexibility). For a two-component system, the two descriptor vectors are concatenated to form $x = [u_A;u_B] \in \mathbb{R}^{32}$, where $u_A,u_B \in \mathbb{R}^{16}$ correspond to the two coformers. This representation is simple and efficient, but it cannot explicitly encode atom-level interaction patterns, which motivates the graph-based MCC-GCN.

Interpretation of descriptor learning for MCC prediction.

In these baselines, the learning problem can be viewed as discovering a mapping from *global, molecule-level* properties to an MCC outcome. For example, descriptor combinations related to polarity and hydrogen-bonding capacity can be informative for salt vs. cocrystal

tendencies, whereas shape and flexibility descriptors can influence packing feasibility and solvate propensity. However, because x summarizes each molecule as a single vector, these models can only represent intermolecular compatibility *implicitly* through nonlinear combinations (RF/MLP) or through a learned boundary in descriptor space (SVM), rather than explicitly modeling atom–atom interaction patterns.

Ordered pair and symmetry consideration.

The concatenation $x = [u_A;u_B]$ encodes an *ordered* pair. In principle, the underlying MCC identity is symmetric under swapping the two coformers. In practice, descriptor-based baselines may address this by enforcing a deterministic ordering rule (e.g., sort by molecular weight) or by augmenting the data with swapped pairs. We report baseline results as in the main manuscript; the goal here is to provide a transparent conceptual foundation for how classical models operate on such descriptors.

3.2.3 Training convention and hyperparameters

All three baselines were implemented with *scikit-learn* (version as reported in the main manuscript)[30]. Consistent with the study setting (and to avoid overfitting by excessive tuning), we used **default hyperparameters** provided by the library for each estimator unless otherwise stated. Therefore, the tables below should be interpreted as *default configurations* rather than tuned best settings.

Why default hyperparameters can be informative in scientific benchmarking.

In many applied pharmaceutical settings, descriptor-based models are used as "quick-to-deploy" screening tools. Using default hyperparameters provides a conservative and reproducible baseline that reflects this usage pattern: it reduces the degrees of freedom in model selection, limits the risk of over-optimistic results due to extensive tuning, and helps isolate whether performance gains arise from the *representation* (graph vs. descriptor) rather than from hyperparameter optimization.

Supplementary Table 8. Default hyperparameter configuration for the Support Vector Machine (SVM) baseline (*scikit-learn*).

Hyperparameter	Value	Description
Estimator	SVC	Support Vector Classifier

Hyperparameter	Value	Description
Kernel	rbf	Radial basis function kernel
C	1.0	Soft-margin penalty (regularization)
γ	scale	RBF width parameter scaling
Class weight	None	No reweighting by default
Probability	False	No Platt scaling by default

Supplementary Table 9. Default hyperparameter configuration for the Random Forest (RF) baseline (*scikit-learn*).

Hyperparameter	Value	Description
Estimator	RandomForestClassifier	Bagged decision trees
n_estimators	100	Number of trees
Criterion	gini	Split quality function
Max depth	None	Expand until pure or constrained
Min samples split	2	Minimum samples to split

Supplementary Table 10. Default hyperparameter configuration for the Multi-Layer Perceptron (MLP) baseline (*scikit-learn*).

Hyperparameter	Value	Description
Estimator	MLPClassifier	Feedforward neural network
Hidden layer sizes	(100,)	One hidden layer by default
Activation	relu	Rectified linear unit
Solver	adam	Adaptive moment estimation
α	0.0001	L2 regularization strength
Initial LR	0.001	Initial learning rate
Max iter	200	Maximum training iterations

3.2.4 Support Vector Machine (SVM)

Core idea. SVM is a maximum-margin classifier. In its simplest linear form, it seeks a hyperplane that separates classes while maximizing the margin, which is associated with improved generalization.

Geometric intuition (margin and support vectors). For a linear decision boundary $w^{\top}x + b = 0$, the signed distance of a point x to the hyperplane is proportional to $w^{\top}x + b$. Maximizing the minimum distance over training points yields a boundary that is "as far as possible" from both classes, making the classifier less sensitive to small perturbations and descriptor noise. Importantly, only a subset of samples—the *support vectors*—determine the final boundary. This behavior is attractive in chemical screening tasks where many samples

may be redundant and where the decision boundary should be controlled by the most informative examples.

Binary soft-margin SVM (primal form). Given labeled samples $\{(x_i,y_i)\}_{i=1}^{n}$ with $y_i \in \{-1, +1\}$, the soft-margin SVM solves

$$\min_{w,b,\xi} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n}\xi_i \quad \text{s.t.} \quad y_i(w^\top x_i + b) \geq 1-\xi_i,\ \xi_i \geq 0, \tag{S11}$$

where w and b define the decision boundary, ξ_i are slack variables allowing non-separable data, and $C > 0$ controls the margin–error trade-off.

Equivalent hinge-loss view. The same objective can be interpreted as minimizing a regularized hinge loss. Defining the margin $m_i = y_i(w^\top x_i + b)$, the hinge loss is $\max(0, 1-m_i)$ and the optimization can be written as

$$\min_{w,b} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n}\max(0, 1-y_i(w^\top x_i + b)). \tag{S12}$$

This emphasizes that SVM penalizes points that fall within the margin band or are misclassified, while points well outside the margin do not contribute to the loss.

Kernel trick and **RBF kernel.** To capture non-linear decision boundaries, SVM can be expressed in a dual form in terms of inner products $\langle x_i, x_j \rangle$. Replacing inner products by a kernel $K(x_i, x_j)$ yields a flexible nonlinear classifier. The RBF kernel is

$$K(x, z) = \exp(-\gamma\|x-z\|_2^2), \tag{S13}$$

where $\gamma > 0$ controls the locality of the similarity: larger γ yields more localized decision boundaries.

Effect of *C* and *γ* (practical intuition). Although we use default hyperparameters in this study, it is useful to interpret what these parameters control. Larger C encourages the model to fit training data more strictly (narrower margin, fewer violations), which may overfit noisy descriptor signals. Smaller C encourages a wider margin and stronger regularization. For the RBF kernel, larger γ makes each support vector influence only a small neighborhood (highly flexible boundary), whereas smaller γ yields smoother, more global boundaries in descriptor space. In pharmaceutical MCC screening, these trade-offs correspond to whether the model behaves as a conservative broad classifier or a highly specific rule-like separator.

Multiclass extension**.** For four-class prediction, practical SVM implementations reduce

multiclass classification to multiple binary subproblems. In *scikit-learn* SVC, the default strategy is one-vs-one (OvO): a separate classifier is trained for each pair of classes, and the final prediction is obtained by voting across these classifiers.

Outputs and probability interpretation. SVM naturally produces a decision score (signed distance in a transformed feature space). Probability estimates can be obtained by an additional calibration step (e.g., Platt scaling), but in SVC this is disabled by default[28]. Therefore, SVM baseline results are best interpreted as *discriminative class predictions* rather than calibrated probabilities.

Algorithm 5 Training procedure for SVM baseline (conceptual)

Require: Descriptor matrix $X \in \mathbb{R}^{n \times d}$, labels y

1. Choose kernel $K(\cdot,\cdot)$ (RBF) and hyperparameters (C, γ)
2. Solve the convex optimization problem to obtain support vectors and coefficients
3. Define decision function $f(x) = \sum_{i \in S} \alpha_i y_i K(x_i, x) + b$
4. **return** Trained SVM model.

SVM training pseudocode.

3.2.5 Random Forest (RF)

Core idea. Random Forest is an ensemble method that aggregates many decision trees trained on randomized versions of the data. It reduces variance relative to a single tree, improves robustness, and typically performs strongly on tabular descriptor features without requiring feature scaling.

Why RF is common in cheminformatics baselines. RF handles mixed-scale and partially redundant descriptors well because splits depend on relative ordering rather than absolute scaling. It can also model non-additive effects through hierarchical splits (e.g., "high polarity AND high HBD/HBA capacity"), which matches the intuition that MCC outcomes are often driven by coupled physicochemical conditions rather than a single descriptor.

Decision trees and impurity. A decision tree recursively partitions the feature space. At each node, the algorithm selects a feature j and split threshold t to maximize the reduction in impurity. For classification, the Gini impurity for a node with class proportions $\{p_c\}_{c=1}^{C}$ is

$$G = 1 - \sum_{c=1}^{C} p_c^2 . \tag{S14}$$

The best split maximizes the impurity decrease

$$\Delta G = G_{\text{parent}} - \left(\frac{n_L}{n} G_L + \frac{n_R}{n} G_R\right), \tag{S15}$$

where L and R denote left/right children and n_L, n_R are their sample counts[32].

Bagging and feature randomness. Each tree is trained on a bootstrap sample of the training data. Additionally, at each split, only a random subset of features is considered (a key source of diversity). The final prediction is obtained by majority vote (classification) or averaging (regression)[33].

Class probabilities. Besides class labels, RF can output class probabilities by averaging the class-frequency distributions at the terminal leaves over all trees. If $p_c^{(t)}(x)$ denotes the proportion of class c in the leaf reached by x in tree t, then the forest probability is

$$\hat{p}_c(x) = \frac{1}{T} \sum_{t=1}^{T} p_c^{(t)}(x). \tag{S16}$$

These probabilities are often useful in screening scenarios for ranking candidate coformers, although they may require calibration for strict probabilistic interpretation.

RF prediction. Let T denote the number of trees and $h_t(x)$ be the predicted class label from tree t. RF predicts

$$\hat{y} = \text{mode}\{h_t(x)\}_{t=1}^{T}. \tag{S17}$$

Interpretability: feature importance (caveat). RF supports feature-importance summaries (e.g., mean decrease in impurity). While such scores provide a coarse indication of which descriptors frequently contribute to splits, they should be interpreted cautiously because correlated descriptors can "share" importance and because impurity-based measures can be biased toward variables with many possible split points.

Algorithm 6 Training procedure for Random Forest baseline (conceptual)

Require: Descriptor matrix X, labels y, number of trees T

1. **For** t =1 to T **do**
2. Draw bootstrap sample $(X^{(t)}, y^{(t)})$ from (X, y)
3. Train a decision tree using randomized feature subsets at each split
4. **end for**
5. **return** Ensemble of trees $\{h_t\}_{t=1}^{T}$

RF training pseudocode.

3.2.6 Multi-Layer Perceptron (MLP)

Core idea. MLP is a feedforward neural network that learns a sequence of nonlinear

transformations from descriptors to class scores. Compared with linear models, MLP can capture nonlinear interactions between descriptors, which is useful when formation depends on multiple coupled physicochemical factors.

Relationship to logistic regression and why it matters. If all activation functions were removed, an MLP would reduce to a linear model similar to multinomial logistic regression. The nonlinear activations (ReLU) allow the model to represent descriptor interactions beyond linear additivity, such as threshold-like behaviors (e.g., "only when both dipole-related descriptors and hydrogen-bonding descriptors exceed certain levels"). This is a plausible abstraction for MCC formation heuristics, where multiple requirements often need to be satisfied simultaneously[34].

Forward model. For an L-layer MLP, the hidden activations are computed iteratively:

$$h^{(0)} = x, \tag{S18}$$

$$h^{(\ell)} = \phi\left(W^{(\ell)} h^{(\ell\text{-}1)} + b^{(\ell)}\right), \quad \ell = 1, \ldots, L\text{-}1, \tag{S19}$$

$$o = W^{(L)} h^{(L\text{-}1)} + b^{(L)}, \tag{S20}$$

where $\phi(\cdot)$ is a nonlinear activation (ReLU in this study). For multiclass classification, outputs are converted to probabilities by softmax:

$$p_c = \frac{\exp(o_c)}{\sum_{k=1}^{C} \exp(o_k)}. \tag{S21}$$

Loss function. For multiclass classification, the cross-entropy loss is

$$\mathcal{L} = -\sum_{i=1}^{n} \sum_{c=1}^{C} \mathbf{1}\,(y_i = c) \log p_{i,c}. \tag{S22}$$

In practice, optimization proceeds by gradient-based methods (e.g., Adam), and regularization may be applied via L_2 penalty [35].

Regularization and convergence (practical notes). MLPs can overfit when training data are limited or labels are noisy. The L_2 penalty (controlled by α in MLPClassifier) discourages excessively large weights, improving stability. Optimization is non-convex, so performance may depend on initialization and solver dynamics; this motivates the use of robust default solvers such as Adam and conservative stopping criteria in standard baseline practice.

Backpropagation (conceptual). Training updates parameters by computing gradients of

$\mathcal{L}$ with respect to each $W^{(\ell)}$ and $b^{(\ell)}$ using the chain rule and then applying an optimizer step.

Algorithm 7 Training procedure for MLP baseline (conceptual)

Require: Training descriptors X, labels y, learning rate η, epochs E

Initialize network parameters $\{W^{(\ell)}, b^{(\ell)}\}$
For e =1 to E **do**
    for all mini-batches (X_b, y_b) do
    Forward pass: compute logits and probabilities
    Compute loss $\mathcal{L}$ (cross-entropy)
    Backward pass: compute gradients via backpropagation
    Update parameters using an optimizer step
    end for
end for
return Trained MLP model

MLP training pseudocode.

3.2.7 Summary: what these baselines can and cannot capture

Strengths. These baselines provide strong references when only global descriptors are available: SVM offers a controlled margin-based boundary, RF offers robust nonlinear partitions with limited preprocessing, and MLP offers flexible nonlinear composition of descriptors.

Limitations relative to graph models. All three models operate on global descriptors and therefore lack explicit access to atom-level interaction motifs (e.g., specific donor–acceptor patterns, localized charge distributions, and substructure complementarity). As a result, they may require more data to generalize across new APIs, and their predictions can be sensitive to representation choices such as descriptor definition, ordering of coformers, and implicit encoding of intermolecular compatibility.

1. Fábián, L. Cambridge Structural Database Analysis of Molecular Complementarity in Cocrystals. Cryst. Growth Des. 9, 1436-1443 (2009).
2. Bruno, I.J. et al. New software for searching the Cambridge Structural Database and visualizing crystal structures. Acta Crystallographica Section B 58, 389-397 (2002).
3. Taylor, R. & Wood, P.A. A Million Crystal Structures: The Whole Is Greater than the Sum of Its Parts. Chem. Rev. 119, 9427-9477 (2019).
4. Aakeröy, C.B. & Salmon, D.J. Building co-crystals with molecular sense and supramolecular sensibility. CrystEngComm 7, 439-448 (2005).
5. Jiang, Y. et al. Coupling complementary strategy to flexible graph neural network for quick discovery of coformer in diverse co-crystal materials. Nat. Commun. 12, 5950 (2021).
6. Ying, R., Bourgeois, D., You, J., Zitnik, M. & Leskovec, J. in Proceedings of the 33rd International Conference on Neural Information Processing Systems Article 829 (Curran Associates Inc., 2019).
7. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011).
8. Grothe, E., Meekes, H., Vlieg, E., ter Horst, J.H. & de Gelder, R. Solvates, Salts, and Cocrystals: A Proposal for a Feasible Classification System. Cryst. Growth Des. 16, 3237-3243 (2016).
9. Aitipamula, S. et al. Polymorphs, Salts, and Cocrystals: What's in a Name? Cryst. Growth Des. 12, 2147-2152 (2012).
10. FDA, A. Pharmaceutical Solid Polymorphism. Chemistry, Manufacturing, and Controls Information, July (2007).
11. Elder, D. ICH Q6A Specifications: test procedures and acceptance criteria for new drug substances and new drug products: chemical substances. ICH quality guidelines: an implementation guide, 433-466 (2017).
12. Chappa, P. et al. Pharmaceutical Solid Form Screening and Selection: Which Form Emerges? Cryst. Growth Des. 25, 4783-4794 (2025).
13. Heller, S.R., McNaught, A., Pletnev, I., Stein, S. & Tchekhovskoi, D. InChI, the IUPAC International Chemical Identifier. J. Cheminf. 7, 23 (2015).
14. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Journal of Chemical Information and Computer Sciences 28, 31-36 (1988).
15. Kipf, T.N. & Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. arXiv:1609.02907 (2016).
16. Jin, S., Sanii, R., Song, B.-Q. & Zaworotko, M.J. Crystal Engineering of Ionic Cocrystals Sustained by the Phenol–Phenolate Supramolecular Heterosynthon. Cryst. Growth Des. 22, 4582-4591 (2022).
17. Schultheiss, N. & Newman, A. Pharmaceutical Cocrystals and Their Physicochemical Properties. Cryst. Growth Des. 9, 2950-2967 (2009).
18. Mohamed, S., Alwan, A.A., Friščić, T., Morris, A.J. & Arhangelskis, M. Towards the systematic crystallisation of molecular ionic cocrystals: insights from computed crystal form landscapes. Faraday Discuss. 211, 401-424 (2018).
19. Bhogala, B.R., Basavoju, S. & Nangia, A. Tape and layer structures in cocrystals of some di- and tricarboxylic acids with 4,4′-bipyridines and isonicotinamide. From binary to ternary cocrystals. CrystEngComm 7, 551-562 (2005).
20. Childs, S.L., Stahly, G.P. & Park, A. The Salt−Cocrystal Continuum: The Influence of Crystal

Structure on Ionization State. Mol. Pharmaceutics 4, 323-338 (2007).
21. Wang, T. et al. Salts, Cocrystals, and Ionic Cocrystals of a “Simple” Tautomeric Compound. Cryst. Growth Des. 18, 6973-6983 (2018).
22. Wood, P.A. et al. Knowledge-based approaches to co-crystal design. CrystEngComm 16, 5839-5848 (2014).
23. Hansen, C. Three dimensional solubility parameter and solvent diffusion coefficient. Importance in surface coating formulation. Doctoral Dissertation (1967).
24. Musumeci, D., Hunter, C.A., Prohens, R., Scuderi, S. & McCabe, J.F. Virtual cocrystal screening. Chem. Sci. 2, 883-890 (2011).
25. Delori, A., Galek, P.T.A., Pidcock, E. & Jones, W. Quantifying Homo- and Heteromolecular Hydrogen Bonds as a Guide for Adduct Formation. Chemistry – A European Journal 18, 6835-6846 (2012).
26. Klamt, A. & Leonhard, K. Validation of the COSMO-RS electrostatics by Monte-Carlo simulations. Fluid Phase Equilib. 261, 162-167 (2007).
27. Cortes, C. & Vapnik, V. Support-vector networks. Mach. Learn. 20, 273-297 (1995).
28. Breiman, L. Random Forests. Mach. Learn. 45, 5-32 (2001).
29. Rumelhart, D.E., Hinton, G.E. & Williams, R.J. Learning representations by back-propagating errors. Nature 323, 533-536 (1986).
30. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825-2830 (2011).
31. Platt, J. Probabilistic Outputs for Support vector Machines and Comparisons to Regularized Likelihood Methods. (1999).
32. Moore II, D.H. Classification and regression trees, by Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. Brooks/Cole Publishing, Monterey, 1984,358 pages, $27.95. Cytometry 8, 534-535 (1987).
33. Breiman, L. Bagging predictors. Mach. Learn. 24, 123-140 (1996).
34. Nair, V. & Hinton, G.E. in Proceedings of the 27th International Conference on International Conference on Machine Learning 807–814 (Omnipress, Haifa, Israel; 2010).
35. Kingma, D.P. & Ba, J. Adam: A Method for Stochastic Optimization. (2017).

4. Additional Results (Confusion Matrix Atlas)

4.1 Reading guide

This section provides a compact "atlas" of confusion matrices used throughout the study. It is intended for side-by-side inspection of error modes across datasets and model families, without introducing any additional experiments beyond those reported in the main manuscript.

Normalization and numbers. All confusion matrices are *row-normalized*: each row (a fixed true label) sums to 100%, which makes class-conditional error patterns directly comparable even when class supports differ. In addition, raw counts are shown inside each cell to indicate the absolute number of samples contributing to each percentage.

Label space and ordering. For the binary task, the labels are *formation* vs. *negative*. For the four-class task, the label space follows the experimental protocol: *negative / salt / cocrystal / solvate*. When comparing panels, focus on (i) diagonal mass (correct classification), and (ii) systematic off-diagonal flows (consistent confusions between specific classes).

Datasets and evaluation settings.

- **CSD test set (in-distribution):** held-out subset drawn from the same CSD-derived distribution used for pretraining.
- **KPXKPR benchmark (out-of-distribution):** experimental evaluation on KPX/KPR pairs, where the target APIs are absent from the CSD-derived training set.

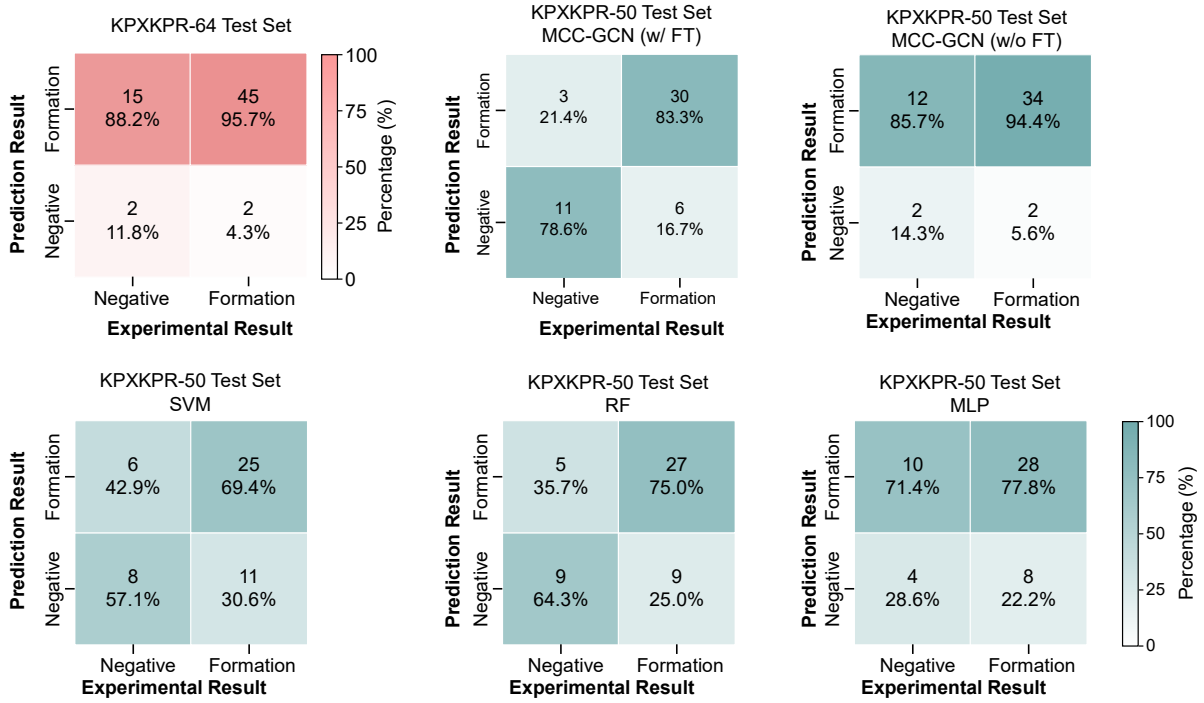
Model abbreviations. MCC-GCN denotes the graph model. **RF/SVM/MLP** denote descriptor-based baselines. For MCC-GCN, **w/o FT** and **w/ FT** indicate without and with fine-tuning (domain adaptation) as defined in the main manuscript.

How to interpret a panel quickly. Because matrices are row-normalized, a "strong" model typically shows (i) high diagonal percentages for each row, and (ii) limited concentration of errors into a single predicted class. In contrast, under distribution shift, models may show (i) a strong bias toward predicting a particular class across many true-label rows, or (ii) persistent confusions between chemically close classes (e.g., salt vs. cocrystal).

4.2 Out-of-distribution benchmark: KPXKPR

4.2.1 Binary crystallization prediction (formation vs. negative)

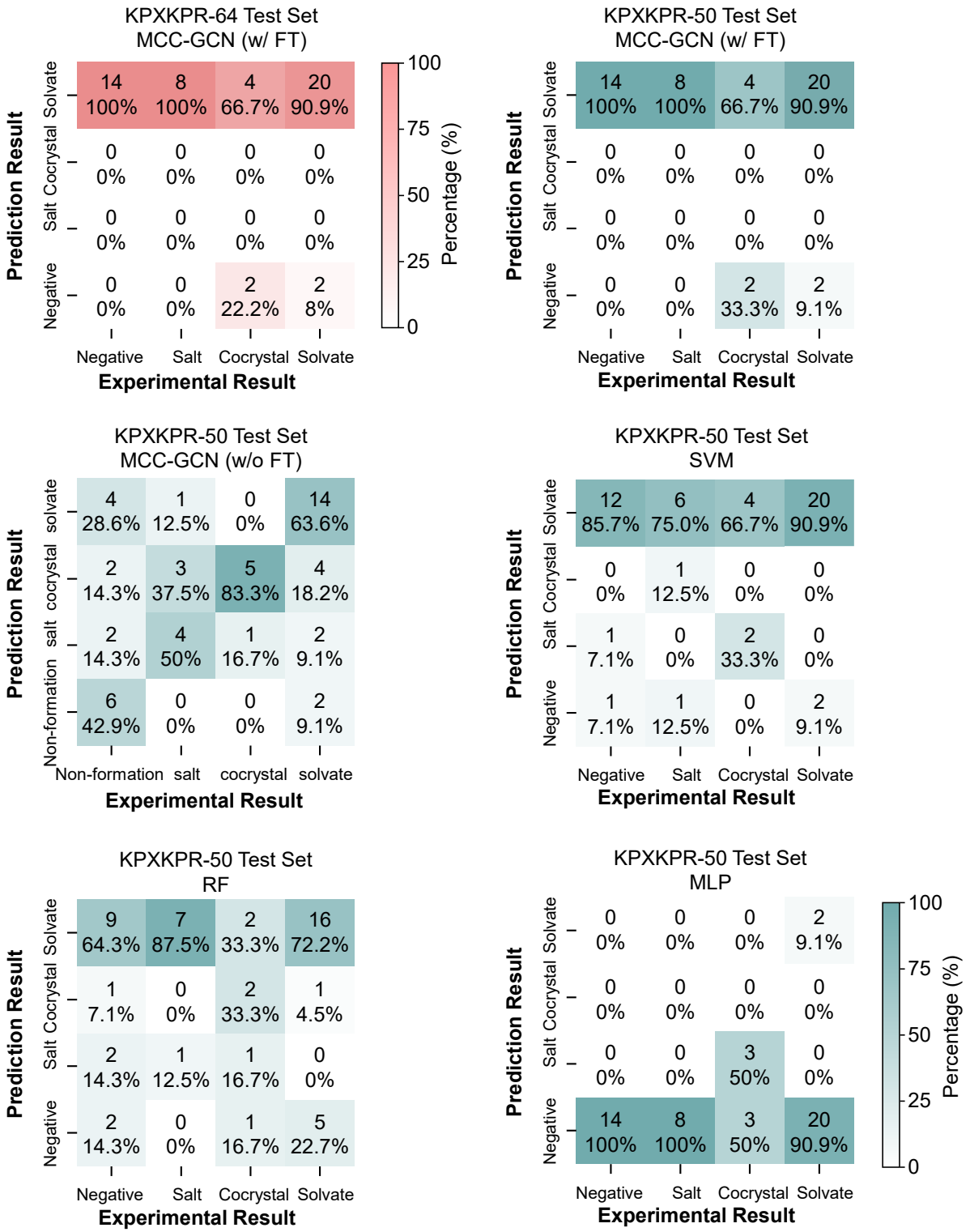
The following multi-panel figure summarizes confusion matrices for the OOD KPXKPR benchmark under the binary crystallization setting. Panels include MCC-GCN (with and without fine-tuning) and descriptor-based baselines on the KPXKPR-50 split; the KPXKPR-64 panel provides the reference performance of MCC-GCN w/o FT on the full experimental benchmark, consistent with the main manuscript.



Supplementary Figure 2. Out-of-distribution (KPXKPR) confusion matrices for the binary crystallization prediction task (formation vs. negative). Percentages indicate class-conditional rates, with columns normalized to sum to 100%.

4.2.2 Four-class crystal-form prediction (cocrystal / salt / solvate / negative)

The following multi-panel figure summarizes confusion matrices for the OOD KPXKPR benchmark under the four-class crystal-form setting, enabling direct inspection of class-specific confusions and the effect of fine-tuning on decision boundaries.

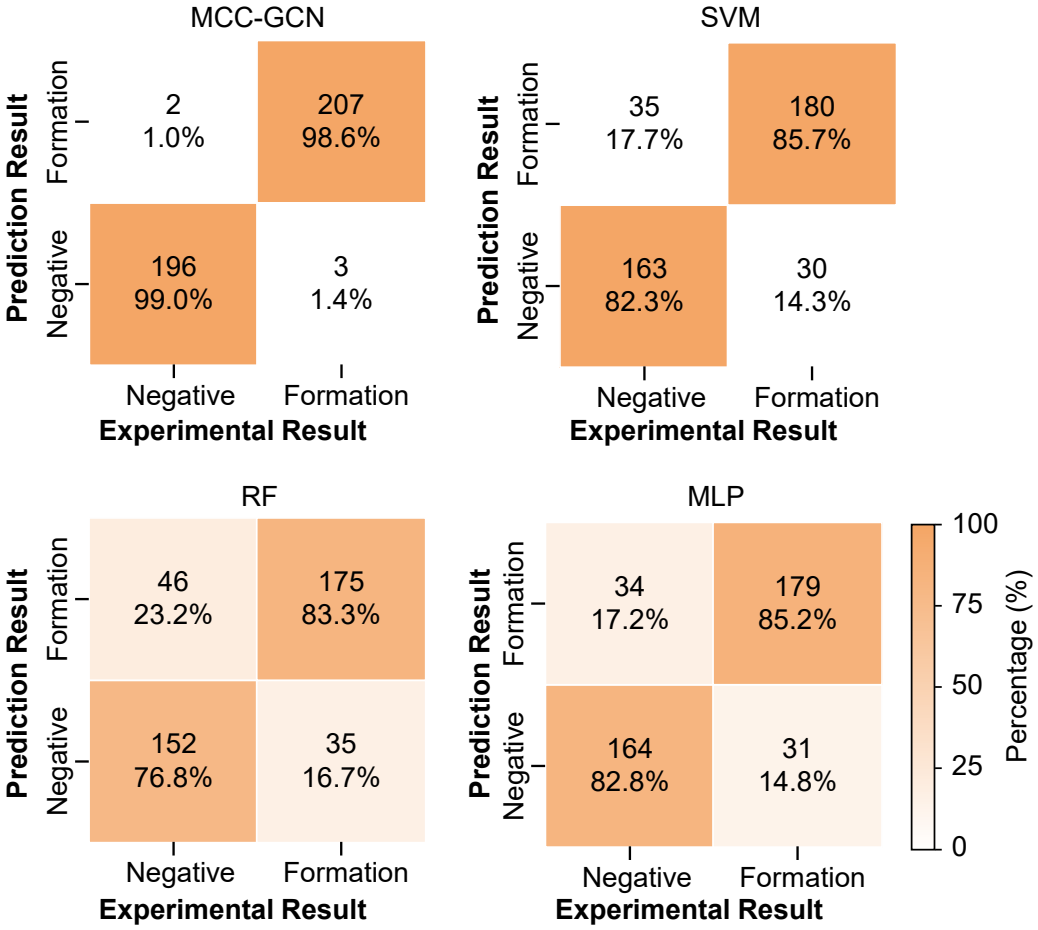


Supplementary Figure 3. Out-of-distribution (KPXKPR) confusion matrices for the four-class crystal-form prediction task. Percentages indicate class-conditional rates, with columns normalized to sum to 100%.

4.3 In-distribution benchmark: CSD test set

4.3.1. Binary crystallization prediction (formation vs. negative)

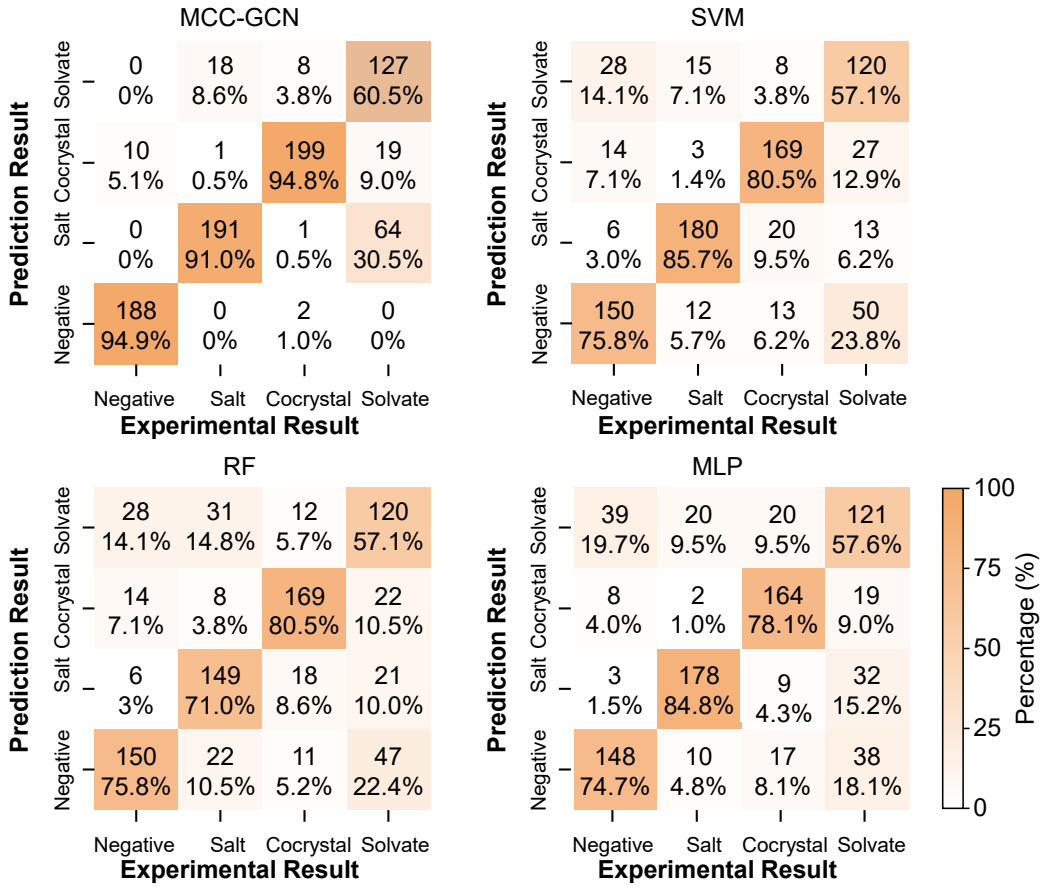
The following multi-panel figure summarizes confusion matrices on the CSD test set for the binary crystallization setting, contrasting MCC-GCN with descriptor-based baselines under an in-distribution evaluation protocol.



Supplementary Figure 4. CSD test set confusion matrices for the binary crystallization prediction task (formation vs. negative). Percentages indicate class-conditional rates, with columns normalized to sum to 100%.

4.3.2. Four-class crystal-form prediction (negative / salt / cocrystal / solvate)

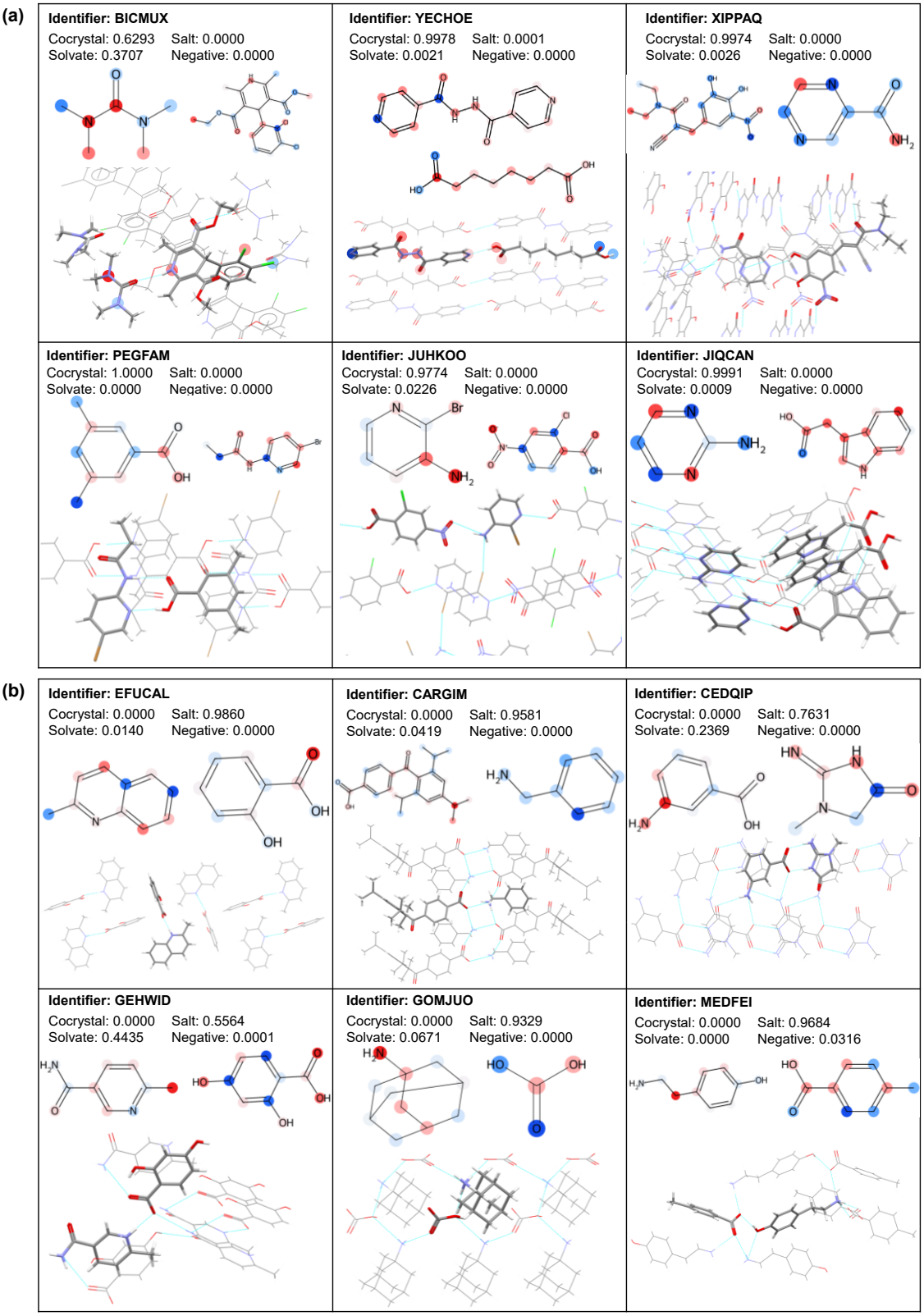
The following multi-panel figure summarizes confusion matrices on the CSD test set for the four-class crystal-form setting.

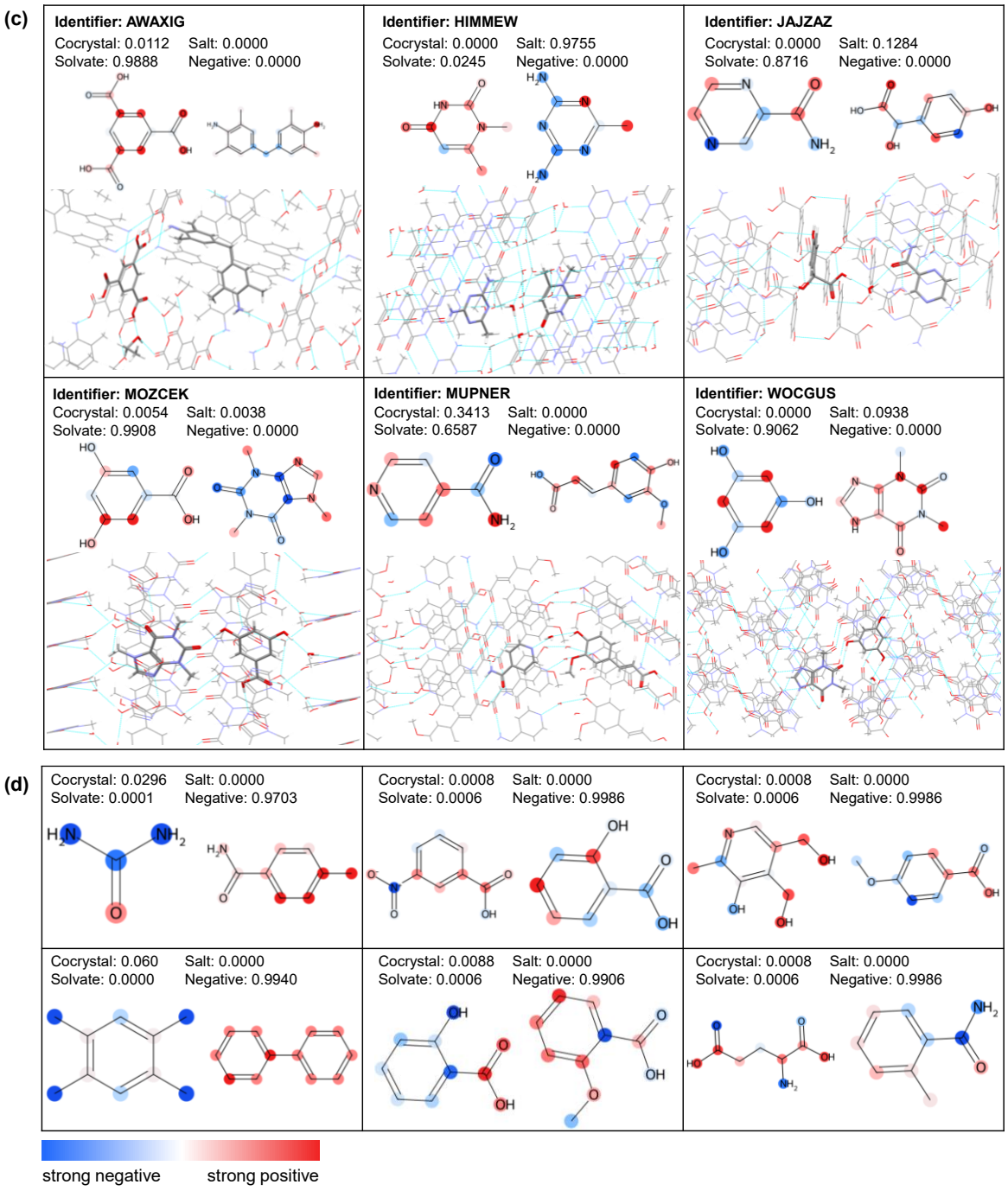


Supplementary Figure 5. CSD test set confusion matrices for the four-class crystal-form prediction task. Percentages indicate class-conditional rates, with columns normalized to sum to 100%.

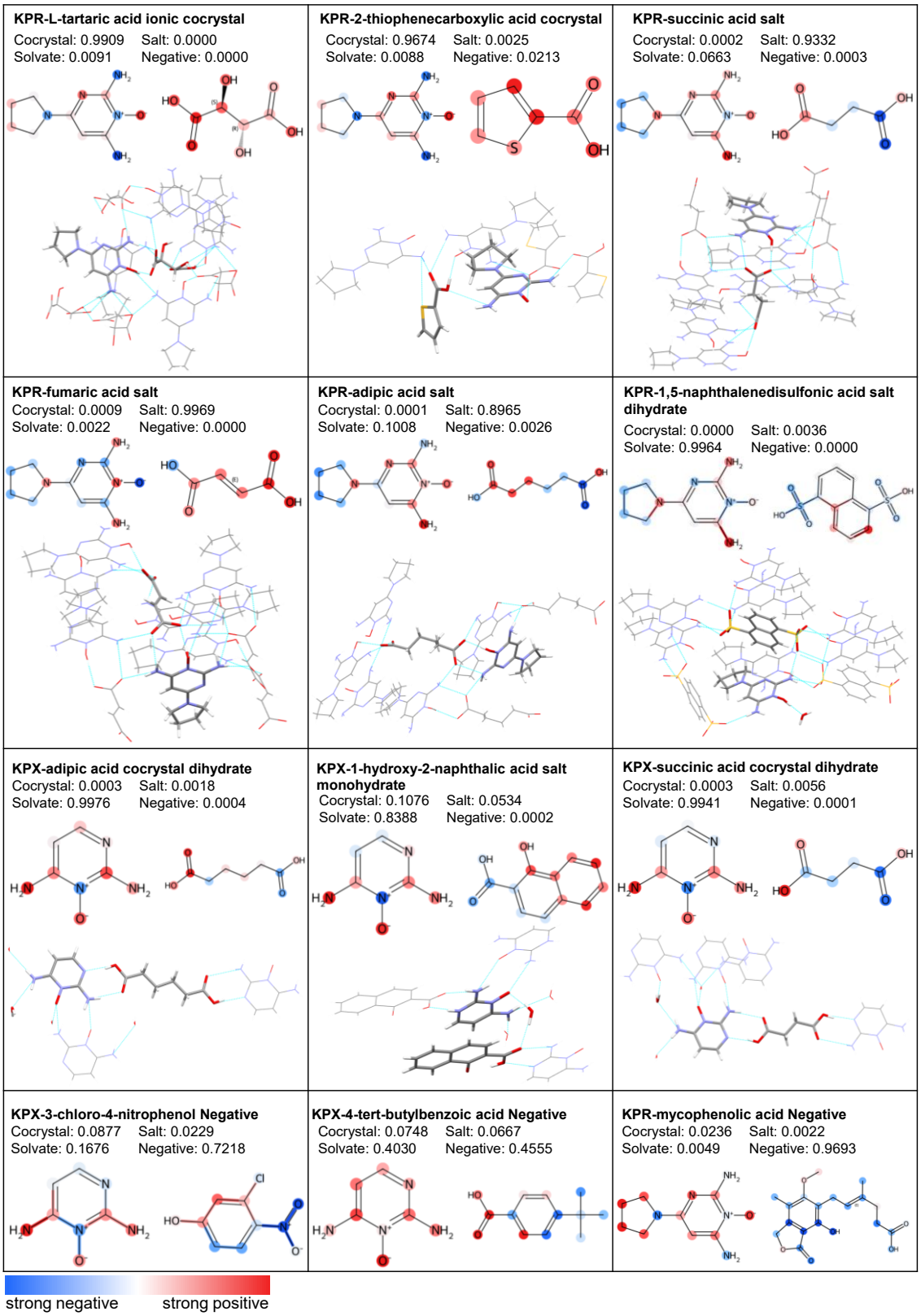
4.4 Interpretability

Interpretability analyses for MCC-GCN (gradient-based attribution) are presented in the main manuscript and following associated supplementary figures.





Supplementary Figure 6. Interpretation of formation via a gradient-based method for MCC cases in CSD, showing atomic contributions (blue: positive; red: negative) compared with the crystal structures displayed by Mercury. The cyan dash line denotes the intermolecular hydrogen bond. (a) Cocrystals including BICMUX, YECHOE, XIPPAQ, PEGFAM, JUHKOO, JIQCAN. (b) Salts including EFUCAL, CARGIM, CEDQIP, GEHWID, GOMJUO, MEDFEI. (c) Solvates including AWAXIG, HIMMEW, JAJZAZ, MOZCEK, MUPNER, WOCGUS. (d) 6 negative pairs.



Supplementary Figure 7. Interpretation of MCC formation of KPX/KPR via a gradient-based method, showing atomic contributions (blue: positive; red: negative) compared with the crystal structures displayed by Mercury. The cyan dash line denotes the intermolecular hydrogen bond.

5. Results of Classical Binary Prediction Models

Supplementary Table 11. Comparison of binary prediction models and experimental results.

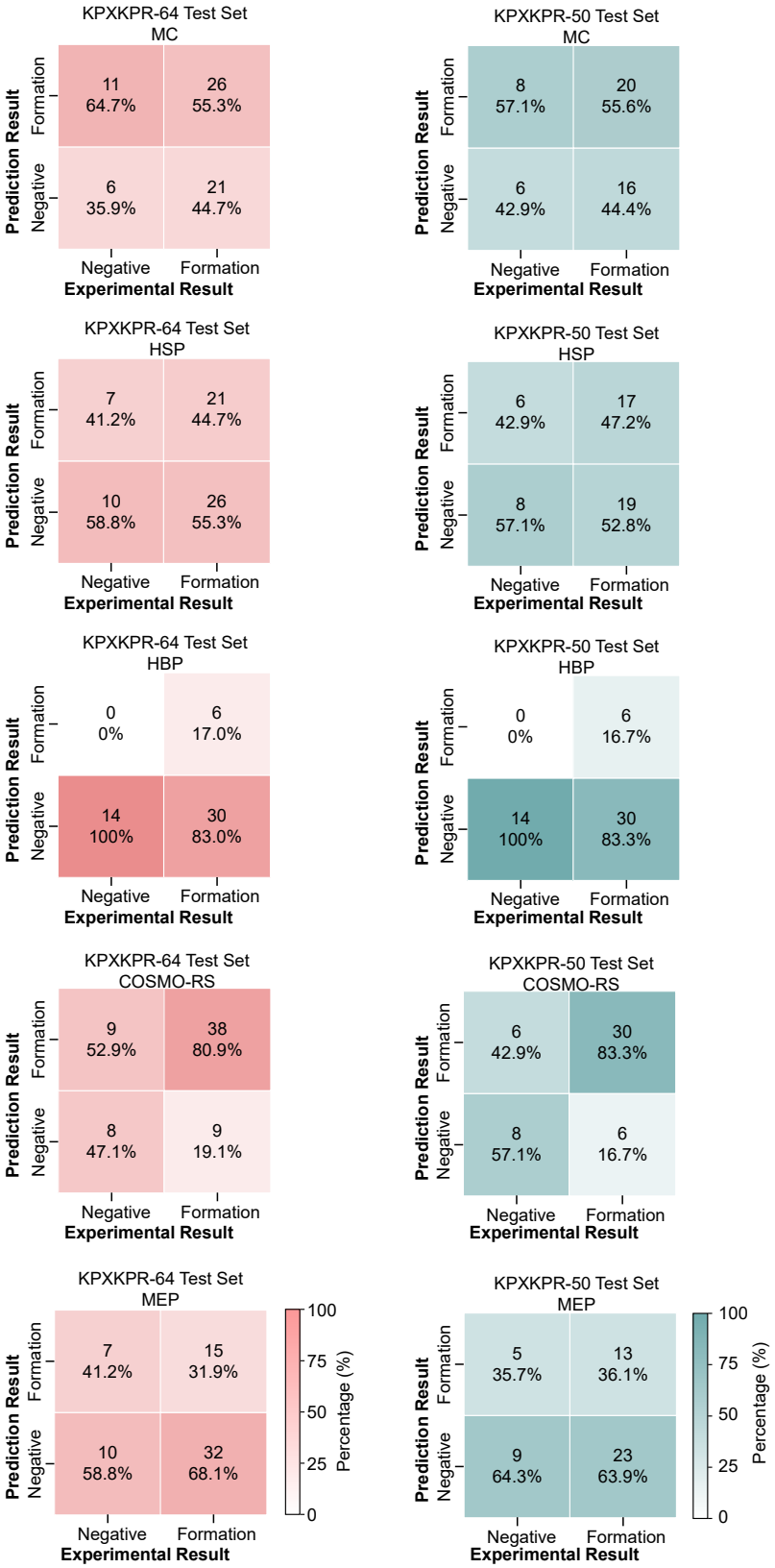
	KPX						KPR					
	MC	HSP	HBP	COSMO-RS	MEP	Exp.	MC	HSP	HBP	COSMO-RS	MEP	Exp.
1,5-naphthalenedisulfonic acid	yes	yes	no	yes	no	yes	yes	yes	no	yes	yes	yes
1-hydroxy-2-naphthoic acid	no	no	no	yes	no	yes	no	yes	no	yes	yes	yes
2,4-dichlorobenzoic acid	no	no	no	no	no	yes	yes	no	no	yes	no	no
2,4-dihydroxybenzoic acid	no	yes	no	yes	yes	yes	no	yes	no	yes	yes	no
2,6-dihydroxybenzoic acid	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
2,6-pyridinedicarboxylic acid	yes	no	no	yes	no	yes	yes	yes	no	yes	no	yes
2-aminoterephthalic acid	yes	no	yes	yes	no	yes	no	yes	yes	yes	yes	yes
2-naphthalenesulfonic acid	no	no	no	yes	no	yes	no	yes	no	yes	no	yes
2-thiophenecarboxylic acid	no	no	no	no	no	yes	yes	no	no	yes	no	yes
3,4-dichlorobenzoic acid	no	no	no	no	no	yes	yes	no	no	yes	no	yes
3,4-dihydroxybenzoic acid	yes	yes	no	yes	yes	yes	yes	yes	no	yes	yes	no
3,5-dichlorobenzoic acid	no	no	no	yes	no	yes	no	no	no	yes	yes	yes
3-chloro-4-nitrophenol	yes	no	no	yes	no	no	yes	yes	no	yes	yes	no
4-aminobenzoic acid	yes	no	no	no	no	yes	yes	yes	no	no	yes	no
4-aminosalicylic acid	yes	yes	no	no	no	no	yes	yes	no	yes	yes	yes
4-hydroxybenzoic acid	yes	no	no	yes	yes	no	yes	yes	no	yes	yes	no
4-hydroxyphenylacetic acid	no	no	no	yes	no	yes	no	yes	no	yes	no	yes
4-methylsalicylic acid	yes	no	no	no	no	yes	yes	yes	no	yes	no	yes
4-tert-butylbenzoic acid	no	no	no	no	no	no	yes	no	no	no	no	yes
5-hydroxyisophthalic acid	yes	no	no	yes	no	yes	no	yes	no	yes	yes	no
adipic acid	no	no	yes	yes	no	yes	no	yes	yes	yes	yes	yes
benzoic acid	no	no	no	no	no	yes	yes	no	no	yes	no	yes
febuxostat	no	no	no	no	no	no	no	no	no	yes	no	yes
fumaric acid	yes	no	no	no	no	yes	yes	yes	no	no	yes	yes
isophthalic acid	yes	no	no	yes	no	no	no	yes	no	yes	yes	yes
L-tartaric acid	yes	yes	no	yes	yes	yes	yes	no	no	yes	yes	yes
mycophenolic acid	no	no	no	no	no	no	yes	no	no	no	no	no
o-toluic acid	no	no	no	no	no	no	yes	no	no	no	no	no
phthalic acid	yes	no	no	yes	no	yes	yes	yes	no	yes	no	yes
salicylic acid	yes	no	yes	yes	no	yes	yes	yes	yes	yes	yes	yes
succinic acid	yes	no	no	yes	no	yes	yes	yes	no	yes	yes	yes
terephthalic acid	yes	no	yes	yes	no	yes	no	yes	yes	yes	yes	yes

Supplementary Table 12. Evaluation of different binary prediction models on KPXKPR-64 dataset.

	MC	HSP	HBP	COSMO-RS	MEP
True Positive	26	21	8	38	15
False Positive	11	7	0	9	7
True Negative	6	10	17	8	10
False Negative	21	26	39	9	32

Supplementary Table 13. Evaluation of different binary prediction models on KPXKPR-50 dataset.

	MC	HSP	HBP	COSMO-RS	MEP
True Positive	20	17	6	30	13
False Positive	6	8	14	8	9
True Negative	8	6	0	6	5
False Negative	16	19	30	6	23



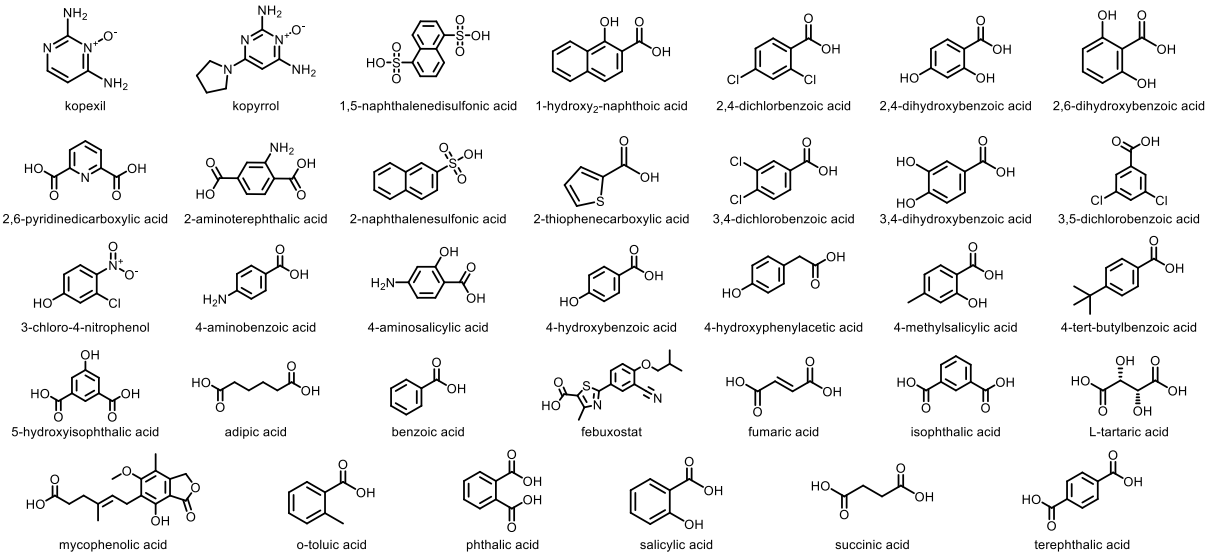
Supplementary Figure 8. Confusion matrixes of classical prediction models (MC, HSP, HBP, COSMO-RS and MEP) evaluated in binary classification task on the KPXKPR-64 test set and KPXKPR-50 test set.

6. Material, KPX and KPR MCC preparation and Solid-State Characterization Methods

KPX (98%), KPR (97%), 2,4-dihydroxybenzoic acid (98%), 2,6-dihydroxybenzoic acid (97%), 2,6-pyridinedicarboxylic acid (99%), 2-naphthalenesulfonic acid (98%), 3,4-dihydroxybenzoic acid (97%), 3,5-dichlorobenzoic acid (98%), 3-chloro-4-nitrophenol (98%), 4-aminobenzoic acid (99%), 4-aminosalicylic acid (98%), 4-methylsalicylic acid (99%), mycophenolic acid (98%), and o-toluic acid (98%) were supplied by Shanghai Macklin Biochemical co. ltd (China). 1,5-naphthalenedisulfonic acid (≥ 99%), 1-hydroxy-2-naphthoic acid (≥ 98%), 2,4-dichlorobenzoic acid (≥ 98%), 2-aminoterephthalic acid (≥ 98%), 2-thiophenecarboxylic acid (≥ 99%), 3,4-dichlorobenzoic acid (≥ 99%), 4-hydroxybenzoic acid (≥ 99%), 4-hydroxyphenylacetic acid (≥ 99%), 4-tert-butylbenzoic acid (≥ 99%), 5-hydroxyisophthalic acid (≥ 98%), febuxostat (≥ 99%), isophthalic acid (≥ 99%), phthalic acid (≥ 99%), and terephthalic acid (≥ 99%) were purchased from Aladdin. Reagent Co., ltd. (China). Adipic acid (99%), benzoic acid (≥ 99.5%), fumaric acid (≥ 99%), L-tartaric acid (≥ 99.5%), salicylic acid (≥ 99.0%), succinic acid (≥ 99.0%), were purchased from Sigma and Aladdin Reagent Company. Methanol (MeOH) and ethanol (EtOH) of analytical grade were supplied by Merck KGaA (Germany). All molecule structures are shown in Supplementary Figure 8.
The purchased KPX is monohydrate, which was dehydrated in a 170 °C oven before KPX MCC preparation. Solvent evaporation method was used to prepare KPX/KPR MCCs. Suitable quantities of KPX/KPR and the selected coformers were dissolved in either methanol or ethanol under vigorous magnetic stirring for 20 minutes. The solution was subsequently filtered through a 0.22 μm nylon membrane to remove impurities. The solvent is removed through evaporation, and the remaining solid is subjected to solid-state characterizations. To obtain single crystals suitable for single crystal X-ray diffraction (SCXRD), the solution was allowed to evaporate slowly at room temperature to facilitate the formation of well-ordered crystalline structures.

Powder X-Ray Diffraction (PXRD) analysis was conducted on a Rigaku SmartLab 9 kW diffractometer (Tokyo, Japan) with a Cu-Kα radiation source (λ = 1.54056 Å) set at 45 kV and 200 mA. The samples were evenly distributed in a quartz holder and scanned across a 2θ range

of 5° to 35° at a step size of 0.02° and a scan speed of 5° per minute. **Differential Scanning Calorimetry (DSC)** was employed to assess thermal properties using a TA DSC 250 instrument (New Castle, DE, USA). Temperature and enthalpy calibrations were performed with an indium reference. Samples (3–5 mg) were carefully weighed into aluminum Tzero pans fitted with perforated lids and heated from 35 °C to 300 °C at 5 °C per minute under a nitrogen flow (50 mL/min) to maintain an inert environment. Thermogravimetric Analysis (TGA) was conducted to examine weight loss; post-solvent evaporation, samples were loaded into platinum pans and heated from ambient temperature to 300 °C at 10 °C per minute in a TGA250 analyzer (TA Instruments, USA). Data collection and analysis were performed using TA TRIOS software (v5.1.1, TA Instruments). **Fourier-transform infrared (FTIR)** spectroscopy measurements were carried out on an ALPHA FTIR spectrophotometer (Bruker, Germany) with a diffuse reflectance attachment. Samples were mixed with dry potassium bromide (KBr) in a 1:100 (w/w) ratio and pressed into 13-mm pellets using a Graseby Specac hydraulic press (Orpington, UK) under 19.6 kN pressure. Spectra were recorded in the mid-IR range (600–3500 cm^{-1}) at 4 cm^{-1} resolution, averaging 64 scans per run. A pure KBr pellet was used for background correction before sample analysis. **SCXRD** data were collected on a Bruker D8 Quest diffractometer with a Mo Kα radiation source (λ = 0.71073 Å). Crystal indexing and data integration were performed using the Difference Vectors method in APEX3 (Bruker AXS, 2019). Absorption corrections were applied via SADABS, and space group determination was conducted using XPREP (integrated in APEX3). Structures were solved by direct methods with SHELXL-2014 and refined via full-matrix least-squares minimization on F^2, applying anisotropic displacement parameters for non-hydrogen atoms. H atoms on C atoms were added using the riding model, while H atoms on O and N atoms were added by picking Q peaks from the Fourier difference map.



Supplementary Figure 9. Molecule structures of the API and coformers used in this study.

7. Solid-State Characterization and Crystal Structure Analysis

The samples were initially analyzed using PXRD and DSC. New characteristic peaks corresponding to crystalline phases were identified in the PXRD patterns, while distinct thermal behaviors associated with these phases were observed in the DSC thermograms. Samples exhibiting potential new crystalline phases were further characterized using additional solid-state techniques, including FTIR, TGA, and SCXRD (when single crystals were available). Herein, PXRD, DSC, TGA, FTIR, and SCXRD results are presented for the confirmed MCCs. For MCCs without suitable single crystals, only PXRD, DSC, TGA, and FTIR data are reported. KPX monohydrate (CCDC No. 2430439), KPX–benzoic acid monohydrate (CCDC No. 2430575), and KPX–salicylic acid monohydrate (CCDC No. 2430576) were reported in our previous study [36]. In cases where cocrystallization failed, only PXRD and DSC results are presented.

36. Deng, Y., Wong, S.N., Chan, W.C.N., Zhang, X. & Chow, S.F. Dissolving microneedles incorporating kopexil multicomponent crystals for improved transdermal delivery. Drug Delivery and Translational Research (2026).

7.1 Crystallographic data and structure refinement parameters

Supplementary Table 14. Crystallographic data and structure refinement parameters of KPX MCCs.

Complex	KPX-1-hydroxy-2-naphthalic acid salt monohydrate	KPX-2,4-dichlorbenzoic acid salt	KPX-2,6-dihydroxybenzoic acid salt monohydrate	KPX-2-aminoterephthalic acid cocrystal dihydrate	KPX-2-naphthalenesulfonic acid salt dihydrate
CCDC no.	2430440	2441565	2430569	2441250	2430570
Empirical formula	$C_{15}H_{16}N_4O_5$	$C_{11}H_{10}Cl_2N_4O_3$	$C_{11}H_{14}N_4O_6$	$C_{16}H_{23}N_9O_8$	$C_{14}H_{18}N_4O_6S$
Formula weight	332.32	317.13	298.26	469.43	370.38
Temperature/K	150.15	296.15	296.15	150.15	296.15
Crystal system	triclinic	monoclinic	monoclinic	triclinic	monoclinic
Space group	$P\bar{1}$	$P2_1/c$	$P2_1/n$	$P\bar{1}$	$P2_1/c$
a/Å	6.7956(18)	13.806	9.659(7)	6.7604(9)	17.7763(13)
b/Å	6.9397(17)	4.957	4.887(3)	6.9066(9)	6.2077(4)
c/Å	17.053(4)	20.454	27.019(17)	12.1059(16)	15.1588(11)
α/°	94.379(7)	90	90	77.328(4)	90
β/°	91.866(8)	101.33	97.24(3)	80.553(4)	95.879(2)
γ/°	108.114(7)	90	90	69.503(3)	90
Volume/Å^3	760.8(3)	1372.5	1265.1(15)	514.20(12)	1664.0(2)
Z	2	4	4	1	4
ρ_{calc}g/cm^3	1.451	1.535	1.566	1.516	1.478
μ/mm^{-1}	0.111	0.485	0.129	0.123	0.235
F (000)	348	648	624	246	776
Radiation	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)
2θ range for data collection/°	6.202 to 50.058	5.51 to 50.14	4.692 to 50.208	6.396 to 50.146	4.608 to 50.056
Index ranges	$-8 \le h \le 8$, $-8 \le k \le 8$, $-19 \le l \le 20$	$-16 \le h \le 16$, $0 \le k \le 5$, $0 \le l \le 24$	$-11 \le h \le 11$, $-5 \le k \le 5$, $-29 \le l \le 31$	$-8 \le h \le 7$, $-8 \le k \le 7$, $-14 \le l \le 14$	$-20 \le h \le 21$, $-7 \le k \le 7$, $-18 \le l \le 17$
Reflections collected	10158	2395	9545	6688	24933
Independent reflections	2650 [R_{int} = 0.1033, R_{sigma} = 0.1320]	2395 [R_{int} = 0, R_{sigma} = 0.0824]	2212 [R_{int} = 0.0584, R_{sigma} = 0.0560]	1802 [R_{int} = 0.0488, R_{sigma} = 0.0578]	2948 [R_{int} = 0.0541, R_{sigma} = 0.0386]
Data/restraints/parameters	2650/1/221	2395/0/171	2212/0/192	1802/1/158	2948/0/232
Goodness-of-fit on F^2	1.026	1.044	1.031	1.096	1.008
Final R indexes [I ≥ 2σ (I)]	R_1 = 0.0866, wR_2 = 0.1205	R_1 = 0.0978, wR_2 = 0.2304	R_1 = 0.0494, wR_2 = 0.1043	R_1 = 0.0760, wR_2 = 0.1548	R_1 = 0.0492, wR_2 = 0.1298
Final R indexes [all data]	R_1 = 0.2053, wR_2 = 0.1610	R_1 = 0.1630, wR_2 = 0.2600	R_1 = 0.0914, wR_2 = 0.1258	R_1 = 0.1248, wR_2 = 0.1786	R_1 = 0.0798, wR_2 = 0.1509
Largest diff. peak/hole / e Å^{-3}	0.24/-0.26	0.40/-0.36	0.23/-0.19	0.25/-0.23	0.19/-0.36

Supplementary Table 14. Crystallographic data and structure refinement parameters of KPX MCCs (continued).

Complex	KPX-2-thiophenecarboxylic acid salt monohydrate	KPX-3,4-dihydroxybenzoic acid cocrystal dihydrate	KPX-3,4-dichlorobenzoic acid cocrystal monohydrate	KPX-4-aminobenzoic acid cocrystal dihydrate	KPX-4-methylsalicylic acid salt monohydrate
CCDC no.	2430571	2430572	2441566	2430580	2430573
Empirical formula	$C_9H_{12}N_4O_4S$	$C_{11}H_{16}N_4O_7$	$C_{11}H_{12}Cl_2N_4O_4$	$C_{11}H_{17}N_5O_5$	$C_{12}H_{15}N_4O_5$
Formula weight	272.29	316.28	335.15	299.29	295.28
Temperature/K	296.15	150.15	296.15	150.15	150.15
Crystal system	monoclinic	triclinic	triclinic	triclinic	triclinic
Space group	$P2_1/n$	$P\bar{1}$	$P\bar{1}$	$P\bar{1}$	$P\bar{1}$
a/Å	6.6664(5)	8.4207(11)	6.745(5)	8.3876(19)	6.7899(7)
b/Å	28.058(2)	9.7134(13)	6.956(5)	9.761(2)	6.8203(7)
c/Å	6.9802(6)	9.9865(13)	16.208(12)	10.056(2)	16.1299(17)
α/°	90	62.967(4)	88.59(3)	118.461(6)	88.746(3)
β/°	111.522(2)	73.274(4)	87.88(3)	91.911(7)	86.426(3)
γ/°	90	73.877(4)	71.10(2)	105.448(7)	69.569(3)
Volume/Å^3	1214.60(17)	686.40(16)	718.8(9)	685.0(3)	698.61(13)
Z	4	2	2	2	2
ρ_{calc}g/cm^3	1.489	1.53	1.548	1.451	1.404
μ/mm^{-1}	0.28	0.129	0.473	0.116	0.111
F (000)	568	332	344	316	310
Radiation	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)
2θ range for data collection/°	5.808 to 50.09	4.662 to 50.19	5.03 to 50.038	4.692 to 50.15	5.06 to 50.096
Index ranges	$-7 \le h \le 7$, $-30 \le k \le 33$, $-8 \le l \le 7$	$-9 \le h \le 10$, $-11 \le k \le 11$, $-11 \le l \le 11$	$-7 \le h \le 8$, $-7 \le k \le 8$, $-19 \le l \le 19$	$-9 \le h \le 9$, $-11 \le k \le 11$, $-11 \le l \le 11$	$-8 \le h \le 8$, $-8 \le k \le 7$, $-19 \le l \le 19$
Reflections collected	19302	8914	8181	9140	9214
Independent reflections	2132 [R_{int} = 0.0884, R_{sigma} = 0.0678]	2415 [R_{int} = 0.0705, R_{sigma} = 0.0867]	2450 [R_{int} = 0.0429, R_{sigma} = 0.0462]	2393 [R_{int} = 0.0522, R_{sigma} = 0.0622]	2423 [R_{int} = 0.0484, R_{sigma} = 0.0548]
Data/restraints/parameters	2132/0/164	2415/0/202	2450/0/190	2393/0/194	2423/0/203
Goodness-of-fit on F^2	1.033	1.012	1.115	1.063	1.043
Final R indexes [I≥2σ (I)]	R_1 = 0.0656, wR_2 = 0.1503	R_1 = 0.0654, wR_2 = 0.1325	R_1 = 0.0720, wR_2 = 0.1841	R_1 = 0.0741, wR_2 = 0.1509	R_1 = 0.0592, wR_2 = 0.1257
Final R indexes [all data]	R_1 = 0.1460, wR_2 = 0.1852	R_1 = 0.1447, wR_2 = 0.1696	R_1 = 0.1083, wR_2 = 0.2078	R_1 = 0.1284, wR_2 = 0.1817	R_1 = 0.1084, wR_2 = 0.1528
Largest diff. peak/hole / e Å^{-3}	0.27/-0.26	0.26/-0.25	0.59/-0.46	0.46/-0.43	0.21/-0.20

Supplementary Table 14. Crystallographic data and structure refinement parameters of KPX MCCs (continued).

Complex	KPX-adipic acid cocrystal dihydrate	KPX-fumaric acid salt dihydrate	KPX-L-tartaric acid salt	KPX-phthalic acid salt	KPX-succinic acid cocrystal dihydrate
CCDC no.	2430574	2430577	2430578	2430594	2430581
Empirical formula	$C_{14}H_{26}N_8O_8$	$C_{12}H_{20}N_8O_8$	$C_8H_{12}N_4O_7$	$C_{12}H_{12}N_4O_5$	$C_{12}H_{22}N_8O_8$
Formula weight	434.43	404.36	276.22	292.26	406.37
Temperature/K	150.15	150.15	296.15	150.15	150.15
Crystal system	triclinic	triclinic	monoclinic	triclinic	triclinic
Space group	$P\bar{1}$	$P\bar{1}$	$P2_1$	$P\bar{1}$	$P\bar{1}$
a/Å	6.874(7)	6.552	7.7393(7)	7.1358(6)	6.5055(11)
b/Å	6.899(6)	7.108	6.9218(6)	7.8448(6)	6.8706(12)
c/Å	11.845(11)	10.556	10.9606(9)	11.9962(9)	11.418(2)
α/°	84.92(2)	107.11	90	91.206(2)	87.552(5)
β/°	79.81(3)	94.92	103.423(2)	98.280(2)	74.056(5)
γ/°	66.71(3)	108.84	90	111.900(2)	68.447(5)
Volume/Å^3	507.7(8)	435.7	571.12(9)	614.53(8)	455.45(14)
Z	1	1	2	2	1
ρ_{calc}g/cm^3	1.421	1.541	1.606	1.579	1.482
μ/mm^{-1}	0.117	0.13	0.142	0.126	0.125
F (000)	230	212	288	304	214
Radiation	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)
2θ range for data collection/°	6.43 to 50.138	6.354 to 50.124	5.412 to 50.19	5.616 to 50.238	6.39 to 50
Index ranges	-8 ≤ h ≤ 8, -8 ≤ k ≤ 6, -14 ≤ l ≤ 12	-7 ≤ h ≤ 7, 8 ≤ k ≤ 8, 12 ≤ l ≤ 12	-9 ≤ h ≤ 8, -8 ≤ k ≤ 7, -12 ≤ l ≤ 13	-8 ≤ h ≤ 8, -9 ≤ k ≤ 9, -14 ≤ l ≤ 13	-7 ≤ h ≤ 7, 8 ≤ k ≤ 7, -13 ≤ l ≤ 13
Reflections collected	6791	1515	9203	8197	6065
Independent reflections	1759 [R_{int} = 0.0405, R_{sigma} = 0.0436]	1515 [R_{int} = 0, R_{sigma} = 0.0384]	1941 [R_{int} = 0.0577, R_{sigma} = 0.0578]	2164 [R_{int} = 0.0480, R_{sigma} = 0.0546]	1582 [R_{int} = 0.0615, R_{sigma} = 0.0672]
Data/restraints/parameters	1759/0/137	1515/0/127	1941/1/175	2164/0/190	1582/0/128
Goodness-of-fit on F^2	1.023	1.089	1.074	1.031	1.02
Final R indexes [I ≥ 2σ (I)]	R_1 = 0.0476, wR_2 = 0.1005	R_1 = 0.0624, wR_2 = 0.1726	R_1 = 0.0434, wR_2 = 0.0831	R_1 = 0.0556, wR_2 = 0.1165	R_1 = 0.0527, wR_2 = 0.0996
Final R indexes [all data]	R_1 = 0.0857, wR_2 = 0.1223	R_1 = 0.0902, wR_2 = 0.1990	R_1 = 0.0777, wR_2 = 0.0966	R_1 = 0.0983, wR_2 = 0.1402	R_1 = 0.1117, wR_2 = 0.1250
Largest diff. peak/hole / e Å^{-3}	0.17/-0.18	0.27/-0.30	0.21/-0.23	0.25/-0.29	0.18/-0.19

Supplementary Table 14. Crystallographic data and structure refinement parameters of KPX MCCs (continued).

Complex	KPX-terephthalic acid salt dihydrate
CCDC no.	2471748
Empirical formula	$C_{16}H_{22}N_8O_8$
Formula weight	454.41
Temperature/K	150.15
Crystal system	triclinic
Space group	$P\bar{1}$
a/Å	6.7876(5)
b/Å	6.7939(5)
c/Å	11.8715(8)
α/°	79.908(2)
β/°	78.985(2)
γ/°	68.788(2)
Volume/Å^3	497.57(6)
Z	1
ρ_{calc}g/cm^3	1.517
μ/mm^{-1}	0.124
F (000)	238
Radiation	MoKα (λ = 0.71073)
2*θ* range for data collection/°	6.476 to 49.97
Index ranges	$-8 \le h \le 8$, $-8 \le k \le 8$, $-14 \le l \le 13$
Reflections collected	6664
Independent reflections	1716 [R_{int} = 0.0402, R_{sigma} = 0.0428]
Data/restraints/parameters	1716/0/145
Goodness-of-fit on F^2	1.038
Final R indexes [I ≥ 2σ (*I*)]	R_1 = 0.0497, wR_2 = 0.1103
Final R indexes [all data]	R_1 = 0.0832, wR_2 = 0.1298
Largest diff. peak/hole / e Å^{-3}	0.25/-0.22

Supplementary Table 15. Crystallographic data and structure refinement parameters of KPR MCCs.

Complex	KPR-1,5-naphthalenedisulfonic acid salt dihydrate	KPR-1-hydroxy-2-naphthalic acid ionic cocrystal	KPR-2,6-dihydroxybenzoic acid ionic cocrystal	KPR-2,6-pyridinedicarboxylic acid ionic cocrystal	KPR-2-aminoterephthalic acid salt
CCDC no.	2430583	2430584	2450451	2430585	2441561
Empirical formula	$C_{26}H_{38}N_{10}O_{10}S_2$	$C_{30}H_{30}N_5O_7$	$C_{23}H_{32}N_{10}O_6$	$C_{30}H_{36}N_{12}O_{10}$	$C_{24}H_{33}N_{11}O_6$
Formula weight	714.78	572.59	544.58	724.71	571.61
Temperature/K	150.15	150.15	296.15	296.15	296.15
Crystal system	triclinic	monoclinic	triclinic	triclinic	monoclinic
Space group	$P\bar{1}$	$P2_1/c$	$P\bar{1}$	$P\bar{1}$	$P2_1/n$
a/Å	10.6136(3)	10.5358(7)	8.0333(9)	12.333(12)	9.154
b/Å	11.8715(3)	29.3949(14)	10.3035(12)	12.427(15)	7.199
c/Å	13.7765(4)	10.1922(5)	17.953(2)	12.695(15)	21.617
α/°	102.9960(10)	90	102.984(3)	78.87(4)	90
β/°	98.8790(10)	118.5900(10)	96.974(3)	66.33(3)	101.91
γ/°	104.2330(10)	90	90.819(3)	68.82(3)	90
Volume/Å^3	1598.55(8)	2771.6(3)	1436.0(3)	1659(3)	1393.8
Z	2	4	2	2	2
ρ_{calc}g/cm^3	1.485	1.372	1.26	1.451	1.362
μ/mm^{-1}	0.239	0.099	0.094	0.112	0.101
F (000)	752	1204	576	760	604
Radiation	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)
2θ range for data collection/°	5.434 to 50.036	4.758 to 50.09	5.114 to 50.006	5.108 to 49.414	4.548 to 52.154
Index ranges	$-12 \le h \le 12$, $-14 \le k \le 14$, $-13 \le l \le 16$	$-12 \le h \le 10$, $-34 \le k \le 33$, $-12 \le l \le 12$	$-9 \le h \le 8$, $-11 \le k \le 12$, $-21 \le l \le 21$	$-14 \le h \le 14$, $-14 \le k \le 14$, $-13 \le l \le 14$	$-11 \le h \le 10$, $-8 \le k \le 8$, $0 \le l \le 26$
Reflections collected	21060	31087	18739	20338	5106
Independent reflections	5585 [R_{int} = 0.0291, R_{sigma} = 0.0317]	4894 [R_{int} = 0.0755, R_{sigma} = 0.0719]	4980 [R_{int} = 0.0550, R_{sigma} = 0.0668]	5552 [R_{int} = 0.0738, R_{sigma} = 0.0870]	2756 [R_{int} = 0.0554, R_{sigma} = 0.0754]
Data/restraints/parameters	5585/0/435	4894/150/420	4980/481/452	5552/76/495	2756/8/192
Goodness-of-fit on F^2	1.038	0.952	1.045	1.025	1.092
Final R indexes [I ≥ 2σ (I)]	R_1 = 0.0547, wR_2 = 0.1265	R_1 = 0.0691, wR_2 = 0.1489	R_1 = 0.1102, wR_2 = 0.2865	R_1 = 0.0740, wR_2 = 0.1442	R_1 = 0.1031, wR_2 = 0.2950
Final R indexes [all data]	R_1 = 0.0740, wR_2 = 0.1405	R_1 = 0.1547, wR_2 = 0.1983	R_1 = 0.1828, wR_2 = 0.3455	R_1 = 0.1505, wR_2 = 0.1886	R_1 = 0.1503, wR_2 = 0.3224
Largest diff. peak/hole / e Å^{-3}	0.76/-0.47	0.22/-0.25	0.46/-0.72	0.33/-0.38	0.32/-0.30

Supplementary Table 15. Crystallographic data and structure refinement parameters of KPR MCCs (continued).

Complex	KPR-2-naphthalenesulfonic acid salt monohydrate	KPR-2-thiophenecarboxylic acid cocrystal	KPR-3,4-dichlorbenzoic acid cocrystal	KPR-3,5-dichlorobenzoic acid salt	KPR-4-aminosalicylic acid ionic cocrystal pentahydrate
CCDC no.	2464486	2430586	2441562	2430587	2430588
Empirical formula	$C_{18}H_{23}N_5O_5S$	$C_{26}H_{34}N_{10}O_6S_2$	$C_{15}H_{17}Cl_2N_5O_3$	$C_{15}H_{17}Cl_2N_5O_3$	$C_{92}H_{148}N_{44}O_{25}$
Formula weight	421.47	646.75	386.23	386.23	2270.54
Temperature/K	150.15	296.15	296.15	273.15	150.15
Crystal system	monoclinic	triclinic	monoclinic	triclinic	monoclinic
Space group	$C2/c$	$P\bar{1}$	$C2/c$	$P\bar{1}$	$C2/c$
a/Å	23.897(4)	10.766(2)	27.631(6)	4.4056(9)	36.9323(15)
b/Å	6.4834(10)	11.979(2)	7.0860(17)	16.984(3)	9.7981(4)
c/Å	26.219(5)	14.237(3)	21.709(5)	22.770(4)	33.9606(15)
α/°	90	74.002(4)	90	87.762(8)	90
β/°	108.763(5)	85.536(5)	126.485(5)	89.823(10)	115.6500(10)
γ/°	90	89.664(5)	90	84.270(9)	90
Volume/Å^3	3846.3(11)	1759.5(6)	3417.4(14)	1693.9(6)	11078.2(8)
Z	8	2	8	4	4
ρ_{calc}g/cm^3	1.456	1.221	1.501	1.515	1.361
μ/mm^{-1}	0.211	0.202	0.406	3.688	0.102
F (000)	1776	680	1600	800	4832
Radiation	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	CuKα (λ = 1.54178)	MoKα (λ = 0.71073)
2θ range for data collection/°	5.598 to 50.024	5.012 to 49.994	4.668 to 50.232	3.884 to 136.284	4.374 to 50.16
Index ranges	$-27 \le h \le 28$, $-7 \le k \le 7$, $-27 \le l \le 31$	$-12 \le h \le 12$, $-14 \le k \le 14$, $-16 \le l \le 16$	$-28 \le h \le 32$, $-8 \le k \le 8$, $-25 \le l \le 22$	$-5 \le h \le 5$, $-20 \le k \le 20$, $-26 \le l \le 27$	$-41 \le h \le 43$, $-11 \le k \le 11$, $-40 \le l \le 35$
Reflections collected	28268	23381	26253	16678	71185
Independent reflections	3381 [R_{int} = 0.0487, R_{sigma} = 0.0309]	6137 [R_{int} = 0.1211, R_{sigma} = 0.1402]	3028 [R_{int} = 0.0542, R_{sigma} = 0.0358]	5917 [R_{int} = 0.1212, R_{sigma} = 0.1427]	9784 [R_{int} = 0.1082, R_{sigma} = 0.0854]
Data/restraints/parameters	3381/0/263	6137/439/490	3028/0/227	5917/43/439	9784/9/719
Goodness-of-fit on F^2	1.085	1.083	1.012	1.037	1.031
Final R indexes [I ≥ 2σ (I)]	R_1 = 0.0429, wR_2 = 0.0882	R_1 = 0.1097, wR_2 = 0.2111	R_1 = 0.0478, wR_2 = 0.1240	R_1 = 0.1239, wR_2 = 0.2997	R_1 = 0.0992, wR_2 = 0.2612
Final R indexes [all data]	R_1 = 0.0606, wR_2 = 0.0972	R_1 = 0.2584, wR_2 = 0.2511	R_1 = 0.0827, wR_2 = 0.1468	R_1 = 0.1932, wR_2 = 0.3447	R_1 = 0.1916, wR_2 = 0.3277
Largest diff. peak/hole / e Å^{-3}	0.18/-0.30	0.92/-0.27	0.28/-0.35	0.81/-0.46	0.80/-0.58

Supplementary Table 15. Crystallographic data and structure refinement parameters of KPR MCCs (continued).

Complex	KPR-4-hydroxyphenylacetic acid salt	KPR-4-tert-butylbenzoic acid salt	KPR-adipic acid salt	KPR-febuxostat salt	KPR-fumaric acid salt
CCDC no.	2430589	2441564	2430590	2430591	2464485
Empirical formula	$C_{16}H_{21}N_5O_4$	$C_{19}H_{27}N_5O_3$	$C_{22}H_{36}N_{10}O_6$	$C_{24}H_{29}N_7O_4S$	$C_{20}H_{30}N_{10}O_6$
Formula weight	347.38	373.45	536.61	511.6	506.54
Temperature/K	150.15	296.15	150.15	150.15	150.15
Crystal system	triclinic	triclinic	triclinic	orthorhombic	triclinic
Space group	$P\bar{1}$	$P\bar{1}$	$P\bar{1}$	$Pna2_1$	$P\bar{1}$
a/Å	8.4519(17)	10.764	11.007(3)	23.6662(14)	10.567(10)
b/Å	8.9030(18)	14.105	11.641(3)	5.2211(2)	10.729(10)
c/Å	11.547(2)	15.947	12.801(4)	41.1856(19)	12.388(12)
α/°	85.107(6)	69.88	81.258(8)	90	68.07(3)
β/°	71.606(6)	85.02	68.333(8)	90	87.83(3)
γ/°	86.944(6)	79.85	62.226(7)	90	65.26(3)
Volume/Å^3	821.2(3)	2237	1348.3(6)	5089.0(4)	1171.5(19)
Z	2	4	2	8	2
ρ_{calc}g/cm^3	1.405	1.109	1.322	1.335	1.436
μ/mm^{-1}	0.104	0.077	0.099	1.505	0.109
F (000)	368	800	572	2160	536
Radiation	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	CuKα (λ = 1.54178)	MoKα (λ = 0.71073)
2θ range for data collection/°	4.594 to 50.148	4.648 to 50.068	4.45 to 50.574	4.29 to 133.684	5.96 to 51.35
Index ranges	$-10 \le h \le 10$, $-10 \le k \le 10$, $-13 \le l \le 13$	$-12 \le h \le 12$, $-15 \le k \le 16$, $0 \le l \le 18$	$-13 \le h \le 13$, $-13 \le k \le 13$, $-15 \le l \le 15$	$-27 \le h \le 28$, $-4 \le k \le 6$, $-48 \le l \le 48$	$-12 \le h \le 12$, $-12 \le k \le 12$, $-15 \le l \le 13$
Reflections collected	10408	7647	20116	34465	15210
Independent reflections	2843 [R_{int} = 0.0668, R_{sigma} = 0.0638]	7647 [R_{int} = 0.0000, R_{sigma} = 0.1538]	4709 [R_{int} = 0.0663, R_{sigma} = 0.0755]	8700 [R_{int} = 0.0952, R_{sigma} = 0.0769]	4208 [R_{int} = 0.0596, R_{sigma} = 0.0685]
Data/restraints/parameters	2843/0/228	7647/336/595	4709/56/367	8700/89/669	4208/2/329
Goodness-of-fit on F^2	1.055	0.972	0.964	1.021	1.045
Final R indexes [I ≥ 2σ (I)]	R_1 = 0.0847, wR_2 = 0.2131	R_1 = 0.1079, wR_2 = 0.2621	R_1 = 0.0879, wR_2 = 0.2198	R_1 = 0.0575, wR_2 = 0.1432	R_1 = 0.0719, wR_2 = 0.1695
Final R indexes [all data]	R_1 = 0.1217, wR_2 = 0.2492	R_1 = 0.2456, wR_2 = 0.3489	R_1 = 0.1714, wR_2 = 0.2887	R_1 = 0.0862, wR_2 = 0.1618	R_1 = 0.1336, wR_2 = 0.2048
Largest diff. peak/hole / e Å^{-3}	0.34/-0.39	0.53/-0.30	0.30/-0.33	0.19/-0.28	0.33/-0.25

Supplementary Table 15. Crystallographic data and structure refinement parameters of KPR MCCs (continued).

Complex	KPR-isophthalic acid ionic cocrystal trihydrate	KPR-L-tartaric acid ionic cocrystal	KPR-phthalic acid salt	KPR-succinic acid salt	KPR-terephthalic acid salt
CCDC no.	2430592	2441253	2441563	2441252	2441251
Empirical formula	$C_{24}H_{38}N_{10}O_9$	$C_{40}H_{64}N_{20}O_{16}$	$C_{16}H_{19}N_5O_5$	$C_{20}H_{34}N_{10}O_6$	$C_{24}H_{34}N_{10}O_6$
Formula weight	610.64	1081.11	361.36	510.57	558.61
Temperature/K	150.15	296.15	296.15	150.15	150.15
Crystal system	triclinic	monoclinic	monoclinic	triclinic	monoclinic
Space group	$P\bar{1}$	$P2_1$	$P2_1/n$	$P\bar{1}$	$P2_1/n$
a/Å	9.695	10.8846(7)	5.1834(13)	10.221(2)	11.7724(9)
b/Å	12.128	18.8661(13)	18.994(5)	10.885(2)	18.3425(13)
c/Å	13.013	12.1753(9)	16.916(4)	12.519(2)	12.7693(9)
α/°	77.11	90	90	69.260(6)	90
β/°	82.04	104.374(2)	96.227(7)	88.831(6)	101.029(2)
γ/°	75.64	90	90	66.997(5)	90
Volume/Å^3	1439.4	2421.9(3)	1655.6(7)	1187.7(4)	2706.4(3)
Z	2	2	4	2	4
ρ_{calc}g/cm^3	1.409	1.482	1.45	1.428	1.371
μ/mm^{-1}	0.11	0.117	0.11	0.108	0.102
F (000)	648	1144	760	544	1184
Radiation	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)	MoKα (λ = 0.71073)
2θ range for data collection/°	4.354 to 50.186	4.498 to 50.042	4.926 to 50.004	4.37 to 50.34	4.852 to 50.068
Index ranges	$11 \le h \le 11$, $14 \le k \le 14$, $15 \le l \le 15$	$-11 \le h \le 12$, $-22 \le k \le 22$, $-14 \le l \le 12$	$-6 \le h \le 5$, $-21 \le k \le 22$, $-18 \le l \le 20$	$-12 \le h \le 12$, $-12 \le k \le 12$, $-14 \le l \le 14$	$-14 \le h \le 12$, $-21 \le k \le 21$, $-15 \le l \le 14$
Reflections collected	5078	20476	14124	26226	23053
Independent reflections	5078 [R_{int} = 0, R_{sigma} = 0.0842]	8366 [R_{int} = 0.0584, R_{sigma} = 0.1022]	2885 [R_{int} = 0.1291, R_{sigma} = 0.1414]	4130 [R_{int} = 0.0591, R_{sigma} = 0.0525]	4767 [R_{int} = 0.0793, R_{sigma} = 0.0774]
Data/restraints/parameters	5078/0/392	8366/177/765	2885/0/239	4130/94/365	4767/35/374
Goodness-of-fit on F^2	0.972	1.016	1.044	1.101	1.015
Final R indexes [I ≥ 2σ (I)]	R_1 = 0.0735, wR_2 = 0.1819	R_1 = 0.0662, wR_2 = 0.1347	R_1 = 0.0812, wR_2 = 0.1354	R_1 = 0.0908, wR_2 = 0.2375	R_1 = 0.0751, wR_2 = 0.1644
Final R indexes [all data]	R_1 = 0.1524, wR_2 = 0.2253	R_1 = 0.1555, wR_2 = 0.1733	R_1 = 0.2093, wR_2 = 0.1834	R_1 = 0.1377, wR_2 = 0.2685	R_1 = 0.1672, wR_2 = 0.2104
Largest diff. peak/hole / e Å^{-3}	0.25/-0.43	0.29/-0.24	0.24/-0.37	0.50/-0.42	0.26/-0.25

Supplementary Table 15. Crystallographic data and structure refinement parameters of KPR MCCs (continued).

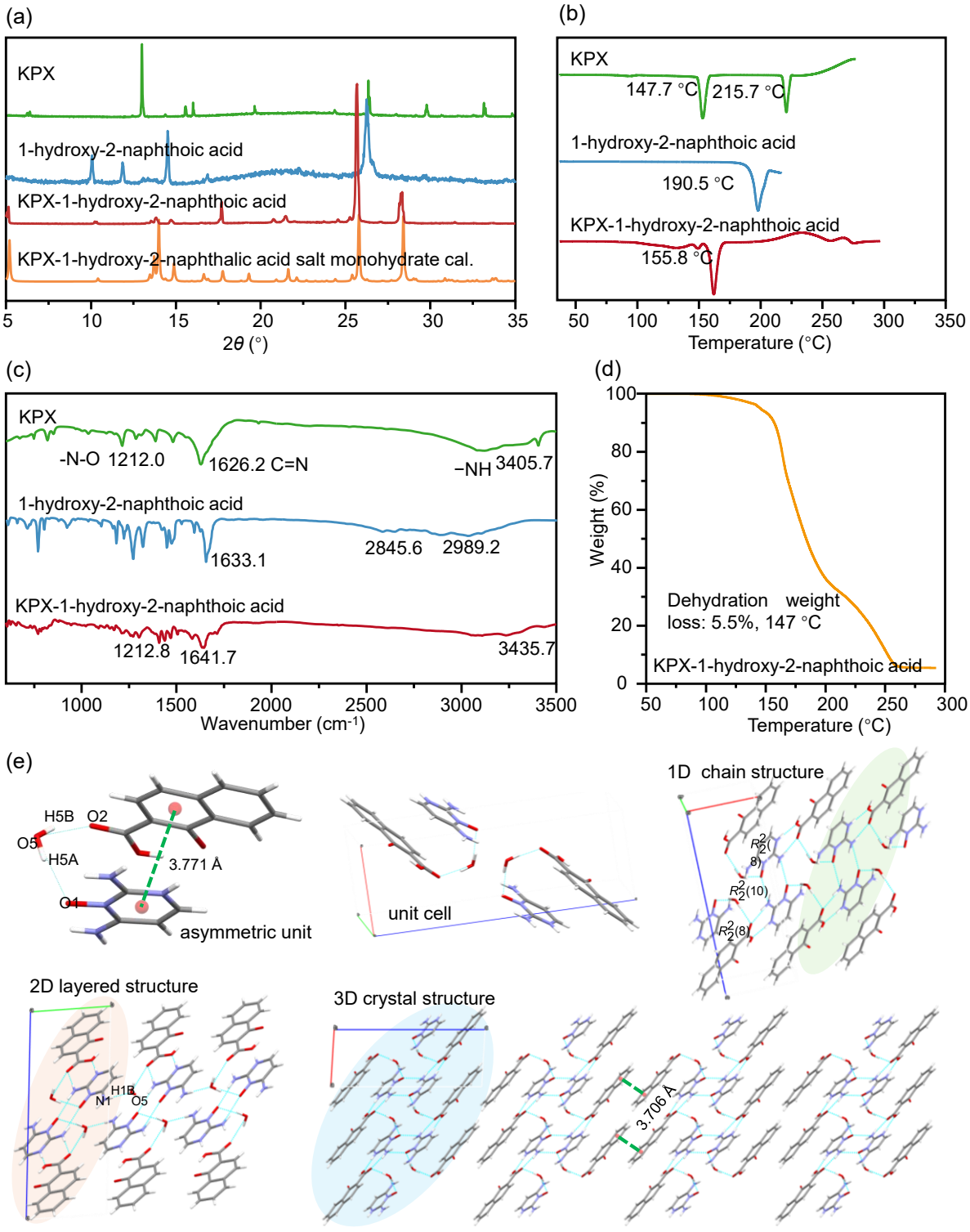
Complex	KPR
CCDC no.	2430582
Empirical formula	$C_8H_{13}N_5O$
Formula weight	195.23
Temperature/K	150.15
Crystal system	orthorhombic
Space group	$Pna2_1$
a/Å	15.9384(7)
b/Å	9.0488(4)
c/Å	13.2399(6)
α/°	90
β/°	90
γ/°	90
Volume/Å^3	1909.50(15)
Z	8
ρ_{calc}g/cm^3	1.358
μ/mm^{-1}	0.097
F (000)	832
Radiation	MoKα (λ = 0.71073)
2θ range for data collection/°	5.112 to 50.122
Index ranges	$-16 \le h \le 18$, $-10 \le k \le 10$, $-15 \le l \le 15$
Reflections collected	20729
Independent reflections	3362 [R_{int} = 0.0577, R_{sigma} = 0.0454]
Data/restraints/parameters	3362/89/292
Goodness-of-fit on F^2	1.06
Final R indexes [I ≥ 2σ (I)]	R_1 = 0.0532, wR_2 = 0.1178
Final R indexes [all data]	R_1 = 0.0919, wR_2 = 0.1412
Largest diff. peak/hole / e Å^{-3}	0.23/-0.26

7.2 KPX-1-hydroxy-2-naphthoic acid salt monohydrate

KPX-1-hydroxy-2-naphthalic acid salt monohydrate crystallizes in the triclinic system with space group $P\bar{1}$ (Z = 2). The asymmetric unit consists of one protonated KPX cation, one deprotonated 1-hydroxy-2-naphthalic acid anion, and one water molecule. Proton transfer occurs from the hydroxy group of 1-hydroxy-2-naphthalic acid to the nitrogen atom of the pyrimidine group in KPX. Within the asymmetric unit, the KPX cation and deprotonated 1-hydroxy-2-naphthalic acid anion are connected to the water molecule via O5–H5A···O1 and O5–H5B···O2 hydrogen bonds, respectively. Additionally, π···π offset stacking is observed between the pyrimidine ring of KPX and the naphthalene ring of the anion (Cg–Cg distance: 3.771 Å, dihedral angle: 8.61°). The unit cell is formed by the asymmetric unit and its centrosymmetrically related counterpart. Two unit cells interact through two $R_2^2(8)$ motifs formed by N4–H4···O3 and N3–H3B···O2 hydrogen bonds and an $R_2^2(10)$ motif formed by N3–H3C···O1 hydrogen bonds, extending along the *a*-axis to generate a 1D chain structure. These 1D chains further pack along the *b*-axis, forming a 2D layered structure stabilized by hydrogen bonds (N1–H1B···O5). Finally, the 2D layers with identical orientation assemble through parallel π···π stacking (Cg–Cg distance: 3.706 Å), completing the 3D crystal structure.

Supplementary Table 16. Hydrogen-bond geometry (Å) of KPX-1-hydroxy-2-naphthalic acid salt monohydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N3–H3C···O1	0.880	2.005	2.755	142.29
N3–H3B···O2	0.880	1.894	2.766	170.68
N4–H4···O3	0.881	1.711	2.589	174.52
N1–H1B···O5	0.880	1.989	2.867	175.76
O5–H5A···O1	0.870	1.883	2.744	170.15
O5–H5B···O2	0.871	2.027	2.869	162.09



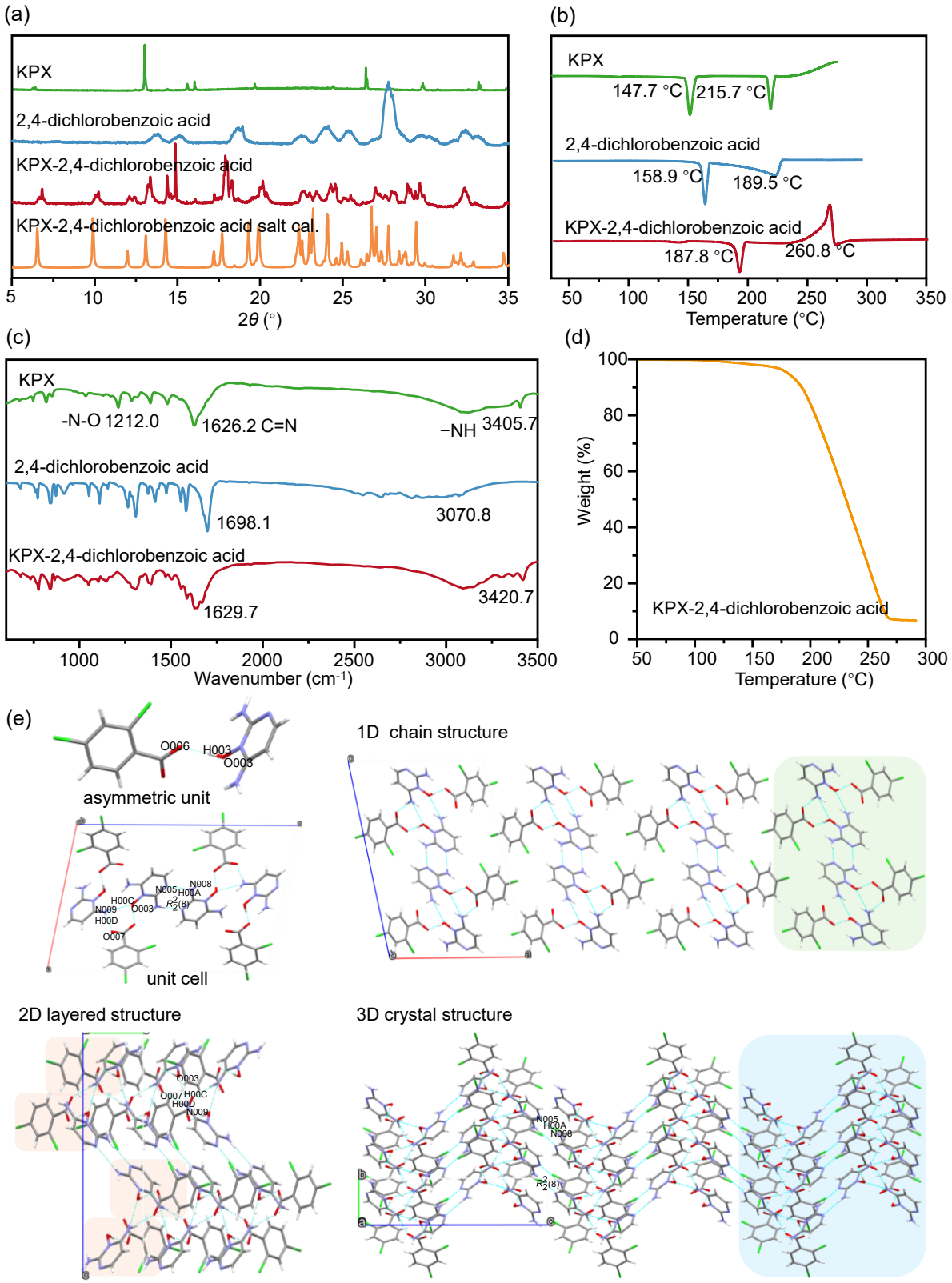
Supplementary Figure 10. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-1-hydroxy-2-naphthalic acid salt monohydrate.

7.3 KPX-2,4-dichlorobenzoic acid salt

KPX-2,4-dichlorobenzoic acid salt crystallizes in the monoclinic system with space group $P2_1/c$ ($Z = 4$). The asymmetric unit consists of one protonated KPX cation and one deprotonated 2,4-dichlorobenzoic acid anion, linked via a O003–H003···O006 hydrogen bond. Proton transfer occurs from the carboxylic acid group of 2,4-dichlorobenzoic acid to the oxygen atom of the pyrimidine N-oxide group in KPX. The asymmetric unit and its centrosymmetrically related counterpart form half of the unit cell through N009–H00C···O003 and N009–H00D···O007 hydrogen bonds. These halves are further connected via an $R_2^2(8)$ motif formed by N008–H00A···N005 hydrogen bonds to complete the unit cell. The unit cells interact through van der Waals forces, extending along the *a*-axis to generate a 1D chain structure. These 1D chains, aligned in the same orientation, pack along the *b*-axis, forming a 2D layered structure stabilized by N009–H00C···O003 and N009–H00D···O007 hydrogen bonds. Finally, the 2D layers assemble into a 3D crystal structure along the *c*-axis through additional $R_2^2(8)$ motifs formed by N008–H00A···N005.

Supplementary Table 17. Hydrogen-bond geometry (Å) of KPX-2,4-dichlorbenzoic acid salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
O003–H003···O006	0.821	1.643	2.462	175.75
N009–H00C···O003	0.861	2.125	2.749	129.02
N009–H00D···O007	0.860	1.965	2.809	166.52
N008–H00A···N005	0.860	2.080	2.939	175.92



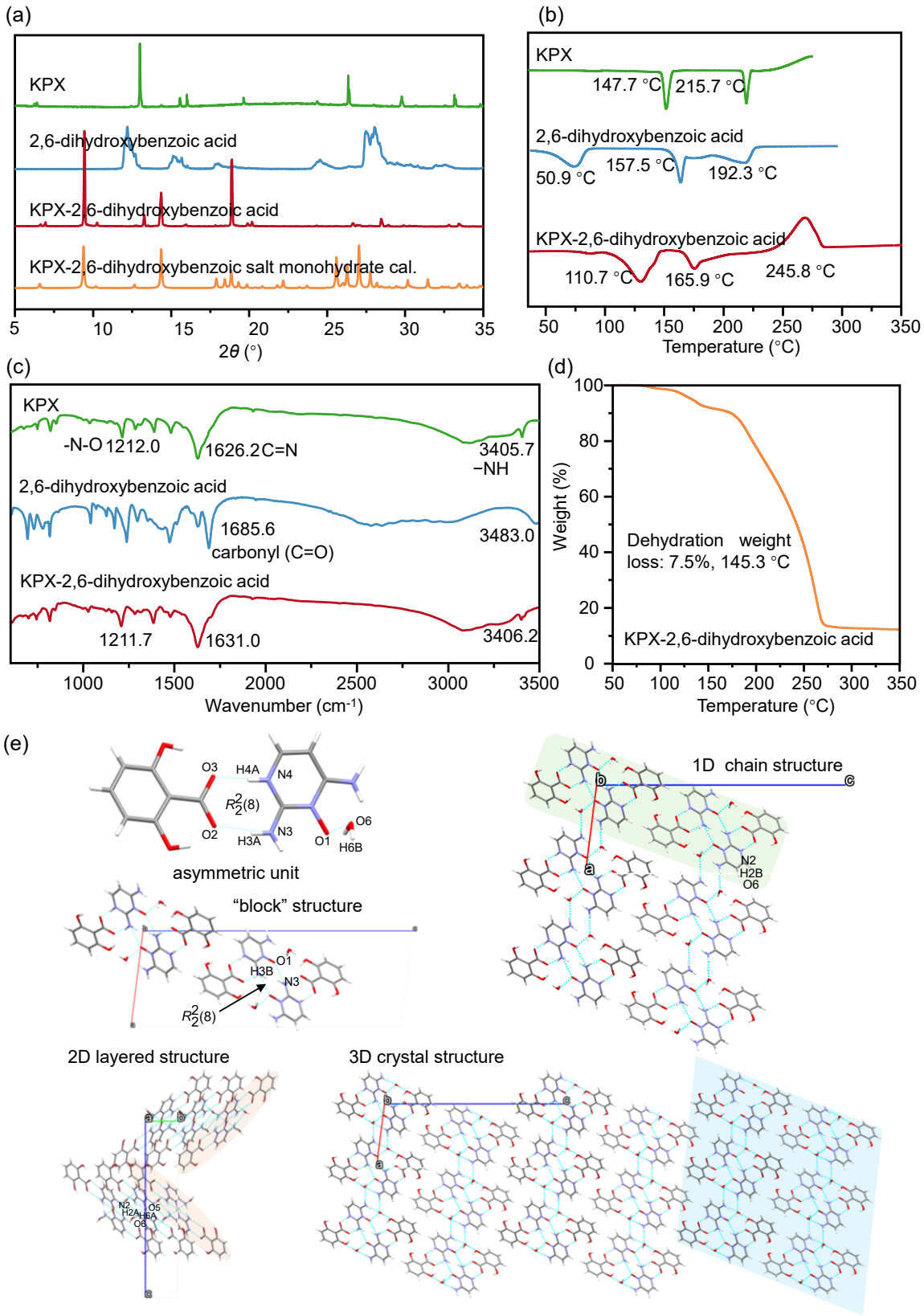
Supplementary Figure 11. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-2,4-dichlorobenzoic acid salt.

7.4 KPX-2,6-dihydroxybenzoic acid salt monohydrate

KPX-2,6-dihydroxybenzoic acid salt monohydrate crystallizes in the monoclinic system with space group $P2_1/n$ ($Z = 4$). The asymmetric unit consists of one protonated KPX cation, one deprotonated 2,6-DHBA anion, and one water molecule. Proton transfer occurs from the carboxylic acid group of 2,6-DHBA to the nitrogen atom of the pyrimidine group in KPX. The KPX cation and 2,6-DHBA anion are linked by an $R_2^2(8)$ motif formed by N4–H4A···O3 and N3–H3A···O2 hydrogen bonds. Additionally, the KPX cation is connected to the water molecule via a O6–H6B···O1 hydrogen bond. The asymmetric unit connects to another asymmetric unit through a second $R_2^2(8)$ motif formed by N3–H3B···O1, forming a "hexamer". These hexamers further assemble into a "block" structure via van der Waals forces after a 4_2 screw-axis operation. Neighboring blocks then interconnect through N2–H2B···O6 hydrogen bonds, extending along the *a*-axis to generate a 1D structure. The 1D structures pack along the *b*-axis, forming a 2D layered network stabilized by O6–H6A···O5 and N2–H2A···O6 hydrogen bonds. Finally, identically oriented 2D layers stack along the *c*-axis through van der Waals interactions, completing the 3D crystal structure.

Supplementary Table 18. Hydrogen-bond geometry (Å) of KPX-2,6-dihydroxybenzoic acid salt monohydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N3–H3B···O1	0.859	2.243	2.927	136.52
N3–H3A···O2	0.859	1.997	2.840	166.47
N4–H4A···O3	0.860	1.719	2.578	175.21
N2–H2A···O6	0.860	2.173	2.950	150.07
O6–H6B···O1	0.850	2.065	2.845	152.17
O6–H6A···O5	0.851	2.010	2.835	163.31
N2–H2B···O6	0.860	1.963	2.800	164.19



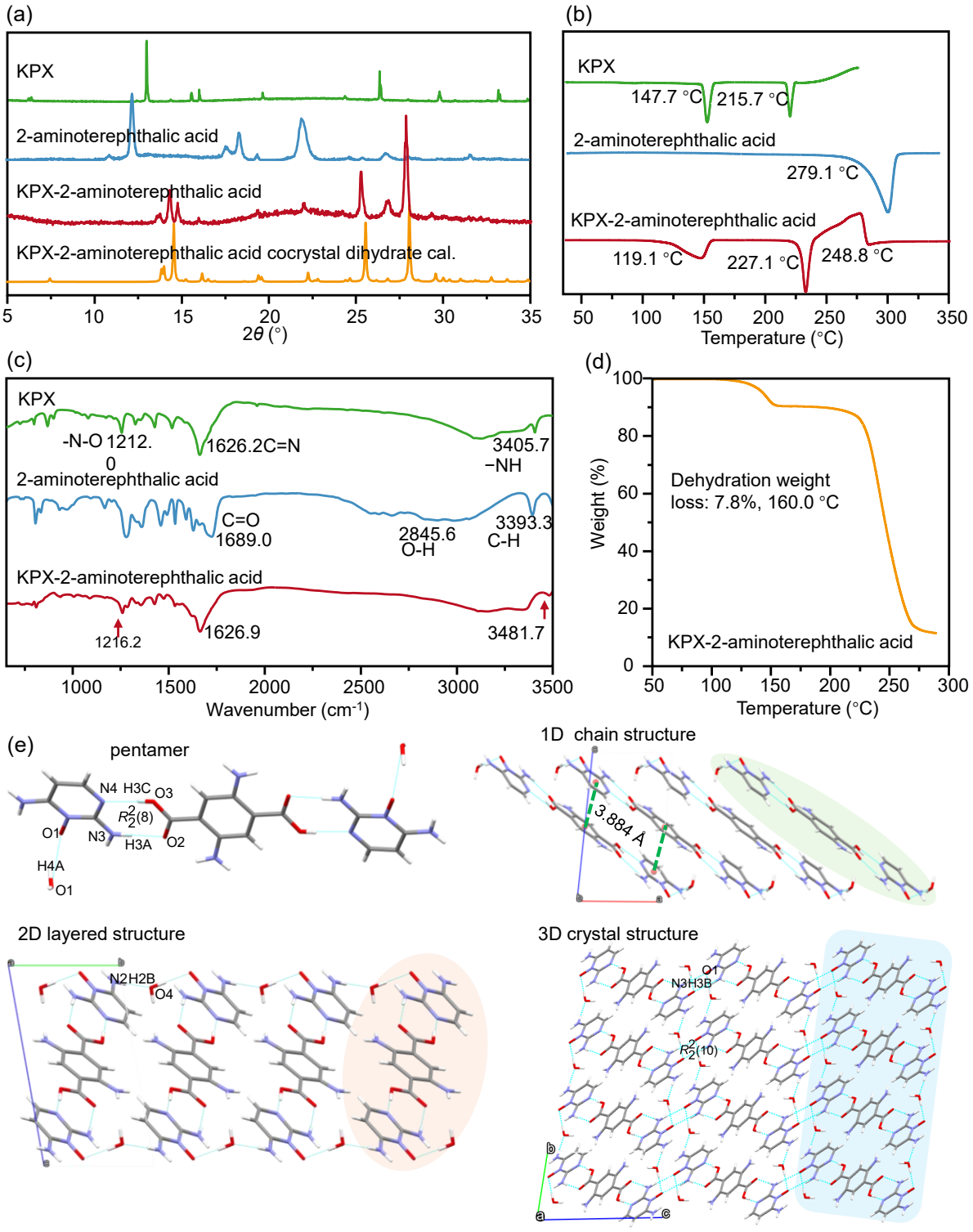
Supplementary Figure 12. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-2,6-dihydroxybenzoic acid salt monohydrate.

7.5 KPX-2-aminoterephthalic acid cocrystal dihydrate

KPX-2-aminoterephthalic acid cocrystal dihydrate crystallizes in the triclinic system with space group $P\bar{1}$ ($Z = 1$). The amino group of 2-aminoterephthalic acid exhibits disorder. The asymmetric unit contains one KPX molecule, one half-deprotonated 2-aminoterephthalic acid molecule, and one water molecule. The 2-aminoterephthalic acid is linked to KPX through an $R_2^2(8)$ motif formed by O3–H3C···N4 and N3–H3A···O2 hydrogen bonds. KPX is further connected to the water molecule via a O4–H4A···O1 hydrogen bond. Additionally, the second carboxylic acid group of 2-aminoterephthalic acid interacts with another KPX molecule in an identical manner, which is similarly linked to another water molecule, resulting in the formation of a pentamer. These pentamers, aligned in the same orientation, pack along the *a*-axis through π···π offset stacking (Cg–Cg distance: 3.884 Å, dihedral angle: 7.85°), generating a 1D structure. The 1D structures then pack along the *b*-axis via N2–H2B···O4 hydrogen bonds, forming a 2D layered network. Finally, the 2D layers assemble into a 3D crystal structure through $R_2^2(10)$ motifs formed by N3–H3B···O1 hydrogen bonds.

Supplementary Table 19. Hydrogen-bond geometry (Å) of KPX-2-aminoterephthalic acid cocrystal dihydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N3–H3A···O2	0.860	1.948	2.813	167.55
N3–H3B···O1	0.881	2.061	2.807	141.94
O3–H3C···N4	0.840	1.811	2.620	161.25
O4–H4A···O1	0.870	1.866	2.695	158.59
N2–H2B···O4	0.880	2.028	2.907	177.71



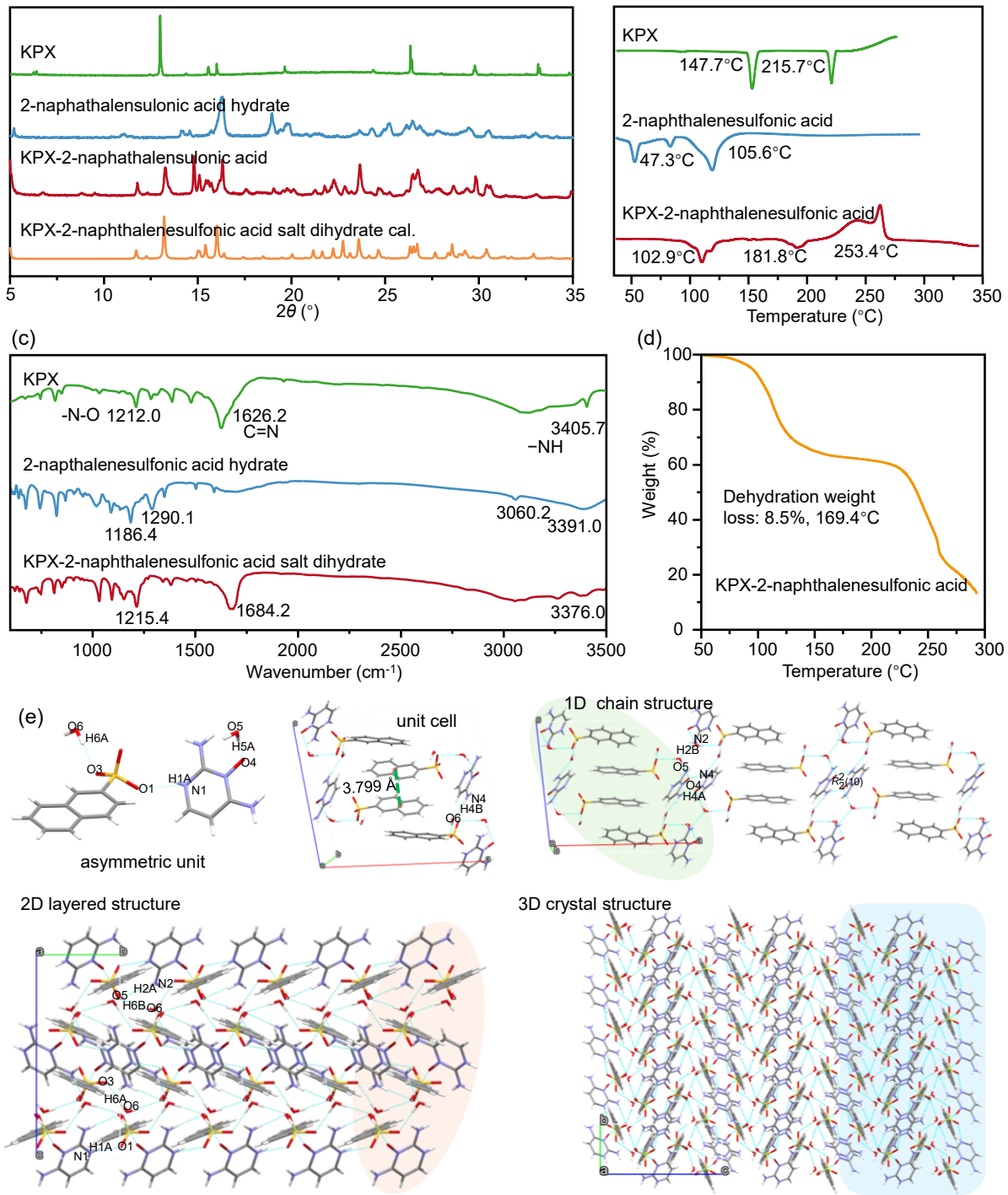
Supplementary Figure 13. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-2-aminoterephthalic acid cocrystal dihydrate.

7.6 KPX-2-naphthalenesulfonic acid salt dihydrate

KPX-2-naphthalenesulfonic acid salt dihydrate crystallizes in the monoclinic system with space group $P2_1/c$ (Z = 4). The asymmetric unit contains one protonated KPX cation, one deprotonated 2-naphthalenesulfonic acid anion, and two water molecules. Proton transfer occurs from the sulfonic acid group of 2-naphthalenesulfonic acid to the nitrogen atom of KPX's pyrimidine ring. In the crystal structure, the 2-naphthalenesulfonic acid anion connects to both the KPX cation through an N1-H1A···O1 hydrogen bond and to one water molecule via an O6-H6A···O3 hydrogen bond. The KPX cation additionally links to the second water molecule through an O5-H5A···O4 hydrogen bond. The asymmetric units combine through a 4_2 screw axis operation involving an N4-H4B···O6 hydrogen bond to form half of the unit cell. These half-unit cells then complete the full unit cell through centrosymmetric operations through π-π stacking interactions (Cg–Cg distance: 3.799 Å, parallel). The unit cells assemble into 1D chains extending along the *a*-axis through N2-H2B···O5 hydrogen bonds and $R_2^2(10)$ motifs formed by N4-H4A···O4 hydrogen bonds. These chains then pack along the *b*-axis to form two-dimensional layers, stabilized by multiple hydrogen bonds including N2-H2A···O5, O6-H6A···O3, O6-H6B···O5, and N1-H1A···O1. Finally, these 2D layers stack into the complete 3D crystal structure through van der Waals interactions between identically oriented layers.

Supplementary Table 20. Hydrogen-bond geometry (Å) of KPX-2-naphthalenesulfonic acid salt dihydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N1–H1A···O1	0.860	1.942	2.715	148.77
O6–H6A···O3	0.850	1.945	2.794	176.71
O6–H6B···O5	0.850	2.035	2.884	176.93
O5–H5A···O4	0.850	1.946	2.793	173.93
O5–H5B···O2	0.850	2.021	2.855	135.61
N4–H4A···O4	0.859	2.009	2.777	146.38
N4–H4B···O6	0.859	1.987	2.835	168.60
N2–H2A···O5	0.860	1.963	2.815	170.53
N2–H2B···O5	0.860	2.265	2.941	135.61



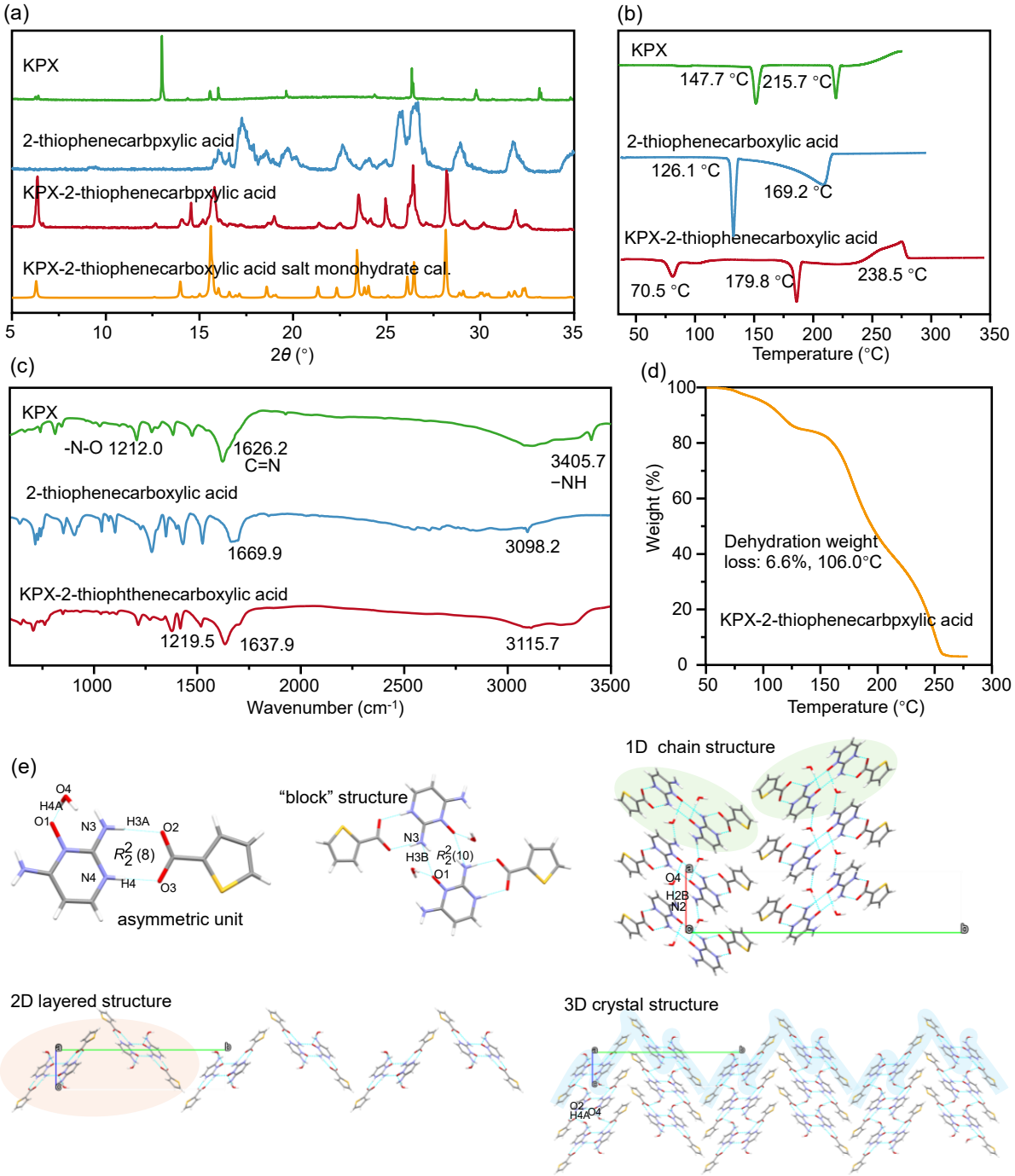
Supplementary Figure 14. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-2-naphthalenesulfonic acid salt dihydrate.

7.7 KPX-2-thiophenecarboxylic acid salt monohydrate

KPX-2-thiophenecarboxylic acid salt monohydrate crystallizes in the monoclinic system with space group $P2_1/c$ ($Z = 4$). The asymmetric unit contains one protonated KPX cation, one deprotonated 2-thiophenecarboxylic acid anion, and one water molecule. Proton transfer occurs from the carboxylic acid group of 2-thiophenecarboxylic acid to the nitrogen atom of the pyrimidine ring in KPX. In the crystal structure, the 2-thiophenecarboxylic acid anion forms an $R_2^2(8)$ hydrogen bond motif with the KPX cation through N4-H4···O3 and N3-H3A···O2 interactions. Additionally, the KPX cation connects to the water molecule via an O4-H4A···O1 hydrogen bond. The asymmetric units combine through a centrosymmetric operation to form a "block" structure, stabilized by an $R_2^2(10)$ motif consisting of N3-H3B···O1 hydrogen bonds. These block structures then assemble into extended chains along the *a*-axis through N2-H2B···O4 hydrogen bonds. Adjacent chains interact through van der Waals forces to complete the 1D structure. The 1D structures subsequently pack along the *b*-axis via van der Waals interactions, forming 2D layers. Finally, these layers stack along the *c*-axis through O4-H4A···O2 hydrogen bonds, completing the three-dimensional crystal structure.

Supplementary Table 21. Hydrogen-bond geometry (Å) of KPX-2-thiophenecarboxylic acid salt monohydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N3–H3B···O1	0.859	2.016	2.754	143.36
N3–H3A···O2	0.860	1.866	2.717	169.99
N4–H4···O3	0.860	1.762	2.623	178.89
N2–H2B···O4	0.859	2.015	2.870	172.98
O4–H4A···O1	0.851	2.266	2.909	174.18
O4–H4A···O2	0.850	1.872	2.719	132.42



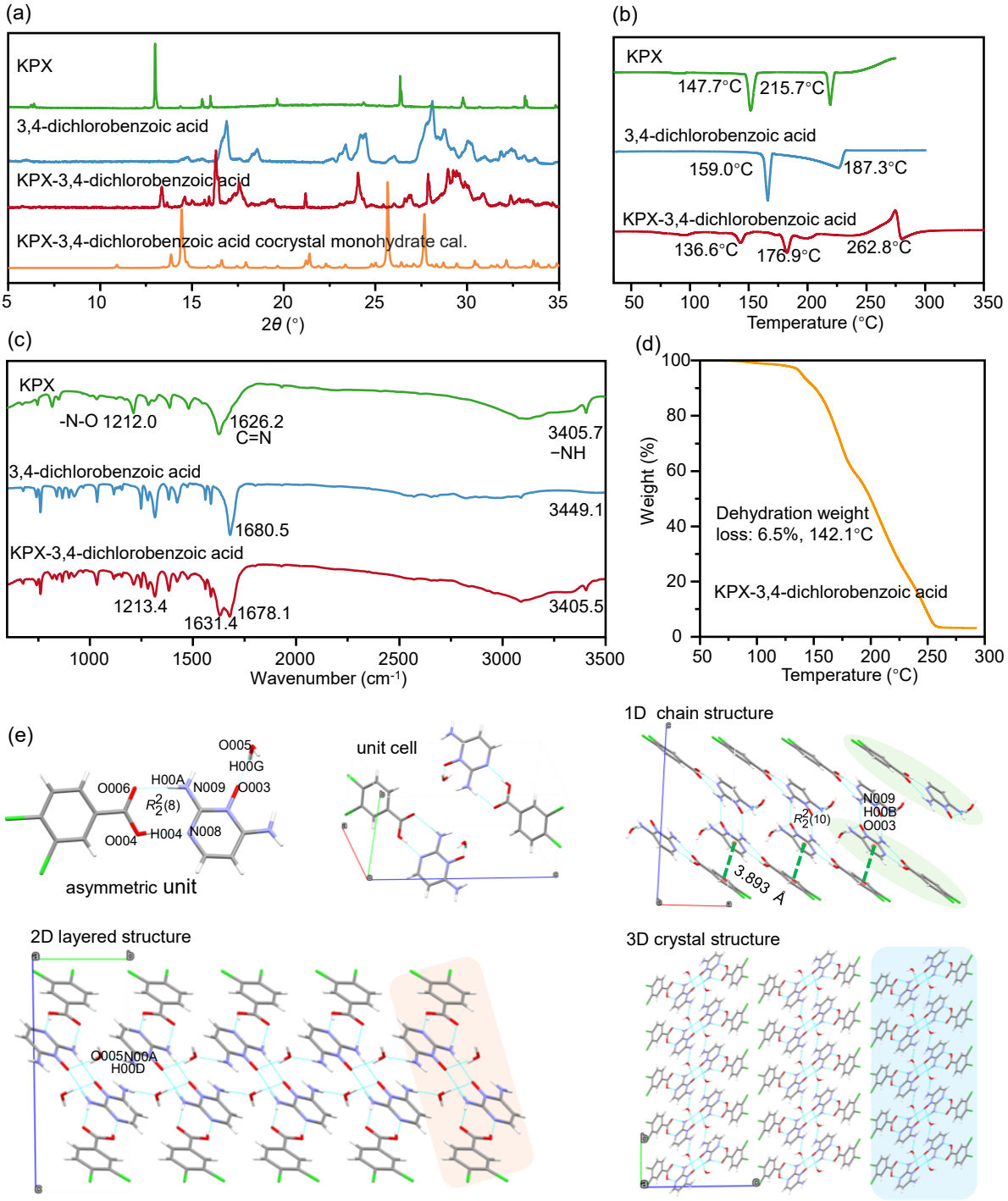
Supplementary Figure 15. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-2-thiophenecarboxylic acid salt monohydrate.

7.8 KPX-3,4-dichlorobenzoic acid cocrystal monohydrate

The KPX-3,4-dichlorobenzoic acid cocrystal monohydrate crystallizes in the triclinic system with space group $P\bar{1}$ (Z = 2). The asymmetric unit consists of one KPX, one 3,4-dichlorobenzoic acid, and one water molecule. In the crystal structure, the 3,4-dichlorobenzoic acid forms an $R_2^2(8)$ hydrogen bond motif with the KPX through N009-H00A···O006 and O004-H004···N008 interactions. The KPX cation additionally connects to the water molecule via an O005-H00G···O003 hydrogen bond. The asymmetric units combine through a centrosymmetric operation to form the complete unit cell. These unit cells then assemble into 1D structures extending along the *a*-axis through two types of interactions: $R_2^2(8)$ motifs formed by N009-H00B···O003 hydrogen bonds and π-π stacking between aromatic rings (centroid-to-centroid distance: 3.893 Å, dihedral angle: 14.33°). The 1D structures subsequently pack along the *b*-axis through N00A-H00D···O005 hydrogen bonds, forming 2D layers. Finally, these layers stack along the *c*-axis via van der Waals interactions, completing the 3D crystal structure.

Supplementary Table 22. Hydrogen-bond geometry (Å) of KPX-3,4-dichlorobenzoic acid cocrystal monohydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
O004-H004···N008	0.820	1.803	2.609	167.05
N009–H00A···O006	0.861	1.974	2.811	164.01
N009–H00B···O003	0.859	2.064	2.777	139.79
O005–H00G···O003	0.849	2.723	2.704	79.77
N00A–H00D···O005	0.889	2.450	2.886	110.53



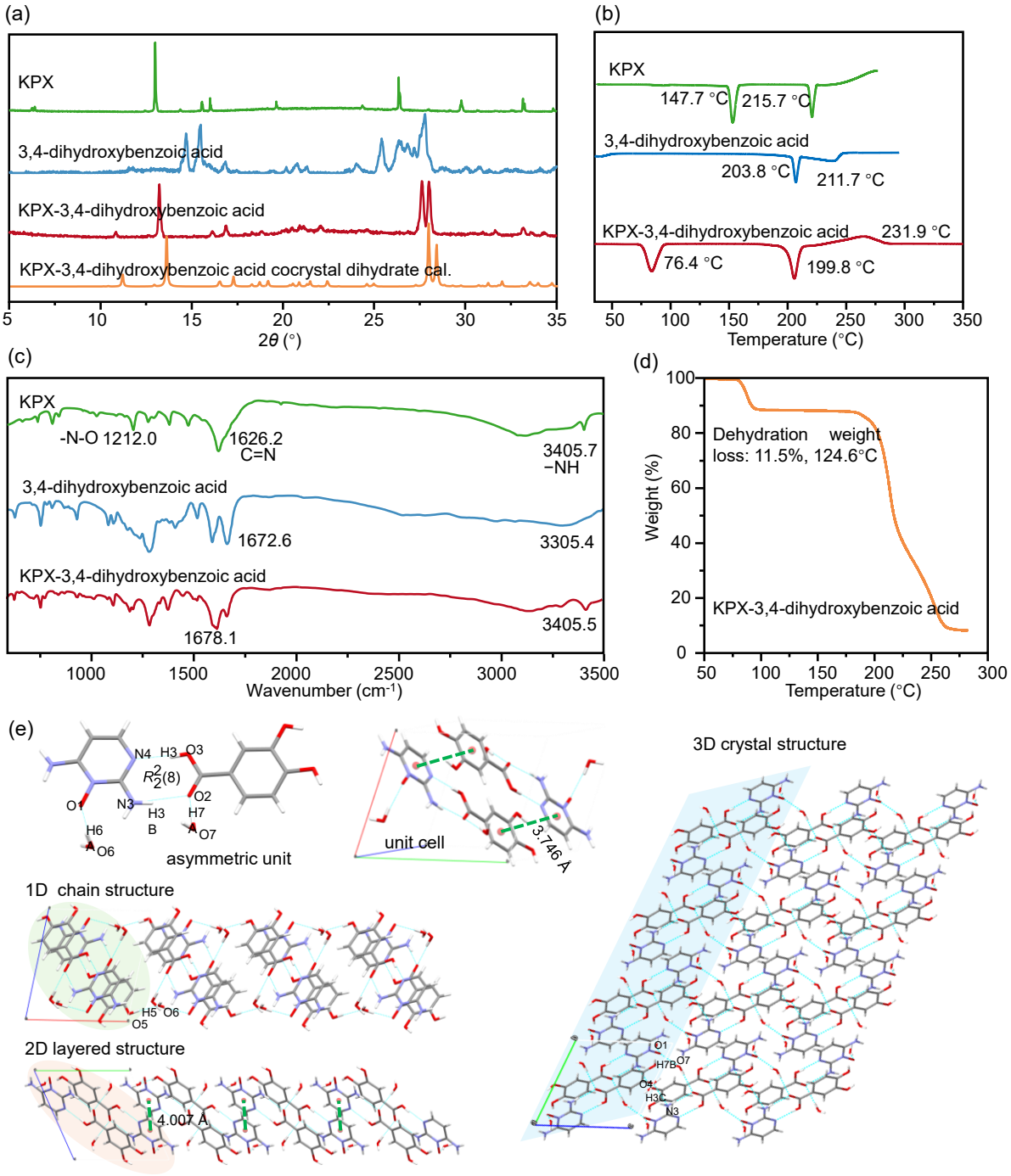
Supplementary Figure 16. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-3,4-dichlorobenzoic acid cocrystal monohydrate.

7.9 KPX-3,4-dihydroxybenzoic acid salt dihydrate

KPX-3,4-dihydroxybenzoic acid cocrystal dihydrate crystallizes in the triclinic system with space group $P\bar{1}$ ($Z = 2$). The asymmetric unit comprises one neutral KPX molecule, one neutral 3,4-dihydroxybenzoic acid molecule, and two water molecules. Within the asymmetric unit, KPX and 3,4-dihydroxybenzoic acid are linked via an $R_2^2(8)$ motif formed by O3–H3···N4 and N3–H3B···O2 hydrogen bonds. KPX is further connected to one water molecule through O6–H6A···O1 hydrogen bond, while 3,4-dihydroxybenzoic acid binds to the other water molecule via O7–H7A···O2 hydrogen bond. The unit cell is generated by the asymmetric unit and its centrosymmetric counterpart. Two asymmetric units interact through offset π···π stacking (Cg–Cg distance: 3.771 Å; dihedral angle: 4.41°) between the pyrimidine ring of KPX and the benzene ring of 3,4-dihydroxybenzoic acid. Adjacent unit cells along the *a*-axis form a 1D chain structure through O5–H5···O6 hydrogen bonds. These 1D chains pack along the *b*-axis, creating a 2D layered structure via parallel π···π stacking between KPX pyrimidine rings (Cg–Cg distance: 4.007 Å). Finally, the 3D crystal structure is stabilized by O7–H7B···O1 and N3–H3C···O4 hydrogen bonds, which connect identically oriented 2D layers.

Supplementary Table 23. Hydrogen-bond geometry (Å) of KPX-3,4-dihydroxybenzoic acid salt dihydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
O3–H3···N4	0.840	1.875	2.710	171.94
N3–H3B···O2	0.880	1.936	2.804	168.54
N3–H3C···O4	0.880	2.308	2.932	127.87
O6–H6A···O1	0.870	1.879	2.743	171.66
O7–H7B···O1	0.870	1.947	2.744	151.77
N2–H2B···O7	0.880	2.127	2.969	160.11
O5–H5···O6	0.840	1.863	2.697	171.07
O7–H7A···O2	0.870	2.058	2.925	173.75



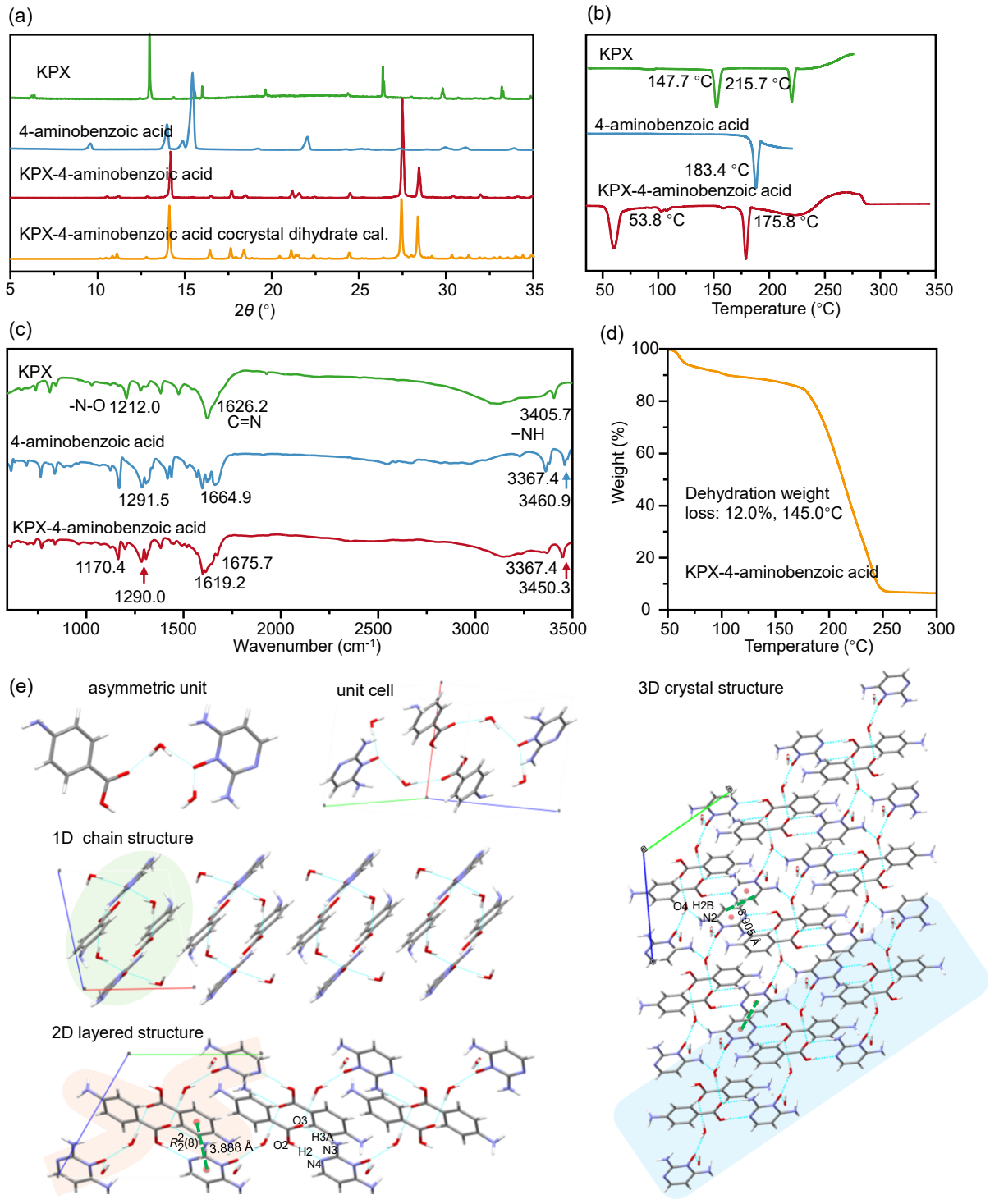
Supplementary Figure 17. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-3,4-dihydroxybenzoic acid salt dihydrate.

7.10 KPX-4-aminobenzoic acid cocrystal dihydrate

The KPX-4-aminobenzoic acid cocrystal dihydrate crystallizes in the triclinic system with the space group $P\bar{1}$ ($Z = 2$). The asymmetric unit consists of one neutral KPX molecule, one neutral 4-aminobenzoic acid molecule, and two water molecules. Within the asymmetric unit, KPX is linked to two water molecules through O4–H4A···O1 and O5–H5C···O1 hydrogen bonds, while 4-aminobenzoic acid is linked to one water molecule via the O4–H4B···O3 hydrogen bond. The unit cell is formed by the asymmetric unit and its centrosymmetrically related counterpart. Unit cells of identical orientation pack through van der Waals interactions, extending along the *a*-axis to generate a 1D chain structure. These 1D chains further assemble along the *b*-axis via $R_2^2(8)$ motifs formed by N3–H3A···O3 and O2–H2···N4 hydrogen bonds and π···π stacking (Cg–Cg distance: 3.888 Å, dihedral angle: 7.52°), resulting in a 2D layered structure. Finally, the 2D layers arrange into a 3D crystal structure through additional π···π interactions (Cg–Cg distance: 3.905 Å, parallel) and N2–H2B···O4 hydrogen bonds.

Supplementary Table 24. Hydrogen-bond geometry (Å) of KPX-4-aminobenzoic acid cocrystal dihydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N3–H3A···O3	0.880	1.981	2.846	167.01
O2–H2···N4	0.840	1.878	2.710	170.40
O4–H4A···O1	0.869	1.954	2.799	163.63
N2–H2B···O4	0.910	2.048	2.942	167.19
O5–H5C···O1	0.870	1.914	2.771	167.82
O4–H4B···O3	0.870	2.060	2.919	168.81



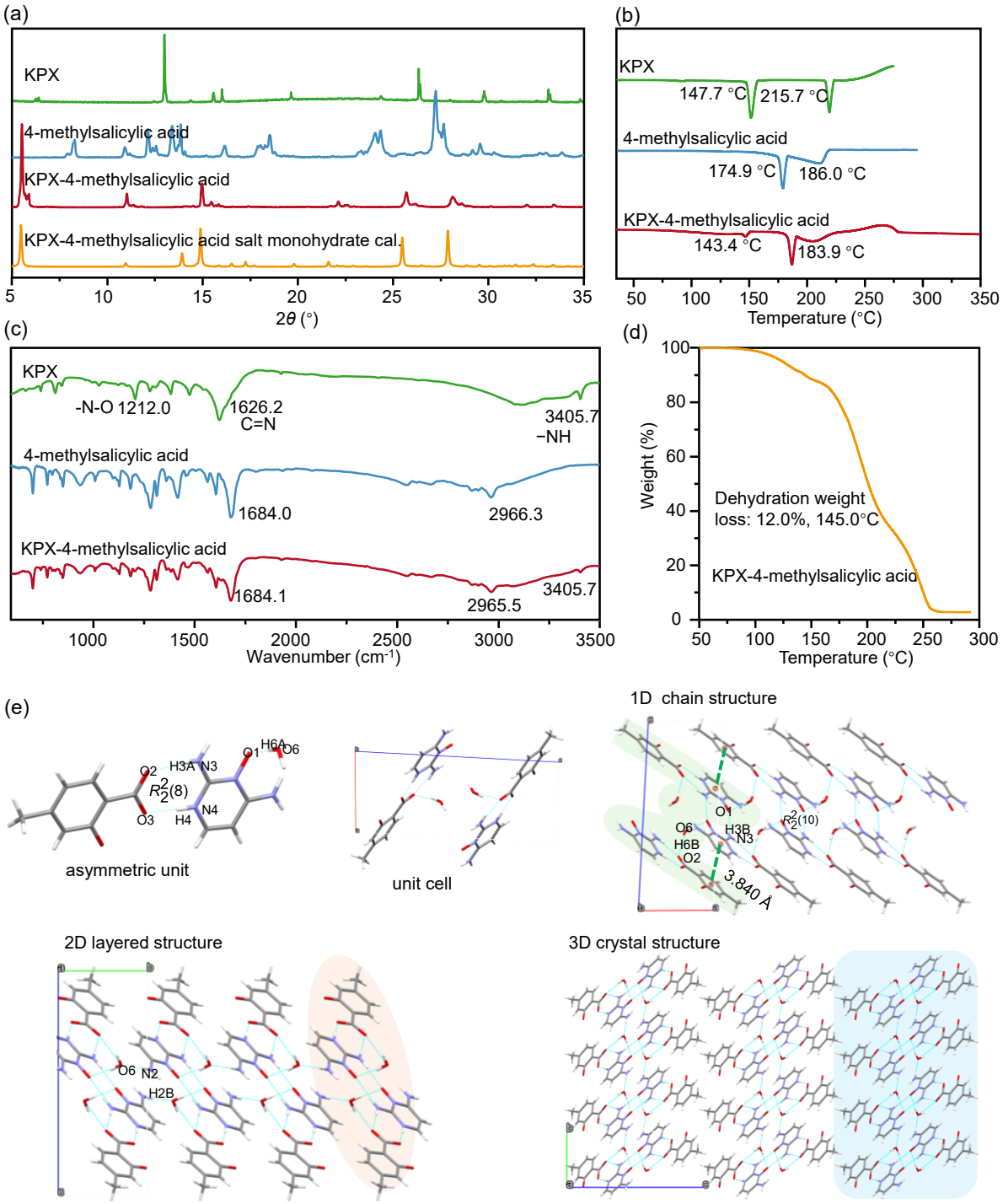
Supplementary Figure 18. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-4-aminobenzoic acid cocrystal dihydrate.

7.11 KPX-4-methylsalicylic acid salt monohydrate

The KPX–4-methylsalicylic acid salt monohydrate crystallizes in the triclinic system with the space group $P\bar{1}$ ($Z = 2$). The asymmetric unit comprises one protonated KPX cation, one deprotonated 4-methylsalicylic acid anion, and one water molecule. Proton transfer occurs from the carboxylic acid group of 4-methylsalicylic acid to the N atom of KPX's pyrimidine ring. The 4-methylsalicylic acid anion and KPX cation are connected via an $R_2^2(8)$ motif, stabilized by N4–H4···O3 and N3–H3A···O2 hydrogen bonds, while the KPX cation further interacts with the water molecule through O6–H6A···O1. The unit cell is formed by the asymmetric unit and its centrosymmetrically related counterpart. Unit cells of identical orientation extend along the *a*-axis, forming a 1D structure through O6–H6B···O2 hydrogen bond, $R_2^2(8)$ motifs formed by N3–H3B···O1, and π···π stacking (Cg–Cg distance: 3.840 Å, dihedral angle: 6.93°). These 1D structures pack along the *b*-axis via N2–H2B···O6 hydrogen bond, generating a 2D layered architecture. Finally, the 2D layers assemble along the *c*-axis through van der Waals interactions, completing the 3D crystal structure.

Supplementary Table 25. Hydrogen-bond geometry (Å) of KPX-4-methylsalicylic acid salt monohydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N3–H3A···O2	0.880	1.913	2.784	170.19
N4–H4···O3	0.879	1.716	2.595	178.89
N3–H3B···O1	0.879	2.026	2.788	144.35
O6–H6A···O1	0.871	1.846	2.714	174.52
O6–H6B···O2	0.870	2.097	2.932	160.35
N2–H2B···O6	0.879	2.003	2.880	175.11



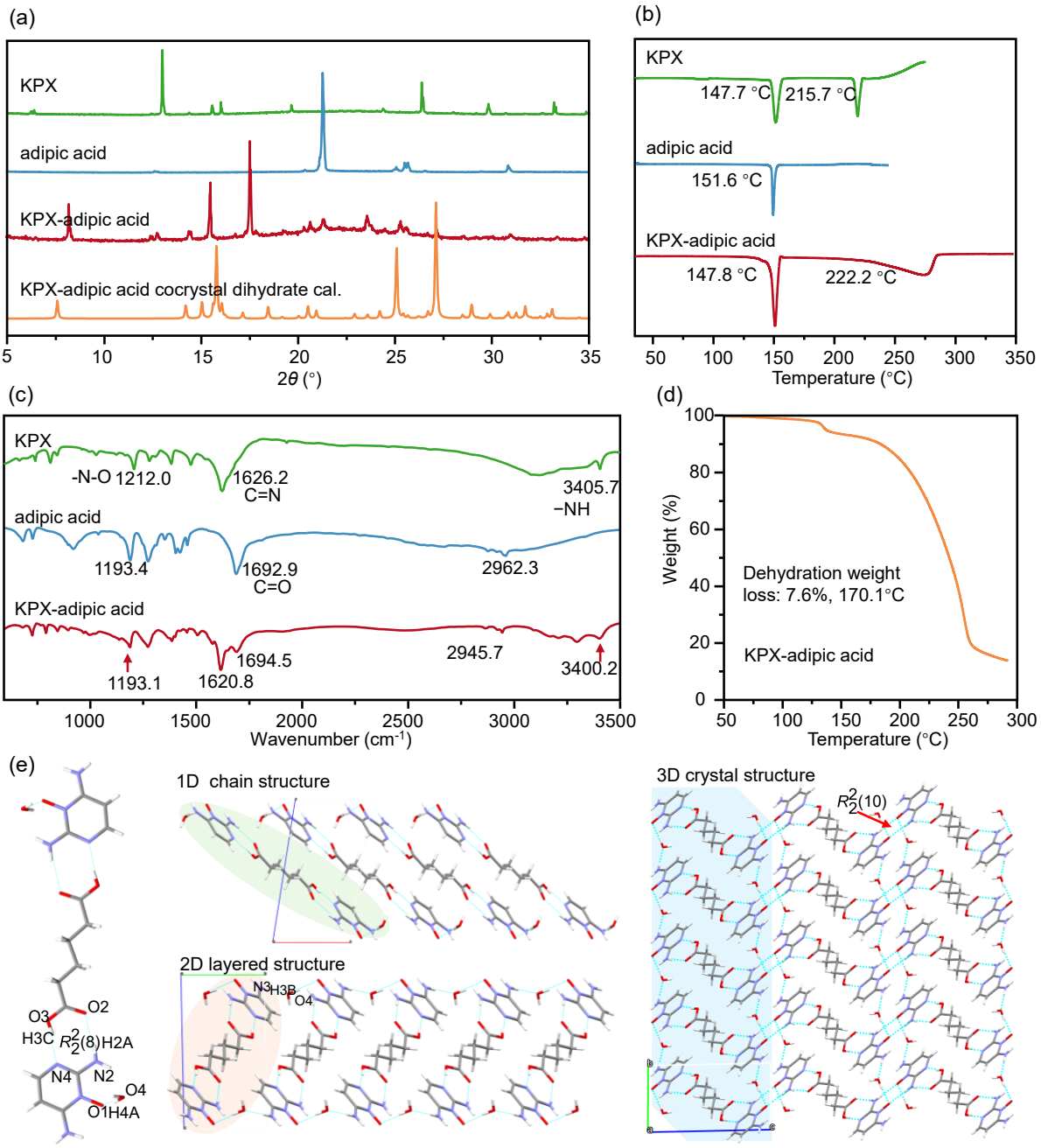
Supplementary Figure 19. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-4-methylsalicylic acid salt monohydrate.

7.12 KPX-adipic acid cocrystal dihydrate

KPX-adipic acid cocrystal dihydrate crystallizes in the triclinic system with space group $P\bar{1}$ (Z = 1). The asymmetric unit contains a characteristic motif formed by one neutral KPX, half neutral adipic acid and one water molecule. The adipic acid is linked to the KPX through an $R_2^2(8)$ motif formed by O3–H3C···N4 and N2–H2A···O2 hydrogen bonds. KPX is linked to one water molecule via the O4–H4A···O1 hydrogen bond. Moreover, the other carboxylic acid group of adipic acid connects to another KPX molecule in the same manner, which is linked to another water molecule in the same manner, forming a pentamer. The pentamers with the same orientation pack along the *a*-axis through van der Waals forces, forming a 1D structure. The connected 1D structures with the same orientation pack along *b*-axis through N3–H3B···O4 hydrogen bonds, giving rise to a 2D layered structure, as depicted in Fig. S4c. Ultimately, the 2D layered structures with the same orientation are arranged along *c*-axis through $R_2^2(10)$ motifs formed by N2–H2B···O1 hydrogen bonds, thereby completing the 3D crystal structure.

Supplementary Table 26. Hydrogen-bond geometry (Å) of KPX-adipic acid cocrystal dihydrate.

	d(D–H)/Å	d(H···A)/Å	d(D···A)/Å	(D–H···A)/°
O3–H3C···N4	0.840	1.796	2.623	168.04
N2–H2A···O2	0.880	2.038	2.902	167.05
O4–H4A···O1	0.870	1.852	2.700	164.20
N3–H3B···O4	0.880	2.051	2.930	175.70
N2–H2B···O1	0.880	2.039	2.787	142.26



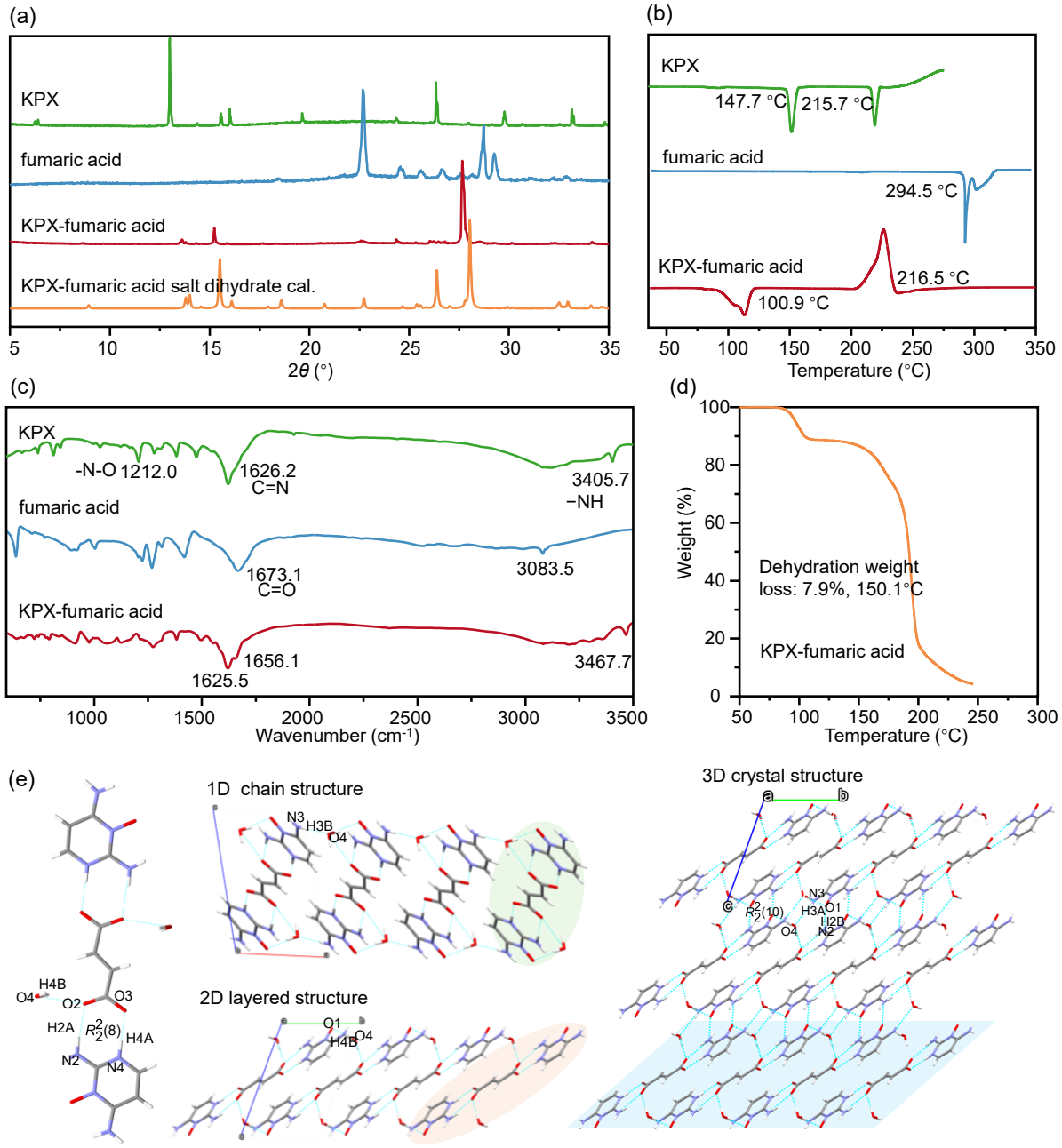
Supplementary Figure 20. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-adipic acid cocrystal dihydrate.

7.13 KPX-fumaric acid salt dihydrate

KPX-fumaric acid salt dihydrate crystallizes in the triclinic system with space group $P\bar{1}$ (Z = 1). The asymmetric unit consists of one protonated KPX cation, half a deprotonated fumaric acid anion, and one water molecule. Proton transfer occurs from the carboxylic acid group of fumaric acid to the pyrimidine nitrogen atom of KPX. The fumaric acid anion connects to the KPX cation through an $R_2^2(8)$ motif stabilized by N2-H2A···O2 and N4-H4A···O3 hydrogen bonds, while the KPX cation additionally binds to a water molecule via O4-H4C···O2 hydrogen bond. The second carboxylic group of the fumaric acid anion similarly links to another KPX cation and water molecule, forming a pentameric unit. These pentamers, all aligned in the same orientation, assemble along the *a*-axis through N3-H3B···O4 hydrogen bonds to create a one-dimensional chain structure. These 1D chains then organize along the *b*-axis via O4-H4B···O1 hydrogen bonds, generating a 2D layered architecture. The complete 3D crystal structure emerges through additional N3-H3A···O4 hydrogen bonds and $R_2^2(10)$ motifs formed by N2-H2B···O1 interactions between adjacent layers along the *c*-axis.

Supplementary Table 27. Hydrogen-bond geometry (Å) of KPX-fumaric acid salt dihydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N2–H2B···O1	0.880	1.992	2.762	145.33
N2–H2A···O2	0.880	1.957	2.823	167.42
N3–H3A···O4	0.880	2.252	3.070	154.51
N3–H3B···O4	0.880	1.989	2.862	171.51
N4–H4A···O3	0.880	1.731	2.611	178.02
O4–H4B···O1	0.870	1.843	2.707	171.52
O4–H4C···O2	0.871	2.192	2.941	143.94



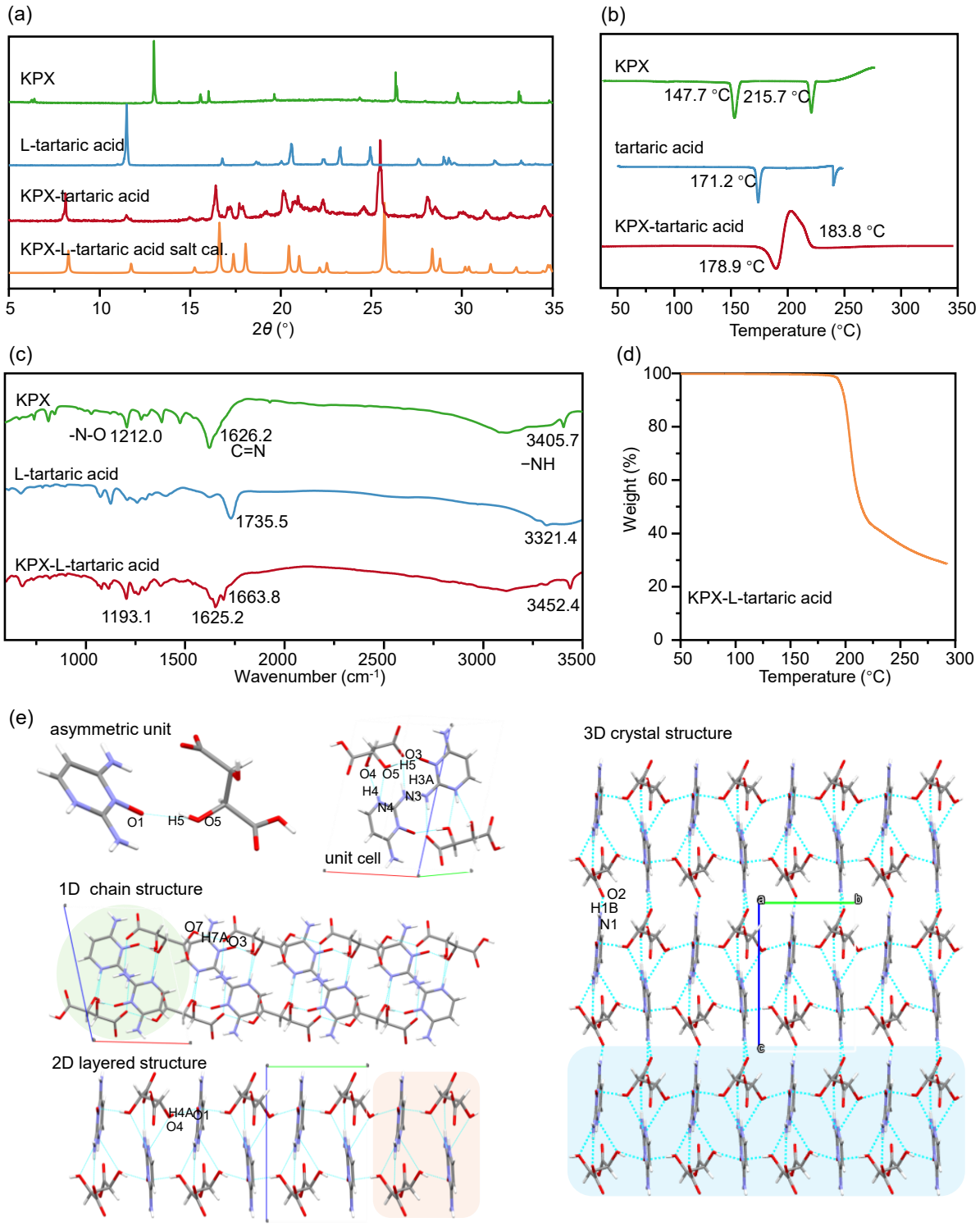
Supplementary Figure 21. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-fumaric acid salt dihydrate.

7.14 KPX-L-tartaric acid salt

KPX-L-tartaric acid salt crystallizes in the monoclinic system with space group $P2_1$ ($Z = 2$). The asymmetric unit comprises one protonated KPX cation and one deprotonated L-tartaric acid anion, with proton transfer occurring from the carboxylic acid group of L-tartaric acid to the pyrimidine nitrogen atom of KPX. The L-tartaric acid anion forms an O5-H5···O1 hydrogen bond with the KPX cation. The unit cell is generated through centrosymmetric operation of the asymmetric unit, stabilized by N4-H4···O4, N4-H4···O5, and N3-H3A···O3 hydrogen bonds. These unit cells, sharing the same orientation, extend along the *a*-axis via O7-H7A···O3 hydrogen bonds to form 1D chain structures. The 1D chains then organize along the *b*-axis through O4-H4A···O1 hydrogen bonds, creating a 2D layered architecture. Final 3D packing is achieved through N1-H1B···O2 hydrogen bonds along the *c*-axis between adjacent layers.

Supplementary Table 28. Hydrogen-bond geometry (Å) of KPX-L-tartaric acid salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
O5–H5···O1	0.820	1.814	2.604	161.29
O4–H4A···O1	0.820	1.864	2.672	168.39
N3–H3A···O3	0.861	1.969	2.809	165.19
N4–H4···O4	0.860	2.176	2.940	147.82
N4–H4···O5	0.860	2.135	2.774	130.73
N1–H1B···O2	0.860	1.951	2.770	158.76
O7–H7A···O3	0.820	1.737	2.555	175.81



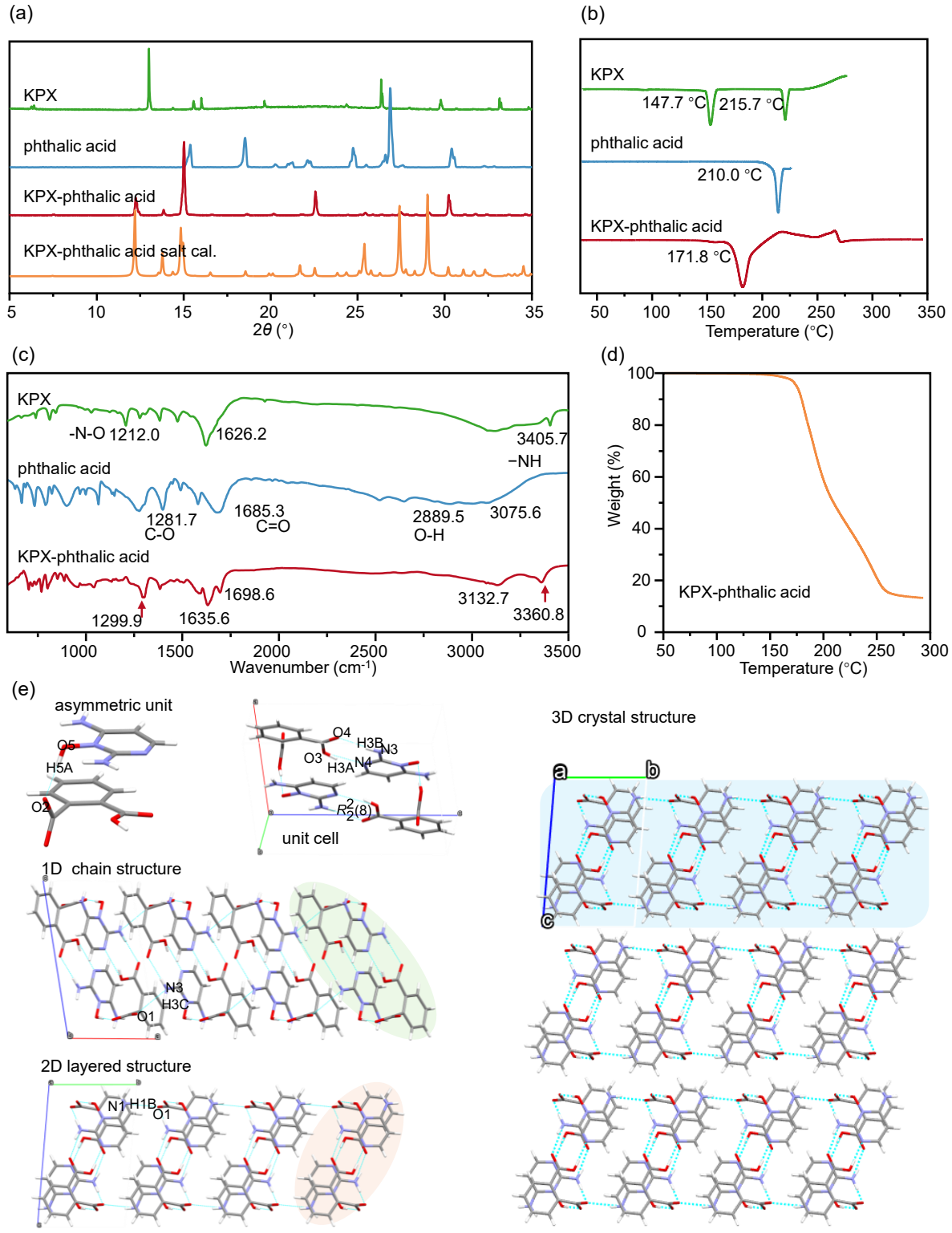
Supplementary Figure 22. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-L-tartaric acid salt.

7.15 KPX-phthalic acid salt

KPX-phthalic acid salt crystallizes in the triclinic system with space group $P\bar{1}$ (Z = 2). The asymmetric unit consists of one protonated KPX cation and one deprotonated phthalic acid anion, connected through an O5-H5A···O2 hydrogen bond. Proton transfer occurs from the carboxylic acid group of phthalic acid to the oxygen atom of KPX's pyrimidine N-oxide group. The unit cell is formed by the asymmetric unit and its centrosymmetric counterpart, stabilized by N3-H3B···O4 and O3-H3A···N4 hydrogen bonds. These unit cells extend along the *a*-axis through N3-H3C···O1 hydrogen bonds, generating 1D chain structures. The 1D chains assemble along the *b*-axis via N1-H1B···O1 hydrogen bonds to form 2D layers. Finally, these 2D layers stack along the *c*-axis through van der Waals interactions to complete the 3D crystal packing.

Supplementary Table 29. Hydrogen-bond geometry (Å) of KPX-phthalic acid salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N1–H1B···O1	0.880	1.964	2.823	164.79
O5–H5A···O2	0.841	1.624	2.459	171.99
N3–H3C···O1	0.879	2.435	2.973	119.96
N3–H3B···O4	0.880	1.971	2.825	162.95
O3–H3A···N4	0.840	1.942	2.726	154.78



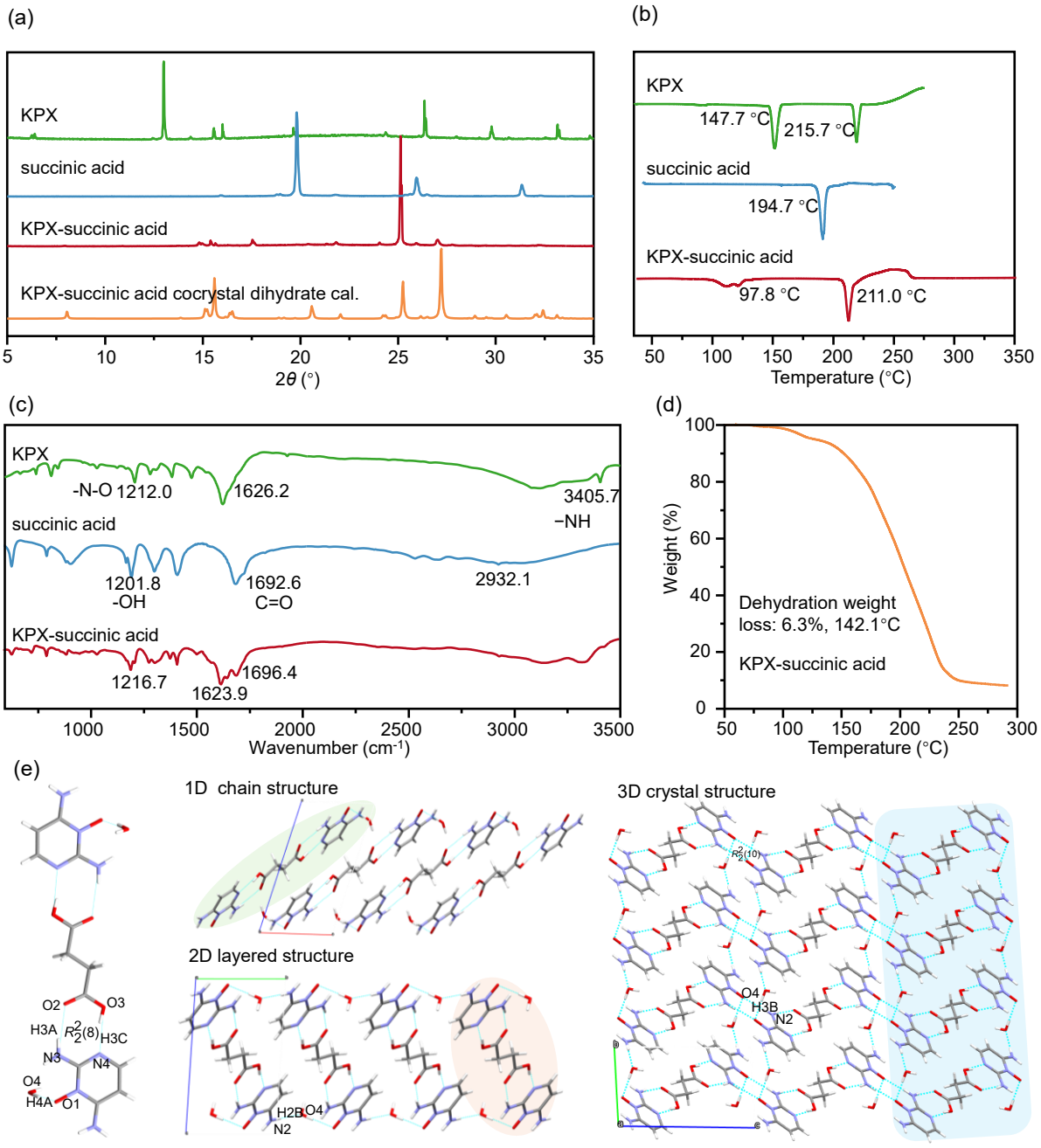
Supplementary Figure 23. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-phthalic acid salt.

7.16 KPX-succinic acid cocrystal dihydrate

KPX-succinic acid cocrystal dihydrate crystallizes in the triclinic system with space group $P\bar{1}$ (Z = 1). The asymmetric unit consists of a characteristic motif formed by one neutral KPX molecule, half a neutral succinic acid molecule, and one water molecule. In this structure, the succinic acid molecule connects to KPX through an $R_2^2(8)$ motif stabilized by O3-H3C···N4 and N3-H3A···O2 hydrogen bonds, while KPX simultaneously binds to a water molecule via the O4-H4A···O1 hydrogen bond. The second carboxylic acid group of succinic acid forms an equivalent connection with another KPX molecule and water molecule, resulting in the formation of a pentameric unit. These pentamers, all oriented identically, assemble along the *a*-axis through van der Waals interactions to generate 1D chain structures. The resulting 1D chains then organize along the *b*-axis via N2-H2B···O4 hydrogen bonds, producing a 2D layered architecture. Finally, these 2D layers arrange along the *c*-axis through $R_2^2(10)$ motifs formed by N3-H3B···O1 hydrogen bonds, completing the 3D crystal packing.

Supplementary Table 30. Hydrogen-bond geometry (Å) of KPX-succinic acid cocrystal dihydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N3–H3B···O1	0.881	2.012	2.784	145.76
N3–H3A···O2	0.880	2.048	2.915	168.43
O3–H3C···N4	0.840	1.804	2.628	166.68
N2–H2B···O4	0.880	1.980	2.858	175.47
O4–H4A···O1	0.870	1.822	2.687	172.11



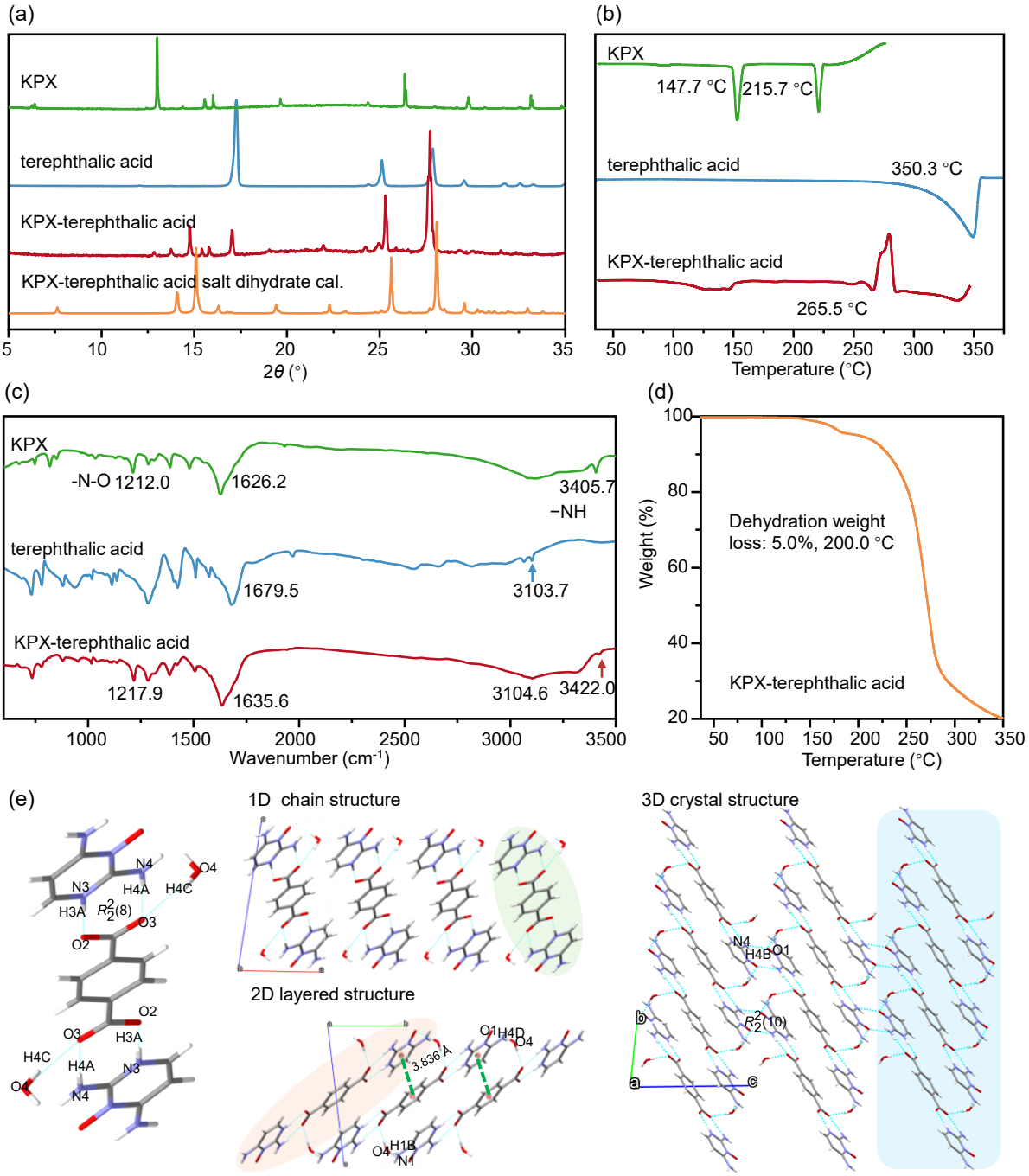
Supplementary Figure 24. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-succinic acid cocrystal dihydrate.

7.17 KPX-terephthalic acid salt dihydrate

KPX-terephthalic acid cocrystal dihydrate crystallizes in the triclinic system with space group $P\bar{1}$ (Z = 1). The asymmetric unit consists of a characteristic motif formed by two protonated KPX cations, one deprotonated terephthalic acid anion, and two water molecules. In this structure, the terephthalic acid anion connects to KPX cations through an $R_2^2(8)$ motif stabilized by N4-H4A···O3 and N3-H3A···O2 hydrogen bonds and connects to water molecules via the O4-H4C···O3 hydrogen bond, resulting in the formation of a pentameric unit. These pentamers, all oriented identically, assemble along the *a*-axis through van der Waals interactions to generate 1D chain structures. The resulting 1D chains then organize along the *b*-axis via π-π stacking interactions (Cg-Cg distance: 3.836 Å, dihedral angle: 6.17°), N1-H1B···O4 and O4-H4D···O1 hydrogen bonds, producing a 2D layered architecture. Finally, these 2D layers arrange along the *c*-axis through $R_2^2(10)$ motifs formed by N4-H4B···O1 hydrogen bonds, completing the 3D crystal packing.

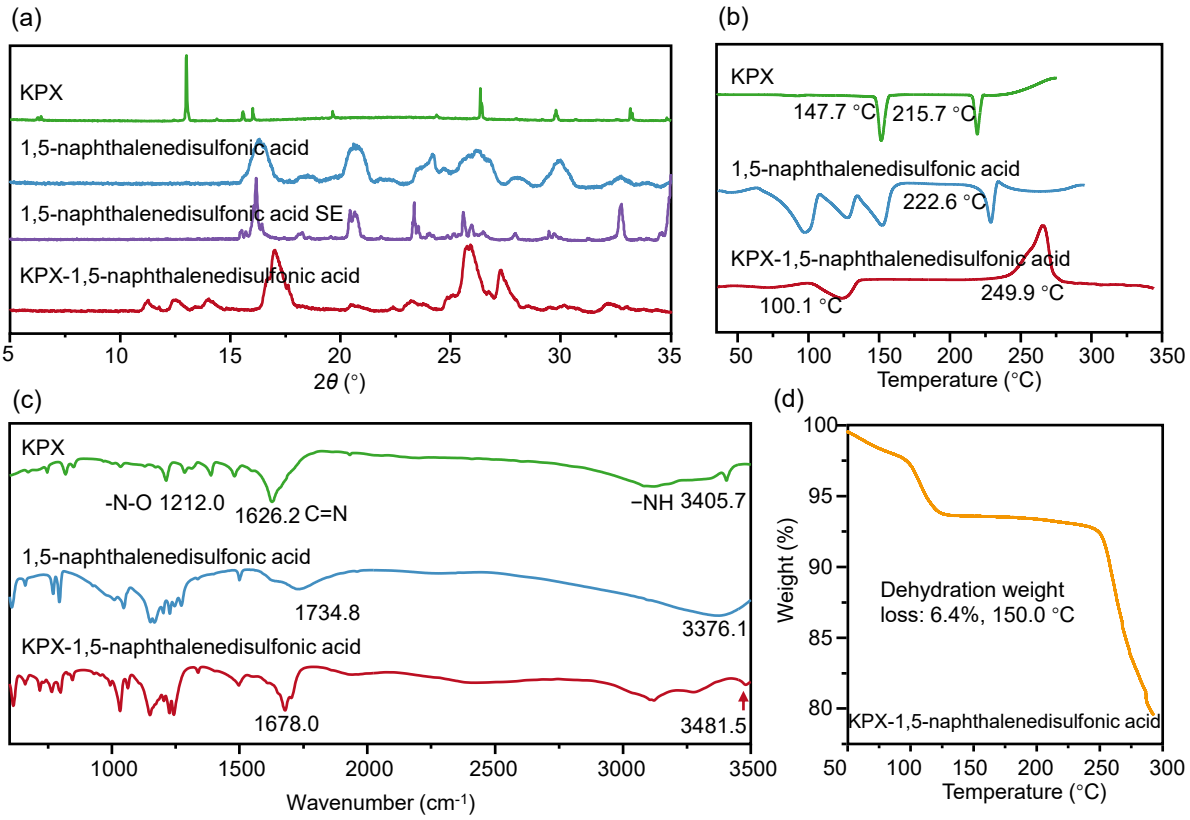
Supplementary Table 31. Hydrogen-bond geometry (Å) of KPX-terephthalic acid cocrystal dihydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N4-H4A···O3	0.880	1.890	2.761	169.01
N3-H3A···O2	0.880	1.894	2.616	177.76
O4-H4C···O3	0.850	2.097	2.936	159.00
N1-H1B···O4	0.881	1.986	2.864	175.11
O4-H4D···O1	0.850	1.901	2.712	169.31
N4-H4B···O1	0.880	2.011	2.761	142.56



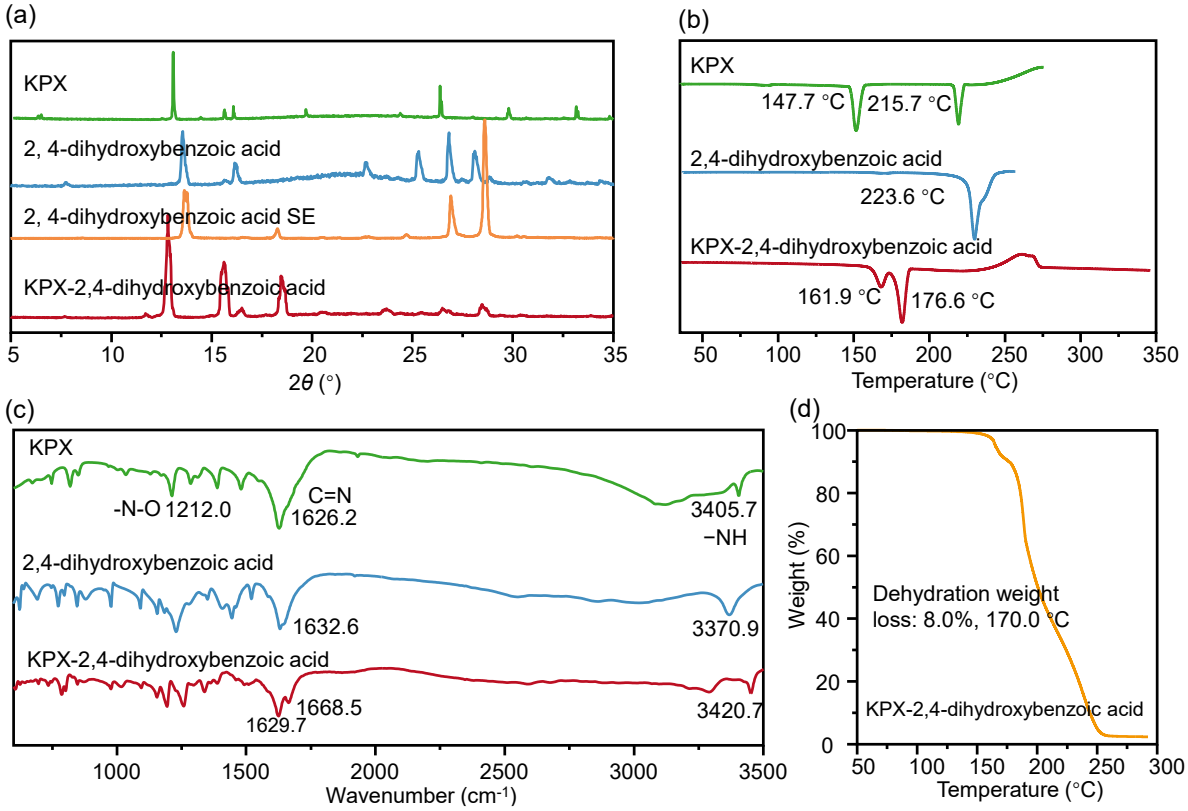
Supplementary Figure 25. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPX-terephthalic acid salt dihydrate.

7.18 KPX-1,5-naphthalenedisulfonic acid hydrate



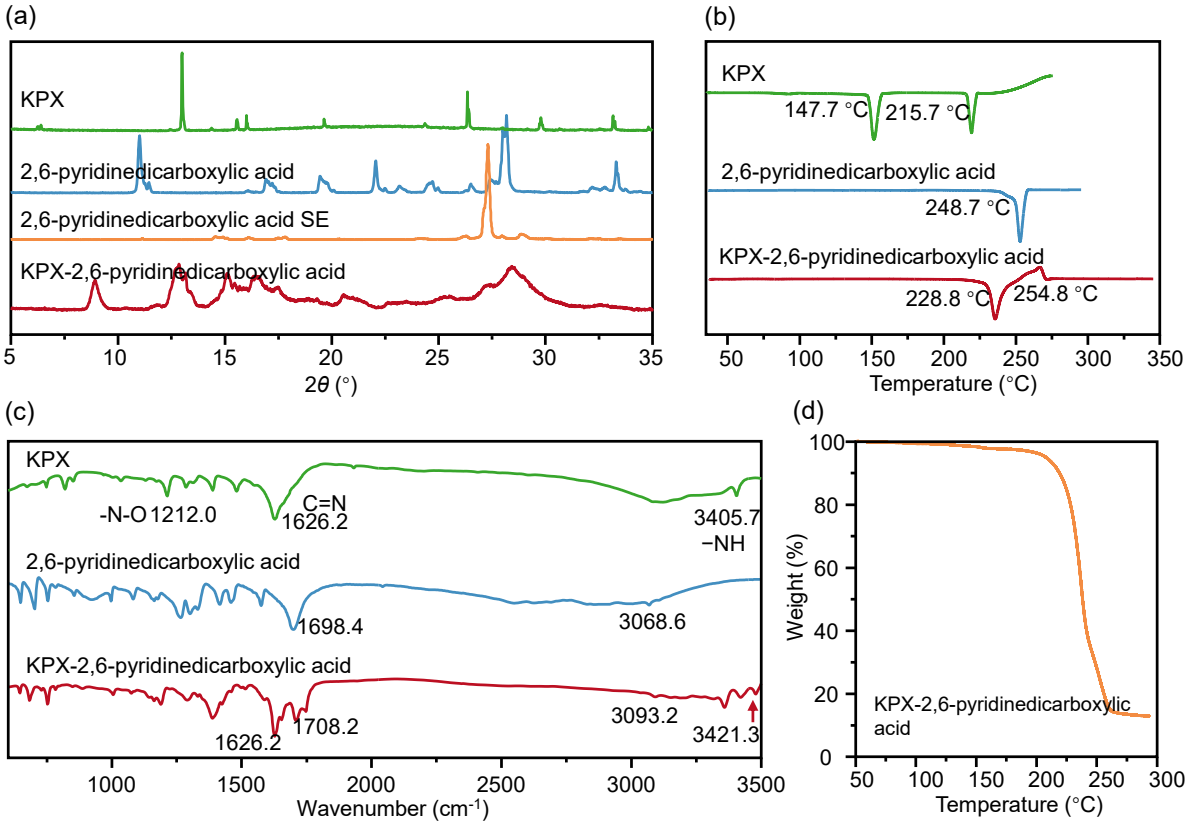
Supplementary Figure 26. Solid-state characterization of starting drugs and KPX-1,5-naphthalenedisulfonic acid hydrate (estimated). (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra and (d) TGA.

7.19 KPX-2,4-dihydroxybenzoic acid hydrate



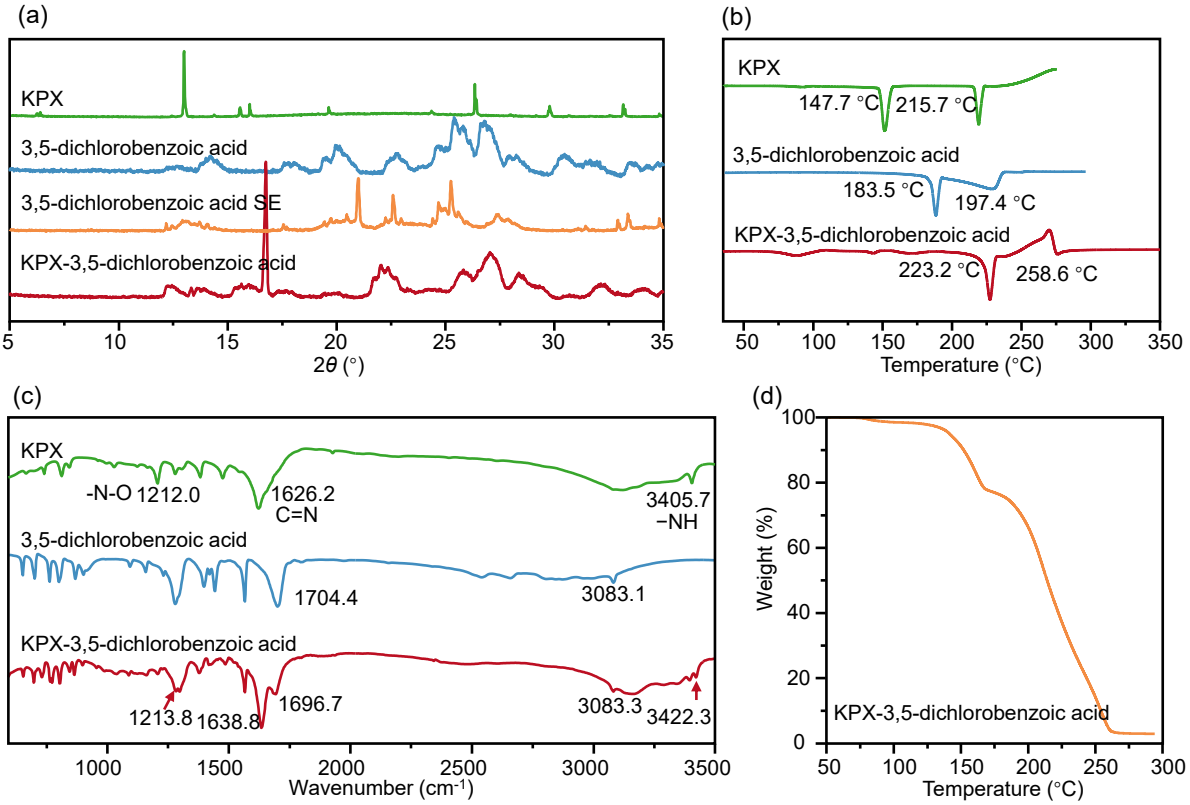
Supplementary Figure 27. Solid-state characterization of starting drugs and KPX-2,4-dihydroxybenzoic acid hydrate (estimated). (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra and (d) TGA.

7.20 KPX-2,6-pyridinedicarboxylic acid cocrystal



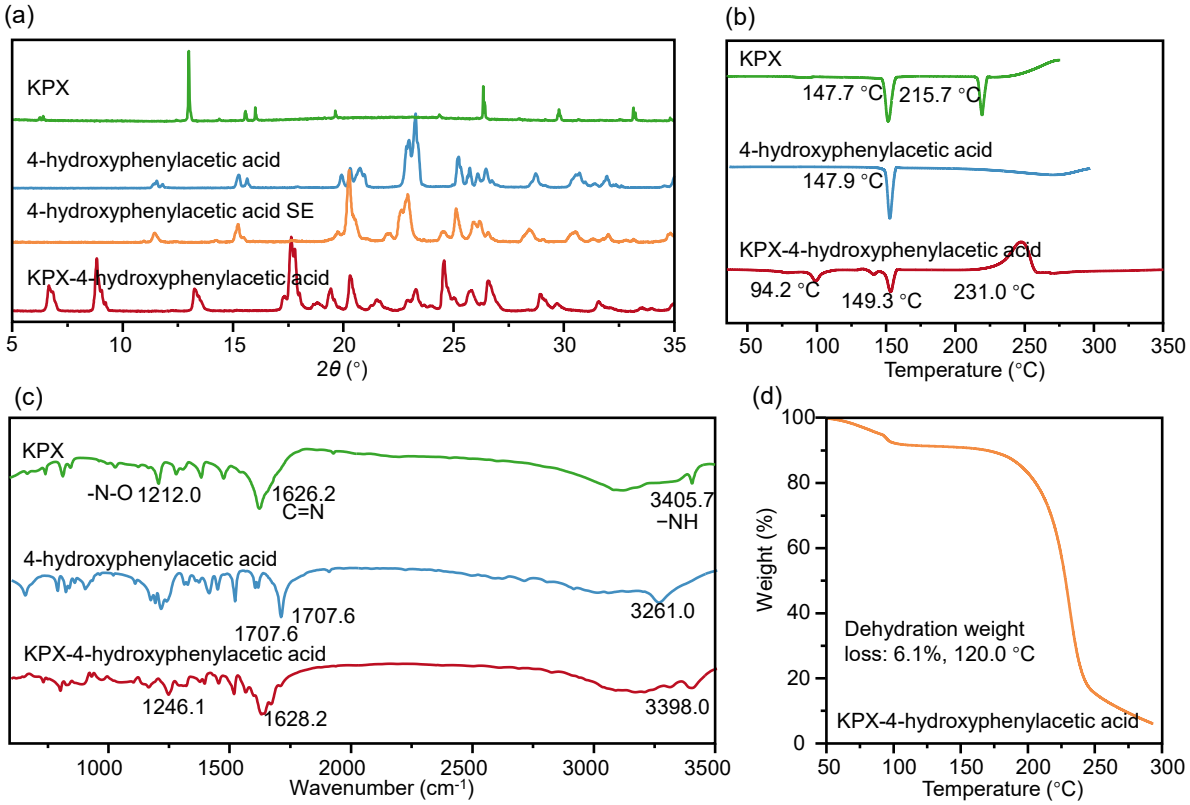
Supplementary Figure 28. Solid-state characterization of starting drugs and KPX-2,6-pyridinedicarboxylic acid cocrystal (estimated). (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra and (d) TGA.

7.21 KPX-3,5-dichlorobenzoic acid cocrystal



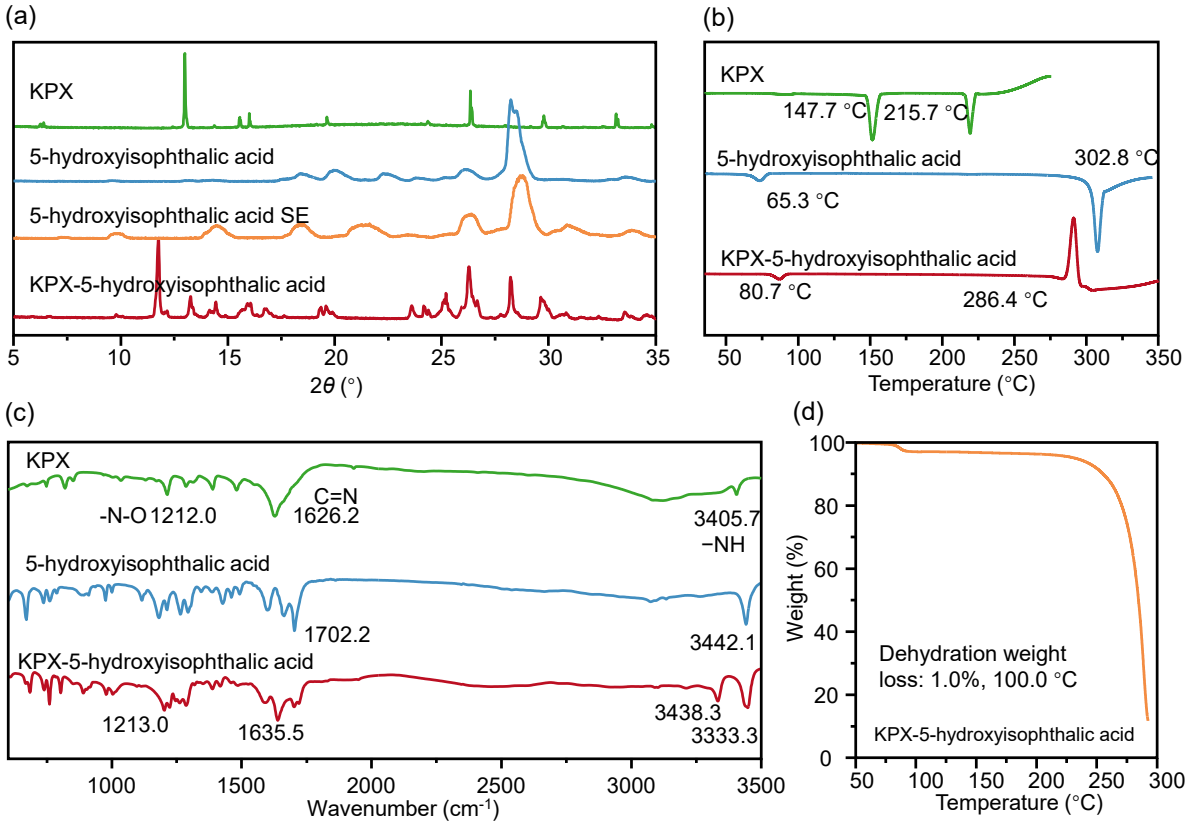
Supplementary Figure 29. Solid-state characterization of starting drugs and KPX-3,5-dichlorobenzoic acid cocrystal (estimated). (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra and (d) TGA.

7.22 KPX-4-hydroxyphenylacetic acid hydrate



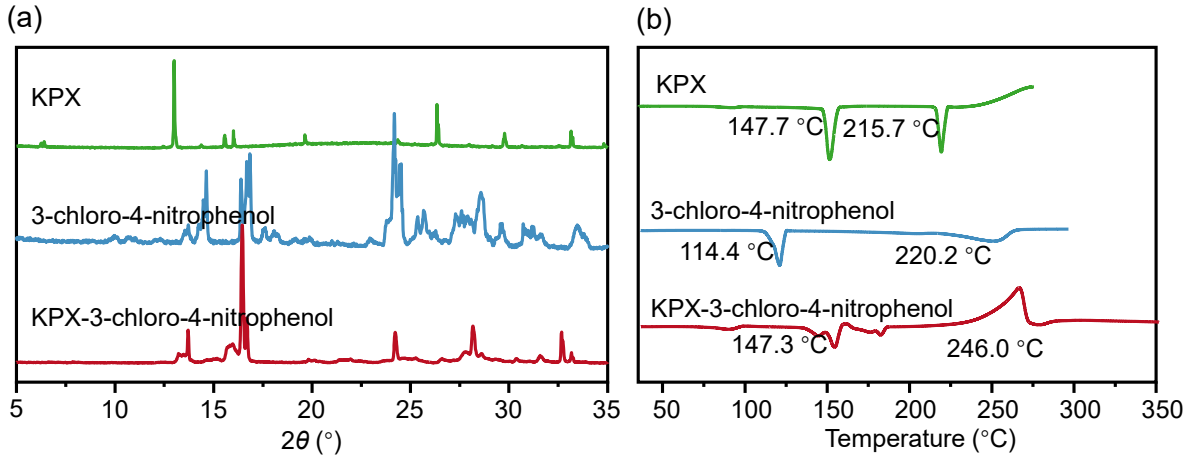
Supplementary Figure 30. Solid-state characterization of starting drugs and KPX-4-hydroxyphenylacetic acid hydrate (estimated). (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra and (d) TGA.

7.23 KPX-5-hydroxyisophthalic acid hydrate



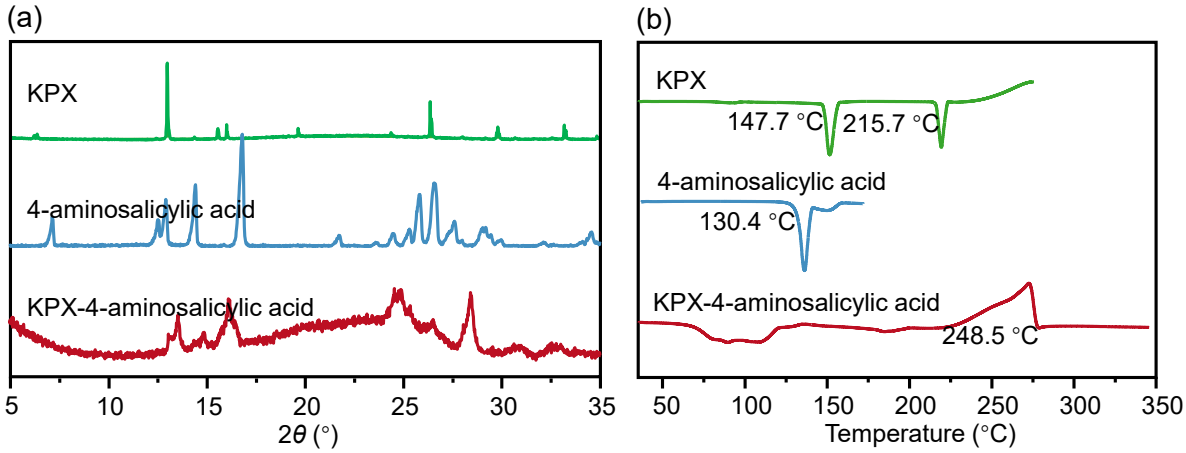
Supplementary Figure 31. Solid-state characterization of starting drugs and KPX-5-hydroxyisophthalic acid hydrate (estimated). (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra and (d) TGA.

7.24 KPX-3-chloro-4-nitrophenol (negative)



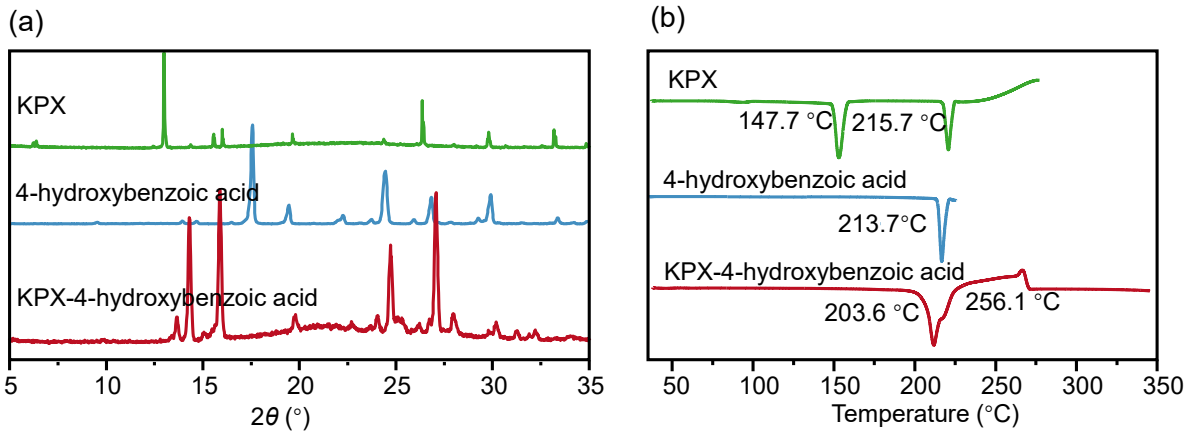
Supplementary Figure 32. (a) PXRD patterns and (b) DSC curves of starting drugs and KPX-3-chloro-4-nitrophenol (negative).

7.25 KPX-4-aminosalicylic acid (negative)



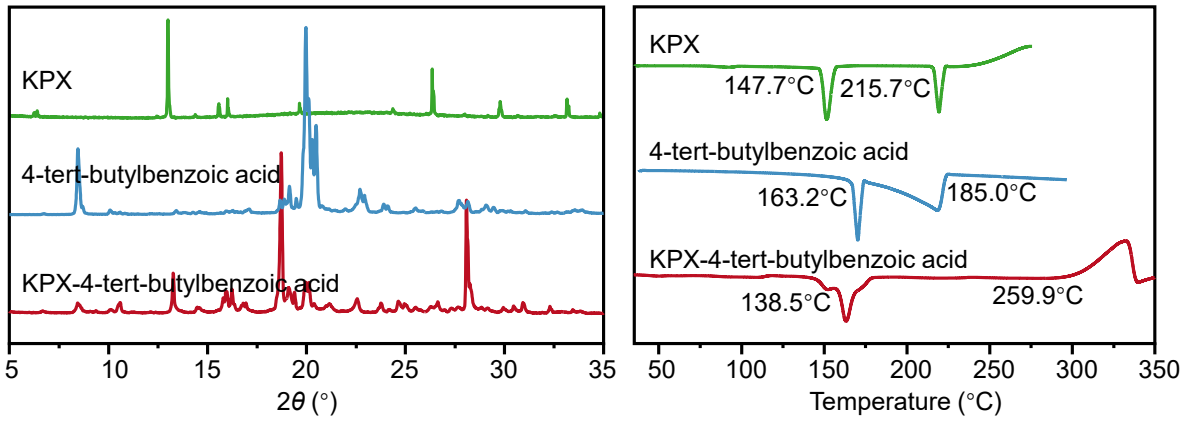
Supplementary Figure 33. (a) PXRD patterns and (b) DSC curves of starting drugs and KPX-4-aminosalicylic acid (negative).

7.26 KPX-4-hydroxybenzoic acid (negative)



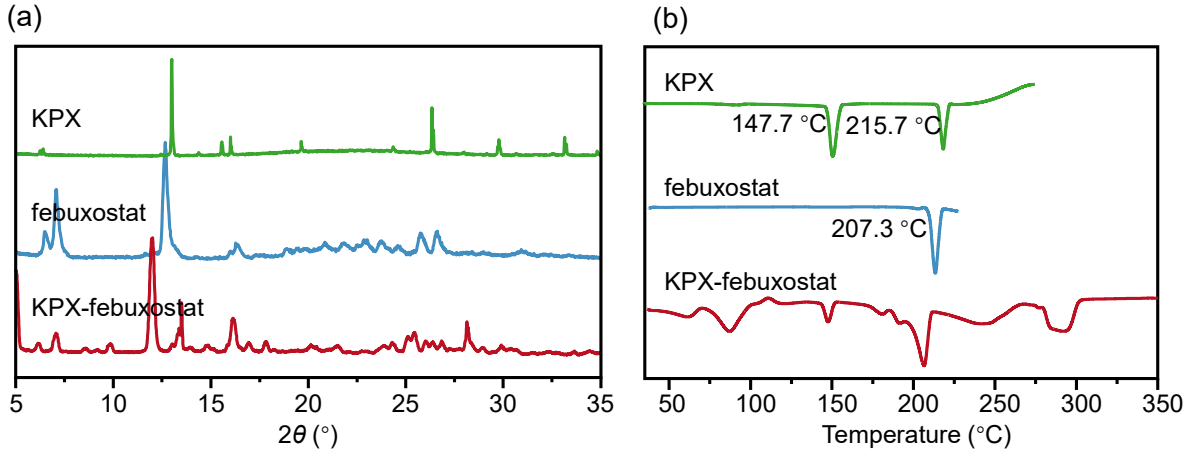
Supplementary Figure 34. (a) PXRD patterns and (b) DSC curves of starting drugs and KPX-4-hydroxybenzoic acid (negative).

7.27 KPX-4-tert-butylbenzoic acid (negative)



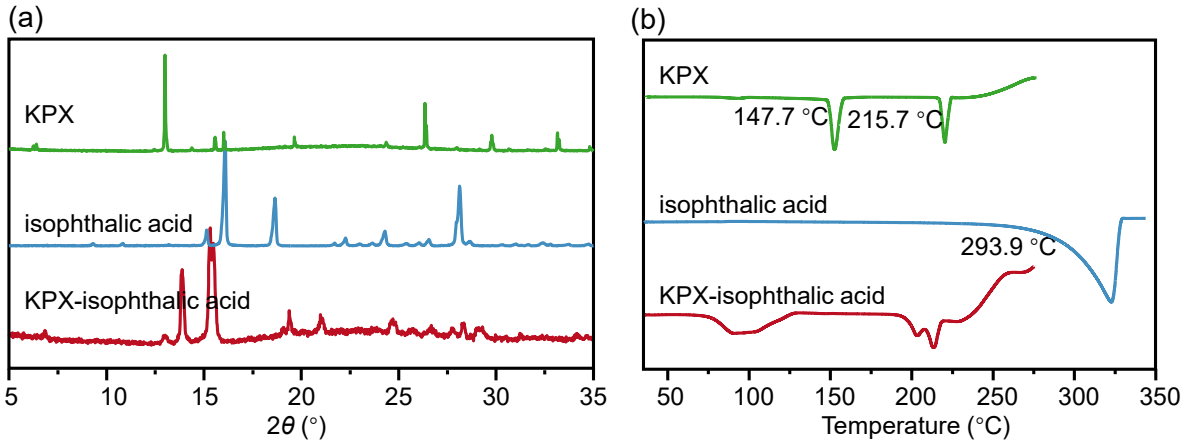
Supplementary Figure 35. (a) PXRD patterns and (b) DSC curves of starting drugs and KPX-4-tert-butylbenzoic acid (negative).

7.28 KPX-febuxostat (negative)



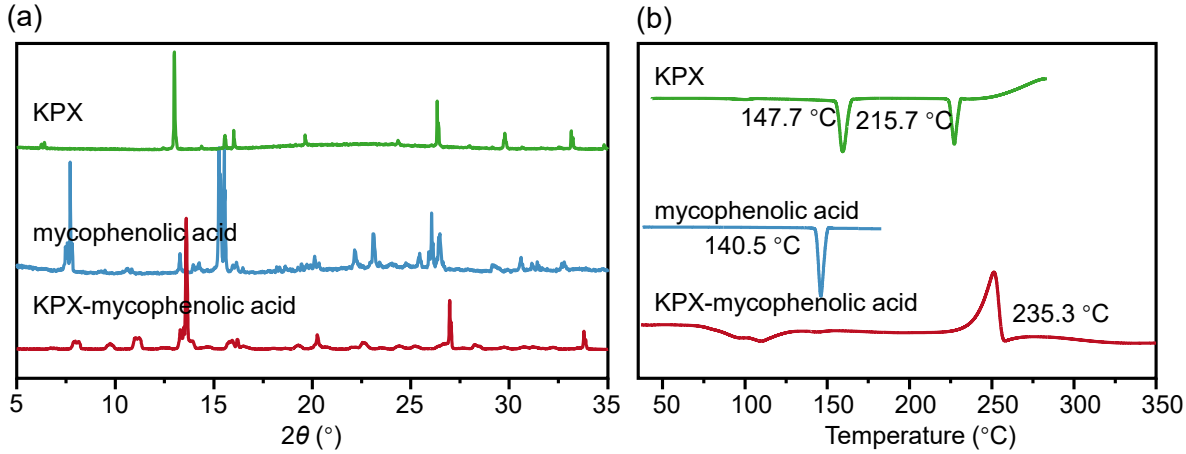
Supplementary Figure 36. (a) PXRD patterns and (b) DSC curves of starting drugs and KPX-febuxostat (negative).

7.29 KPX-isophthalic acid (negative)



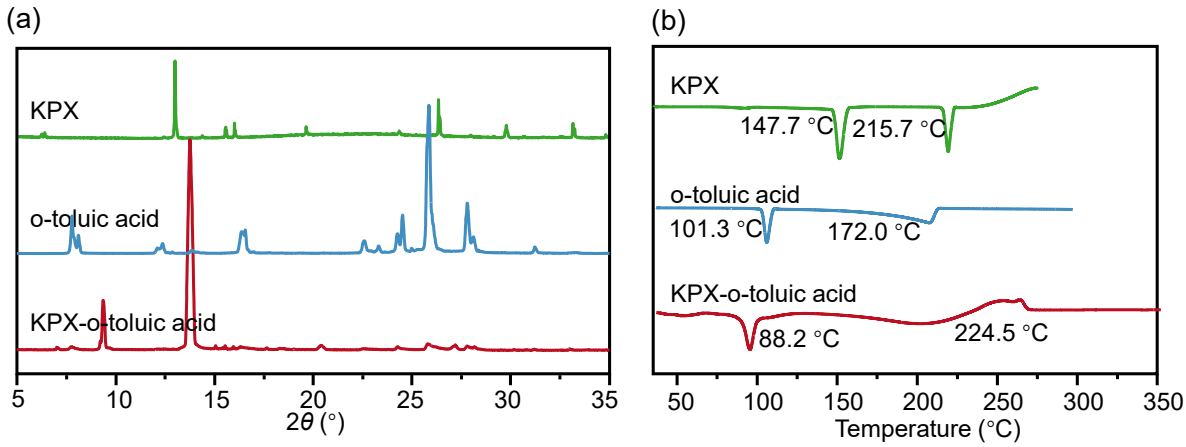
Supplementary Figure 37. (a) PXRD patterns and (b) DSC curves of starting drugs and KPX-isophthalic acid (negative).

7.30 KPX-mycophenolic acid (negative)



Supplementary Figure 38. (a) PXRD patterns and (b) DSC curves of starting drugs and KPX-mycophenolic acid (negative).

7.31 KPX-o-toluic acid (negative)



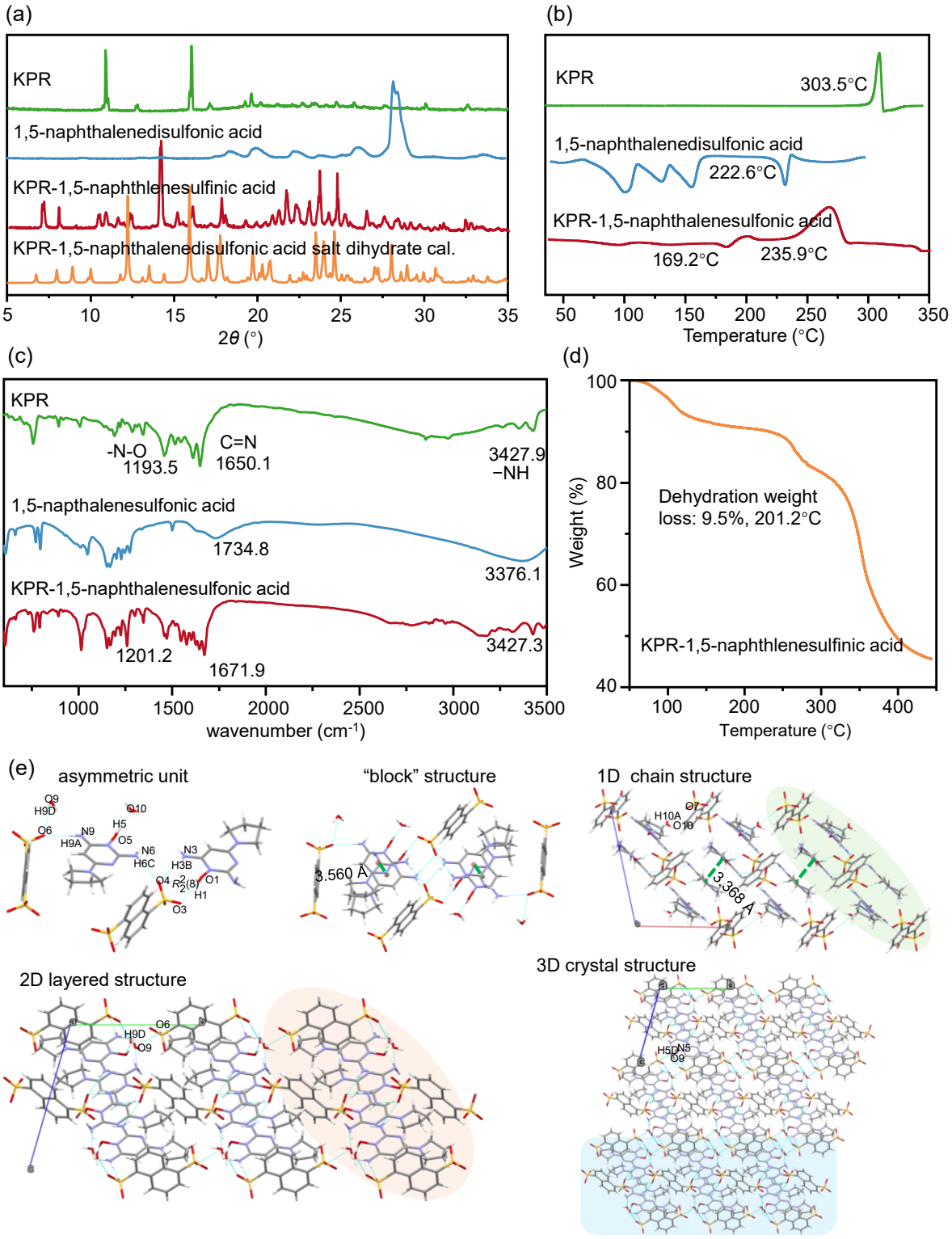
Supplementary Figure 39. Solid-state characterization of starting drugs and KPX-o-toluic acid hydrate (estimated). (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra and (d) TGA.

7.32 KPR-1,5-naphthalenedisulfonic acid salt dihydrate

KPR-1,5-naphthalenedisulfonic acid salt dihydrate crystallizes in the triclinic system with space group $P\bar{1}$ (Z = 2). The asymmetric unit contains two protonated KPR cations, two deprotonated 1,5-naphthalenedisulfonic acid anions, and two water molecules, with proton transfer occurring from the sulfonic acid groups to the oxygen atoms of KPR's pyrimidine N-oxide groups. In the crystal structure, the first 1,5-naphthalenedisulfonic acid anion connects to two KPR cations through an $R_2^2(8)$ motif formed by N3-H3B···O4 and O1-H1···O3 hydrogen bonds, with additional stabilization from N6-H6C···O4 hydrogen bond. Simultaneously, the second KPR cation interacts with the other sulfonic acid anion via N9-H9A···O6 hydrogen bond and with a water molecule through O5-H5···O10 hydrogen bond, while the second sulfonic acid anion links to the remaining water molecule via O9-H9D···O6 hydrogen bond. The asymmetric units combine through centrosymmetric operations, forming "block" structures stabilized by π-π stacking interactions (Cg-Cg distance: 3.560 Å, dihedral angle: 5.34°). These blocks extend along the a-axis through O10-H10A···O7 hydrogen bonds and additional parallel π-π stacking (Cg-Cg distance: 3.368 Å), generating one-dimensional chain structures. Subsequent organization along the b-axis occurs through O9-H9D···O6 hydrogen bonds, producing two-dimensional layered architecture. The three-dimensional crystal structure is ultimately achieved through interlayer N5-H5D···O9 hydrogen bonds.

Supplementary Table 32. Hydrogen-bond geometry (Å) of KPR-1,5-naphthalenedisulfonic acid salt dihydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N3–H3B···O4	0.880	2.187	2.917	139.97
O1–H1···O3	0.840	1.799	2.614	162.86
N5–H5D···O9	0.909	2.194	2.931	137.59
N6–H6C···O4	0.880	2.033	2.871	158.74
N9–H9A···O6	0.879	2.016	2.894	176.45
O5–H5···O10	0.841	1.742	2.575	170.39
O10–H10A···O7	0.871	2.405	2.921	119.34
O9–H9D···O6	0.870	2.252	2.337	124.54
O9–H9C···O7	0.870	2.209	2.870	132.59



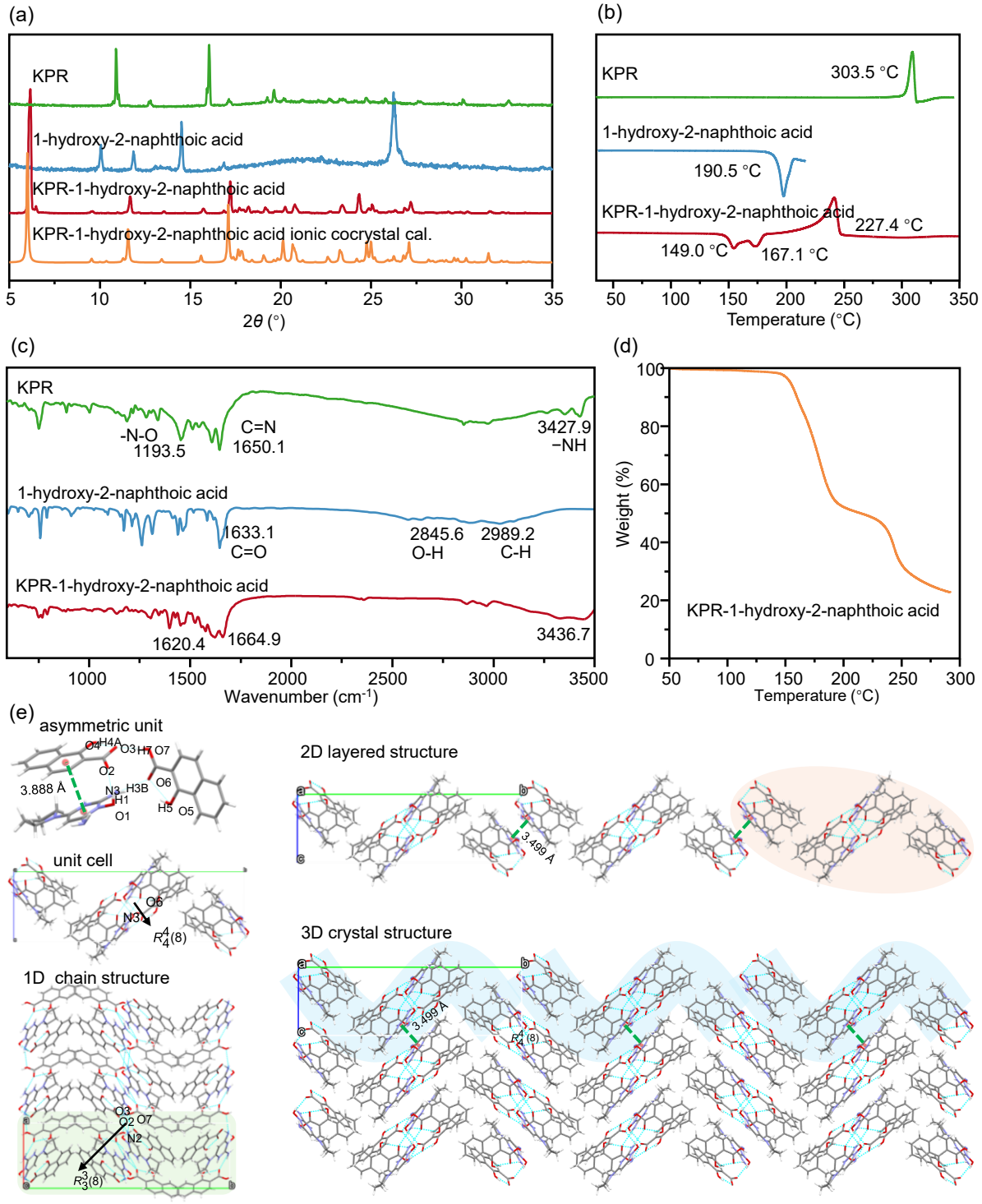
Supplementary Figure 40. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-1,5-naphthalenedisulfonic acid salt dihydrate.

7.33 KPR-1-hydroxy-2-naphthalic acid ionic cocrystal

KPR-1-hydroxy-2-naphthalic acid ionic cocrystal crystallizes in the monoclinic system with space group $P2_1/c$ ($Z = 4$), exhibiting disorder in the piperidine ring of KPR, where the major occupancy component was used for structural determination. The asymmetric unit forms a trimeric assembly comprising one protonated KPR cation, one deprotonated 1-hydroxy-2-naphthalic acid anion, and one neutral 1-hydroxy-2-naphthalic acid molecule, with proton transfer occurring from the carboxylic acid group to the pyrimidine N-oxide oxygen. In this structure, the neutral 1-hydroxy-2-naphthalic acid serves as a bridge between ionic components, forming N3-H3B···O6 and O7-H7···O3 hydrogen bonds with the KPR cation and anion respectively. Additional stabilization comes from an O1-H1···O2 hydrogen bond and π-π stacking interaction (Cg–Cg distance: 3.629 Å, dihedral angle: 9.67°) between the cation and anion. Through centrosymmetric operations, these trimers organize into $R_4^4(8)$ motifs via N3-H3B···O6 and N3-H3A···O6 hydrogen bonds, with four such units constituting the complete unit cell. The crystal architecture develops through successive dimensional organization. Along the *a*-axis, unit cells interconnect via $R_3^3(8)$ motifs formed by N2-H2B···O7, N2-H2A···O2, and O7-H7···O3 hydrogen bonds to create 1D chains. These chains then assemble along the *b*-axis through parallel π-π stacking interactions (Cg–Cg distance: 3.499 Å) to form 2D layers. The 3D crystal structure emerges through additional π-π stacking (Cg–Cg distance: 3.499 Å, parallel) and $R_4^4(8)$ motifs formed by N3-H3B···O6 and N3-H3A···O6 hydrogen bonds between adjacent layers.

Supplementary Table 33. Hydrogen-bond geometry (Å) of KPR-1-hydroxy-2-naphthalic acid ionic cocrystal.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
O1–H1···O2	0.841	1.834	2.652	164.00
N2–H2A··O2	0.884	2.143	2.902	143.63
N3–H3A···O6	0.881	2.226	3.084	164.54
N3–H3B···O6	0.880	2.133	2.979	161.13
N2–H2B··O7	0.884	2.541	3.110	122.73
O4–H4A···O3	0.840	1.830	2.568	145.83
O7–H7··O3	0.839	1.681	2.519	176.41
O5–H5··O6	0.839	1.826	2.583	145.72



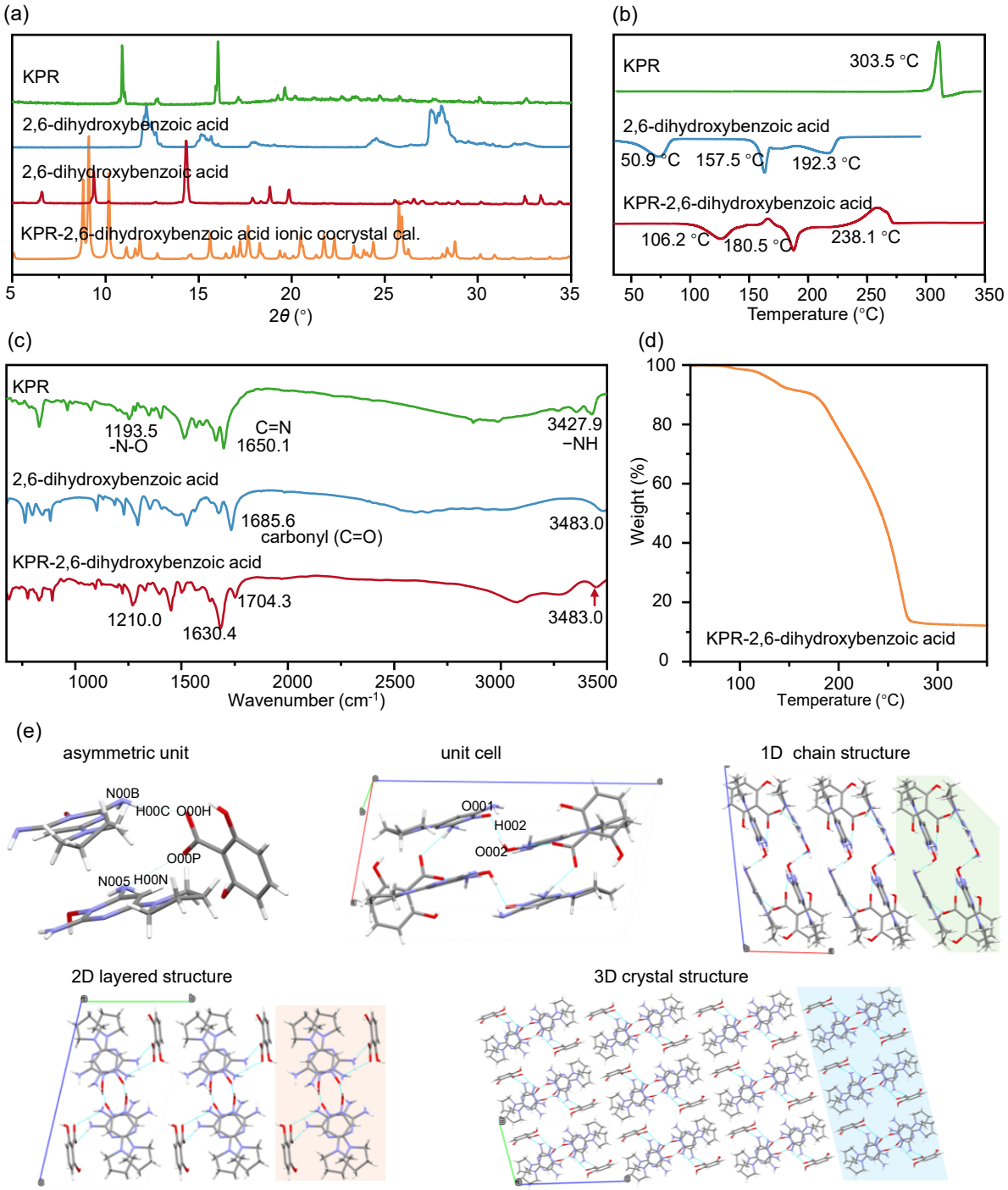
Supplementary Figure 41. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-1-hydroxy-2-naphthalic acid ionic cocrystal.

7.34 KPR-2,6-dihydroxybenzoic acid ionic cocrystal

The KPR-2,6-dihydroxybenzoic acid ionic cocrystal crystallizes in the triclinic system with space group $P\bar{1}$ ($Z = 2$). Both the piperidine ring of KPR and 2,6-dihydroxybenzoic acid exhibit disorder, with the major occupancy component used for structural analysis. The asymmetric unit comprises one protonated KPR cation, one deprotonated 2,6-dihydroxybenzoic acid anion, and one neutral KPR molecule. Proton transfer occurs from the carboxylic acid group of 2,6-dihydroxybenzoic acid to the oxygen atom of the pyrimidine N-oxide in KPR. The 2,6-dihydroxybenzoate anion forms hydrogen bonds with the KPR cation (N005–H00N···O00P) and the neutral KPR (N00B–H00C···O00H). Centrosymmetric operation links the asymmetric units via an O002–H002···O001 hydrogen bond, generating the unit cell. These unit cells then interconnect along the *a*-axis through van der Waals interactions, forming a 1D structure. The 1D structures further pack along the *b*-axis via van der Waals forces, resulting in a 2D layered arrangement. Finally, identically oriented 2D layers stack along the *c*-axis through additional van der Waals interactions, completing the 3D crystal architecture.

Supplementary Table 34. Hydrogen-bond geometry (Å) of KPR-2,6-dihydroxybenzoic acid ionic cocrystal.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N005–H00N···O00P	0.863	2.472	2.948	115.52
N00B–H00C···O00H	0.863	2.042	2.895	169.20
O002–H002···O001	0.820	1.639	2.451	170.06



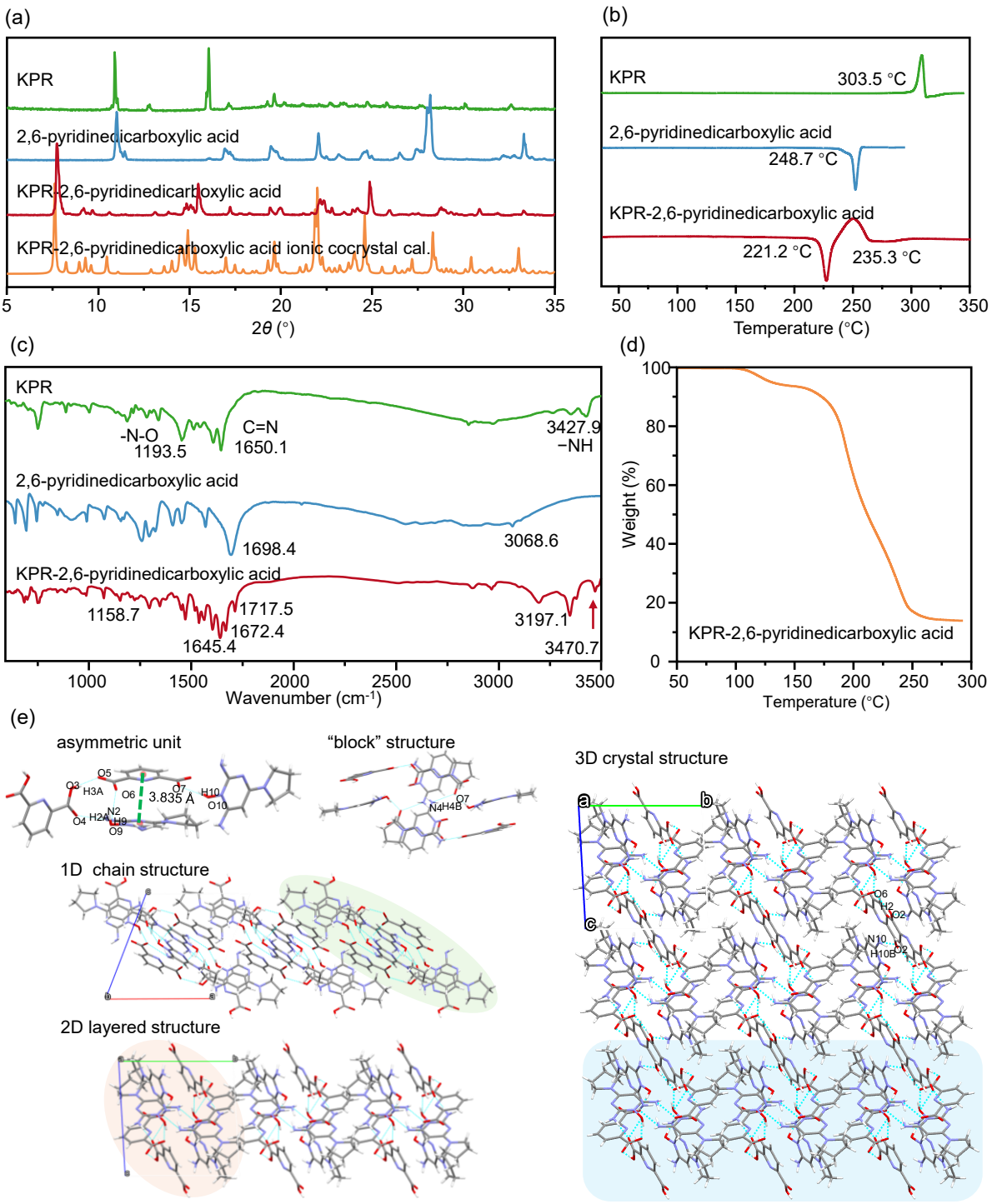
Supplementary Figure 42. Solid-state characterization of starting drugs and KPR-2,6-dihydroxybenzoic acid ionic cocrystal. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra and (d) TGA.

7.35 KPR-2,6-pyridinedicarboxylic acid ionic cocrystal

KPR-2,6-pyridinedicarboxylic acid ionic cocrystal crystallizes in the triclinic system with space group $P\bar{1}$ ($Z = 2$). The structure features disorder in the piperidine ring of KPR, with the major occupancy component being used for structural analysis. The asymmetric unit consists of two protonated KPR cations, one fully deprotonated 2,6-pyridinedicarboxylic acid anion, and one neutral 2,6-pyridinedicarboxylic acid molecule, with proton transfer occurring from the carboxylic acid groups to the oxygen atoms of the pyrimidine N-oxide moieties in KPR. In the crystal packing, the deprotonated anion forms hydrogen bonds with one KPR cation (O9-H9···O6) and the neutral acid molecule (O3-H3A···O5), while also engaging in π-π stacking interactions with the KPR cation (interplanar distance: 3.835 Å, dihedral angle: 9.02°). The same KPR cation additionally connects to the neutral acid molecule through an N2-H2A···O4 hydrogen bond. Centrosymmetric operation of the asymmetric unit generates a "block" structure through N4-H4B···O7 hydrogen bonds. These structural blocks assemble along the *a*-axis through van der Waals interactions to form 1D chains. Subsequent organization along the *b*-axis, again mediated by van der Waals forces, produces 2D layers. The complete 3D architecture is achieved through interlayer O2-H2···O6 and N10-H10B···O2 hydrogen bonds.

Supplementary Table 35. Hydrogen-bond geometry (Å) of KPR-2,6-pyridinedicarboxylic acid ionic cocrystal.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
O10–H10···O7	0.819	1.661	2.461	164.96
O3–H3A···O5	0.820	1.728	2.540	170.06
O2–H2···O6	0.820	1.824	2.640	172.94
N2–H2A···O4	0.860	2.132	2.902	148.84
N10–H10B···O2	0.866	2.373	2.985	127.97
O9–H9···O6	0.819	1.980	2.706	147.00
N4–H4B···O7	0.865	2.297	2.955	132.90



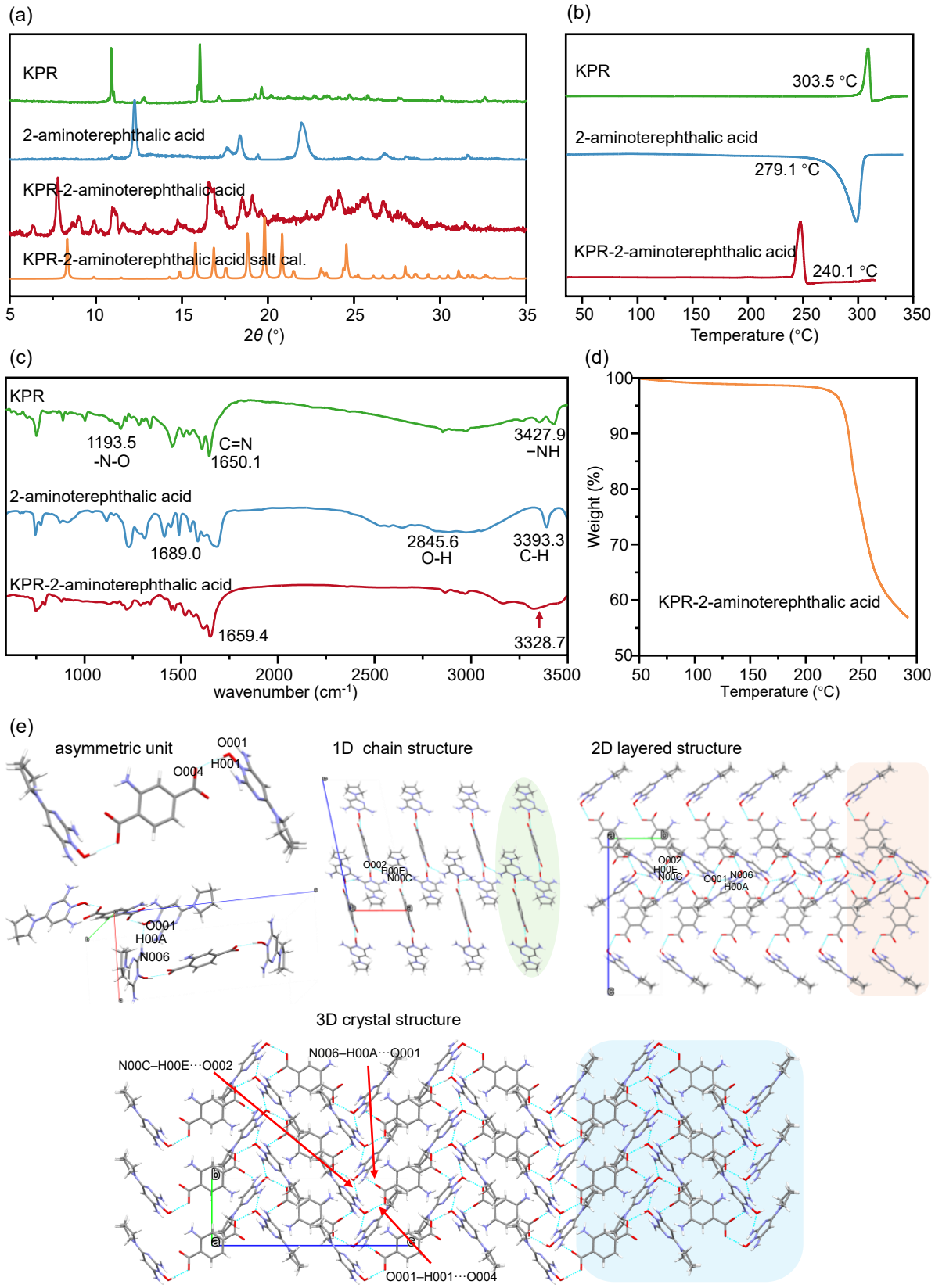
Supplementary Figure 43. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-2,6-pyridinedicarboxylic acid ionic cocrystal.

7.36 KPR-2-aminoterephthalic salt

KPR-2-aminoterephthalic acid salt crystallizes in the monoclinic system with space group $P2_1/n$ ($Z = 2$). The amino group of 2-aminoterephthalic acid exhibits disorder, with the major occupancy component used for crystal structure analysis. The asymmetric unit consists of two protonated KPR cations and one deprotonated 2-aminoterephthalic acid anion. Protons are transferred from the carboxylic acid groups of 2-aminoterephthalic acid to the oxygen atoms of the pyrimidine N-oxide in KPR. The 2-aminoterephthalic acid anion forms a O001–H001···O004 hydrogen bond with KPR cations. Additionally, the asymmetric unit connects to another asymmetric unit via N006–H00A···O001 hydrogen bonds. These interactions extend along the *a*-axis through N00C–H00E···O002 hydrogen bonds, forming a 1D chain structure. The 1D chains, aligned in the same orientation, pack along the *b*-axis, generating a 2D layered structure stabilized by hydrogen bonds (N00C–H00E···O002 and N006–H00A···O001). Finally, the 2D layers with opposite orientations stack along the *c*-axis through hydrogen bonds (N00C–H00E···O002, N006–H00A···O001, and O001–H001···O004), completing the 3D crystal structure.

Supplementary Table 36. Hydrogen-bond geometry (Å) of KPR-2-aminoterephthalic acid salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N00C–H00E···O002	0.859	2.007	2.881	155.52
N006–H00A···O001	0.860	2.039	2.888	167.79
O001–H001···O004	0.819	1.688	2.494	168.84



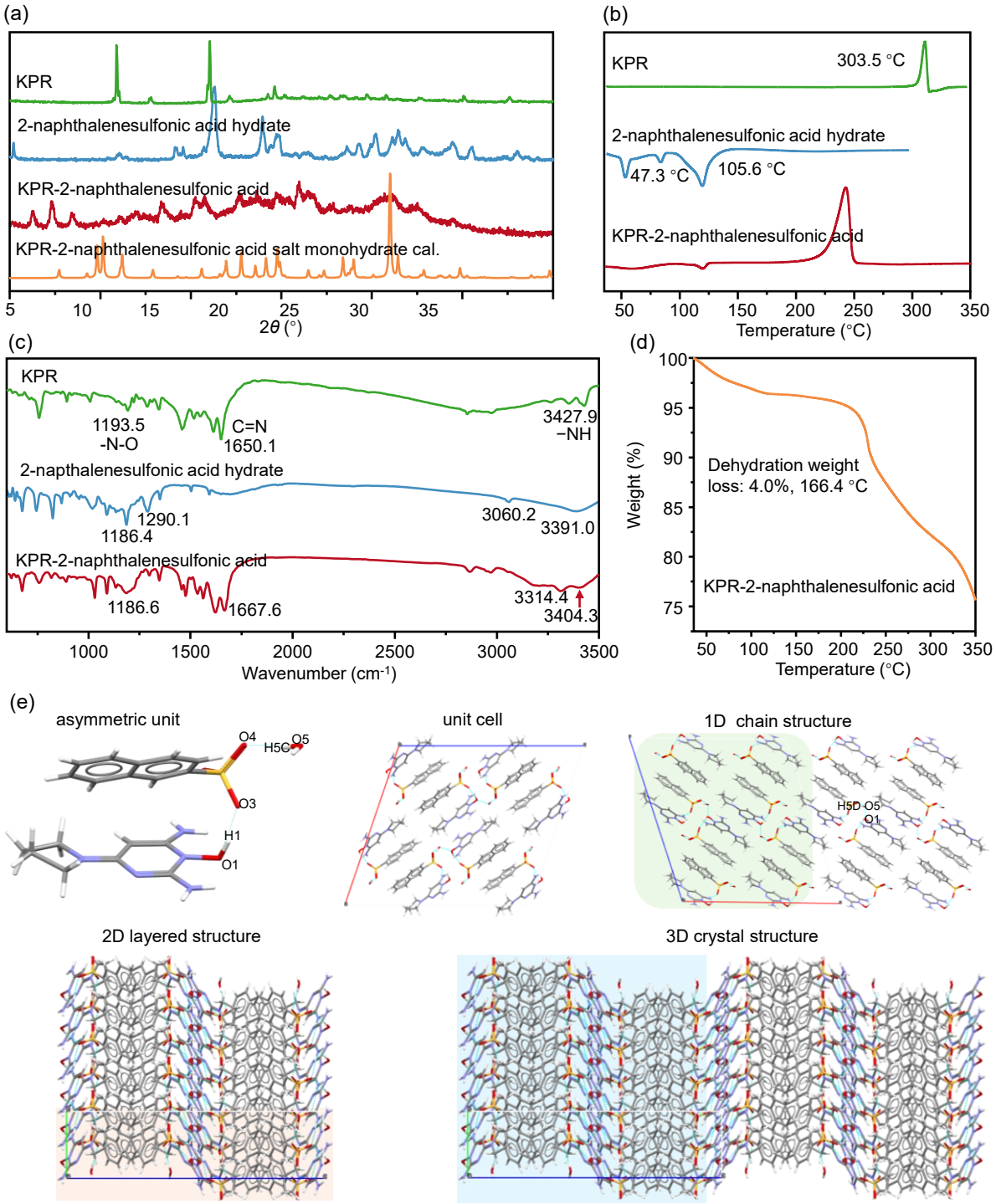
Supplementary Figure 44. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-2-aminoterephthalic acid salt.

7.37 KPR-2-naphthalenesulfonic acid salt monohydrate

KPR-2-naphthalenesulfonic acid salt monohydrate crystallizes in the monoclinic system with space group *C*2/*c* (*Z* = 8). The asymmetric unit contains one protonated KPR cation, one deprotonated 2-naphthalenesulfonic acid anions, and two water molecules, with proton transfer occurring from the sulfonic acid group to the oxygen atoms of KPR's pyrimidine N-oxide groups. In the crystal structure, the 2-naphthalenesulfonic acid anion connects to the KPR cation and water molecule through an O1-H1···O3 hydrogen bond and an O5-H5C···O4 hydrogen bond, respectively. Eight asymmetric units form a unit cell. The unit cells extend along the a-axis through van der Waals interactions and O5-H5D···O1 hydrogen bonds, generating 1D chain structures. Subsequent organization along the *b*-axis occurs through N5-H5B···O3 and O1-H1···O3 hydrogen bonds, producing 2D layered architecture. The 3D crystal structure is ultimately achieved through interlayer O5-H5D···O1 hydrogen bonds.

Supplementary Table 37. Hydrogen-bond geometry (Å) of KPR-2-naphthalenesulfonic acid salt monohydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
O1-H1···O3	0.840	1.798	2.624	167.27
O5-H5C···O4	0.869	2.005	2.870	173.71
O5-H5D···O1	0.811	2.588	2.837	97.55
N5-H5B···O3	0.880	2.063	2.880	154.00



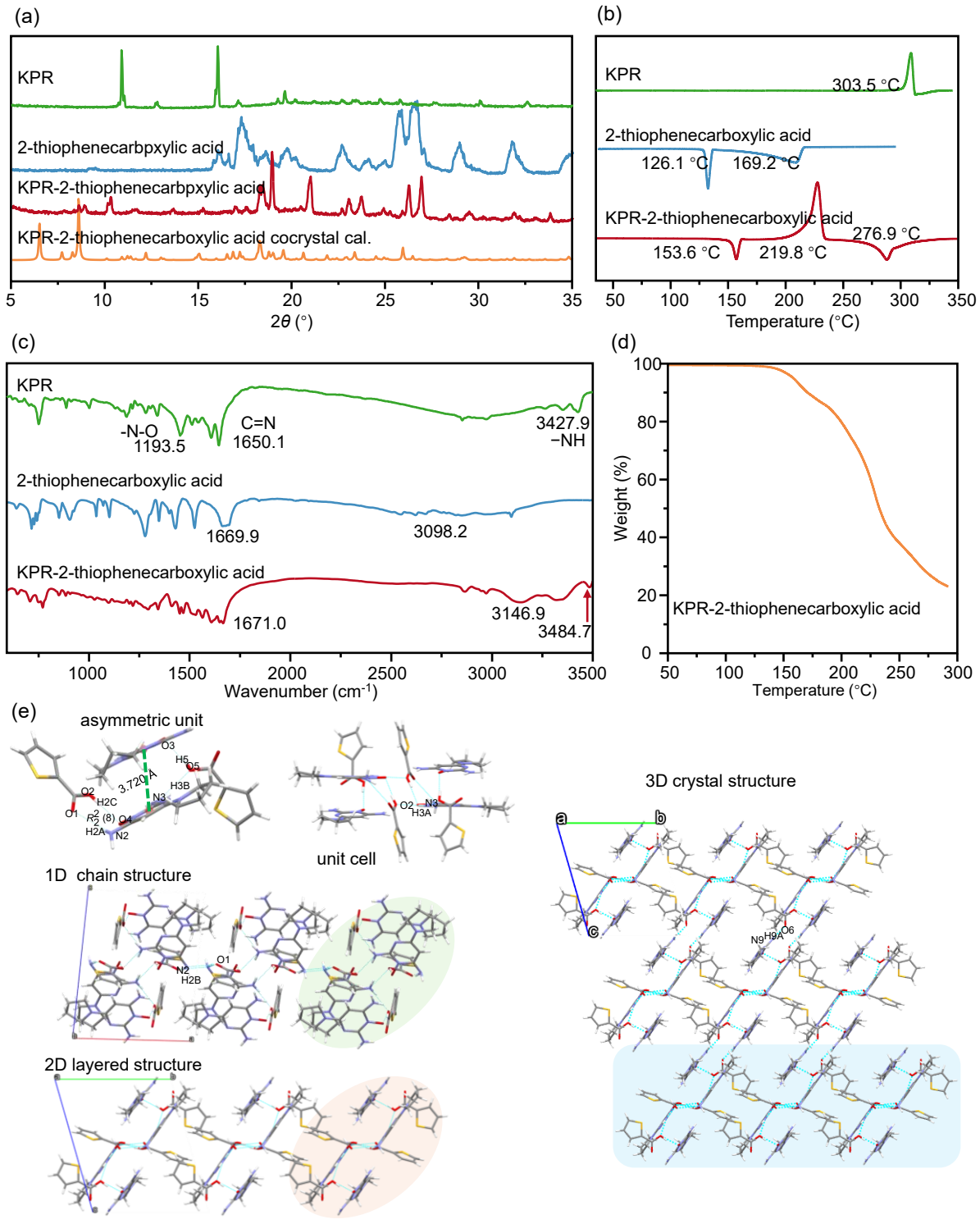
Supplementary Figure 45. (a) PXRD patterns and (b) DSC curves of starting drugs and KPR-2-naphthalenesulfonic acid (negative).

7.38 KPR-2-thiophenecarboxylic acid cocrystal

KPR-2-thiophenecarboxylic acid cocrystal crystallizes in the triclinic system with space group $P\bar{1}$ ($Z = 2$) The piperidine ring of KPR and the thiophene ring of 2-thiophenecarboxylic acid exhibit disorder, with the major occupancy component used for structural analysis. The asymmetric unit comprises two KPR molecules and two 2-thiophenecarboxylic acid molecules. One KPR molecule is linked to two 2-thiophenecarboxylic acid molecules via the N3–H3B···O5 hydrogen bond, as well as an $R_2^2(8)$ motif formed by O2–H2C···O4 and N2–H2A···O1 hydrogen bonds. The other KPR molecule interacts with the first KPR and one 2-thiophenecarboxylic acid molecule through π···π stacking (Cg–Cg distance: 3.720 Å, dihedral angle: 9.18°) and O5–H5···O3 hydrogen bond, respectively. The asymmetric unit and its centrosymmetrically related counterpart are connected via the N3–H3A···O2 hydrogen bond, forming the unit cell. These unit cells further interact through N2–H2B···O1 hydrogen bonds, extending along the *a*-axis to generate a 1D chain structure. The 1D chains, aligned in the same orientation, pack along the *b*-axis through van der Waals interactions, resulting in a 2D layered structure. Finally, these 2D layers are interconnected along the *c*-axis via N9–H9A···O6 hydrogen bonds, completing the 3D crystal architecture.

Supplementary Table 38. Hydrogen-bond geometry (Å) of KPR-2-thiophenecarboxylic acid cocrystal.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N2–H2A···O1	0.861	2.143	2.920	149.97
O2–H2C···O4	0.820	1.711	2.524	171.15
N3–H3A···O2	0.880	2.080	2.909	156.71
N9–H9A···O6	0.880	2.013	2.878	166.92
O5–H5···O3	0.821	1.666	2.469	165.71
N3–H3B···O5	0.879	1.995	2.842	161.21
N2–H2B···O1	0.862	2.738	2.840	74.49



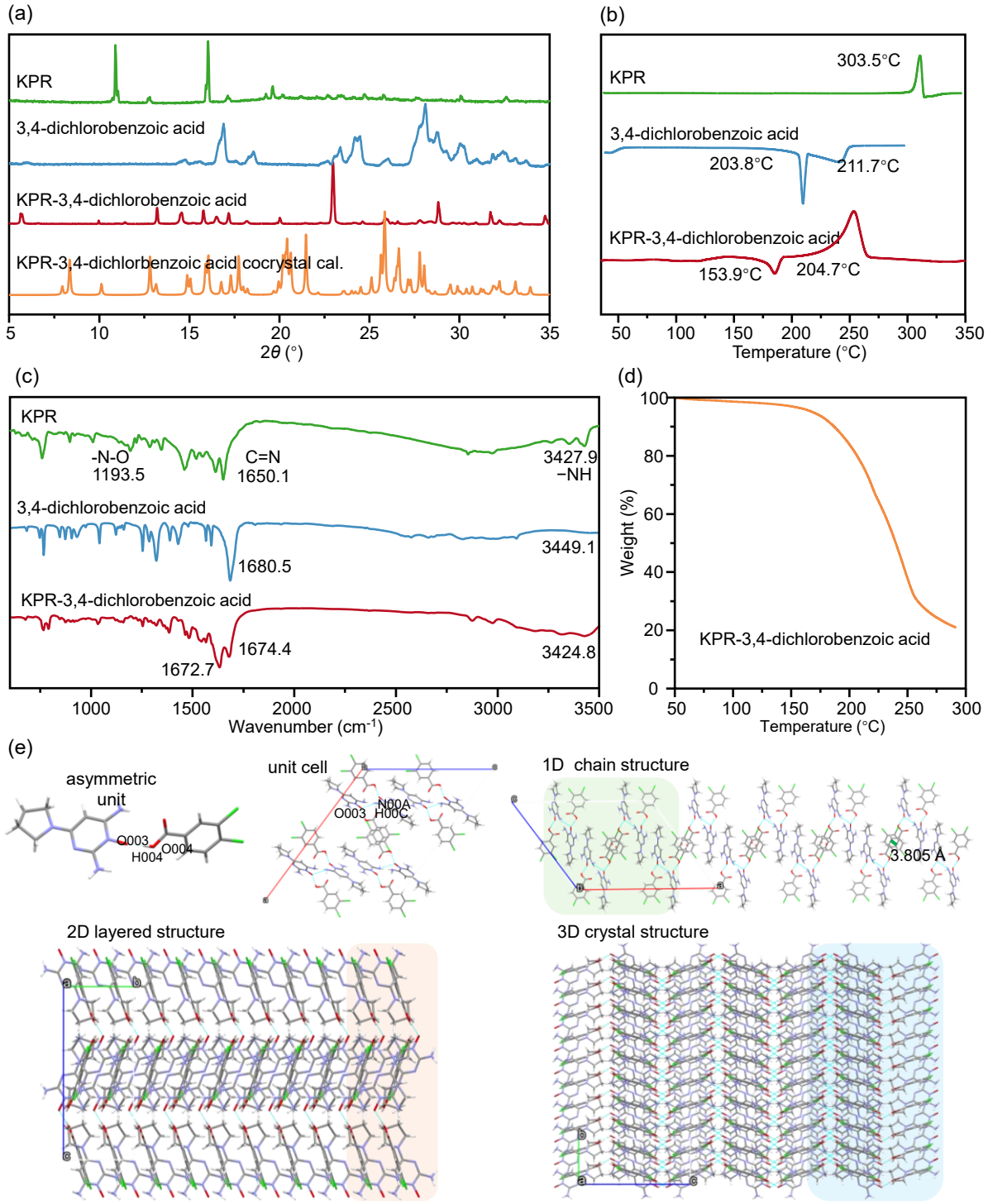
Supplementary Figure 46. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-2-thiophenecarboxylic acid cocrystal.

7.39 KPR-3,4-dichlorobenzoic acid cocrystal

KPR-3,4-dichlorobenzoic acid cocrystal crystallizes in the monoclinic system with the space group *C*2/*c* (*Z* = 8). The asymmetric unit consists of one KPR molecule and one 3,4-dichlorobenzoic acid molecule, linked via the O004–H004···O003 hydrogen bond. The asymmetric unit and its symmetry-equivalent counterpart are connected through the N00A–H00C···O003 hydrogen bond, forming the unit cell. Eight asymmetric units assemble into the unit cell. Unit cells of the same orientation interact via π···π stacking (Cg–Cg distance: 3.805 Å, parallel), extending along the *a*-axis to generate a 1D chain structure. These 1D chains pack along the *b*-axis through van der Waals interactions, resulting in a 2D layered structure. Finally, the 2D layers with opposite orientations are interconnected along the *c*-axis via O004–H004···O003 and N00A–H00C···O003 hydrogen bonds, completing the 3D crystal architecture.

Supplementary Table 39. Hydrogen-bond geometry (Å) of KPR-3,4-dichlorobenzoic acid cocrystal.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
O004–H004···O003	0.820	1.647	2.465	175.42
N00A–H00C···O003	0.863	2.465	2.919	113.58



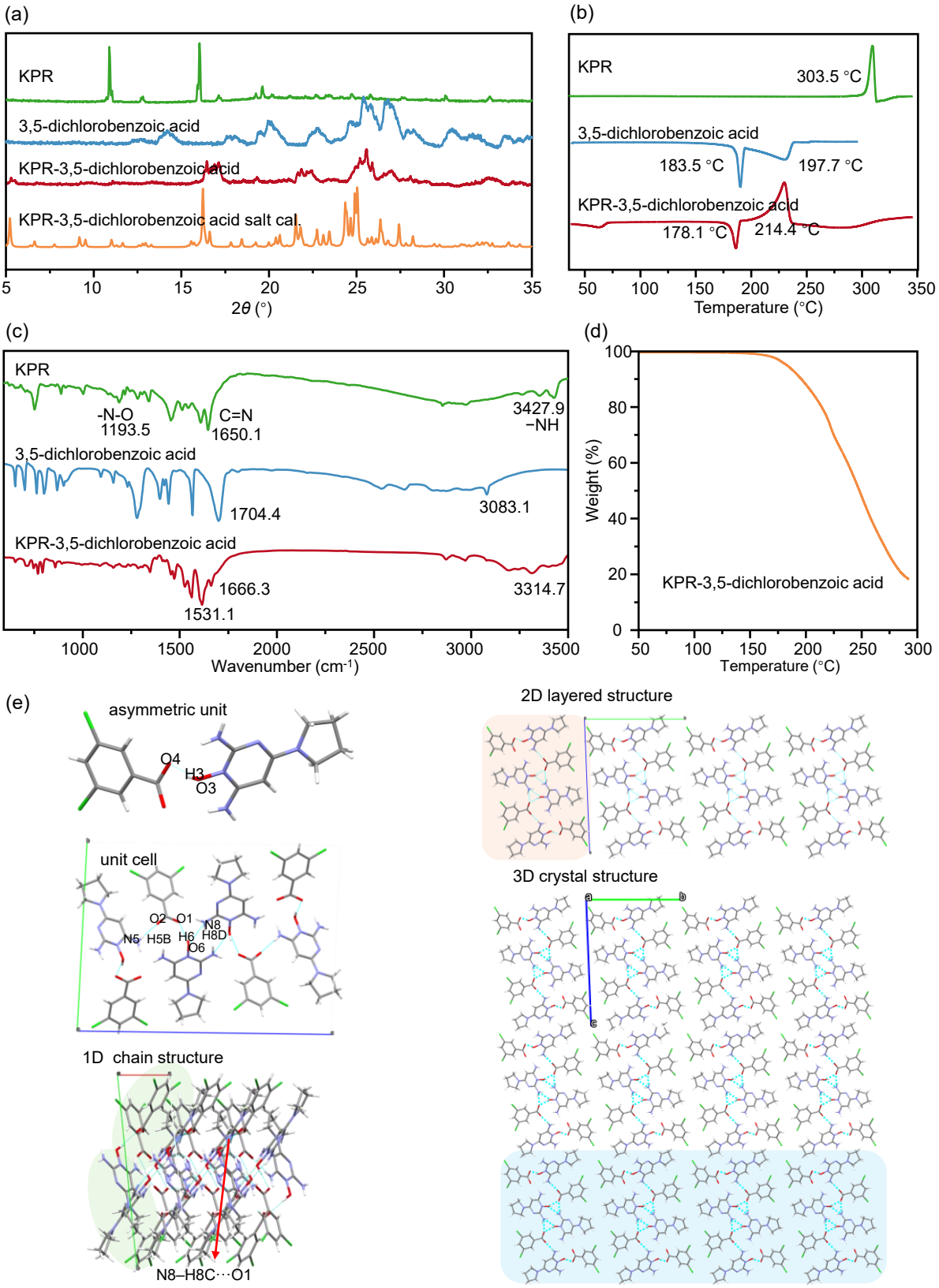
Supplementary Figure 47. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-3,4-dichlorobenzoic acid cocrystal.

7.40 KPR-3,5-dichlorobenzoic acid salt

KPR–3,5-dichlorobenzoic acid salt crystallizes in the triclinic system with the space group $P\bar{1}$ ($Z = 4$). The piperidine ring of KPR exhibits disorder, with the major occupancy component used for structural analysis. The asymmetric unit consists of one protonated KPR cation and one deprotonated 3,5-dichlorobenzoic acid anion. Proton transfer occurs from the carboxylic acid group of 3,5-dichlorobenzoic acid to the O atom of the pyrimidine N-oxide in KPR. The KPR cation and 3,5-dichlorobenzoic acid anion are linked via O3–H3···O4 or O6–H6···O1 hydrogen bonds. The asymmetric unit and its symmetry-related counterpart (generated by a 4_2 screw axis operation) are connected through N5–H5B···O2 and N8–H8D···O6 hydrogen bonds, forming the unit cell. Unit cells of the same orientation interact via hydrogen bond N8–H8C···O1, extending along the *a*-axis to generate a 1D chain structure. These chains pack along the *b*-axis through van der Waals interactions, resulting in a 2D layered structure. Finally, the 2D layers with opposite orientations assemble along the *c*-axis via van der Waals forces, completing the 3D crystal architecture.

Supplementary Table 40. Hydrogen-bond geometry (Å) of KPR-3,5-dichlorobenzoic acid salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N5–H5B···O2	0.860	2.034	2.891	174.70
O6–H6···O1	0.821	1.660	2.455	162.33
N8–H8C···O1	0.860	2.117	2.888	148.92
O3–H3···O4	0.820	1.646	2.449	166.11
N8–H8D···O6	0.861	2.292	2.993	138.68



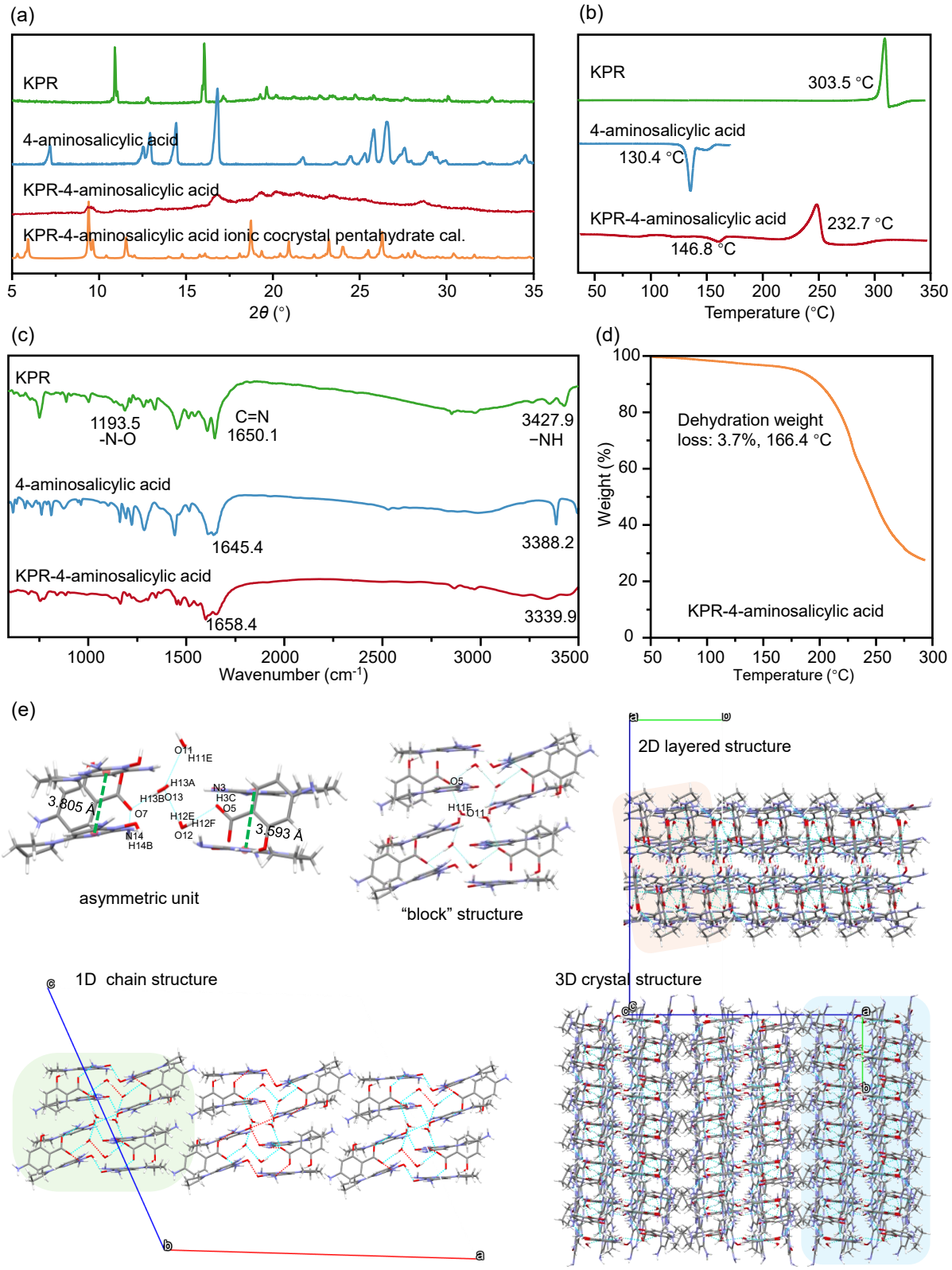
Supplementary Figure 48. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-3,5-dichlorobenzoic acid salt.

7.41 KPR-4-aminosalicylic acid ionic cocrystal pentahydrate

KPR-4-aminosalicylic acid ionic cocrystal trihydrate crystallizes in the monoclinic system with space group *C*2/*c* (*Z* = 4). The asymmetric unit contains two neutral KPX molecules, two protonated KPX cations, two deprotonated 4-aminosalicylic acid anions, and three water molecules. Protons transfer from the carboxylic acid group of 4-aminosalicylic acid to the oxygen atom of the pyrimidine N-oxide in KPX. Two neutral KPX molecules are linked by π···π stacking (Cg–Cg distance: 3.805 Å, dihedral angle: 3.36°). Similarly, two protonated KPX cations exhibit π···π stacking (Cg–Cg distance: 3.593 Å, dihedral angle: 3.22°). One 4-aminosalicylic acid anion connects to a KPX cation via N14–H14B···O7 and to a water molecule via O13–H13B···O7, while the other anion interacts with a KPX molecule through N3–H3C···O5 and with a water molecule via O12–H12F···O5. The water molecules form a network through O12–H12E···O13 and O13–H13A···O11 hydrogen bonds. Adjacent asymmetric units associate via O11–H11F···O5, forming a "block" structure. These blocks, oriented in the same direction, extend along the *a*-axis through van der Waals interactions, generating a 1D chain. The chains then pack along the *b*-axis, forming a 2D layered structure stabilized by O9–H9C···O1, N2–H2D···O5, and N7–H7A···O4 hydrogen bonds. Finally, the 3D crystal structure is completed by the antiparallel arrangement of these layers, linked by O10–H10···O10 and N20–H20B···O8 interactions along the *c*-axis.

Supplementary Table 41. Hydrogen-bond geometry (Å) of KPR-4-aminosalicylic acid ionic cocrystal pentahydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N7–H7A···O4	0.883	2.903	3.027	73.65
O9–H9C···O1	0.841	1.602	2.423	164.51
N3–H3C···O5	0.879	2.129	2.933	151.63
N2–H2D···O5	0.885	2.278	2.947	132.24
O11–H11F···O5	0.869	2.140	2.971	169.87
O12–H12F···O5	0.871	2.050	2.917	173.24
N14–H14B···O7	0.883	2.712	2.964	97.76
O10–H10···O10	0.839	2.753	2.984	97.67
N20–H20B···O8	0.880	2.272	2.850	123.20
O13–H13B···O7	0.869	2.034	2.869	160.84
O13–H13A···O11	0.870	2.618	2.872	98.03
O12–H12E···O13	0.871	2.024	2.869	163.33



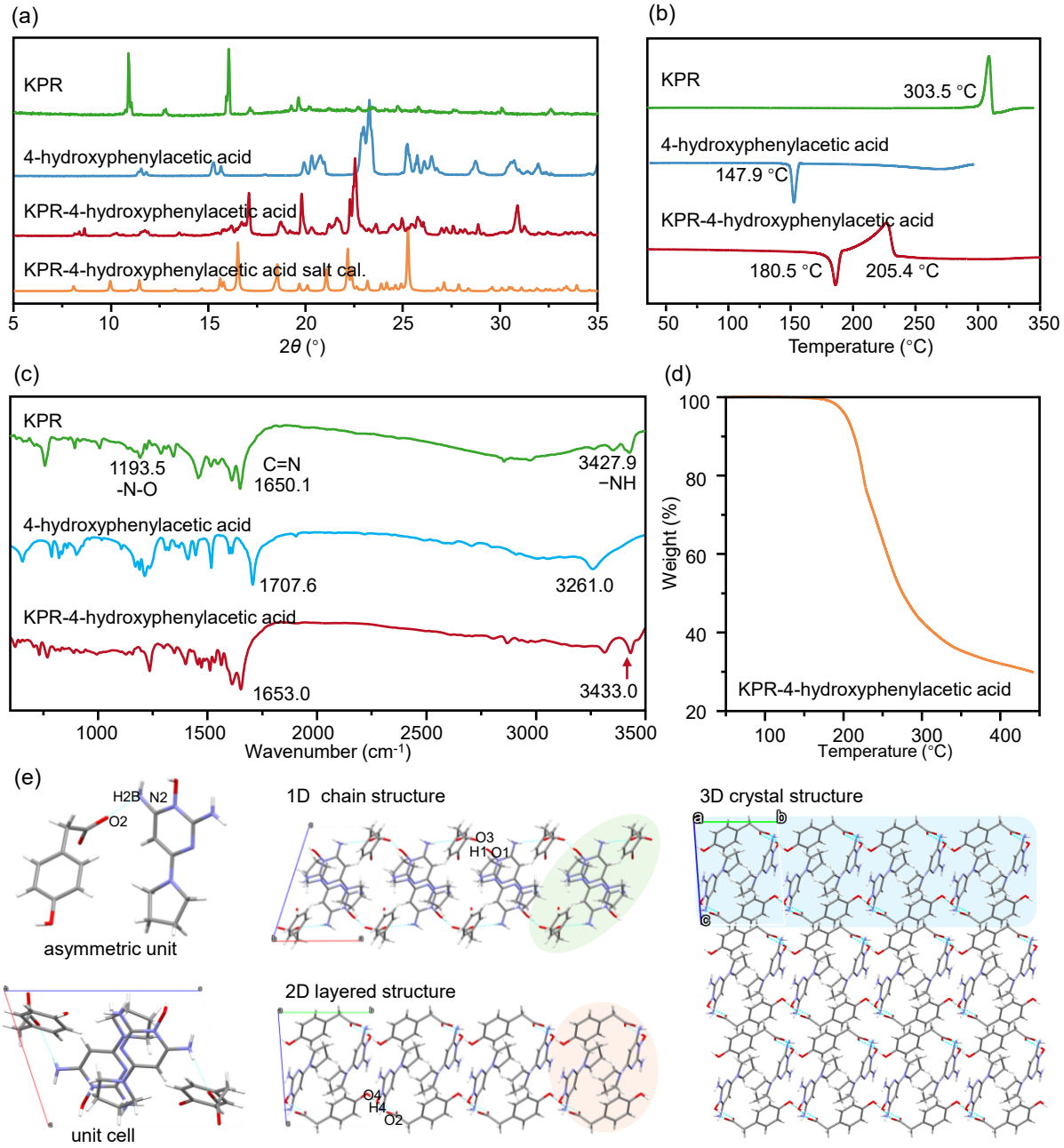
Supplementary Figure 49. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-4-aminosalicylic acid ionic cocrystal pentahydrate.

7.42 KPR-4-hydroxyphenylacetic acid salt

KPR-4-hydroxyphenylacetic acid salt crystallizes in the triclinic system with space group $P\bar{1}$ ($Z = 2$). The asymmetric unit consists of one KPR cation and one 4-hydroxyphenylacetic acid anion. The KPR cation forms a N2–H2B···O2 hydrogen bond with the anion, and proton transfer occurs from the carboxylic acid group of 4-hydroxyphenylacetic acid to the oxygen atom of the pyrimidine N-oxide in KPR. Adjacent asymmetric units, related by centrosymmetry, associate via van der Waals interactions, forming the unit cell. These unit cells further interconnect through O1–H1···O3 hydrogen bonds, extending along the *a*-axis to generate a 1D chain structure. Identically oriented chains pack along the *b*-axis, stabilized by O4–H4···O2 hydrogen bonds, to yield a 2D layered architecture. Finally, these 2D layers stack uniformly along the *c*-axis via van der Waals forces, completing the 3D crystal structure.

Supplementary Table 42. Hydrogen-bond geometry (Å) of KPR-4-hydroxyphenylacetic acid salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
O1–H1···O3	0.840	1.646	2.470	165.90
N2–H2B···O2	0.880	2.523	2.918	108.01
O4–H4···O2	0.840	1.823	2.656	170.74



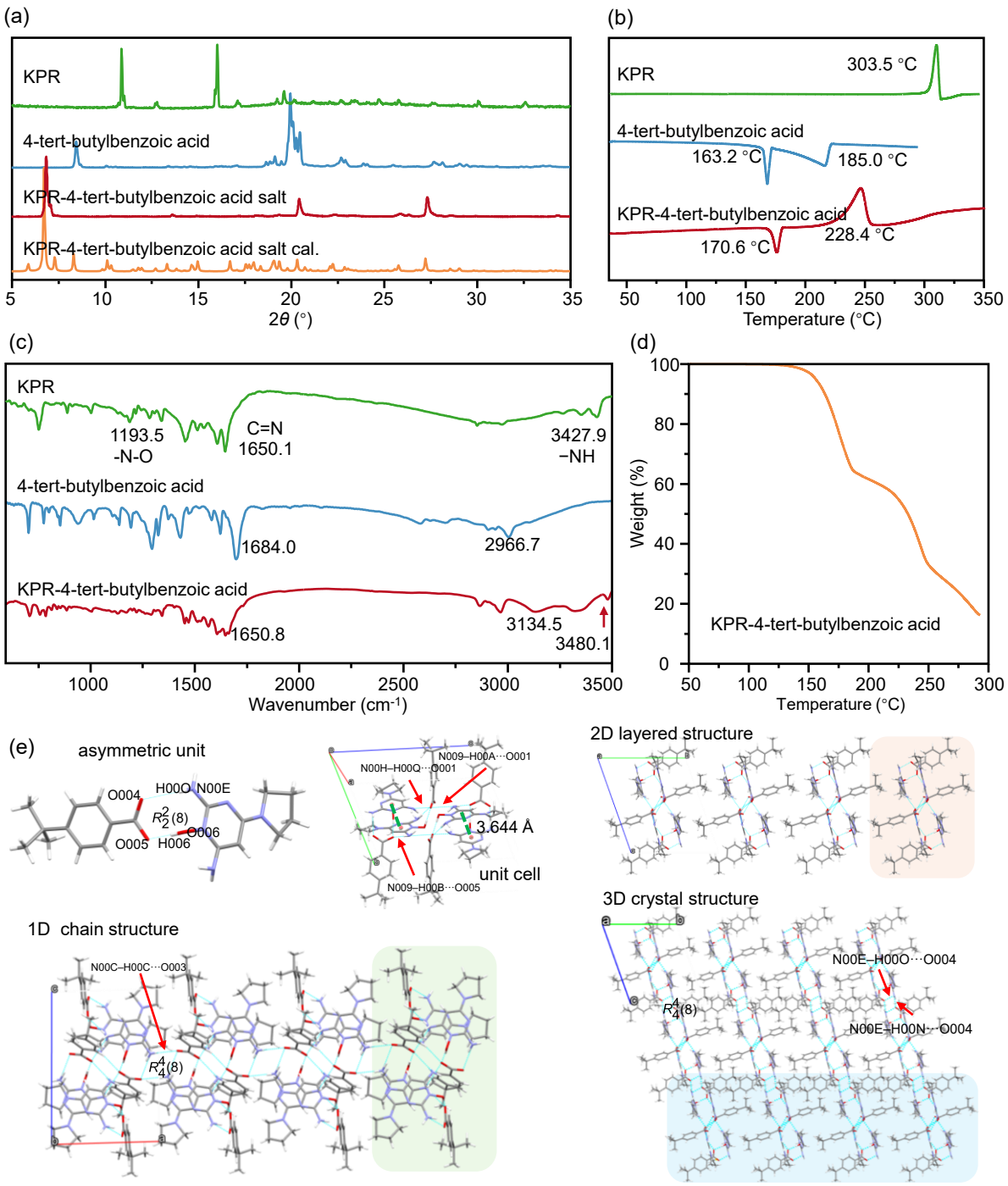
Supplementary Figure 50. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-4-hydroxyphenylacetic acid salt.

7.43 KPR-4-tert-butylbenzoic acid salt

KPR-4-tert-butylbenzoic acid salt crystallizes in the triclinic system with space group $P\bar{1}$ (Z = 4). Disorder is observed in both the piperidine ring of KPR and the tert-butyl group of the coformer, with the major occupancy component being used for structural analysis. The asymmetric unit consists of one protonated KPR cation and one deprotonated 4-tert-butylbenzoic acid anion, which are connected through a cyclic $R_2^2(8)$ motif formed by N00E–H00O···O004 and O006–H006···O005 hydrogen bonds. Centrosymmetrically related asymmetric units are linked through N00H–H00Q···O001 and N009–H00B···O005 hydrogen bonds and π···π stacking (Cg–Cg distance: 3.644 Å, dihedral angle: 8.43°). These units further associate via N009–H00A···O001 hydrogen bond to form the unit cell. The crystal structure develops through hierarchical interactions. Along the *a*-axis, unit cells interact through $R_4^4(8)$ motifs formed by N00C–H00C···O003 hydrogen bonds, generating 1D chains. These chains then pack along the *b*-axis through van der Waals interactions to form 2D layers. Finally, the 3D framework is completed by the arrangement of these layers along the *c*-axis, mediated by additional $R_4^4(8)$ motifs involving N00E–H00O···O004 and N00E–H00N···O004 hydrogen bonds.

Supplementary Table 43. Hydrogen-bond geometry (Å) of KPR-4-tert-butylbenzoic acid salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N00C–H00C···O003	0.866	2.131	2.984	168.42
N009–H00A···O001	0.871	2.292	2.947	132.05
N00H–H00Q···O001	0.865	2.297	2.965	134.18
N009–H00B···O005	0.872	2.008	2.880	177.84
N00E–H00O···O004	0.860	2.271	2.907	130.92
O006–H006···O005	0.821	1.626	2.447	178.33



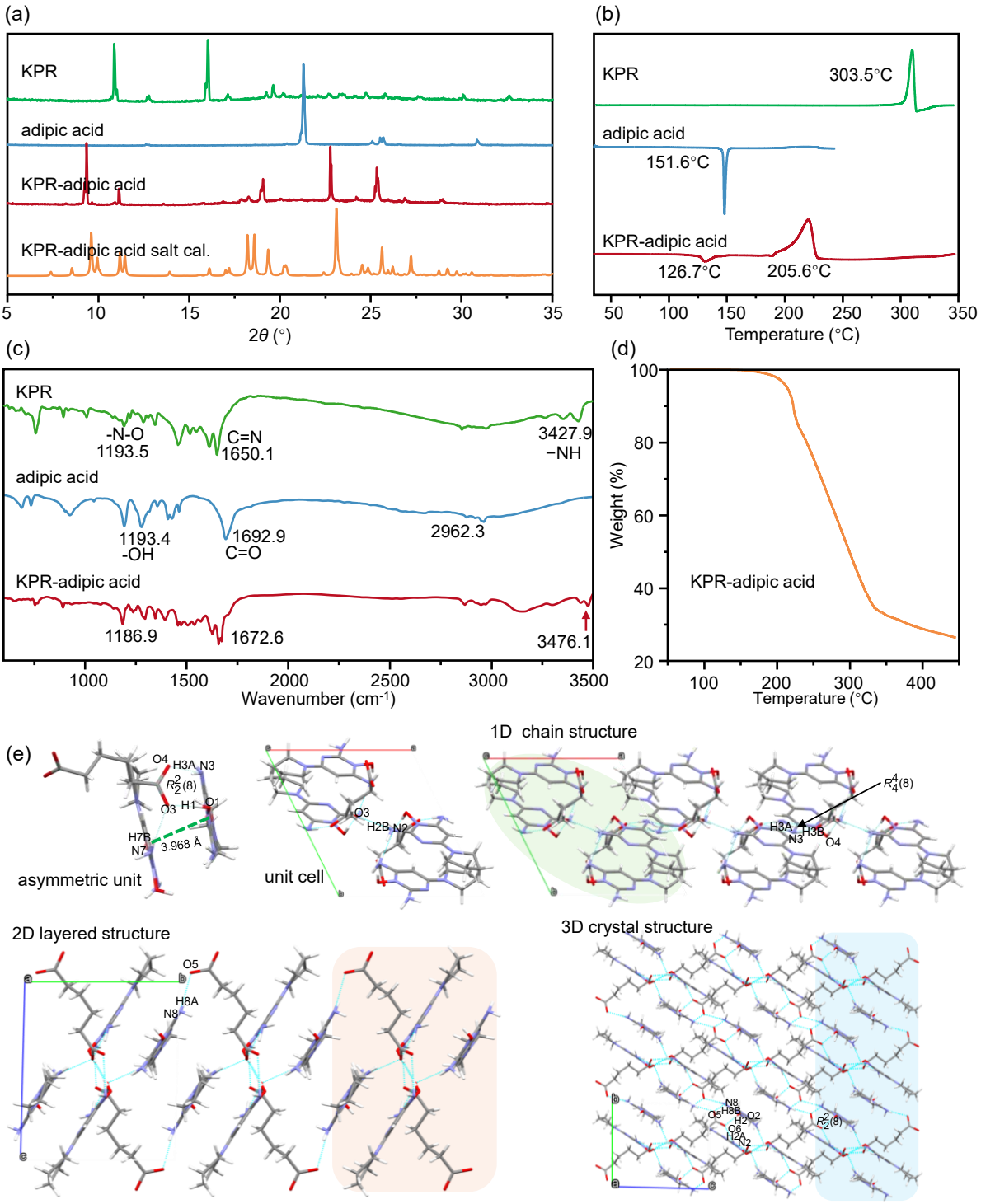
Supplementary Figure 51. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-4-tert-butylbenzoic acid salt.

7.44 KPR-adipic acid salt

KPR-adipic acid salt crystallizes in the triclinic system with space group $P\bar{1}$ (Z = 2). The piperidine ring of KPR exhibits disorder, and the major occupancy component was used for crystal structure analysis. The asymmetric unit consists of two protonated KPR cations and one deprotonated adipic acid anion. Protons are transferred from the carboxylic acid groups of adipic acid to the oxygen atoms of the pyrimidine N-oxide in KPR. The adipic acid anion connects to two KPR cations via the N7–H7B···O3 hydrogen bond, as well as an $R_2^2(8)$ motif formed by O1–H1···O3 and N3–H3A···O4 hydrogen bonds. The two KPR cations are further linked by π···π stacking interactions (Cg–Cg distance: 3.968 Å, dihedral angle: 4.73°). The asymmetric unit and an adjacent asymmetric unit are connected through the N2–H2B···O3 hydrogen bond, forming the unit cell. Unit cells with the same orientation pack along the *a*-axis via an $R_2^2(8)$ motif generated by N3–H3A···O4 and N3–H3B···O4 hydrogen bonds, resulting in a 1D structure. These 1D structures, aligned in the same orientation, further pack along the *b*-axis through N8–H8A···O5 hydrogen bonds, yielding a 2D layered structure. Finally, the 2D layered structures arrange into a 3D crystal framework through N2–H2A···O6 hydrogen bonds and $R_2^2(8)$ motifs formed by N8–H8B···O5 and O2–H2···O6 hydrogen bonds.

Supplementary Table 44. Hydrogen-bond geometry (Å) of KPR-adipic acid salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N3–H3A···O4	0.887	2.019	2.874	161.50
N3–H3B···O4	0.887	2.058	2.903	158.94
O1–H1···O3	0.841	1.650	2.489	175.39
N2–H2A···O6	0.882	2.046	2.906	164.74
N2–H2B···O3	0.881	2.305	2.983	133.80
N7–H7B···O3	0.880	2.090	2.953	166.77
N8–H8A···O5	0.880	2.153	2.965	153.01
N8–H8B···O5	0.880	2.877	2.988	74.26
O2–H2···O6	0.840	1.630	2.469	177.71



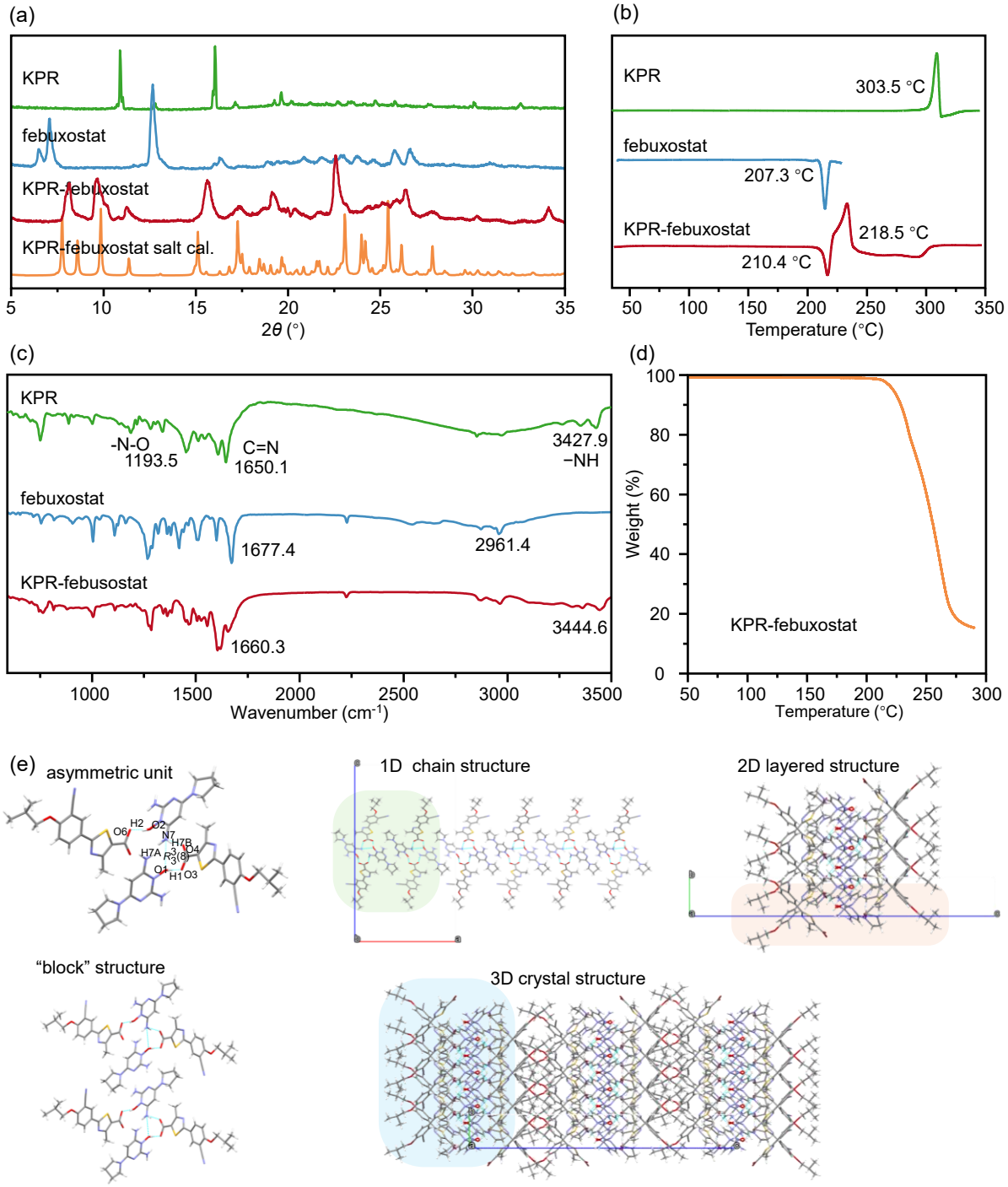
Supplementary Figure 52. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-adipic acid salt.

7.45 KPR-febuxostat salt

KPR-febuxostat salt crystallizes in the orthorhombic system with space group $Pna2_1$ ($Z = 8$). The piperidine ring of KPR exhibits disorder, and the major occupancy component was used for crystal structure analysis. The asymmetric unit contains two protonated KPR cations and two deprotonated febuxostat anions. A proton is transferred from the carboxylic acid group of febuxostat to the oxygen atom of the pyrimidine N-oxide in KPR. Two KPR cations and one febuxostat anion are connected via an $R_3^3(8)$ motif formed by N7–H7A···O1, N7–H7B···O4, and O1–H1···O3 hydrogen bonds, which further links to the second febuxostat anion through the O2–H2···O6 hydrogen bond. The asymmetric unit and a symmetry-related unit assemble via van der Waals interactions, forming a "block" structure. Two identically oriented "block" structures interact through van der Waals forces, extending along the *a*-axis to generate a 1D chain structure. These 1D chains, aligned in the same orientation, pack along the *b*-axis through N3–H3B···O7 hydrogen bonds, resulting in a 2D layered structure. Finally, the 2D layers arrange in opposite orientations along the *c*-axis via van der Waals interactions, completing the 3D crystal structure.

Supplementary Table 45. Hydrogen-bond geometry (Å) of KPR-febuxostat salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N7–H7B···O4	0.880	1.924	2.787	166.54
N7–H7A···O1	0.880	2.694	2.955	98.52
O1–H1···O3	0.841	1.625	2.461	172.45
N3–H3B···O7	0.880	2.946	2.803	164.09
O2–H2···O6	0.839	1.628	2.461	171.45



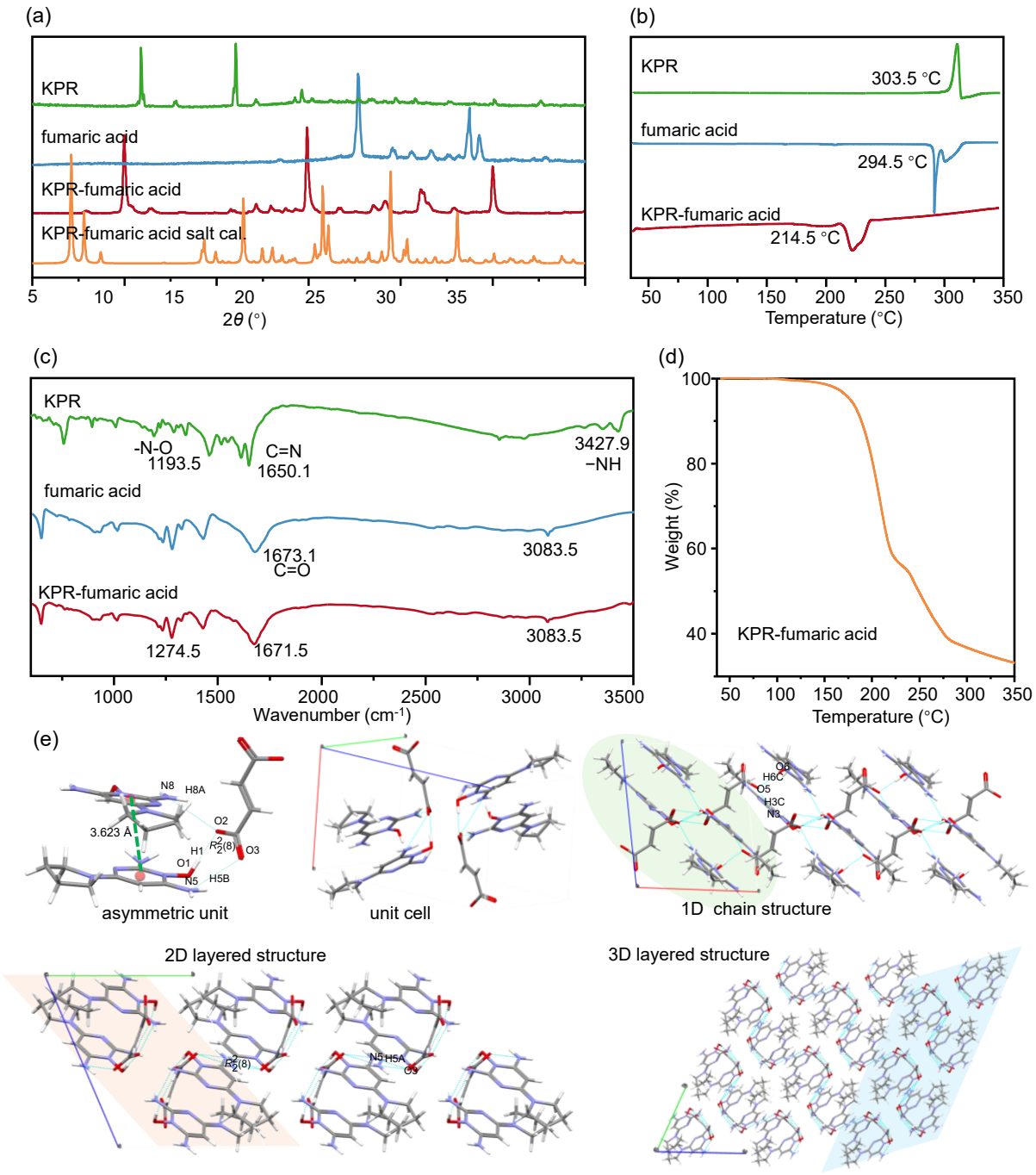
Supplementary Figure 53. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA and (e) molecular packing arrangements of KPR-febuxostat salt.

7.46 KPR-fumaric acid salt

KPR-fumaric acid salt crystallizes in the triclinic system with space group $P\bar{1}$ (Z = 2). The asymmetric unit contains two protonated KPR cations and one deprotonated fumaric acid anion. Protons are transferred from the carboxylic acid groups of fumaric acid to the oxygen atoms of the pyrimidine N-oxide in KPR. The succinate anion connects to one KPR cation through an $R_2^2(8)$ motif formed by N5-H5B···O3 and O1-H1···O2 hydrogen bonds, and to the other KPR cation via an N8-H8A···O2 hydrogen bond. The two protonated KPR cations are further linked by π-π stacking interactions (Cg-Cg distance: 3.623 Å, dihedral angle: 6.94°). Asymmetric units complete the full unit cell through centrosymmetric operations through van der Waals. These unit cells propagate along the *a*-axis through N3-H3C···O5 and O6-H6C···O5 hydrogen bonds, forming a 1D chain structure. Parallel-aligned 1D chains stack along the *b*-axis through $R_2^2(8)$ motifs formed by N5-H5A···O3 and N5-H5B···O3 hydrogen bonds, creating a 2D layered structure. Finally, these 2D layers arrange in opposite orientations along the *c*-axis through N6-H6A···O4 hydrogen bonds, completing the 3D crystal framework.

Supplementary Table 46. Hydrogen-bond geometry (Å) of KPR-fumaric acid salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N5-H5B···O3	0.881	2.269	2.953	134.44
N5-H5A···O3	0.880	2.011	2.862	162.06
O1-H1···O2	0.840	1.674	2.513	177.95
N3-H3C···O5	0.881	1.995	2.867	170.05
O6-H6C···O5	0.839	1.639	2.476	175.02
N8-H8A···O2	0.880	2.235	2.984	142.84
N6-H6A···O4	0.879	2.006	2.879	171.99



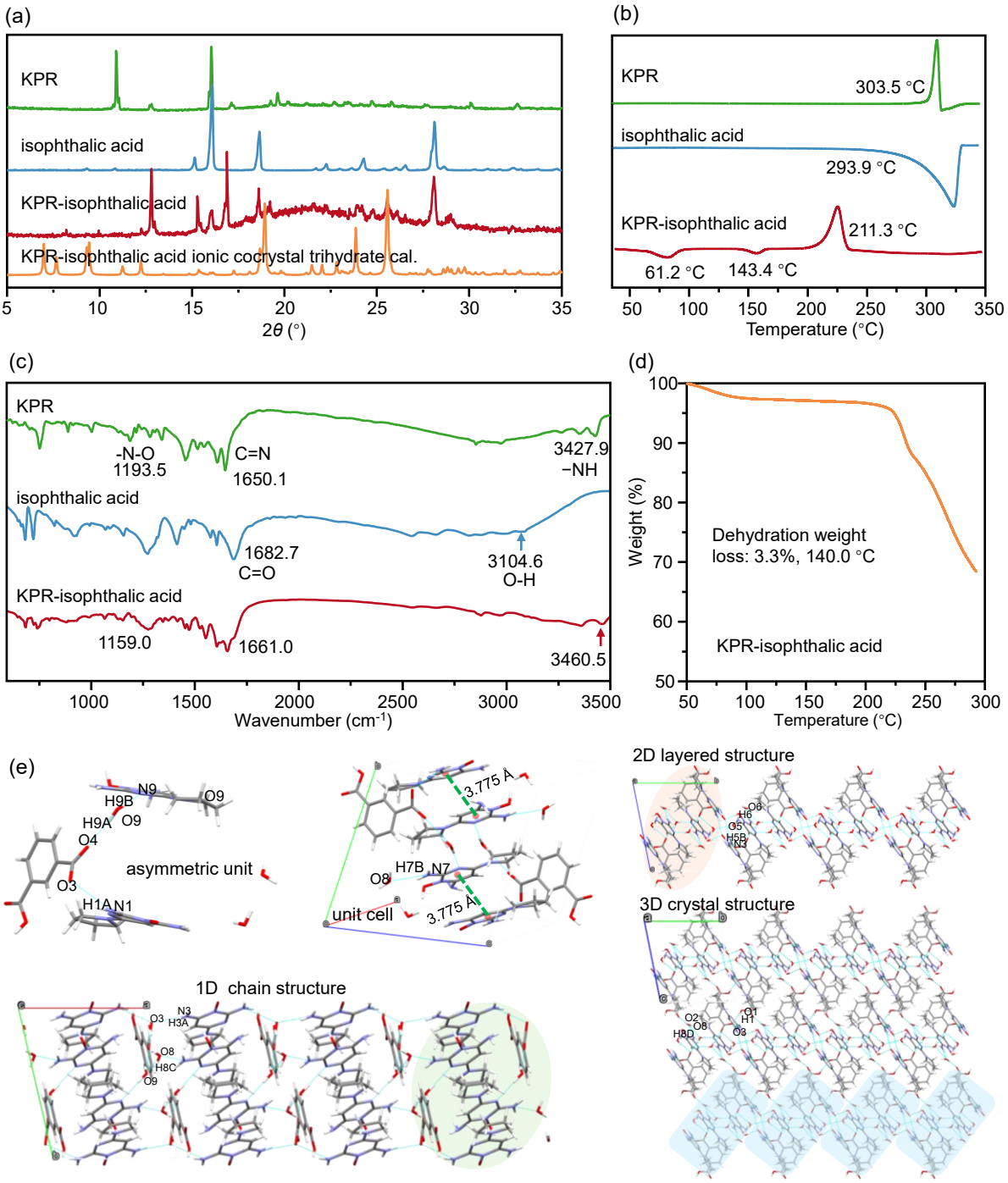
Supplementary Figure 54. (a) PXRD patterns and (b) DSC curves of starting drugs and KPR-fumaric acid (negative).

7.47 KPR-isophthalic acid ionic cocrystal trihydrate

KPR-isophthalic acid ionic cocrystal trihydrate crystallizes in the triclinic system with space group $P\bar{1}$ ($Z = 4$). The asymmetric unit consists of one protonated KPR cation, one deprotonated isophthalic acid anion, one neutral KPR molecule, and three water molecules. A proton is transferred from the carboxylic acid group of isophthalic acid to the oxygen atom of the pyrimidine N-oxide in KPX. The isophthalic acid anion forms hydrogen bonds with the neutral KPR (N1–H1A···O3) and one water molecule (O9–H9A···O4), while the water molecule further connects to the neutral KPR via O9–H9B···N9 hydrogen bond. The asymmetric unit interacts with a symmetry-related unit through a N7–H7B···O8 hydrogen bond and π···π stacking (Cg–Cg distance: 3.775 Å, parallel). Subsequently, the unit cells extend along the *a*-axis via hydrogen bonds (N3–H3A···O3, O7–H7C···O4, and O8–H8C···O9), forming a 1D chain structure. These 1D chains pack along the *b*-axis through additional hydrogen bonds (N3–H3B···O5 and O6–H6···O5), generating a 2D layered structure. Finally, the 2D layers arrange along the *c*-axis into a 3D framework through hydrogen bonds (O8–H8D···O2 and O1–H1···O3), completing the crystal structure.

Supplementary Table 47. Hydrogen-bond geometry (Å) of KPR-isophthalic acid ionic cocrystal trihydrate.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N1–H1A···O3	0.880	2.200	2.986	148.53
N3–H3A··O3	0.880	2.073	2.916	160.18
O1–H1···O3	0.840	1.786	2.612	167.27
O7–H7C···O4	0.870	2.017	2.872	167.32
O8–H8D···O2	0.870	1.976	2.837	169.97
O9–H9A···O4	0.871	1.960	2.778	156.17
N3–H3B··O5	0.881	2.158	2.938	147.37
O6–H6···O5	0.840	1.602	2.441	176.30
N7–H7B···O8	0.880	2.294	2.945	130.71
O9–H9B···N9	0.870	2.013	2.882	176.16
O8–H8C···O9	0.869	2.180	2.768	124.59



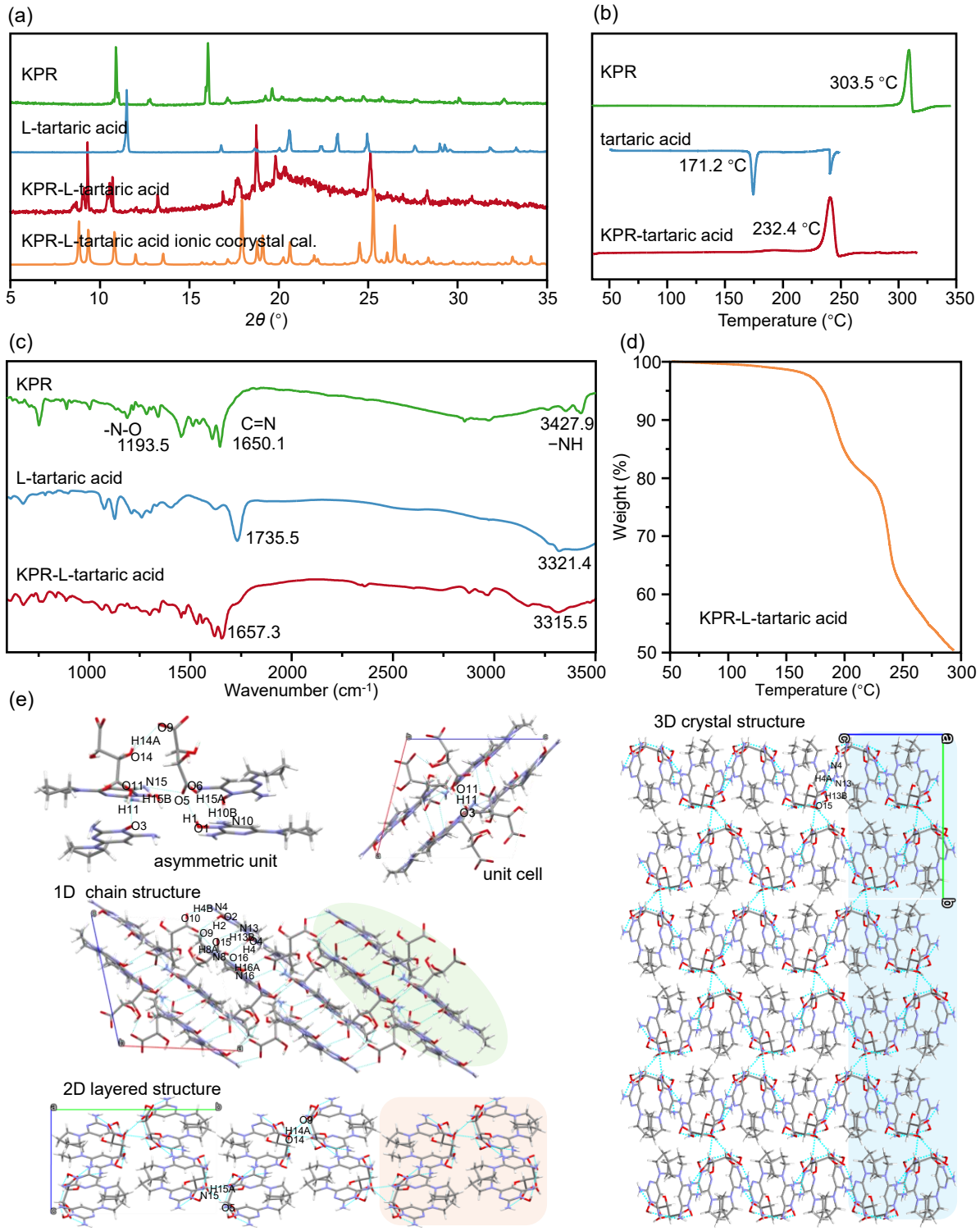
Supplementary Figure 55. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-isophthalic acid ionic cocrystal trihydrate.

7.48 KPR-L-tartaric acid ionic cocrystal

KPR-L-tartaric acid ionic cocrystal crystallizes in the triclinic system with space group $P\bar{1}$ ($Z = 2$). The piperidine ring of KPR exhibits disorder, and the major occupancy component was used for crystal structure analysis. The asymmetric unit consists of one neutral KPX molecule, three protonated KPR cations, one fully deprotonated L-tartaric acid anion, and one half-deprotonated L-tartaric acid anion. Protons are transferred from the carboxylic acid groups of L-tartaric acid to the oxygen atoms of the pyrimidine N-oxide in KPR. The half-deprotonated L-tartaric acid anion forms hydrogen bonds with the neutral KPR (O11–H11···O3), one KPR cation (N15–H15B···O11), and the fully deprotonated L-tartaric acid anion (O14–H14A···O9). The fully deprotonated L-tartaric acid anion connects to two KPR cations via a hydrogen bond (N15–H15A···O5) and an $R_2^2(8)$ motif formed by N10–H10B···O6 and O1–H1···O5. Additionally, the neutral KPR molecule and one protonated KPR cation are linked to the other two KPR molecules through π···π stacking interactions (Cg–Cg distances 3.521 Å and 3.634 Å; dihedral angles 9.47° and 12.01°, respectively). The asymmetric unit and its centrosymmetrically related counterpart assemble into the unit cell via hydrogen bonds O11–H11···O3 and N18–H18B···O6. These unit cells extend along the *a*-axis through multiple hydrogen bonds (N4–H4B···O10, O2–H2···O9, N8–H8A···O9, N13–H13B···O15, O4–H4···O16, and N16–H16B···O16), forming a 1D chain structure. The 1D chains, aligned in the same orientation, pack along the *b*-axis through hydrogen bonds (N15–H15A···O5 and N14–H14A···O9), generating a 2D layered structure. Finally, the 2D layers arrange into a 3D framework along the *c*-axis via hydrogen bonds (N4–H4A···O15 and N13–H13B···O15), completing the crystal structure.

Supplementary Table 48. Hydrogen-bond geometry (Å) of KPR-L-tartaric acid ionic cocrystal.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N4–H4A···O15	0.880	2.051	2.911	165.44
N4–H4B···O10	0.880	2.180	2.861	133.94
O2–H2···O9	0.839	1.762	2.600	176.10
N10–H10B···O6	0.880	2.259	2.788	118.51
O1–H1···O5	0.839	1.732	2.535	169.37
N8–H8A···O9	0.880	2.085	2.931	161.03
N15–H15B···O11	0.866	1.944	2.847	168.21
N13–H13B···O15	0.881	2.310	2.891	123.54
O4–H4···O16	0.840	1.745	2.575	169.31
N15–H15A···O5	0.868	2.299	2.997	125.10
O11–H11···O3	0.819	1.671	2.453	159.02
N16–H16B···O16	0.865	2.087	2.910	158.58
N18–H18B···O6	0.862	2.795	2.790	80.84
O14–H14A···O9	0.841	2.068	2.811	146.96



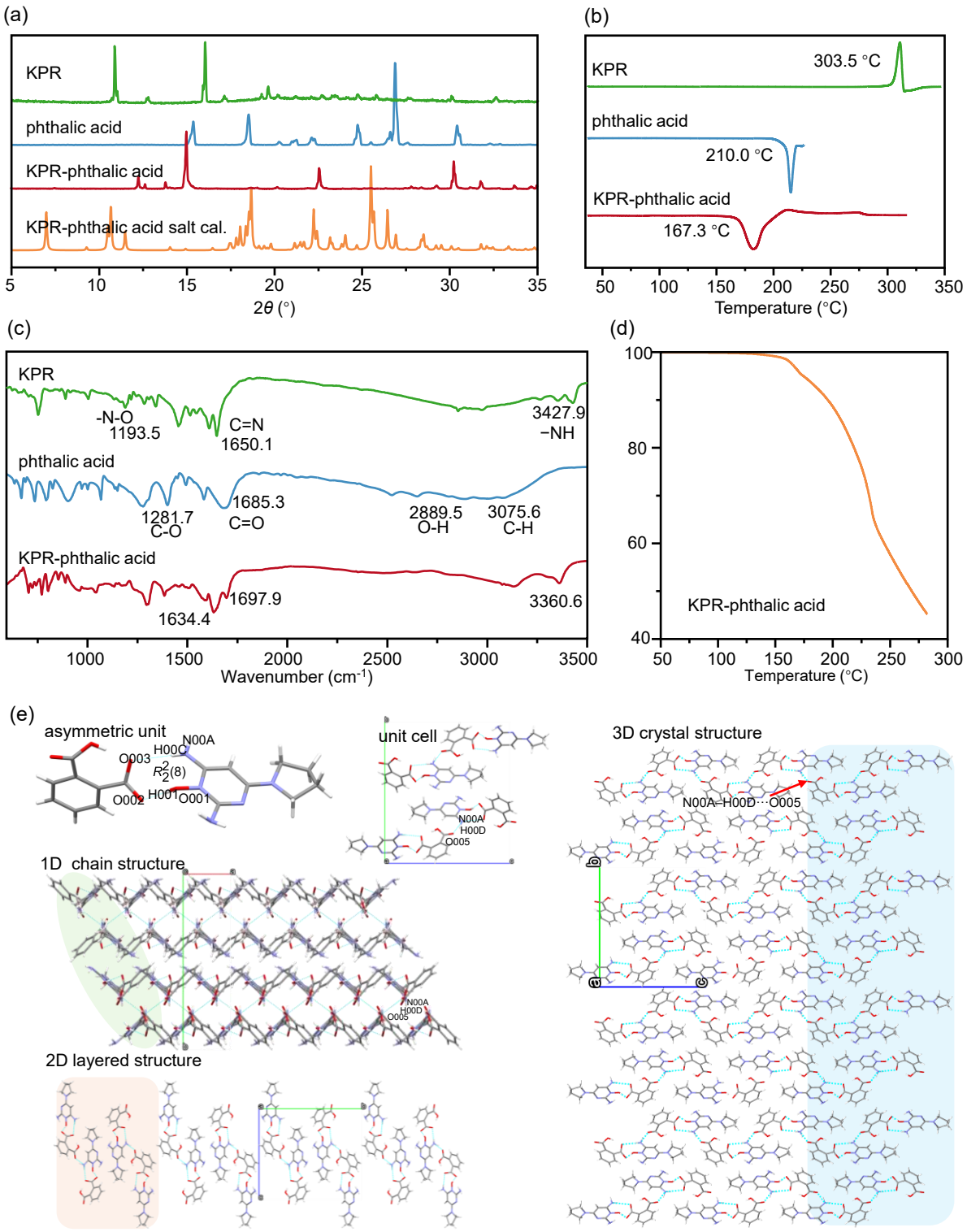
Supplementary Figure 56. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-L-tartaric acid ionic cocrystal.

7.49 KPR-phthalic acid salt

KPR-phthalic acid salt crystallizes in the triclinic system with space group $P\bar{1}$ (Z = 4). The asymmetric unit consists of one protonated KPR cation and one deprotonated phthalic acid anion, connected through an $R_2^2(8)$ motif formed by N00A-H00C···O003 and O001-H001···O002 hydrogen bonds. Centrosymmetrically related asymmetric units associate via N00A-H00D···O005 hydrogen bonds, with these dimeric units further connecting through van der Waals interactions to form the unit cell. The unit cells propagate along the *a*-axis through N00A-H00D···O005 hydrogen bonds, generating a 1D chain structure. These parallel-aligned chains stack along the *b*-axis via van der Waals forces, forming a 2D layered structure. Finally, the layers interconnect along the *c*-axis through N00A-H00D···O005 hydrogen bonds, completing the 3D crystal framework.

Supplementary Table 49. Hydrogen-bond geometry (Å) of KPR-phthalic acid salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N00A–H00C···O003	0.864	2.224	2.972	144.82
O001–H001···O002	0.820	1.798	2.617	178.42
N00A–H00D···O005	0.865	2.428	2.911	115.76



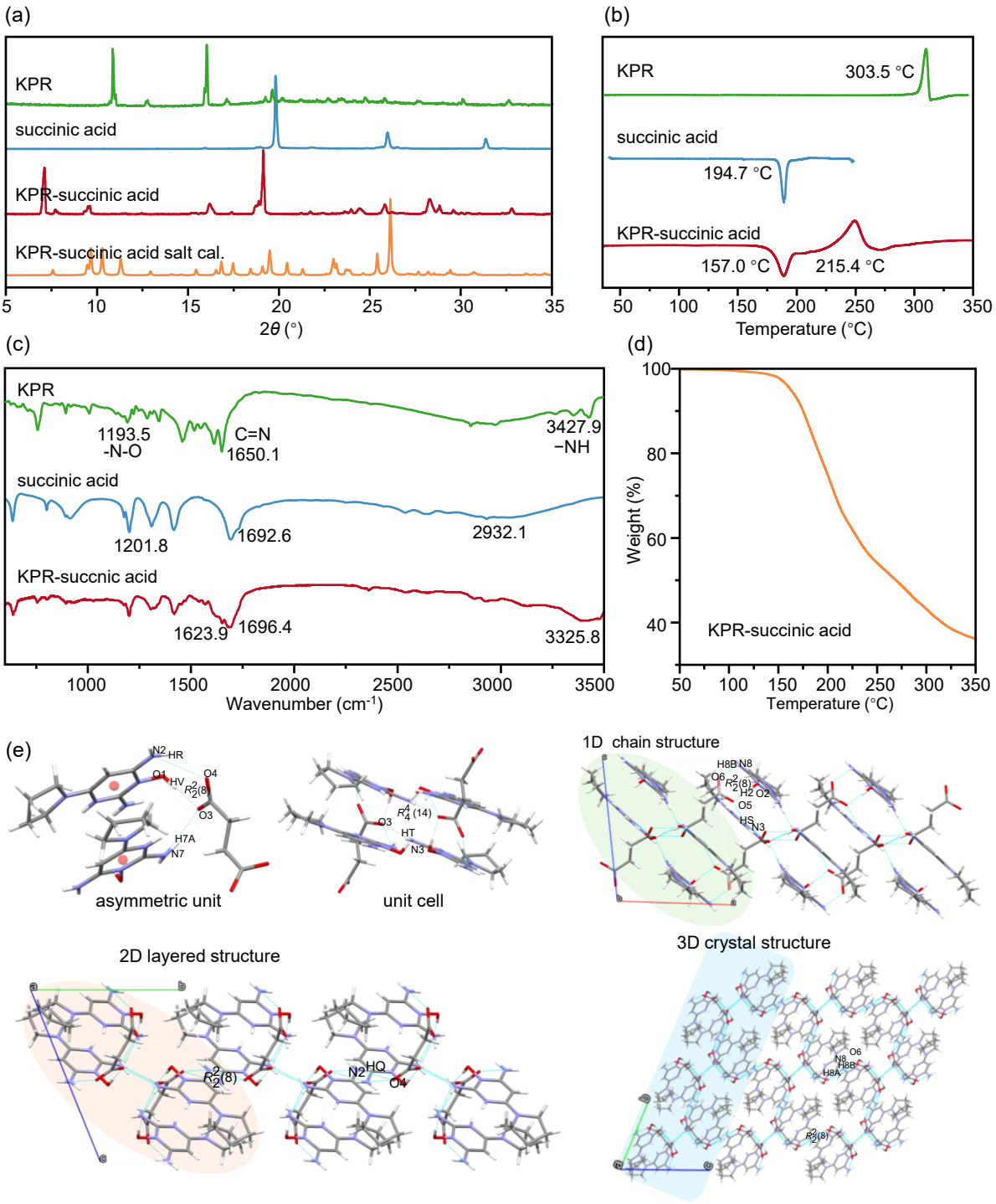
Supplementary Figure 57. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-phthalic acid salt.

7.50 KPR-succinic acid salt

KPR-succinic acid salt crystallizes in the triclinic system with space group $P\bar{1}$ ($Z = 2$). The piperidine ring of KPR exhibits disorder, with the major occupancy component used for crystal structure analysis. The asymmetric unit contains two protonated KPR cations and one deprotonated succinic acid anion. Protons are transferred from the carboxylic acid groups of succinic acid to the oxygen atoms of the pyrimidine N-oxide in KPR. The succinate anion connects to one KPR cation through an $R_2^2(8)$ motif formed by N2-HR···O4 and O1-HV···O3 hydrogen bonds, and to the other KPR cation via an N7-H7A···O3 hydrogen bond. The two protonated KPR cations are further linked by π-π stacking interactions (Cg-Cg distance: 3.696 Å, dihedral angle: 8.08°). Centrosymmetrically related asymmetric units associate through an $R_4^4(14)$ motif formed by N3-HT···O3 and O1-HV···O3 hydrogen bonds, generating the unit cell. These unit cells propagate along the *a*-axis through N3-HS···O5 hydrogen bonds and $R_2^2(8)$ motifs formed by N8-H8B···O6 and O2-H2···O5 hydrogen bonds, forming a 1D chain structure. Parallel-aligned 1D chains stack along the *b*-axis through $R_2^2(8)$ motifs formed by N2-HQ···O4 and N2-HR···O4 hydrogen bonds, creating a 2D layered structure. Finally, these 2D layers arrange in opposite orientations along the *c*-axis through $R_2^2(8)$ motifs formed by N8-H8A···O6 hydrogen bonds, completing the 3D crystal framework.

Supplementary Table 50. Hydrogen-bond geometry (Å) of KPR-succinic acid salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N3–HS···O5	0.880	2.051	2.870	154.42
N2–HQ···O4	0.881	2.009	2.843	157.56
N2–HR···O4	0.880	2.294	2.899	125.90
O1–HV···O3	0.840	1.685	2.525	178.81
N3–HT···O3	0.881	2.162	2.965	151.32
N8–H8A···O6	0.881	2.063	2.889	164.20
N8–H8B···O6	0.879	2.302	2.950	130.60
O2–H2···O5	0.840	1.648	2.484	173.26
N7–H7A···O3	0.884	2.220	2.969	142.35



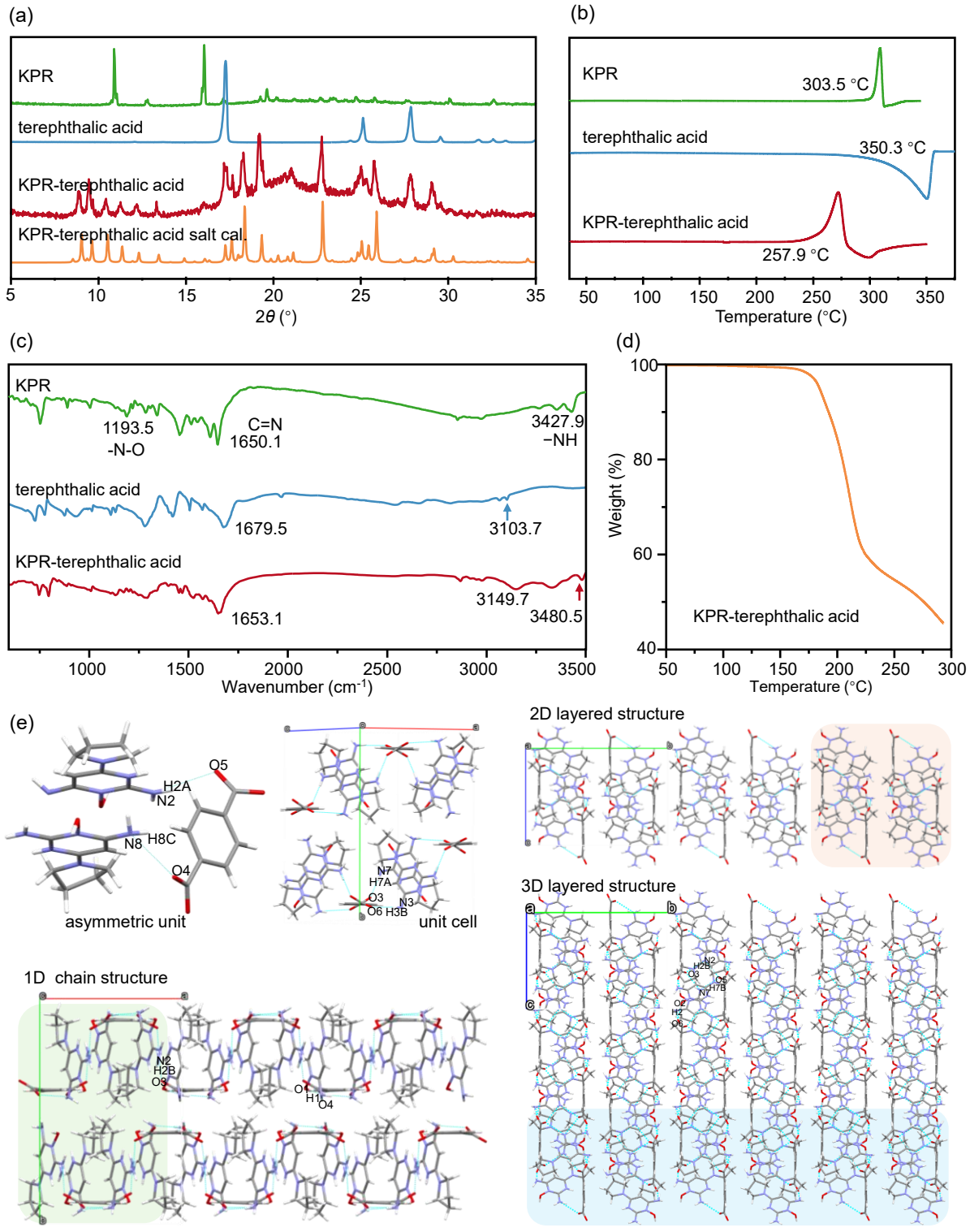
Supplementary Figure 58. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-succinic acid salt.

7.51 KPR-terephthalic acid salt

KPR-terephthalic acid salt crystallizes in the monoclinic system with space group $P2_1/n$ (Z = 4). The piperidine ring of KPR exhibits disorder, with the major occupancy component used for crystal structure analysis. The asymmetric unit contains two protonated KPX cations and one deprotonated terephthalic acid anion. The terephthalic acid anion connects to two KPX cations through N8–H8C···O4 and N2–H2A···O5 hydrogen bonds. Protons are transferred from the carboxylic acid groups of terephthalic acid to the oxygen atoms of the pyrimidine N-oxide in KPX. The asymmetric unit and a symmetry-related unit associate via N7–H7A···O3 and N3–H3B···O6 hydrogen bonds, forming a "block" structure. Two such blocks constitute the unit cell. The unit cells propagate along the *a*-axis through N2–H2B···O3 and O1–H1···O4 hydrogen bonds, generating a 1D chain structure. These parallel-aligned 1D chains stack along the *b*-axis via van der Waals interactions, forming a 2D layered structure. Finally, the 2D layers arrange into a 3D framework along the *c*-axis through O2–H2···O6, N2–H2B···O3, and N7–H7B···O5 hydrogen bonds, completing the crystal structure.

Supplementary Table 51. Hydrogen-bond geometry (Å) of KPR-terephthalic acid salt.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N2–H2B···O3	0.879	2.160	2.921	144.52
O1–H1···O4	0.841	1.661	2.499	174.10
N3–H3B···O6	0.881	2.010	2.883	170.54
N2–H2A···O5	0.880	2.126	2.964	158.88
N8–H8C···O4	0.880	2.420	2.971	120.98
N7–H7B···O5	0.880	2.177	2.935	144.10
O2–H2···O6	0.880	2.177	2.935	177.05
N7–H7A···O3	0.880	2.148	2.984	158.48



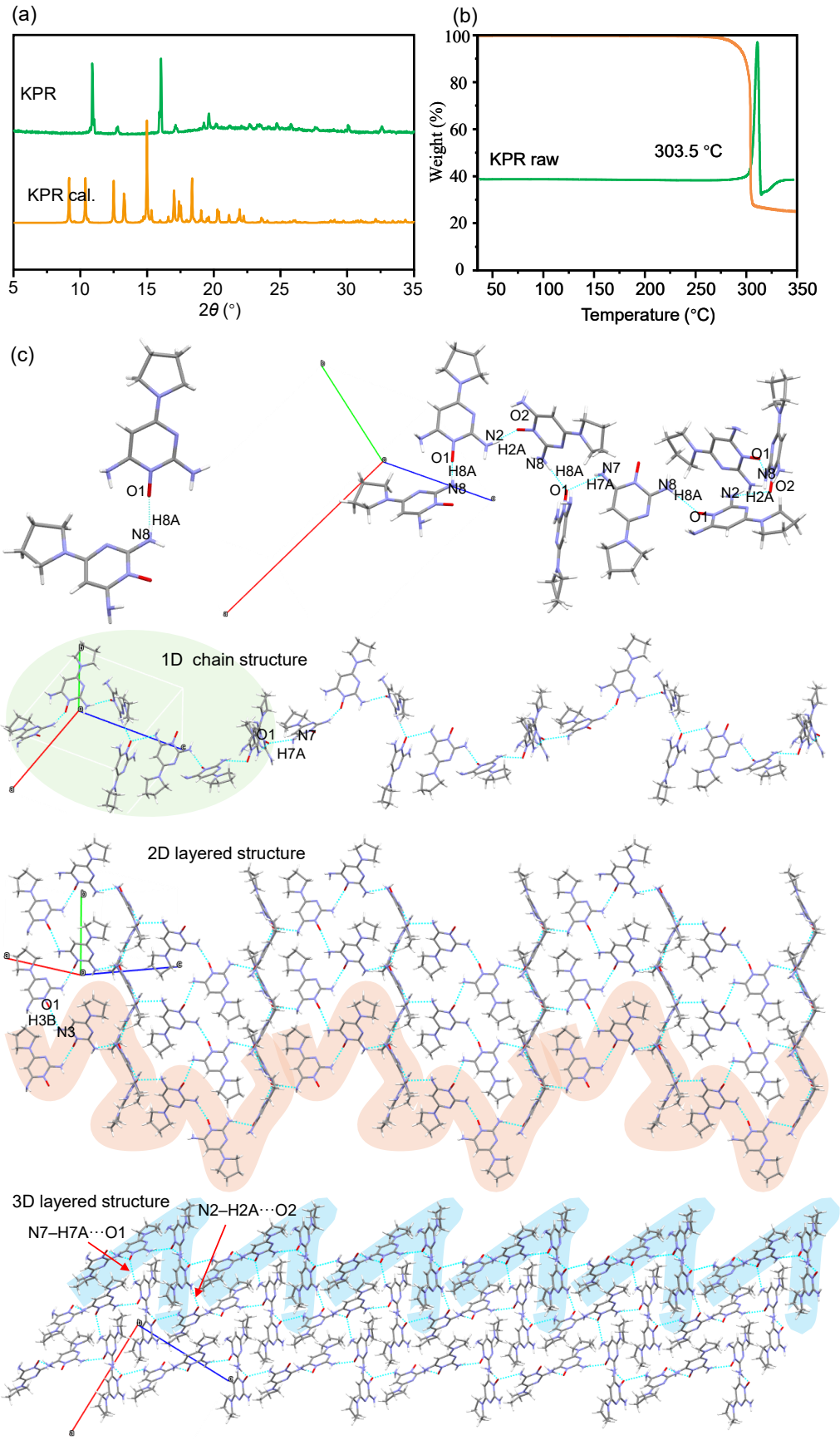
Supplementary Figure 59. Solid-state characterization. (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra, (d) TGA, (e) molecular packing arrangements of KPR-terephthalic acid salt.

7.52 KPR

KPR crystallizes in the orthorhombic system with space group $Pna2_1$ ($Z = 8$). The piperidine ring exhibits disorder, with the major occupancy component used for crystal structure analysis. The asymmetric unit contains a dimer of two KPR molecules connected by an N8-H8A···O1 hydrogen bond. These dimers associate via N2-H2A···O2 hydrogen bonds to form tetramers, which further connect through N7-H7A···O1 hydrogen bonds to generate "block" structures. Adjacent blocks interconnect through additional N7-H7A···O1 hydrogen bonds, creating a 1D undulating chain structure extending along the [011] direction. These chains pack along the *a*-axis through N3-H3B···O2 hydrogen bonds, forming a 2D layered structure. Finally, the 2D layers arrange into a 3D framework through N7-H7A···O1 and N2-H2A···O2 hydrogen bonds along the [010] direction, completing the crystal structure.

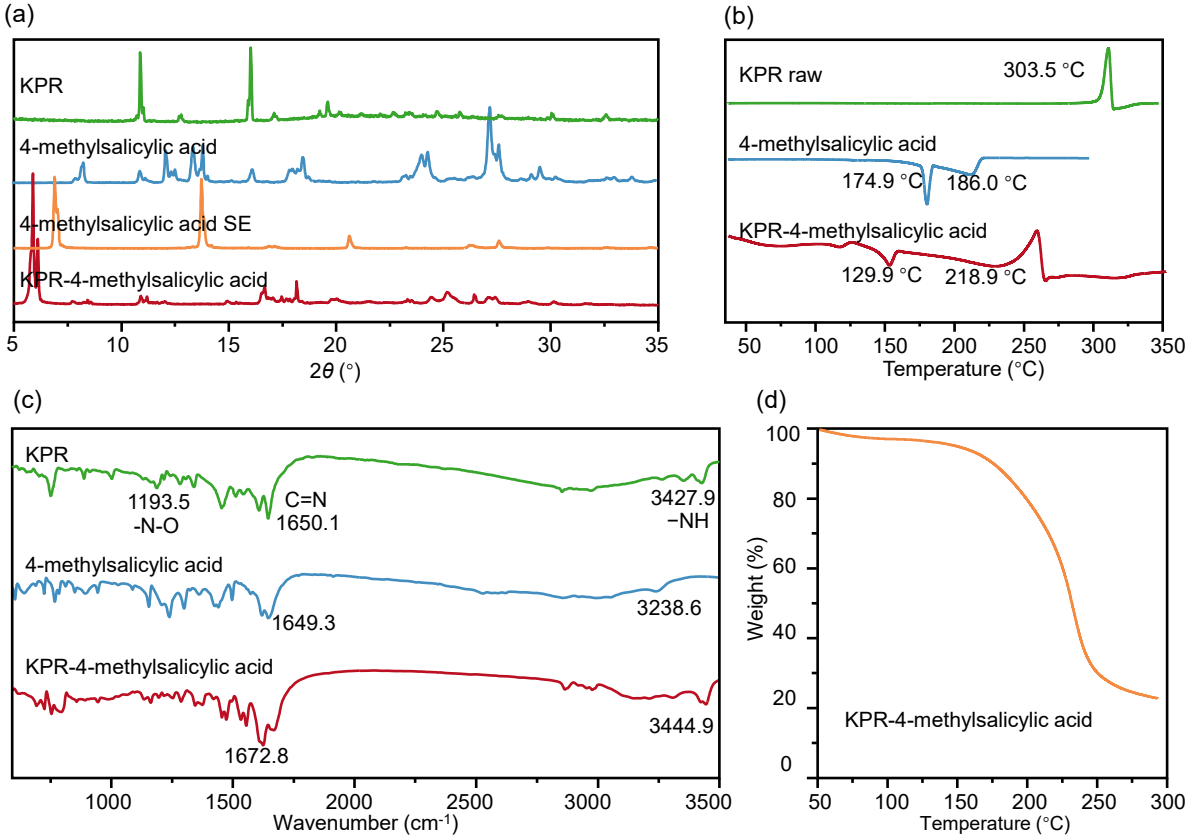
Supplementary Table 52. Hydrogen-bond geometry (Å) of KPR.

	d(D–H)/Å	*d*(H···A)/Å	*d*(D···A)/Å	(D–H···A)/°
N2–H2A···O2	0.881	1.987	2.787	150.26
N8–H8A···O1	0.881	1.866	2.742	173.00
N7–H7A···O1	0.883	2.881	2.885	81.41
N3–H3B···O2	0.880	2.666	2.927	98.38



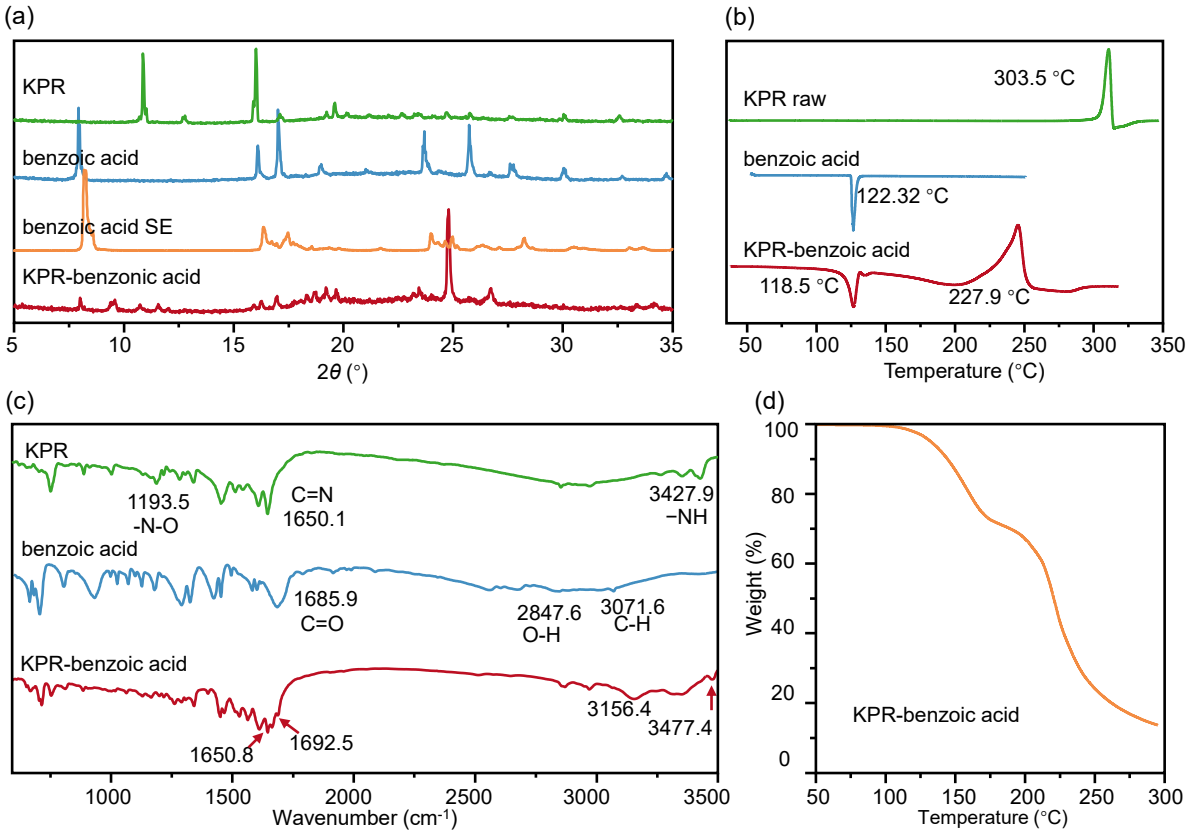
Supplementary Figure 60. Solid-state characterization. (a) PXRD patterns, (b) DSC and TGA curves, and (c) molecular packing arrangements of KPR.

7.53 KPR-4-methylsalicylic acid hydrate



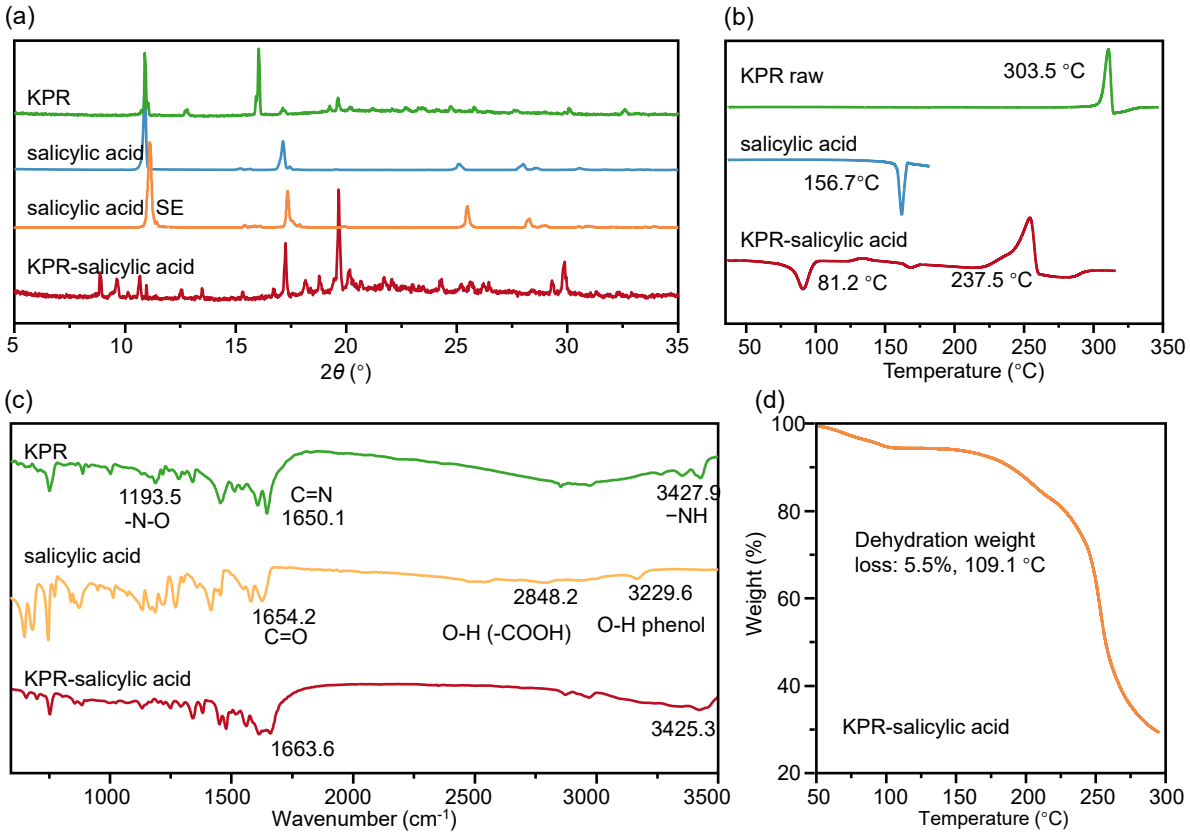
Supplementary Figure 61. Solid-state characterization of starting drugs and KPR-4-methylsalicylic acid hydrate (estimated). (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra and (d) TGA.

7.54 KPR-benzoic acid cocrystal



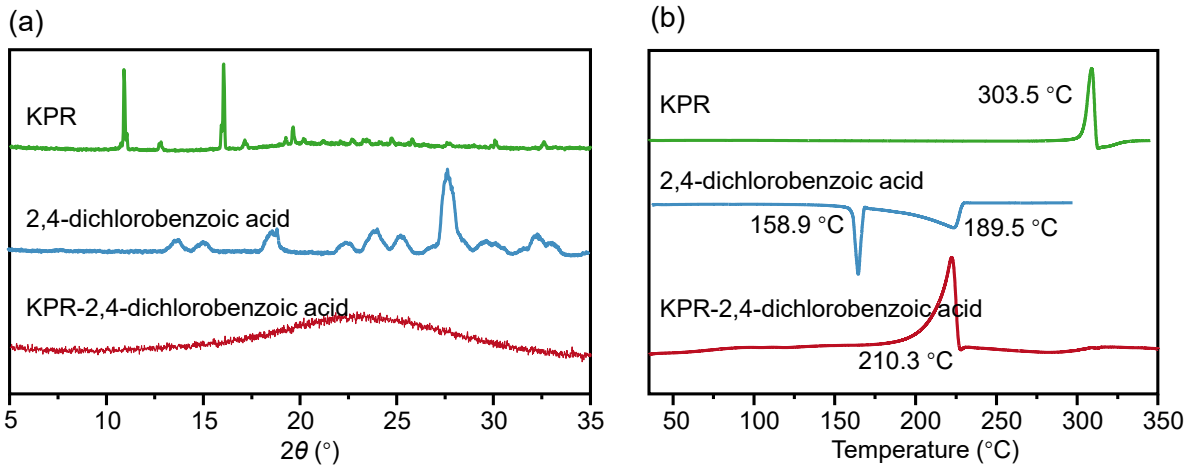
Supplementary Figure 62. Solid-state characterization of starting drugs and KPR-benzoic acid cocrystal (estimated). (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra and (d) TGA.

7.55 KPR-salicylic acid hydrate



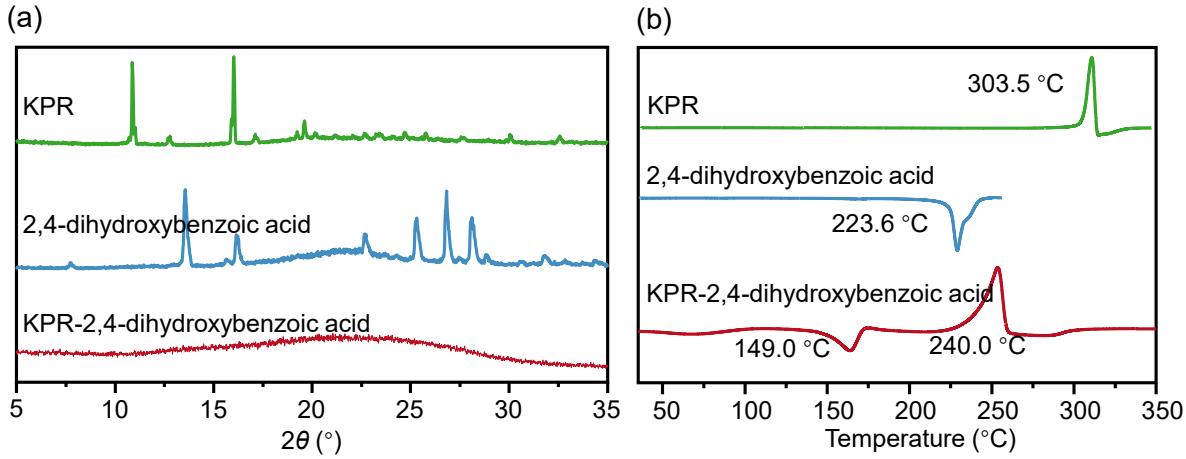
Supplementary Figure 63. Solid-state characterization of starting drugs and KPR-salicylic acid hydrate (estimated). (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra and (d) TGA.

7.56 KPR-2,4-dichlorobenzoic acid (negative)



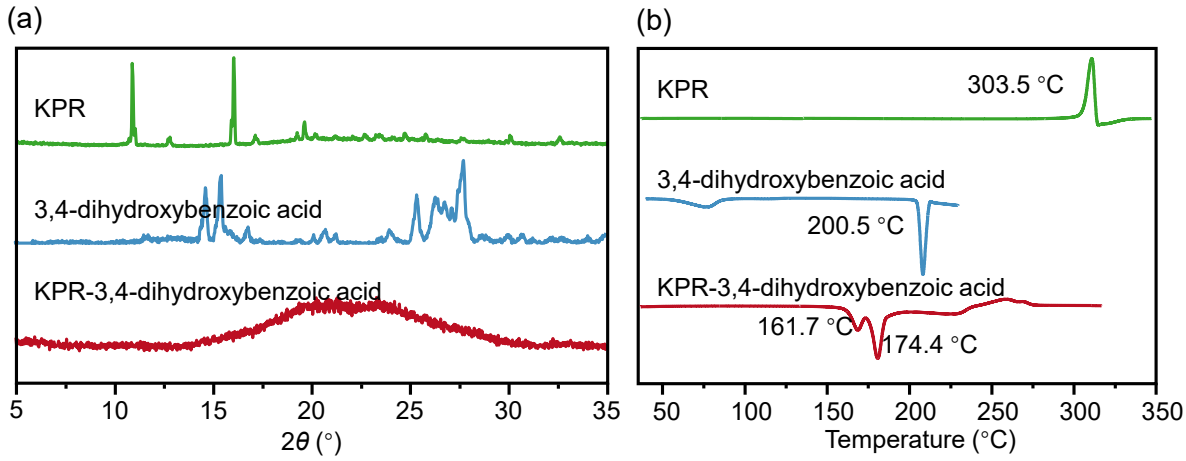
Supplementary Figure 64. (a) PXRD patterns and (b) DSC curves of starting drugs and KPR-2,4-dichlorobenzoic acid (negative).

7.57 KPR-2,4-dihydroxybenzoic acid (negative)



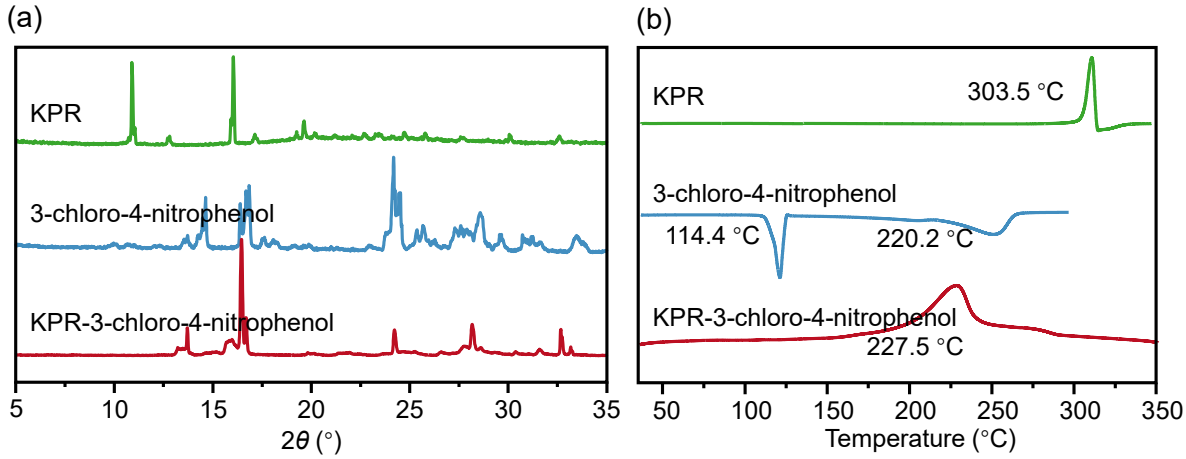
Supplementary Figure 65. (a) PXRD patterns and (b) DSC curves of starting drugs and KPR-2,4-dihydroxybenzoic acid (negative).

7.58 KPR-3,4-dihydroxybenzoic acid (negative)



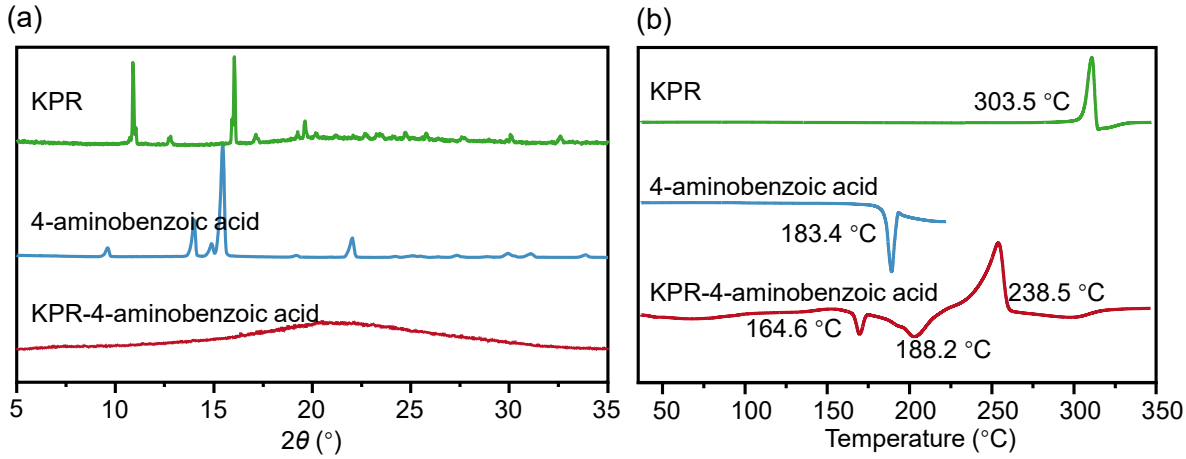
Supplementary Figure 66. (a) PXRD patterns and (b) DSC curves of starting drugs and KPR-3,4-dihydroxybenzoic acid (negative).

7.59 KPR-3-chloro-4-nitrophenol (negative)



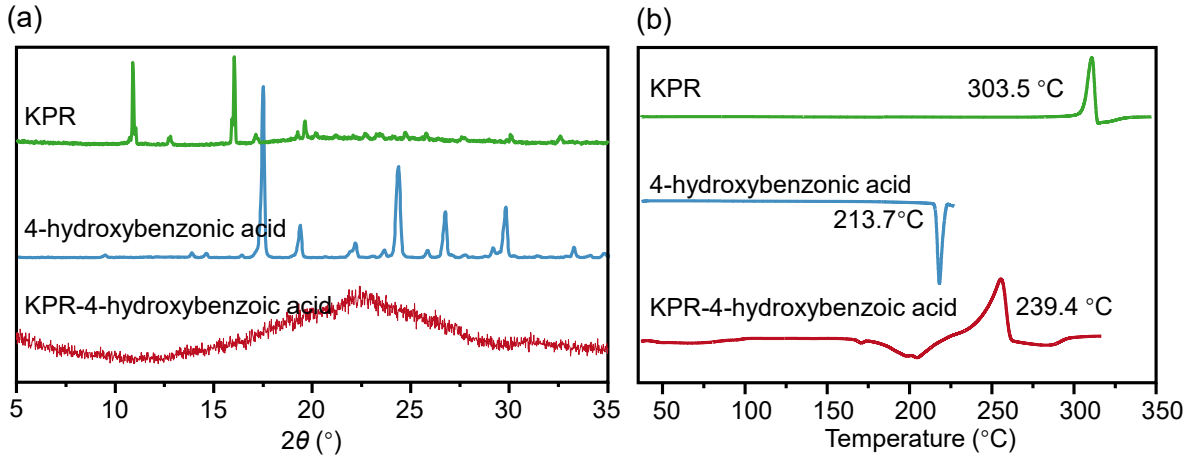
Supplementary Figure 67. (a) PXRD patterns and (b) DSC curves of starting drugs and KPR-3-chloro-4-nitrophenol (negative).

7.60 KPR-4-aminobenzoic acid (negative)



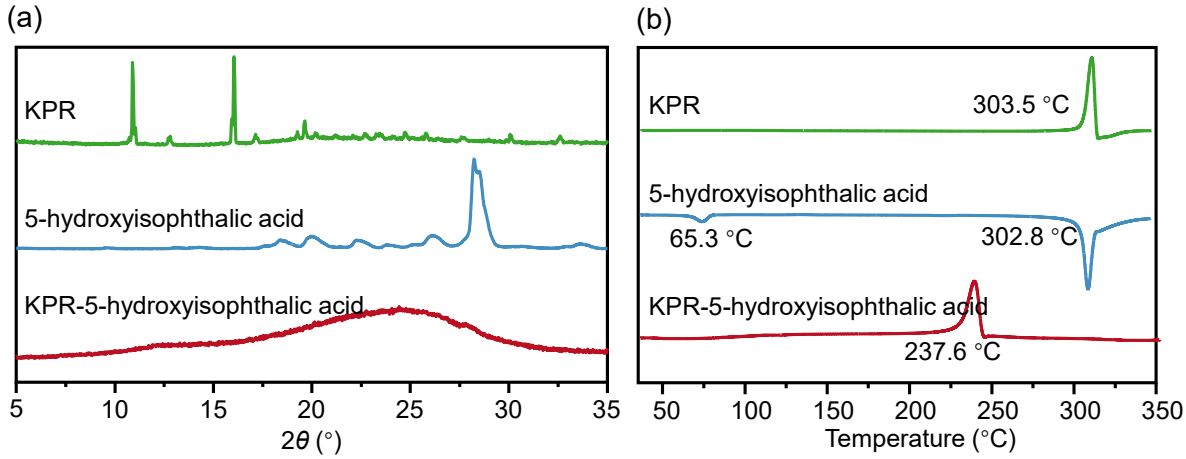
Supplementary Figure 68. (a) PXRD patterns and (b) DSC curves of starting drugs and KPR-4-aminobenzoic acid (negative).

7.61 KPR-4-hydroxybenzoic acid (negative)



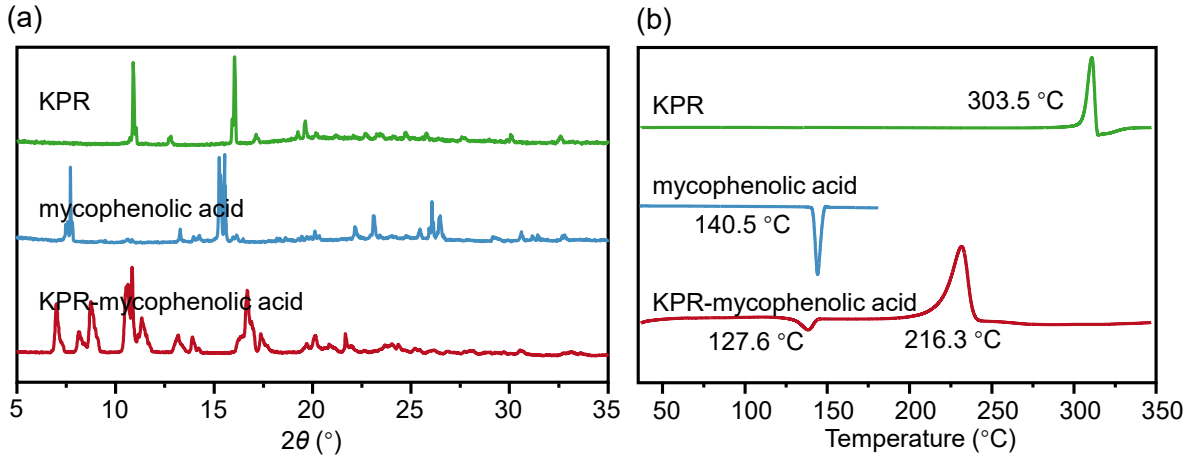
Supplementary Figure 69. (a) PXRD patterns and (b) DSC curves of starting drugs and KPR-4-hydroxybenzoic acid (negative).

7.62 KPR-5-hydroxyisophthalic acid (negative)



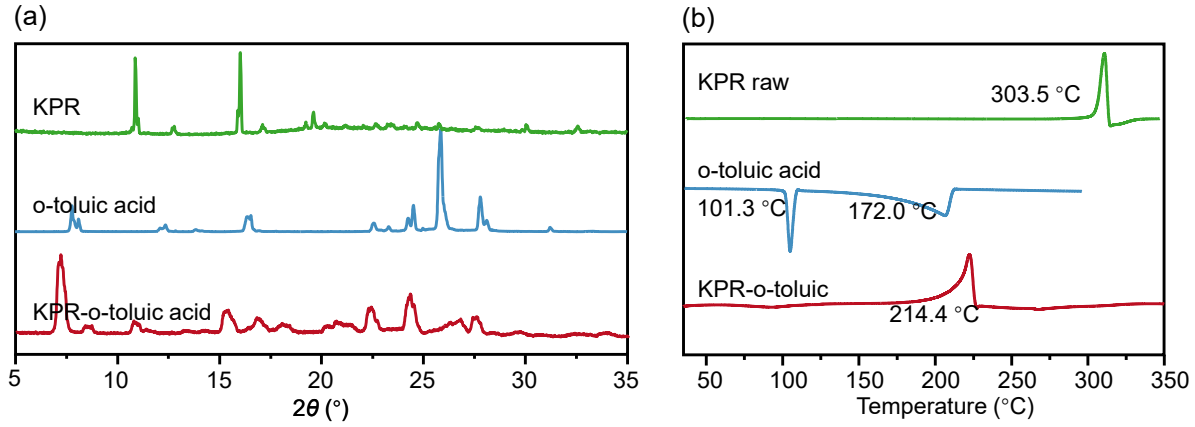
Supplementary Figure 70. (a) PXRD patterns and (b) DSC curves of starting drugs and KPR-5-hydroxyisophthalic acid (negative).

7.63 KPR-mycophenolic acid (negative)



Supplementary Figure 71. (a) PXRD patterns and (b) DSC curves of starting drugs and KPR-mycophenolic acid (negative).

7.64 KPR-o-toluic acid (negative)



Supplementary Figure 72. Solid-state characterization of starting drugs and K KPR-o-toluic acid salt (estimated). (a) PXRD patterns, (b) DSC curves, (c) FTIR spectra and (d) TGA.