





Cite this: *Phys. Chem. Chem. Phys.*,
2023, 25, 29266

Using singular value decomposition to analyze drug/ β -cyclodextrin mixtures: insights from X-ray powder diffraction patterns†

Kanji Hasegawa, Satoru Goto, * Chihiro Tsunoda, Chihiro Kuroda, Yuta Okumura, Ryosuke Hiroshige, Ayako Wada-Hirai, Shota Shimizu, Hideshi Yokoyama and Tomohiro Tsuchida 

The article discusses the use of mathematical models and linear algebra to understand the crystalline structures and interconversion pathways of drug complexes with β -cyclodextrin (β -CD). It involved the preparation and analysis of mixtures of indomethacin, diclofenac, famotidine, and cimetidine with β -CD using techniques such as differential scanning calorimetry (DSC), X-ray powder diffraction (XRPD), and proton nuclear magnetic resonance ($^1\text{H-NMR}$). Singular value decomposition (SVD) analysis is used to identify the presence of different polymorphs in the mixtures of these drugs and β -CD, determine interconversion pathways, and distinguish between different forms. In general, linear algebra or artificial intelligence (AI) is used to approximate the contribution of distinguishable entities to various phenomena. We expected linear algebra to completely reveal all eight entities present in the diffractogram dataset. However, after performing the SVD procedure, we found that only six independent basis functions were extracted, and the entities of the INM α -form and the CIM B-form were not included. It is considered that this is due to that data processing is limited to revealing only six or seven independent factors, as it is a small world. The authors caution that these may not always reproduce or approach reality in complicated real-world situations.

Received 12th June 2023,
Accepted 18th September 2023

DOI: 10.1039/d3cp02737f

rsc.li/pccp

1. Introduction

Several fields require investigation of polymorphism to regulate crystallization phenomena. The formation of polymorphs involves various factors, including physical conditions such as temperature, humidity, and pressure, chemical conditions (solvents, additives and others), and preparation/storage processes. In materials science, crystals' optical effects are applied to the development of optical communications and devices, while the study of the crystalline structure is useful for the development of new materials.^{1–3} In earth science, we can understand the origin of the earth from the formation and structure of crystals.^{4–6} The crystallizing phenomenon is particularly important in pharmaceutical, cosmetic, and food sciences.^{7–10} A compound has different crystal polymorphs, known or unknown, and there may be differences in their physicochemical properties and efficacy.^{11–13} It is necessary to control crystallization to maintain high quality at various steps in the manufacturing process.^{7,14}

The crystal habit (*i.e.*, crystal structure, such as cutting direction and crystallinity) can influence crystal formation and can be useful in analyzing the effects of adjustments such as heating, cooling, pressurization, the addition of semi-solidifying agents, and additives. A crystal habit control technology is used for this purpose.¹⁵ Since pharmaceuticals are manufactured and stored under various conditions, the crystal habit of active pharmaceutical ingredients (APIs) may change even if their crystal polymorph does not change. Differences in crystal habit, as well as polymorphism of APIs, can affect the physical and chemical properties and stability of pharmaceuticals.^{16,17} For example, different crystallizing behavior may result in different solubility of APIs. An unstable crystal may improve solubility, but a stable crystal may decrease solubility.^{18,19} Furthermore, differences in crystallizing behavior may affect APIs' pulverization-ability, compressibility, water retention, and emulsification properties.²⁰ To maintain pharmaceutical quality, it is important to pay attention to changes in crystallizing behavior and create an appropriate manufacturing and storage environment. The present study's aim is quantitatively to distinguish and classify the crystalline states (polymorphism and habits) of pure or complex active APIs.

Shimada and Tateuchi focused on the importance of the effects of basic drugs on the physical properties of indomethacin.^{21,22}

Faculty of Pharmaceutical Sciences, Tokyo University of Science, 2641 Yamazaki, Noda, Chiba 278-8510, Japan. E-mail: s.510@rs.tus.ac.jp

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d3cp02737f>

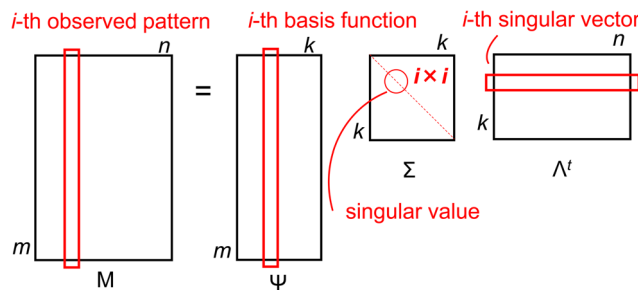


A quantitative structure–activity relationship study of the solubility of indomethacin (INM) showed that the value of the partition coefficient of basic drugs and the decrease in melting point (ΔT_m) due to physical mixing influence the solubility of INM.²¹ In a recent study, the radar chart analysis for Lipinski's drug-likeness was used to search for candidate compounds as useful drugs for sulfa drug-pyrrole complexes.²³ Kasai and Shiono *et al.* found that famotidine (FAM) and cimetidine (CIM), which have similar structures to acidic drugs, show marked differences in their behavior.²⁴ Tsunoda *et al.* have discussed the individual intermolecular interactions of these basic drugs with acidic drugs.²⁵ Therefore, indomethacin and diclofenac were selected as low-solubility acidic drugs with the aim of clarifying the difference in the effects of FAM and CIM.

Explainable artificial intelligence (XAI) refers to the development of AI systems that can provide understandable and transparent explanations of their recognition and decision-making processes.^{26–32} IBM Watson uses various techniques to achieve XAI, such as explaining how a particular decision was made using a set of predefined rules, responding to user queries with relevant explanations through natural language processing, generating visualizations to help users understand the logic behind the decision, allowing interactive feedback during the decision-making process, and more.³³ Incorporating transparency and interpretability aims to increase human trust in AI systems. However, it is noteworthy that providing explanations alone does not guarantee trust or acceptance of the system's decisions.^{34,35} To build trust in AI systems, it is crucial to consider a range of ethical and social considerations and tailor the explanations and interactions provided by the system to the specific needs and context of the users.^{36,37} While IBM Watson can provide a large amount of information, it is worth noting that the number of sources surveyed by the system does not prove the objective authenticity of scientific findings.³⁸ The validity of scientific discoveries depends on the quality of evidence and the reproducibility of results, rather than popularity or the number of sources surveyed. In scientific research, scientists use inductive or deductive logic. Particularly, experimental chemists presuppose the deductive components consisting of chemically/physically pure entities (namely, molecular species, resonance structures, racemic compounds, and others) or analytically indistinguishable molecular assemblies to verify the validity of their findings.

In our previous work, we explored a novel approach to analytical chemistry that involved using linear algebra and basis functions derived from various spectroscopic techniques, such as Fourier-transform infrared (FTIR),³⁹ circular dichroism (CD),⁴⁰ ultraviolet/visible light (UV/vis),^{41,42} electron paramagnetic resonance (ESR),^{43,44} and fluorescence (FL).^{45,46} This approach enabled us to reconstruct observed data by applying a similar technique to analyze diffraction patterns obtained from X-ray powder diffractometry (XRPD).^{47,48}

In the present study, we assumed that the observable diffraction patterns were linear combinations of latent patterns for unknown molecular entities or distinguishable complexes (independently adding the newly synthesized or



Scheme 1 Outline of singular value decomposition (SVD).

derived complex), which we arranged into a matrix **M** based on the experimental conditions. Using singular value decomposition (SVD) on the **M** matrix, we obtained a basis function matrix **Ψ**, a diagonal matrix **Σ** of singular values, and the transposed matrix **Λ^t** of singular vectors: **M** = **ΨΣΛ^t**, where the *i*-th vector in the **Ψ** matrix represents a latent pattern for an assumed entity, the *i*-th singular value in the **Σ** matrix indicates the statistical variance in the contribution of the corresponding entity, and the *j*-th singular vector in the **Λ^t** matrix corresponds to the vector descriptor used to represent each observed pattern using the linear combination coefficients for the basis functions and singular values (see Scheme 1). Thus, any arbitrary spectrum observed can be rationally reproduced by a linear combination of the confirmable basis functions. Refer to our previous publications for a more detailed explanation of the mathematical protocol used for the SVD procedure.^{39–46}

Based on this reflection, we reported that an approach with a melting entropy-phase diagram indicated whether inclusion complexation with HP-β-CD succeeded for nifedipine or nifedipine hydrochloride.⁴⁹ Different polymorphs were found in a mixture of urea and naphthalene, indicating that entropy plays an important role.⁴⁸ We recently applied the melting entropy-phase diagram approach to determine the efficiencies of the inclusion complexation of INM with HP-β-CD for mixtures prepared using the physical mixture (PM) method and those prepared using the solution mixture (SM) method.⁵⁰ We considered that these complexes contained independent molecules with single polymorphs. Therefore, for transformation among the polymorphs, we analyzed their DSC thermograms using rough approximations.

This linear algebra approach may not provide the same level of advanced prediction and decision-making capabilities as AI, while it does offer chemists a comprehensive way to identify the latent pure entities that contribute to the apparent patterns of various states.^{34–38} This approach offers a more explainable and traceable (but quite simple) alternative to deep learning procedures that typically rely on complicated mathematical approaches. The induced basis functions help us understand why an unknown combination of components is expected to derive its own XRPD pattern. Each basis function corresponds to a curated pure molecular entity or analytically distinguishable complex. The decision-making process involved in arriving at these expectations is similar to the deductive thinking used by chemists, with the logic built using a linear combination of observable basis functions. While



the predicted patterns may appear complex, these methods are only slightly different from general analyses that rely on calibration and measurement (it means about procedures, not evaluation of data).

2. Experimental section

2.1. Materials

DCF acid and its sodium salt were purchased from Tokyo Chemical Industry (Tokyo, Japan). INM, FAM, CIM, and β -cyclodextrin (β -CD; a cyclic heptamer of glucosides) were obtained from Fujifilm Wako Pure Chemical Corporation (Osaka, Japan). All other commercially available materials and solvents were of analytical reagent grade.

Two mixing methods were used to produce the API/ β -CD mixtures: physical mixing (PM) and solvent mixing (SM). The PM method involved kneading the mixed API and β -CD with an agate mortar and pestle. Neat API and plain β -CD samples were used after grinding, similar to the PM preparation. The PM-prepared mixtures were produced as an equimolar mixture of the API and β -CD. Meanwhile, the SM method involved mixing an ethanol solution of completely dissolved API and β -CD, removing the ethanol by evaporation, and then drying in a vacuum desiccator at 298 K overnight to obtain a powder. The SM-prepared mixtures were provided in molar ratios (mole fraction of API) of 1:0 (100%), 3:1 (75%), 2:1 (67%), 1:1 (50%), 1:2 (33%), 1:3 (25%), and 0:1 (0%). Because the SM-prepared INM/ β -CD mixture induced the α -form of INM,⁵⁰ the neat INM dried in a vacuum was used as the γ -form and the α -form crystal was yielded by recrystallization with ethanol, evaporating overnight at a temperature of 278 K.

2.2. Thermal analysis of neat API and API mixture with β -CD

Differential scanning calorimetry (DSC) was used to measure the thermal behavior of neat API and API mixture with β -CD. The DSC instrument (DSC8230, Rigaku Co., Tokyo, Japan) was used to scan the samples from 303 K to 453 K at a rate of 5.0 or 10.0 K min⁻¹, under a nitrogen gas flow of 30 mL min⁻¹. The scanning range for DCF was 303–463 K. T_m , which is the melting start temperature, was obtained using Thermo Plus 2 software (Rigaku Co., Tokyo, Japan) by measuring the intersection of the baseline extension and the maximum slope point of the peak. If the thermogram showed a simple endothermic peak, the area enclosed by the endothermic curve and the baseline was converted to the total melting enthalpy ($\Delta_{\text{fus}}H$) for a given mass amount of the component. The total melting entropy ($\Delta_{\text{fus}}S$) was simultaneously approximated using the quotient of $\Delta_{\text{fus}}H$ divided by T_m , according to the classical definition of Clausius.

2.3. Nuclear magnetic resonance spectroscopy

The ¹H-NMR measurements were performed at 400 MHz using LA-400 (JEOL Ltd). The samples were dissolved in either deuterated DMSO (DMSO-*d*₆) or deuterated water (D₂O) at a temperature of 298 K. The chemical shift of the solvent is used as a reference point (DMSO-*d*₆ 2.500 ppm, D₂O 4.800 ppm). The

ppm of the solvent was calibrated with tetramethylsilane (TMS) set as the origin.

2.4. XRPD diffraction of API and API mixture with β -CD

XRPD pattern measurements were used to identify the polymorph of neat API and API mixture with β -CD. RINT 2000 instrument (Rigaku Co., Tokyo, Japan) was used with a Cu K α radiation source at a voltage of 40 kV and a current of 40 mA, filtered by the Ni filter to produce monochromatic radiation. The X-ray irradiation was performed using the parallel-beam method in a 2 by theta (2θ) range from 5 to 40 degrees at a scanning rate of 0.02 degree steps. 2θ from 5 to 40 degrees is customary in pharmaceutical analysis.⁵¹ Angles smaller than 5 degrees fall within the range of X-ray Small-Angle Scattering (SAXS) measurements, which offer distinct information from the molecular-level details obtained from XRPD.⁵² Whereas, the use of high theta (higher angle than 40 degrees) data is used to obtain resolution at a small distance and to obtain total scattering profiles (such as pair distribution functions, *i.e.*, PDF, or similar Diff(*r*) functions), that are very useful in materials science. This could more easily be achieved with molybdenum, silver radiation, or synchrotron. In this study, by focusing on such commonly used measurement parameters, we have avoided unnecessary complexities in interpretation. The spectra were presented as the average of five scans, and the scanning sequences were conducted in triplicate or more times. The samples were crushed in an agate mortar and pestle, and then the powders were mixed. The diffractograms obtained by XRPD were published in the ESI,† TXT (406 K) and TXT (304 K), which were described in comma separated values (CSV).

To identify the polymorph, we compared the observed diffractograms of single-crystal structures of the API with reference ones. They were obtained by converting 3D coordinates using the reflex module of powder diffraction on BIOVIA/ACCELRYs Materials Studio 2022 (Dassault Systems) and calculating the Miller indices of conspicuous peaks. The 3D crystalline coordinates were retrieved from the Cambridge Crystallographic Data Centre (CCDC).

For INM, we surveyed the most stable γ -form and metastable α -form from the CCDC reference codes INDMET (1972) and INDMET04 (2011), respectively. DCF was found in three polymorphs: the most stable HD2-form retrieved as SIKLIH (1990), metastable HD1-form as SIKLIH02 (1997), and metastable HD3-form as SIKLIH04 (2001).

CIM contained four forms: the A-form as CIMETD02 (1979), the B-form as CIMETD06 (2019), the C-form as CIMETD04 (2013), and the z-form as CIMETD01 (1984). FAM obtained two forms: its A-form as FOGVIG01 (1989) and B-form as FOGVIG02 (2002).

As previously reported, the CIM complexes with β -CD and γ -CD (the cyclic octamer of glucosides) formed the polymorph of channel form, while the non-complexed CDs and the CIM complex with α -CD (the cyclic hexamer of glucosides) induced the structure of the cage form, which was confirmed with the simulated pattern from BCDEXD03 (1994).⁵¹ The channel form was verified with the diffraction angles of the *para*-hydroxybiphenyl complex with β -CD from the reference code OFAXID (2007).



2.5. Singular value decomposition (SVD) procedure applied to diffractograms generated by XRPD analysis

The observed i -th diffractogram $\{\vec{\Phi}_i | 1 \leq i \leq n\}$ of the sample was represented as an m -dimensional vertical vector measured at the specific 2θ . Since 2θ is measured over a range of 5–40 degrees with an interval of 0.02 degrees, the value of m is 1751. Matrix \mathbf{M} was composed of a horizontal sequence of vectors from the first diffractogram vector and the i -th diffractogram vectors, with an $m \times n$ rectangular matrix defined as eqn (1):

$$\mathbf{M} = \begin{pmatrix} \vec{\Phi}_1 & \vec{\Phi}_2 & \dots & \vec{\Phi}_n \end{pmatrix} \\ = \begin{pmatrix} \phi_1(2\theta_1) & \phi_2(2\theta_1) & \dots & \phi_n(2\theta_1) \\ \phi_1(2\theta_2) & \phi_2(2\theta_2) & \dots & \phi_n(2\theta_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(2\theta_m) & \phi_2(2\theta_m) & \dots & \phi_n(2\theta_m) \end{pmatrix} \quad (1) \\ \in \mathbb{R}^{m \times n}$$

Let \mathbf{M} and \mathbf{M}^t be real and transposed matrices, respectively. Their products $\mathbf{M}^t\mathbf{M}$ and $\mathbf{M}\mathbf{M}^t$ become orthogonal matrices, and the rows of Ψ and Λ are the left and right singular vectors, respectively. The matrices describing \mathbf{M} can be transformed into eqn (2):

$$\mathbf{M} = \Psi \Sigma \Lambda^t \\ = \Psi \begin{pmatrix} \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_\rho) & O_{\rho, n-\rho} \\ O_{m-\rho, \rho} & O_{m-\rho, n-\rho} \end{pmatrix} \begin{pmatrix} \vec{\lambda}_1 & \vec{\lambda}_2 & \dots & \vec{\lambda}_n \end{pmatrix}^t \\ \approx \Psi \begin{pmatrix} \sigma_1 \vec{\lambda}_1 & \sigma_2 \vec{\lambda}_2 & \dots & \sigma_r \vec{\lambda}_r \end{pmatrix} \quad (2)$$

The diagonal matrix Σ contains the diagonal elements $\{\sigma_i | 1 \leq i \leq \rho\}$ that have positive real values ordered in a descending order. These elements represent the singular values, which indicate dispersion. The i -th column of the orthogonal matrix Λ is the coefficient vector corresponding to the singular value σ_i , and vector $\vec{\lambda}_i$ is called a specific singular vector. The rows of the matrix Ψ are denominated as basis function vectors.^{39–48} The principal component vector \vec{w}_i is the coefficient vector $\vec{\lambda}_i$ multiplied by the corresponding singular value σ_i .

$$\mathbf{M} = \Psi \begin{pmatrix} \sigma_1 \vec{\lambda}_1 & \sigma_2 \vec{\lambda}_2 & \dots & \sigma_r \vec{\lambda}_r \end{pmatrix} \\ = \Psi \begin{pmatrix} \vec{w}_1 & \vec{w}_2 & \dots & \vec{w}_r \end{pmatrix} \quad (3)$$

Matrix Ψ has rows that are the basis function vectors. In this study, we applied SVD to a matrix of 1751×35 and 1751×26 (2θ : 5–40 degrees) diffractogram data. Matrix Ψ has rows that are the basis function vectors. From the diagram representing the logarithm of the singular value in descending order against the index corresponding to the documental diffractogram, we practically determined the dimensionality, *i.e.*, the minimum

dimensionality of the basic functions required to reproduce the vector space of the documental diffractogram. It might be practically negligible, with a singular value less than several hundredths of the highest singular value of the first principal components. In other words, components with such too small singular values hardly contribute to the reproduced spectrum. Owing to the dimensionality r selected by applying this criterion instead of the mathematical rank ρ , the yielded principal components approximately reproduce the vector space including the documental diffractogram as the j -th feature vector \vec{x}_j composed of the i -th elements $x_{i,j}$:

$$x_{i,j} \approx \sum_{k=1}^r \psi_{i,k} \sigma_{k,j} \lambda_{k,j} \\ = \psi_{i,1} \sigma_{1,j} \lambda_{1,j} + \psi_{i,2} \sigma_{2,j} \lambda_{2,j} + \dots + \psi_{i,r} \sigma_{r,j} \lambda_{r,j} \quad (4)$$

Principal component analysis (PCA)⁴⁷ was performed on $\sigma_i \lambda_i$ extracted by SVD. The 3D plots of principal component (PC) vectors are shown in the Supplementary Movie (ESI[†]).

3. Results

3.1. Preparation of acidic API complexes with β -CD and their diffractograms

The SM-prepared INM/ β -CD mixtures at molar ratios of 3:1, 2:1, 1:1, 1:2, and 1:3, the SM-treated INM (at a molar ratio of 1:0), the neat INM (γ -form), the α -form INM crystal, the SM-treated β -CD (0:1), and the plain β -CD were prepared as described in the Experimental section. Their DSC thermograms are shown in Fig. 1A. The thermogram of the SM-prepared INM/ β -CD equimolar mixture contained two endothermic troughs at the dropping temperature of 428 K and 434 K and their intermedial peak. The left and right troughs were corresponding to the signal of the neat INM and that of the α -form INM crystal, respectively. The intermedial peak indicated the phase transition from the metastable α -form to the most stable γ -form.⁵⁰ The SM-prepared INM/ β -CD equimolar mixture showed the broad signal of the cage form of β -CD at about 380–400 K and the sharp signal of the α -form INM at the dropping temperature of 428 K. Compared to the thermogram of the PM-prepared INM/ β -CD equimolar mixture, the INM/ β -CD inclusion complex was partially obtained, but the crystalline INM of α -form would remain.

The corresponding results were obtained in the XRPD diffractograms in Fig. 2A. Signals from both the α -INM crystal and the plain β -CD were observed in the diffractogram of the SM-prepared INM/ β -CD equimolar mixture. The diffractogram of neat INM was confirmed as the most stable γ -form with the simulated pattern in Fig. S2A (ESI[†]). Its signals did not appear any patterns of the SM-prepared mixture at various molar ratios and the PM-prepared equimolar mixture.

In Fig. 1B, the DSC thermograms of the DCF/ β -CD mixtures show the endothermic signal at a dropping temperature of about 445–447 K except for that of the plain β -CD (0:1).⁴⁶ Their XRPD diffractograms are shown in Fig. 2B. The diffractogram



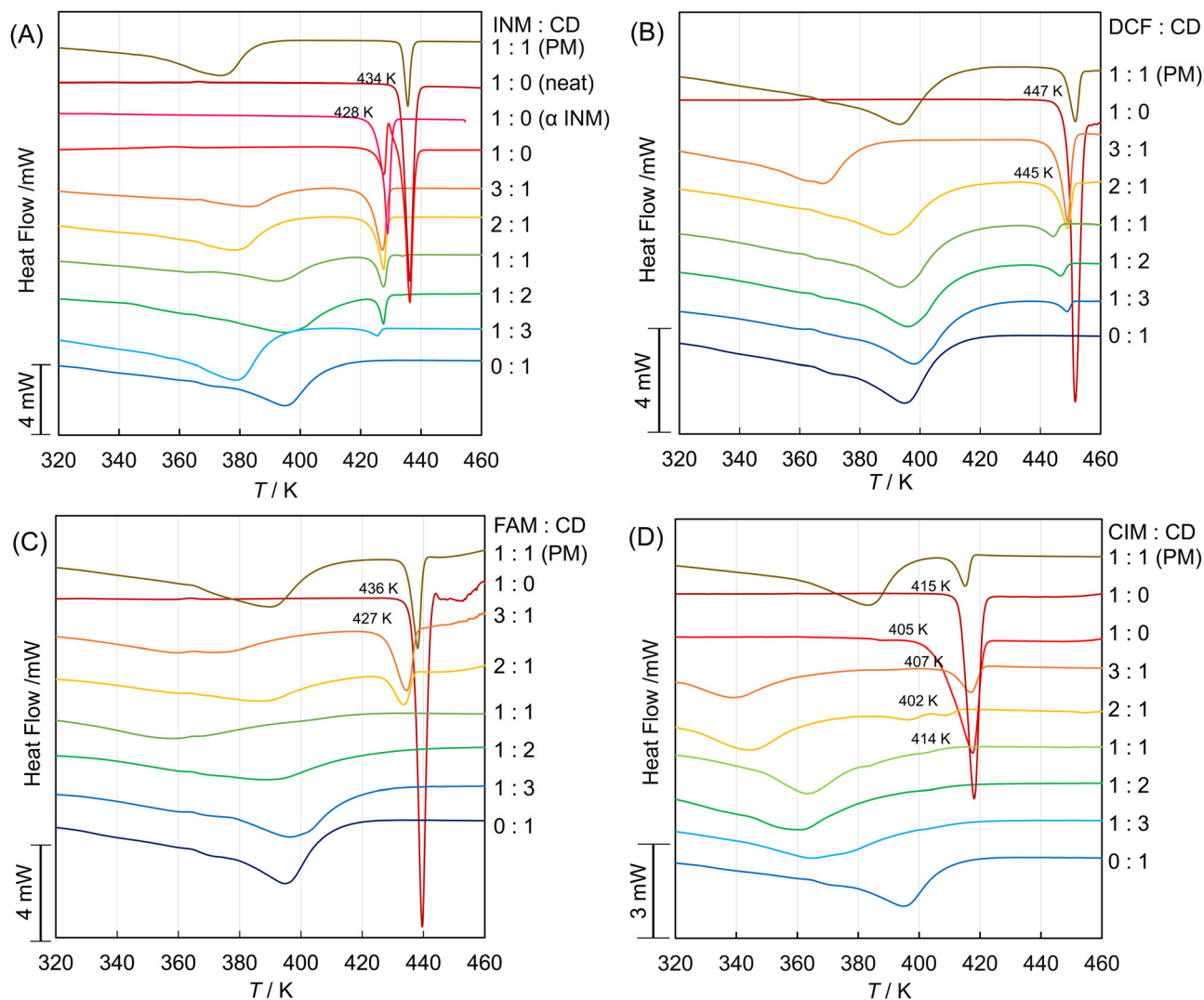


Fig. 1 The DSC thermograms of APIs and their mixtures: (A) those of PM-treated INM/ β -CD (brown), neat INM (burgundy), α -INM (magenta), the SM-prepared INM/ β -CD mixtures at various molar ratios, and neat β -CD (indigo); (B) those of PM-treated DCF/ β -CD (brown), neat DCF (burgundy), the SM-prepared DCF/ β -CD mixtures at various molar ratios, and neat β -CD (indigo); (C) those of PM-treated FAM/ β -CD (brown), neat FAM (burgundy), the SM-prepared FAM/ β -CD mixtures at various molar ratios, and neat β -CD (indigo); and (D) those of PM-treated CIM/ β -CD (brown), neat CIM (burgundy), the SM-prepared CIM/ β -CD mixtures at various molar ratios, and neat β -CD (indigo). The neat INM has the polymorph of γ -form, while the crystal of α -form was obtained via evaporation of ethanol solution.

of neat DCF was verified as the most stable form HD2 with the simulated patterns in Fig. S2B (ESI[†]). Its signals were observed in the patterns of the DCF/ β -CD mixtures. It indicated that the HD2 crystal was included in the SM-prepared mixtures at various molar ratios and the PM-prepared equimolar mixture.

The ¹H-NMR spectra of plain β -CD, neat DCF, and their equimolar mixtures in DMSO-*d*₆ and D₂O were measured, as shown in Fig. S3A–C and S4A–C (ESI[†]). Although the signals of DCF showed few differences in the absence and presence of β -CD in deuterated DMSO, they shifted from neat DCF to its β -CD mixture in deuterated water. The doublet and double-triplet signals in the 2,6-dichloroaniline moiety and the double-triplet signal of the para-positioned proton in the phenylacetic acid moiety shifted to a lower magnetic field at about 0.01 ppm. This shift verified that the 2,6-dichloroaniline moiety partially intruded into the shielded interior cavity of β -CD. Since the

H5 multiplet signal in the interior cavity of β -CD was superimposed onto the adjacent H6 double-doublet signal, its accurate chemical shift and coupling constants were unable to be determined. The H3 triplet signal in the interior cavity of β -CD provided only an insignificant difference in chemical shift because of the thermal perturbation of seven glucoside units attenuating the shift to 1/7 intensity as a time average.

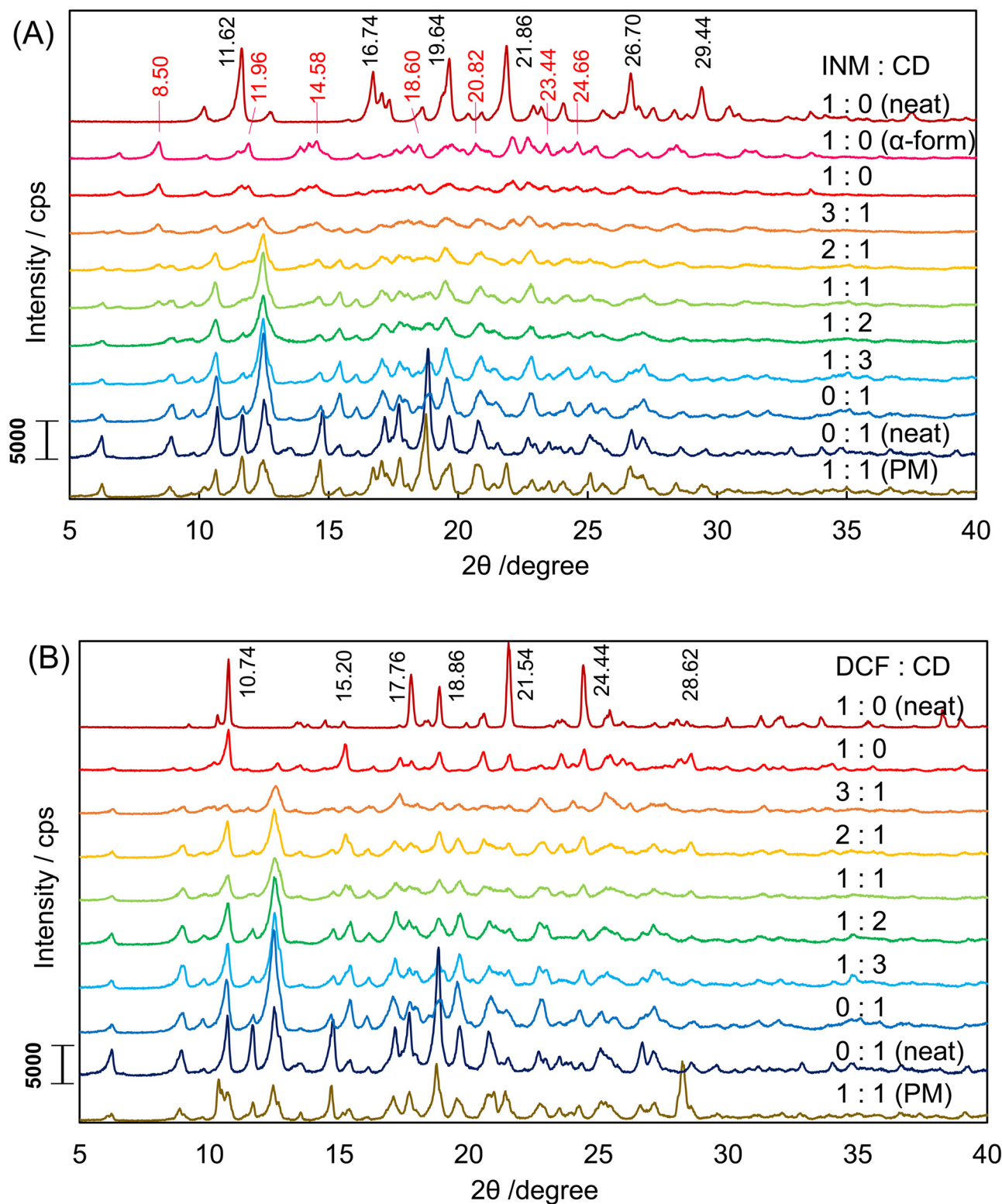
3.2. Preparation of basic API complexes with β -CD and their diffractograms

The DSC thermograms of FAM and CIM are shown in Fig. 1C and D. Their equimolar mixtures with β -CD demonstrated no endothermic signals that appeared in the thermograms of the neat FAM and the neat CIM. The XRPD diffractograms of the SM-prepared FAM/ β -CD mixture at various molar ratios, the PM-prepared FMA/ β -CD equimolar mixture, the neat FAM,



and the plain β -CD are shown in Fig. 2C. The SM-prepared DCF/ β -CD equimolar mixture had the halo pattern. This is interpreted as indicating that the crystal structure was completely destroyed by the formation of an inclusion complex with cyclodextrin. The diffractograms of the neat DCF and the DCF

mixtures at molar ratios of 1:0, 3:1, and 2:1 were confirmed as containing the A-form of FAM with the patterns shown in Fig. S2C (ESI[†]). On the other hand, the diffractograms of the DCF mixtures at molar ratios of 1:2, 1:3, and 0:1 had the signals corresponding to those obtained from the plain β -CD



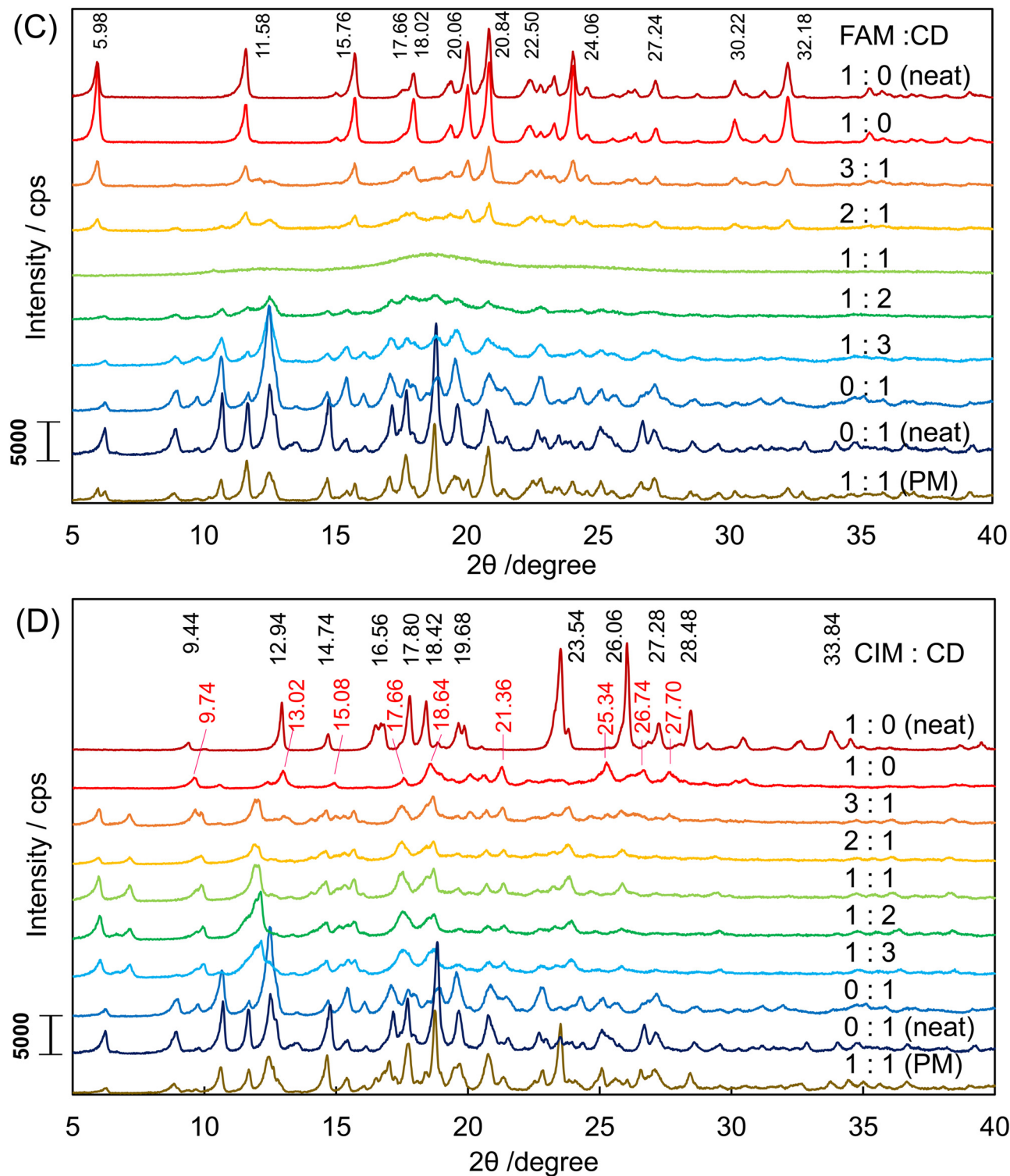


Fig. 2 The diffractograms of APIs and their mixtures: (A) those of neat INM (burgundy), α -INM (magenta), the SM-prepared INM/ β -CD mixtures at various molar ratios, neat β -CD (indigo), and PM-treated INM/ β -CD (brown); (B) those of neat DCF (burgundy), the SM-prepared DCF/ β -CD mixtures at various molar ratios, neat β -CD (indigo), and PM-treated DCF/ β -CD (brown); (C) those of neat FAM (burgundy), the SM-prepared FAM/ β -CD mixtures at various molar ratios, neat β -CD (indigo), and PM-treated FAM/ β -CD (brown); and (D) those of neat CIM (burgundy), the SM-prepared CIM/ β -CD mixtures at various molar ratios, neat β -CD (indigo), and PM-treated CIM/ β -CD (brown). The neat INM has the polymorph of γ -form, while the crystal of α -form was obtained via evaporation of ethanol solution.



(neat CD). It indicated that the DCF/ β -CD mixtures were stoichiometrically formed in the inclusion complex at a molar ratio of 1 : 1.

The $^1\text{H-NMR}$ spectra of the neat FAM and the FAM/ β -CD equimolar mixtures in deuterated DMSO and deuterated water were measured, as shown in Fig. S3D, E, S4D and E (ESI †). The signals of FAM shifted from neat FAM to the mixture in deuterated water. The singlet signal of 4-(2-guanydy-1,3-thiazolyl)-methylene protons shifted to a lower magnetic field at about 0.10 ppm. The triplet signal of the 2-positioned aliphatic proton in the 1-aminosulphonylimino-propylamine moiety shifted to a lower magnetic field at about 0.03 ppm. As the signals of amino-protons vanished in deuterated water, the orientation of FAM intruding into β -CD was not specified. It suggested that even the sophisticated measurements (COSY and ROESY) are not expected to determine the interaction between FAM and β -CD in protic solvents. We speculate that the complexed FAM molecule penetrates the β -CD ring.

As shown in Fig. 2D, the diffractogram of the neat CIM was obtained, confirmed as the A-form from comparison to the simulated patterns in Fig. S2D (ESI †). But the SM-prepared CIM had its diffractogram inconsistent with that of the A-form, considering that it transformed to the B-form, which contains signals at 2θ angles of 9.60, 12.38, 14.88, 17.52, 18.38, 21.24, 25.12, 26.48, and 27.50 degrees in Fig. S2D (ESI †). The SM-prepared CIM/ β -CD mixture at a molar ratio of 3 : 1 maintained the B-form signals. The greater the molar ratio of β -CD in the SM-prepared CIM mixtures, the more the B-form signals diminished. The diffractograms of the SM-prepared CIM/ β -CD mixtures contained no halo pattern and were different from that of the plain β -CD (neat). As shown in Fig. 2A–C, the diffractograms of the plain β -CD (neat) and the SM-treated β -CD (0 : 1) are similar to each other. Their signals accorded to the simulated pattern of the cage form from CCDC 3D coordinates, as shown in Fig. S2E (ESI †). Both of the plain β -CD (neat) and the SM-treated β -CD (0 : 1) were relationships of crystalline habits, which contained signals at the same angle but with different intensities.

The SM-prepared CIM/ β -CD equimolar mixtures' diffractogram corresponded to the channel form's simulated pattern from CCDC 3D coordinates, as shown in Fig. S2F (ESI †). Namely, the CIM/ β -CD complex formed the polymorph different from the plain β -CD.⁵¹ Such channel form was not found in the INM/ β -CD, DCF/ β -CD, and FAM/ β -CD mixtures. The signal at the two by theta angle of 12.02 degrees would be of Millar's indices of $hkl = 002$ (corresponding to the signal at the two by theta angle of 12.18 degrees in the reference diffractogram of the channel form). This third axis C of crystalline coordinates was coincident with the rotational symmetry axis of the cylindrical β -CD molecule. The intensity of the diffraction at this angle increased, depending on the decrease of CIM or the increase of β -CD in the mixtures. Hence, it appeared that the crystal of cylindrical β -CD grew along the rotational symmetry axis with the increasing proportion of β -CD. We speculate that a small amount of the CIM molecule acts as a glue among the β -CD molecules to accumulate the channel cylinders.

Eventually, eight diffractograms of the distinguishable polymorphs were obtained as the γ -/ α -forms of INM, the HD2 form

of DCF, the A-form of FAM, the A/B-forms of CIM, and the cage-/channel-forms of β -CD in the XRPD measurements of 35 as shown in Fig. 2A–D. In the present study, we dealt with these diffractograms to correspond to the information reproducing the individual chemical entities.

3.3. The basis functions of SVD analysis for diffractograms of the API/ β -CD complexes

To distinguish the deductive components of chemical entities, *i.e.*, eight polymorphs (and their crystalline habits, by chance), found in the API/ β -CD mixtures of 35, their diffractograms combined as column vectors in the matrix \mathbf{M} were treated by the procedure of the SVD method. The obtained singular values σ_i in the orthogonal matrix were plotted according to the rank on the common logarithm grid sorted in descending order, as shown in Fig. 3A. The first component showed the highest contribution of 33.7%. The cumulative contribution for the seventh component grew to 71.3%, and that for the eighth one to 75.1%. As the eighth σ plot seemed to be isolated from the adjacent plots of the previous and subsequent components, we retained not only the case of eight components but also that of seven components to be analyzed ($r = 7, 8$).

Fig. 3B shows their basis functions as the singular vectors $\{\Psi_i | 1 \leq i \leq r\}$, in which distinctive peaks were observed. The individual basis functions are shown in Fig. S5 (ESI †). In Fig. S5A (ESI †), the first basis function Ψ_1 has negative peaks and its signals were agreeable with the diffractogram of the cage form in the plain β -CD. In general, the first basis function corresponded to the average of all samples, so it often included little individual specificity. In Fig. S5B and C (ESI †), the second basis function Ψ_2 has positive peaks, which corresponded to the diffractogram of the β -CD cage-form, and negative peaks, which corresponded to the XRPD pattern of FAM. Since the intensity of the negative peaks of the Ψ_1 is almost half of that of the negative peaks of the Ψ_2 , the angle of the extracted pattern of the β -CD cage-form was about -120 degrees ($\cos -120^\circ = -0.5$) to the plane of the Ψ_1 and was about $+150$ degrees ($\cos 150^\circ = -0.866$) to the perpendicular plane of the Ψ_2 . It meant that the plane of the pattern for the β -CD cage form made a dihedral angle of about 150 degrees to the plane of the pattern for FAM.

In Fig. S5D and E (ESI †), the third basis function Ψ_3 has positive peaks, which corresponded to the diffractogram of the CIM A-form, and negative peaks, which corresponded to the XRPD pattern of FAM. In Fig. S5F (ESI †), the fourth basis function Ψ_4 has positive peaks, which corresponded to the diffractogram of the INM γ -form, and no negative peaks. In Fig. S5G and H (ESI †), the fifth basis function Ψ_5 has positive peaks, which corresponded to the diffractogram of the β -CD channel form, and negative peaks, which corresponded to the XRPD pattern of FAM. In Fig. S5I (ESI †), the sixth basis function Ψ_6 has negative peaks, which corresponded to the diffractogram of the CIM B-form. In Fig. S5J (ESI †), the seventh basis function Ψ_7 has negative peaks, which corresponded to the diffractogram of the DCF HD2-form. In Fig. S5K (ESI †), the eighth basis function Ψ_8 gave no distinctive peaks resembling



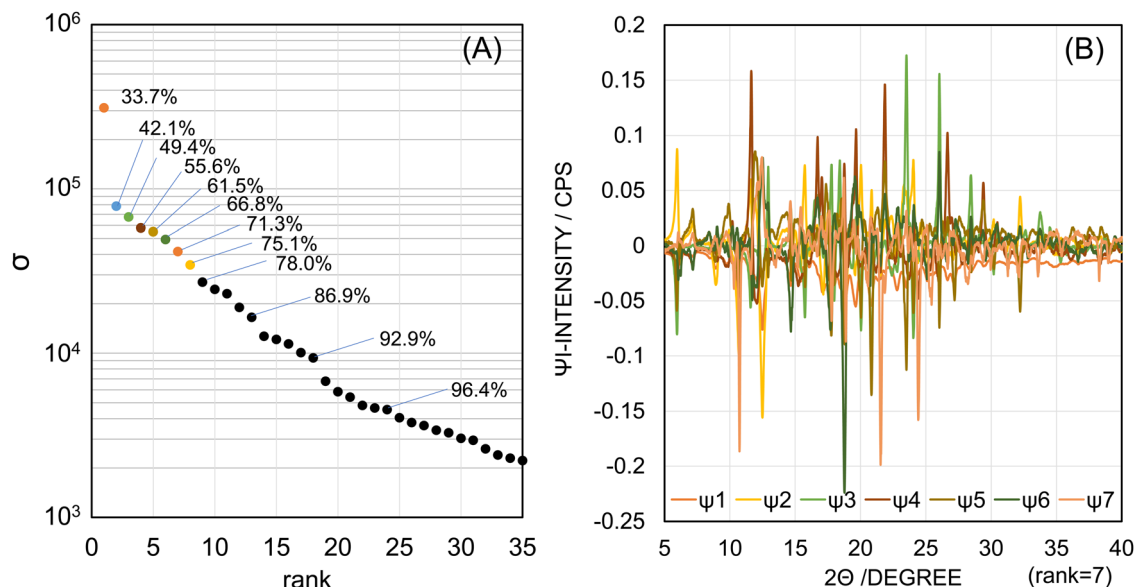
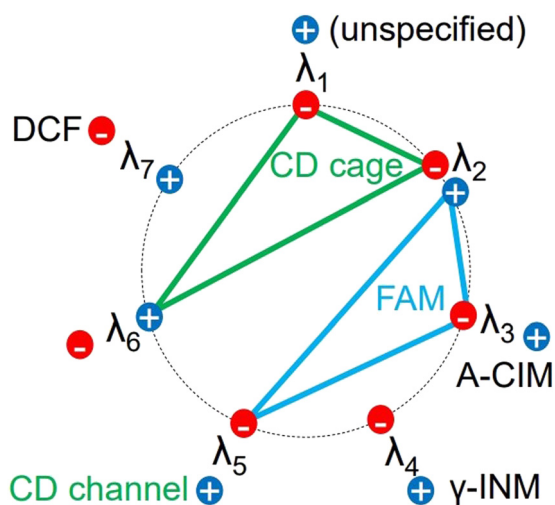


Fig. 3 The SVD analysis for the matrix with 35 diffractograms of the API/ β -CD mixtures: (A) the singular values and (B) the orthogonal basis functions of the obtained components.



Scheme 2 Summarized scheme of the obtained components by the SVD analysis for the matrix with 35 diffractograms of the API/ β -CD mixtures.

the diffractograms of any polymorphs comparatively to other basis functions (Ψ_1 – Ψ_7). This illustrated Ψ_8 to be not essential to reproduce the observed patterns. Therefore, we determined $r = 7$.

The correspondence of the basis functions to the observed diffractograms of the polymorphs is summarized in Scheme 2. The diffractogram of the β -CD cage-form was divided into the negative peaks of Ψ_1 and Ψ_2 as described above. In this regard, the signals for the β -CD cage form were not yet observed in the positive peaks in Ψ_6 . The diffractogram of FAM seemed to be divided into the positive peaks of Ψ_2 and the negative peaks of Ψ_3 and Ψ_5 . From the above results, we found no basis functions indicating the INM α -form. As shown in Fig. S6 (ESI[†]), the signals depending on the INM α -form were not found in Ψ_5 and other basis functions.

3.4. The coefficient vectors for diffractograms of the API/ β -CD complexes

The basis functions presented the hyperdimensional “rulers” to translate the observed diffractograms to the linear combination of the pattern elements that correspond to the consisting entities. Subsequently, to verify the actual representations with the specific coefficient vectors in the right singular matrix, Fig. 4 shows the components $\{\lambda_{ij}|1 \leq i \leq r, 1 \leq j \leq n\}$ of the j -th specific coefficient vector plotted as a function of the mole fractions of the APIs for the SM-prepared API/ β -CD mixtures. The first component λ_1 (hereafter, we omit subscript j) monotonically increased in all the API/ β -CD mixtures. The second component λ_2 changed parallel to λ_1 , but that in the FAM/ β -CD mixture increased beyond zero of the ordinate in Fig. 4C. It reflected that the positive peaks in the second basis function Ψ_2 represented the diffractogram of FAM. In Fig. 4D, λ_2 increased in the CIM/ β -CD mixture, which has the channel form of β -CD. It indicated that the cage form of β -CD vanished in the CIM/ β -CD mixtures.

In association, the fifth component λ_5 showed a positive plateau just for the CIM/ β -CD mixture in Fig. 4D. It was reasonable because the positive peaks of the fifth basis function Ψ_5 represented the diffractogram of the channel form of β -CD. As the negative peaks of Ψ_5 corresponded to the pattern of FAM, λ_5 decreased in the high mole fraction of FAM in Fig. 4C. Suppose that the pattern of FAM was considered to be divided into the basis functions Ψ_5 (negative peaks), Ψ_3 (negative peaks), and Ψ_2 (positive peaks), it was comprehensible that the components λ_3 and λ_2 simultaneously decreased and increased, respectively.

The third component λ_3 might have positive peaks indicating the contribution of the CIM A-form, concurrently with its negative peaks reflecting the diffractogram of FAM. In Fig. 4D,



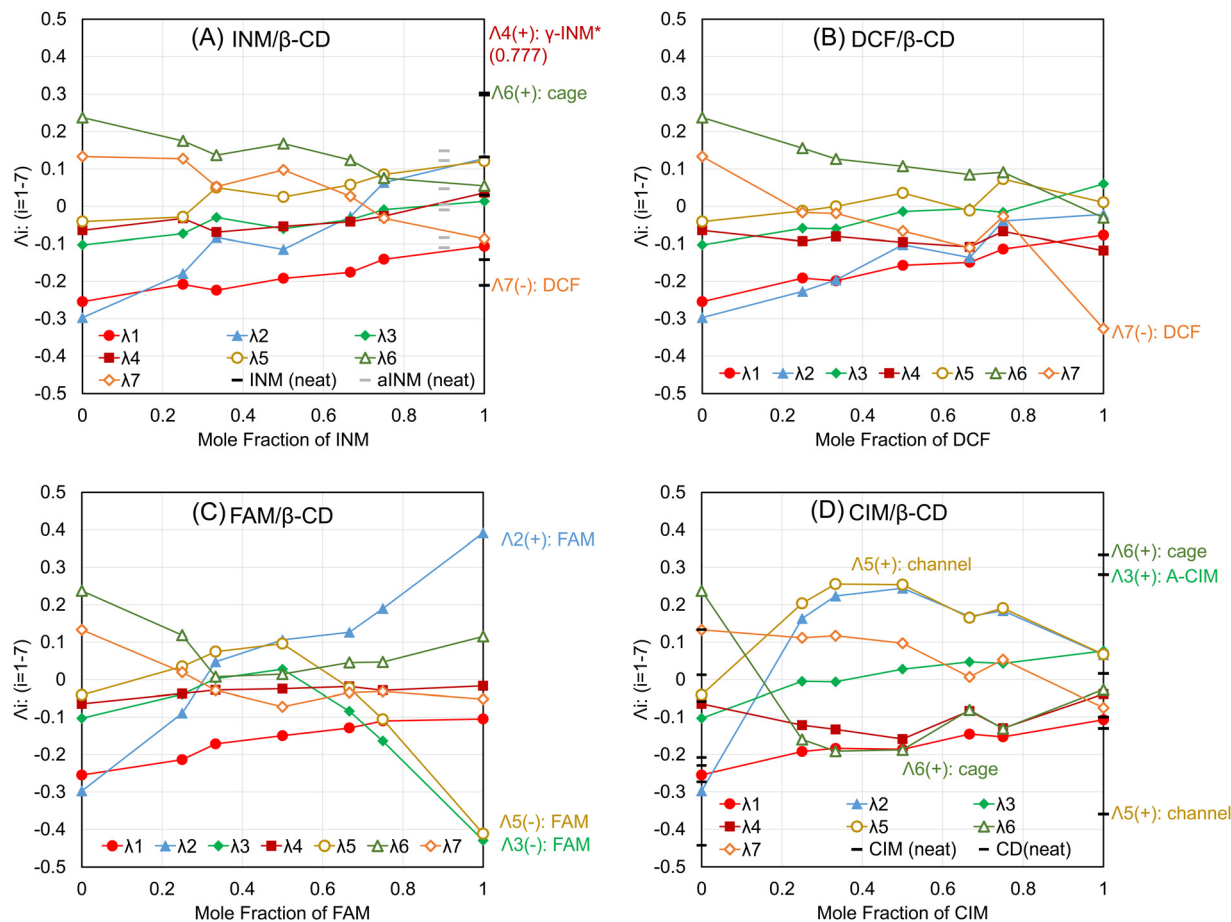


Fig. 4 The specific coefficient vectors for the (A) INM/ β -CD mixtures, (B) DCF/ β -CD mixtures, (C) FAM/ β -CD mixtures, and (D) CIM/ β -CD mixtures.

the specific coefficient vector of the neat CIM (A-form) was expressed as bars at a mole fraction of 100%, overlapped with the plots of the vector components of the SM-treated CIM (B-form). The λ_3 of the neat CIM was a positively large contribution. The fifth component λ_5 and the sixth component λ_6 of the neat CIM were negatively and positively large, respectively. Assume the positive peaks of λ_5 corresponded to the channel form of β -CD and the positive peaks of λ_6 demonstrated the cage form of β -CD; these two components λ_5 and λ_6 would represent that the neat CIM included neither channel nor cage forms of β -CD. On the other hand, the vector components of the SM-treated CIM (B-form) resulted in small values under ± 0.1 on the ordinary axis. In Fig. S5I (ESI[†]), the sixth basis function Ψ_6 had the negative peaks of four corresponding to the diffractogram of the CIM B-form, but the intensities of its matched peaks were all small. It was reviewed to contain little evidence that λ_6 correlated to the diffractogram of the CIM B-form (denied in Scheme 2). The results of the SM-treated CIM in Fig. 5D suggested the positive peaks of Ψ_6 to partially represent the contribution of the cage form of β -CD. The positive peaks at the angles of 12.54 and 19.64 degrees were considered to share the contribution of the diffractogram of the cage form of β -CD, as shown in Fig. S5I (ESI[†]).

The fourth and seventh components λ_4 and λ_7 were characterized as representing the contributions of the INM γ -form

and DCF, respectively. λ_4 nearly unchanged depending on the mole fraction of APIs in the SM-prepared mixtures, as shown in Fig. 4A–D. Its large value appeared in the neat INM (γ -form), the specific components of which overlapped with those of the SM-treated INM plotted at a mole fraction of 100% in Fig. 4A. As its value of the λ_4 of the neat INM was 0.777, its plot bar was beyond the ordinate axis region. It was concluded that the basis function Ψ_4 and the specific coefficient component λ_4 independently expressed the diffractogram and its contribution to the INM γ -form. λ_7 showed a little positive value at the lower mole fraction of API in the SM-prepared mixtures. The large value of λ_7 was obtained in the neat DCF positioned at a mole fraction of 100% in Fig. 4B. As the negative peaks of Ψ_7 corresponded to the diffractogram of DCF, the neat DCF showed a negatively large value. As shown in Fig. S5J (ESI[†]), the positive peaks at angles of 12.48 and 18.72 degrees were obtained, contributing very slightly to representing the cage form of β -CD.

The results are summarized in Scheme 2, but we were unable to obtain the basis function corresponding to the diffractograms of the INM α -form and the CIM B-form in the eight independent polymorphs. In Fig. 4A, the specific coefficient components of the recrystallized α -form of INM are expressed with gray bars at a mole fraction of 90%. All absolute intensities of the components were less than ± 0.15 , so no basis



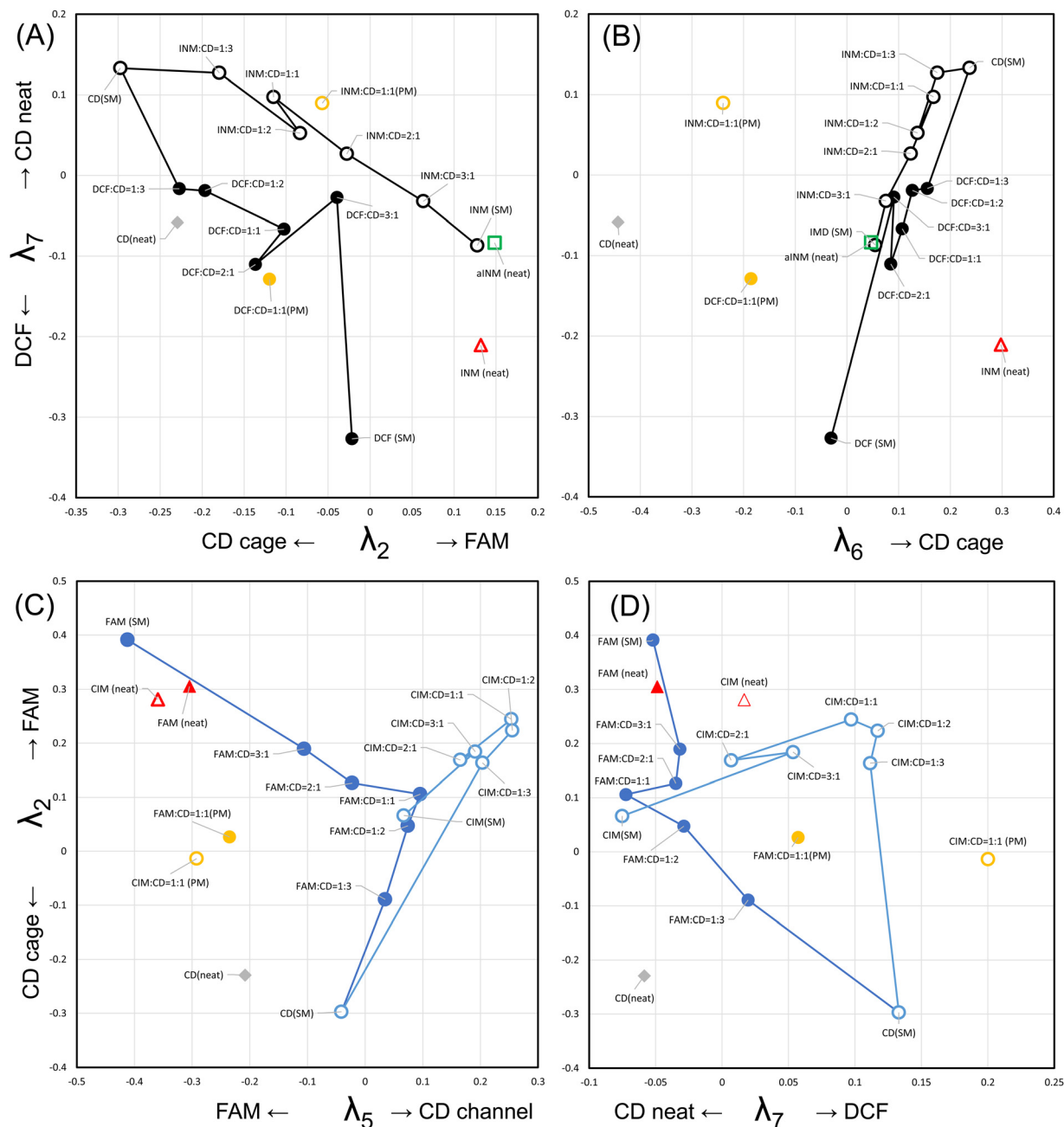


Fig. 5 The trajectories of the interconversion pathways for the (A) DCF/ β -CD and INM/ β -CD mixtures and (B) FAM/ β -CD and CIM/ β -CD mixtures. See Supplementary Movies, MP4 (7273 K) and MP4 (8521 K) (ESI[†]) of the three dimensional matrix.

functions reflected the diffractogram of the INM α -form. Furthermore, no basis functions were obtained also as a corresponding pattern to the diffractogram of the CIM B-form. Similar analyses were performed for $r = 8$, but the full set of basis functions corresponding to the diffractograms of eight polymorphs could not be obtained. In Fig. 3A, the components with the higher σ in the rank of 13 ($r = 13$) showed a cumulative contribution of 86.99%, and those in the rank of 18 ($r = 18$) showed a cumulative contribution of 92.9%. By assembling such components, the represented patterns were

refined to the corresponding diffractograms. However, the obtained variations of basis functions were not improved. The use of more basis functions to represent the experimental diffractograms could increase the reproducibility of the generated XRPD patterns. However, chemists do not expect the observation of experimental data to be refined. The aim should be to clarify the factors that independently make up the phenomena. In the present work, the number of independent entities was six, which includes the INM γ -form, DCF, FAM, the CIM A-form, the β -CD cage form, and the β -CD channel form.



3.5. The trajectory of the interconversion pathway in diffractograms of the API/ β -CD complexes

Although we cannot directly recognize the hyperdimensional vectors and their polytope, projecting them onto a three-dimensional space can help us imagine their geometrical properties.^{52–54} To achieve this, we decided on the combination of three axes in the projected space based on the wide distributions (standard deviations) of the plots along the axes (see Fig. 5A and B). Also, the 3D plots of PC vectors obtained by performing PCA on $\sigma_i\lambda_i$ extracted by SVD are shown in the Supplementary Movie, MP4 (7273 K) (ESI[†]). We visualized the linear polygonal lines of the crystalline states in the projected Cartesian space and traced the interconversion pathway from the state of SM-treated β -CD to that of the α -form of INM (open circles).⁵⁴

However, as strains of the coefficient vectors were conspicuous in the projected diagrams (see Fig. 4A and B), the observed linearity might be hard to distinguish, and the pathways seemed to meander. Therefore, in lower-dimensional projections, the strains would be often emphasized, just as a 3D perfect circle looks like an ellipse on a planar projection. The coordinates for the DCF mixture were $\lambda_2 = -0.062$, $\lambda_6 = -0.059$, and $\lambda_7 = -0.799$, and its plot was located 0.46 units under the plot of the SM-treated DCF on the λ_7 coordinate. This distance between the plots for the SM-treated DCF and the neat DCF was considered to reflect their crystalline habits, which were confirmed in Fig. 2B. The interconversion pathway of DCF/ β -CD mixtures (closed circles) would sprawl from the plot for the plain β -CD (as a beginning) to that for the SM-treated DCF (as a destination).

In Fig. 5C and D, the characteristic bending lines of pathways for the FAM and CIM mixtures were observed in the Cartesian space consisting of the λ_5 , λ_7 , and λ_2 axes. The 3D plots of PC vectors obtained by performing PCA are shown in the Supplementary Movie, MP4 (8521 K) (ESI[†]). The interconversion pathway of the FAM/ β -CD mixtures prepared with SM (closed circles) lay on the λ_5 - λ_2 diagram plane in Fig. 5C. The plot for the plain CD was located at the λ_2 coordinate of -0.3 , and that for the equimolar mixture of FAM was located at the λ_2 coordinate of $+0.1$. The interconversion pathway progressed directly to the region of high λ_2 and low λ_5 , depending on the molar ratio of FAM, and then progressed to the plot of the SM-treated FAM. Although the pathway from the equimolar mixture to the SM-treated FAM appeared to curve towards zero on the λ_7 axis in Fig. 5D, the bent at the plot for the equimolar mixture was more prominent. This bending trajectory in the hyperdimensional space is referred to as the V-shape.

Meanwhile, the interconversion pathway of the CIM/ β -CD mixture prepared with SM (open circles) lay on the λ_2 - λ_7 diagram plane, with the equimolar mixture acting as a corner. The plot of plain β -CD was shared with the pathway for the FAM mixtures, and the pathway lay along the rough trajectory from the plot of the plain β -CD to the corner. The continuous trajectory bent from the corner and spanned a border of $\lambda_7 = 0$, then meandered, reaching the plot for the SM-treated CIM. The trajectory of the CIM/ β -CD mixtures can also be considered

to take the V-shape if details are not taken into account. The verified trajectories in Fig. 5, *i.e.*, the linear shapes for the INM and DCF mixtures and the V-shapes for the FAM and CIM mixtures, reveal the apparent variations of the diffractograms in Fig. 2.

In Fig. 5C, the V-shaped trajectory of the interconversion pathway from the plain β -CD to the SM-treated FAM *via* their equimolar mixture spanned the λ_2 - λ_5 diagram plane. In Fig. 5D, the trajectory from the plain β -CD to the SM-treated CIM *via* their equimolar mixture spanned the λ_2 - λ_7 diagram plane, being perpendicular to that for the β -CD to FAM. In Fig. 5C and D, the distance between the equimolar FAM/ β -CD mixture and the SM-treated CIM is significantly close. The differences from the seven-dimensional vector of the equimolar FAM/ β -CD mixture (at the center in Fig. 4C) to that of the SM-treated CIM (at the right edge in Fig. 4D) included the λ_1 of $+0.05$ and the λ_3 of -0.05 . However, the similarity between the diffractograms of the equimolar FAM/ β -CD mixture in Fig. 2C (halo pattern) and the SM-treated CIM in Fig. 2D (assigned to the B-form of CIM in Fig. S2D, ESI[†]) cannot be recognized. The basis functions of Ψ_1 , Ψ_2 , and Ψ_3 seemed to involve the peaks related to neither the halo pattern nor the CIM B-form. The configuration space of the interconversion pathways of API/ β -CD mixtures was reproduced in the seven-dimensional vector space, but we obtained no diffractograms of the CIM B-form as one of the entities.

3.6. SVD analysis for diffractograms of the INM mixtures of its polymorphs and those with β -CD

Based on our analysis of the diffractogram dataset, we found that the analyzed samples contain both γ - and α -forms of INM and A- and B-forms of CIM. However, using SVD analysis, we were able to distinguish between the presence of the γ -form of INM and the A-form of CIM. Our analysis identified six basis functions as independent factors (chemical entities) that can be used to identify these forms. However, we were unable to identify other states, such as the α -form of INM and the B-form of CIM. If any machine learning procedure was used, any unidentified states, such as the INM α -form and the CIM B-form, could be labeled with identification. Whether using linear algebra or machine learning (with any functional or any multiple-layered function), the mathematical mapping of the states and the interconversion can be processed, and there would be no essential difference. The remaining states can be reproduced because alternative factors (opposite poles) can be transformed by each other using linear algebra. While this interpretation seems reasonable, it may not be universal. For example, the channel forms of β -CD obtained in the CIM/ β -CD mixtures were also alternatives to the cage form of β -CD. However, we can recognize the basis function for the channel form and the combination of those for the cage form.

To determine whether a single basis function or dual functions are provided by such an alternative pair (opposite poles) of chemical entities, we conducted an SVD analysis for both the combination of the α - and γ -forms of INM and that of the plain β -CD and the INM/ β -CD complex. We measured the mixtures of



α - and γ -crystals of INM at various molar ratios and merged them with those of the SM-prepared INM/ β -CD mixtures, resulting in 26 samples. The SVD analysis was conducted, and the results are summarized in Fig. 6. The first five higher ranks showed a cumulative contribution of 85.4%.

In Fig. S7 (ESI[†]), the first basis function Ψ_1 represented the diffractogram of the INM γ -form. The second basis function Ψ_2 corresponded to the diffractogram of the plain β -CD and the INM γ -form. The third basis function Ψ_3 matched the diffractogram of the plain β -CD and the INM α -form. The fourth basis function Ψ_4 matched the diffractogram of the plain β -CD and the INM/ β -CD complex with a crystalline habit different from the plain β -CD. The fifth and sixth basis functions were not consistent with either the diffractograms of INM forms or those of β -CDs.

In Fig. 6, the specific coefficient components λ_i were plotted as a function of the mole fraction of INM in the SM-prepared

INM/ β -CD mixtures. λ_2 and λ_5 decreased, while λ_4 increased depending on the molar ratio of β -CD. These changes were attributed to the positive peaks of Ψ_2 and negative peaks of Ψ_4 correlated to the diffractogram of the plain β -CD (cage form). The assignment for Ψ_5 was ambiguous, but λ_5 seemed to be responsible for the amount of plain β -CD. λ_3 increased depending on the molar ratio of INM, reflecting that the diffractogram of the INM α -form corresponds to the positive peaks of Ψ_3 . As always, λ_1 was probably proportional to the average intensity of all peaks. The peaks of neat β -CD were shown with gray bars at a mole fraction of 0%. λ_4 for neat β -CD was the highest because the positive peaks of Ψ_4 corresponded to neat β -CD. The peaks of the PM-prepared equimolar mixture were expressed with bars at a mole fraction of 50%. The reason why λ_5 for the PM-treated equimolar mixture was the lowest might be that Ψ_5 was responsible for the amount of plain β -CD. Fig. 6D shows the coefficient components for the α - and γ -forms of INM mixtures.

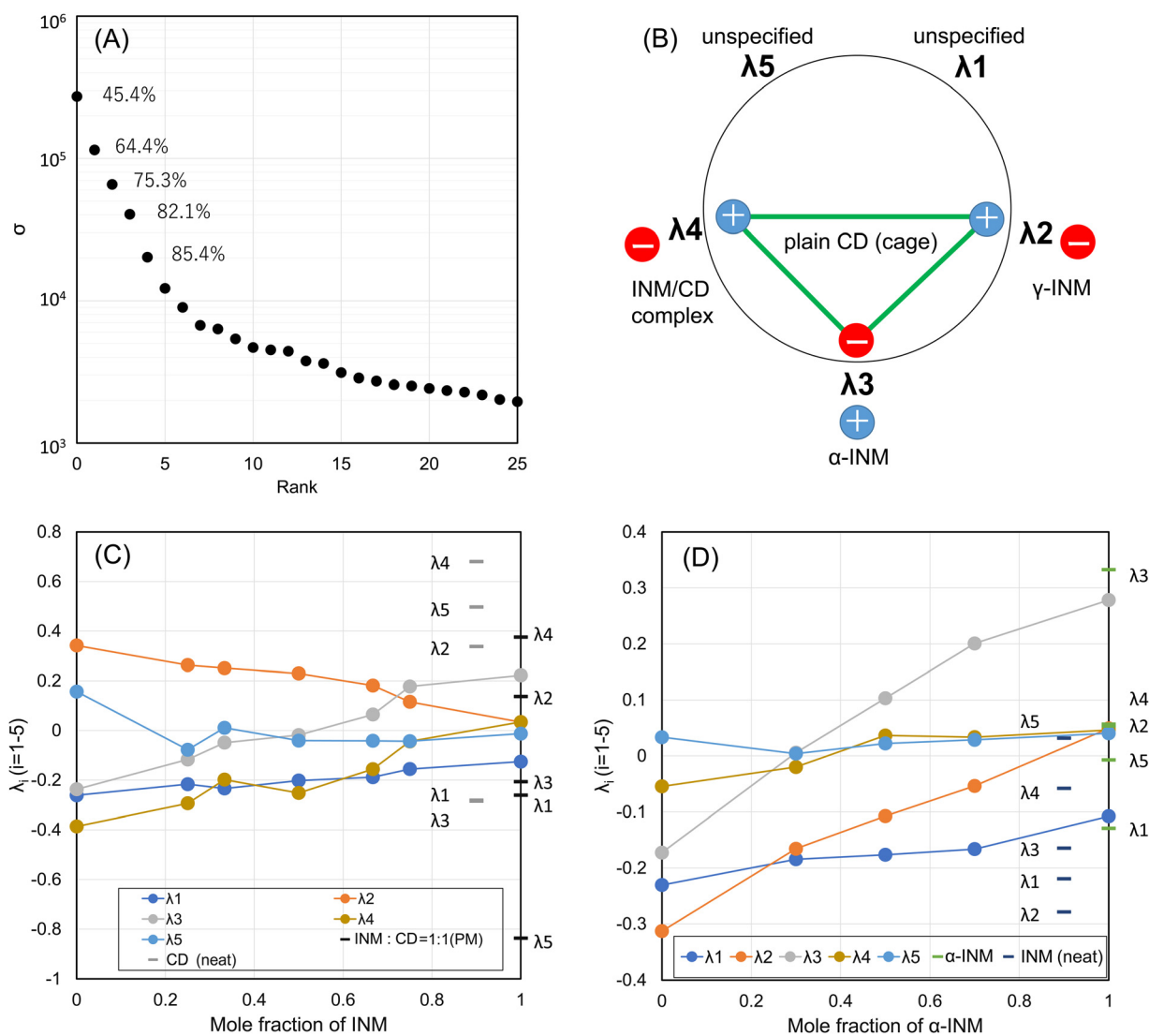


Fig. 6 The SVD analysis for the matrix with 26 diffractograms of the INM mixtures of α - and γ -crystals and the INM/ β -CD mixtures: (A) the singular values and (B) the summarized scheme of the obtained components. The specific coefficient vectors for the (C) INM/ β -CD mixtures and (D) INM mixtures of α - and γ -crystals.



It suggested that λ_2 and λ_3 corresponded to γ - and α -forms, respectively. λ_4 and λ_5 did not respond to the molar ratio of INM polymorphs.

3.7. The trajectory of the interconversion pathway in diffractograms of the INM mixtures

Fig. 7A and B show views from both sides in a Cartesian space made up of the λ_2 , λ_3 , and λ_4 axes of the polytope of the interconversion pathway in an INM mixture of its polymorphs and those with β -CD at various molar ratios. The Supplementary Movie, MP4 (7146 K) (ESI[†]) shows views from both sides with the PC1, PC2, and PC3 axes, which correspond to the four-dimensionally rotating body of the projected three-dimensional space in Fig. 7A and B. The interconversion pathways of INM/ β -CD mixtures and mixtures of α - and γ -polymorphs move in meandering but single-direction trajectories when viewed from any direction in the hyperspace. In Fig. 7B, these trajectories overlap. In other figures, it can be seen that both trajectories share an intersection at the plot corresponding to the INM α -form (around the zero λ_2 coordinate). From this starting point, the trajectory for INM/ β -CD mixtures moves in the direction of the plot for plain β -CD (positive PC1 (λ_2) coordinate and PC3 (λ_4) coordinate with a large distribution range in the movie file), while the trajectory for INM mixtures of polymorphs traces the plot for the INM γ -form (negative PC1 (λ_2) coordinate and PC3 (λ_4) coordinate with a small distribution range in the movie file). The degree of λ_4 is thought to reflect the transition between the crystalline habits of β -CD, regardless of the amount and polymorphs of INM. Therefore, the current SVD analysis can distinguish between two opposing concepts: the crystalline habits of β -CD in the cage form and the α - or γ -polymorphs of INM.

4. Discussion

In 1967, social psychologist Stanley Milgram claimed ‘six degrees of separation’, corresponding to his well-known and many times criticized experiments about a social network structure of human relationships. His hypothesis was subsequently called a small-world phenomenon.⁵⁵ In 1998, sociologist Watts and mathematician Strogatz found that the addition of a small number of random links shortened the distance between any two vertices, compared with a regular lattice in the network.⁵⁶ They accounted for it as a small-world phenomenon.

In our linear algebra study, the relationships among the X-ray powder diffractograms of the mixtures of INM and β -CD showed the state of elements linked with direct tie lines. Meanwhile, the addition of DCF, FAM, and CIM caused to loss of the basis function for the diffractions of the INM α -form. To explain the observed diffractogram, those of the individual states (polymorphs and crystalline habits) are required, although the results of linear algebra analysis can reproduce the diffractograms constituted a combination of selected components (Fig. S8, ESI[†]). Tie lines in the network of the interconversion pathways would be shortened (reduction of dimensionality) and the reproduction was achieved with a small number of basis functions. We speculate it occurred due to one of the small-world phenomena.

The interconversion pathway (φ_i) in the observable instrumental quantity is essentially defined as a linear combination of differential equations and basis functions, as shown in eqn (5):

$$\varphi_i = \sum_{j=1}^r \frac{\partial \alpha_{i,j}}{\partial x_j} \Psi_j \approx \sum_{j=1}^r \omega_{i,j} \Psi_j \quad (5)$$

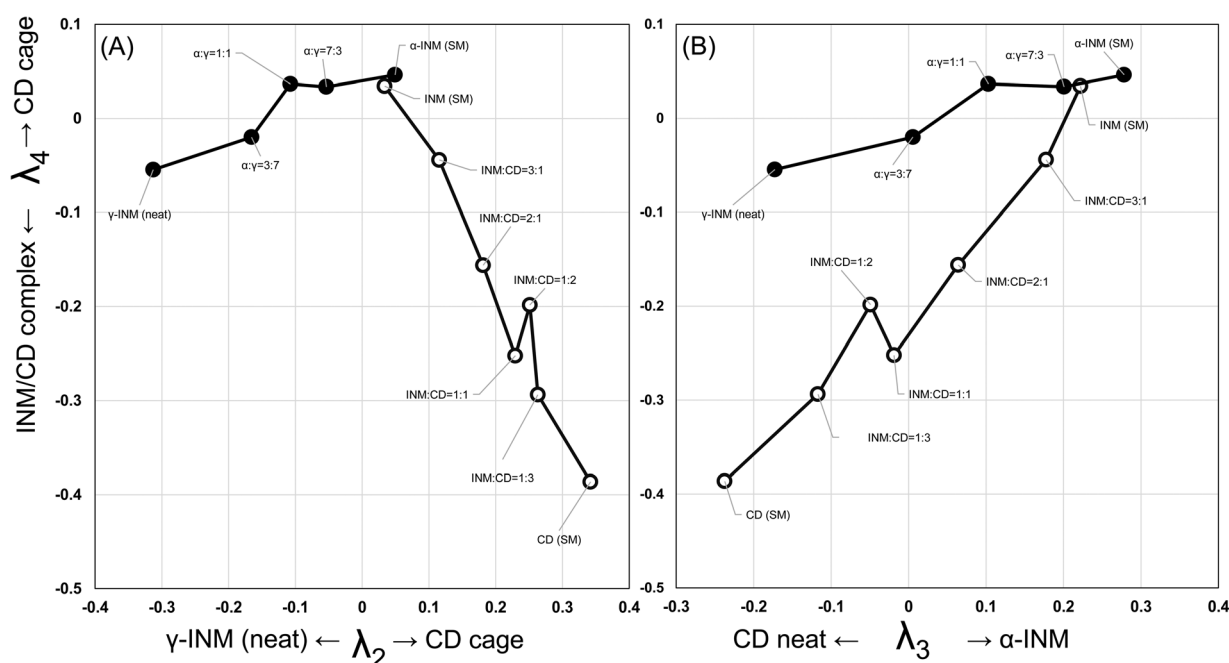


Fig. 7 The trajectories of the interconversion pathways for the INM mixtures of the α - and γ -crystals and the INM/ β -CD mixtures, projected on the λ_2 - λ_4 plane (A) and the λ_3 - λ_4 plane (B). See Supplementary Movie, MP4 (7146 K) (ESI[†]) of the three dimensional matrix.



where the specific spectrum of the j -th entity (x_j) is represented by the basis function Ψ_j . The quantitative coefficient of Ψ_j in the sample mixture is expressed as a ratio, which increases depending on the mole fraction of the entity x_j . If the entities x_1 and x_2 interact with each other, their coefficients are proportional to the mole fractions of x_1 and x_2 . If the mixing of entities x_1 and x_2 results in a new entity x_3 , the resulting mixture contains the remaining amounts of x_1 and x_2 , as well as the quantitatively produced amount of x_3 . In this case, the coefficients can be approximated to be proportional to the amounts of the distinguishable entities. As the above approximations, linear algebra is used in the present study, and any multiple-layered functions or any functionals can be adopted in deep learning studies.

Configuration space is a mathematical space that describes the nodes (states) and edges (interconversion pathways) of all parts of a system.^{54,57–59} It is typically represented by a hyper-dimensional space, where each dimension corresponds to a degree of freedom of the system. By analyzing it, analysts can determine the range of possible diversity and regulate the pathway that the assembled materials should follow to reach a particular state (existence).^{40–42,54} One of the challenges of working with configuration space is that it can be very complex and high-dimensional, especially for systems with many degrees of freedom. In such cases, it may be necessary to use techniques to analyze the configuration space and regulate state interconversion. It is important to note that alternative factors can be transformed into each other using linear algebra. For example, if entities x_1 and x_2 interact to produce entities x_3 and x_4 , respectively, the allowed configuration space would have four dimensions, consisting of two independent axes for the interconversion from x_1 to x_3 and from x_2 to x_4 , and two independent axes for the sum of x_1 and x_3 and that of x_2 and x_4 . However, since the sum of x_1 and x_3 and that of x_2 and x_4 can be represented by their molar ratio under stoichiometric interaction, the dimensionality reduces to three. This experimental setup corresponds to the analysis of diffractograms of mixtures such as the α - and γ -polymorphs and the INM/ β -CD mixtures. In the SVD analysis, their particular states and interconversion resulted in two degrees of freedom, meaning that the configuration space was two-dimensional. This configuration space, including the V-shaped trajectory shown in Fig. 6, is a two-dimensional manifold in a three-degree-of-freedom space. The reduction in dimensionality (from 3 to 2) is due to the dependence of INM on the α -polymorph as long as it is present in the INM/ β -CD complex. As a result, it was not possible to obtain the state of the γ -form of INM included in β -CD, indicating that not all entities present can be extracted from experiments or experiences.

Regarding our analysis of the dataset of diffractograms of API/ β -CD mixtures, we found that the samples contain both γ - and α -forms of INM and A- and B-forms of CIM. However, the SVD analysis was able to distinguish between the presence of the γ -form of INM and the A-form of CIM, identifying six basis functions as independent factors (chemical entities). Nevertheless, some unidentified states, such as the α -form of INM and the B-form of CIM, remain. If a machine learning

procedure was applied, any unidentified states, such as the INM α -form and the CIM B-form, could be labeled with identification. The mathematical mapping of both the states and interconversion can be processed using linear algebra or machine learning with any multiple-layered functions or functionals, so there would be no essential difference. The remaining states can be reproduced because alternative factors (opposite poles) can be transformed into each other using linear algebra. The addition of DCF, FAM, and CIM resulted in highly clustered entities; yet the interconversion pathways between them were shortened. This interpretation seems reasonable, but it may not be universal. For example, the channel forms of β -CD obtained in the CIM/ β -CD mixtures were also alternatives to the cage form of β -CD.

We expected linear algebra to completely reveal all eight entities present in the diffractogram dataset. However, after performing the SVD procedure, we found that only six independent basis functions were extracted, and the entities of the INM α -form and the CIM B-form were not included. This could be explained by information about these hidden entities being shared by other basis functions. On the other hand, this could also indicate that obtaining the full set of basis functions consisting of all diffractograms is not possible, and there is a limit to the amount of information or knowledge that can be processed. It is considered that this is due to the fact that data processing is limited to revealing only six or seven independent factors, as it is a small world.

The diffractogram of the NM α -form as an independent entity was hidden in the SVD analysis of the API/ β -CD mixtures, while it was recovered in that of the mixtures of α -/ γ -polymorphs and the INM/ β -CD mixtures. Suppose the upper limit of the number of dependent parameters was seven (or the number of connections six). It suggested that linear algebra should be limited to applying the dataset which involves seven or fewer independent factors. The X-ray diffractograms depend on the crystalline periodic features of the mixed powdered solids, so they are inappropriate to the continuous figures as the response in the scientific experiments. Because the change observed in diffractograms is only dependent on the quantities of component entities, the interpretation of data is simple and independent. However, to study the relationship between the complexity of data and recognizability in linear algebra approaches, other observations (FTIR, for example) as new subjects should be considered.

5. Conclusion

In the present study, we measured and analyzed the X-ray powder diffractograms of APIs, β -CD, and their mixtures. APIs and β -CD transformed their polymorphs and crystalline habits depending on their compositions, but their transformations cannot randomly occur. The practical observations of solid states were unattainable to the number of the capable components (chemical entities and undistinguishable complexes) although we expected to extract their individual diffractograms (basis functions) using linear algebra. Nevertheless, the INM



α -form and the CIM B-form were able to be reproduced as the linear combinations of these basis functions. The projection of the interconversion pathway in the hyperdimensional space allows us to visualize the relationship between the polymorphs or crystalline habits. We highlighted that the alternative factors can be transformed into each other using linear algebra but cannot be extracted as deductive entities from experiments or experiences. The authors further discussed the limitations of experimental chemists in verifying the validity of their findings without presupposing the deductive components consisting of chemical/physical pure entities, while AI may recognize such entities with the assistance of machine learning. They concluded by discussing the limitations of XAI in providing a deductive explanation and obtaining independent deductive entities from experiments and experiences.

Author contributions

Kanji Hasegawa: investigation, visualization, writing – original draft. Satoru Goto: investigation and supervision. Chihiro Tsunoda: review and editing. Chihiro Kuroda: review and editing. Yuta Okumura: review and editing. Ryosuke Hiroshige: review and editing. Ayako Wada-Hirai: review and editing. Shota Shimizu: review and editing. Hideshi Yokoyama: review and editing. Tomohiro Tsuchida: review and editing.

Conflicts of interest

There are no conflicts to declare.

References

- 1 K. Sakoda, *Optical Properties of Photonic Crystals*, Springer Series in Optical Sciences, Springer, Berlin, 2nd edn, 2005, vol. 80.
- 2 B. T. Hoga, E. Kovalska, M. F. Craciun and A. Baldycheva, *J. Mater. Chem. C*, 2017, **5**, 11185–11195.
- 3 T. Kato, M. Gupta, D. Yamaguchi, K.-P. Gan and M. Nakayama, *Bull. Chem. Soc. Jpn.*, 2021, **94**, 357–376.
- 4 K. D. Putirka and F. J. Tepley, *Minerals, Inclusions and Volcanic Processes, Reviews in Mineralogy and Geochemistry*, Mineralogical Society of America, Virginia USA, 2008, vol. 69.
- 5 N. L. Bowen, *J. Geol.*, 1919, **27**, 393–430.
- 6 T. Mikouchi, M. Komatsu, K. Hagiya, K. Ohsumi, M. E. Zolensky, V. Hoffmann, J. Martinez, R. Hochleitner, M. Kaliwoda, Y. Terada, N. Yagi, M. Takata, W. Satake, Y. Aoyagi, A. Takenouchi, Y. Karouji, M. Uesugi and T. Yada, *Earth, Planets Space*, 2014, **66**, 82.
- 7 J. Orehek, D. Teslić and B. Likozar, *Org. Process Res.*, 2021, **25**, 16–42.
- 8 R. W. Hartel, *Annu. Rev. Food Sci. Technol.*, 2013, **4**, 277–292.
- 9 K. Sato, *Crystallization of Lipids: Fundamentals and Applications in Food, Cosmetics, and Pharmaceuticals*, John Wiley & Sons, 2018.
- 10 F. Artusio and R. Pisano, *Int. J. Pharm.*, 2018, **547**, 190–208.
- 11 B. Rodríguez-Spong, C. P. Price, A. Jayasankar, A. J. Matzger and N. Rodríguez-Hornedo, *Adv. Drug Delivery Rev.*, 2004, **56**, 241–274.
- 12 K. Greco and R. Bogner, *Mol. Pharmaceutics*, 2010, **7**, 1406–1418.
- 13 T. Van Duong, D. Lüdeker, P.-J. Van Bockstal, T. De Beer, J. Van Humbeeck and G. Van den Mooter, *Mol. Pharmaceutics*, 2018, **15**, 1037–1051.
- 14 S. K. Jha, S. Karthika and T. K. Radhakrishnan, *Resour.-Effic. Technol.*, 2017, **3**, 94–100.
- 15 C.-S. Su, C.-Y. Liao and W.-D. Jheng, *Chem. Eng. Technol.*, 2014, **38**, 181–186.
- 16 A. Heinz, C. J. Strachan, K. C. Gordon and T. Rades, *J. Pharm. Pharmacol.*, 2009, **61**, 971–988.
- 17 A. M. Healy, Z. A. Worku, D. Kumar and A. M. Madi, *Adv. Drug Delivery Rev.*, 2017, **117**, 25–46.
- 18 S. Sareen, G. Mathew and L. Joseph, *Int. J. Pharm. Invest.*, 2012, 212–217.
- 19 D. J. Good and N. Rodríguez-Hornedo, *Cryst. Growth Des.*, 2010, **10**, 1028–1032.
- 20 K. Imamura, M. Nomura, K. Tanaka, N. Kataoka, J. Oshitani, H. Imanaka and K. Nakanishi, *J. Pharm. Sci.*, 2010, **99**, 1452–1463.
- 21 Y. Shimada, R. Tateuchi, H. Chatani and S. Goto, *J. Mol. Struct.*, 2018, **1155**, 165–170.
- 22 R. Tateuchi, N. Sagawa, Y. Shimada and S. Goto, *J. Phys. Chem. B*, 2015, **119**, 9868–9873.
- 23 M. Gümüş, Ş. N. Babacan, Y. Demir, Y. Sert, İ. Koca and İ. Gülçin, *Arch. Pharm.*, 2022, **355**, e2100242.
- 24 T. Kasai, K. Shiono, Y. Otsuka, Y. Shimada, H. Terada, K. Komatsu and S. Goto, *Int. J. Pharm.*, 2020, **590**, 119841.
- 25 C. Tsunoda, S. Goto, R. Hiroshige, T. Kasai, Y. Okumura and H. Yokoyama, *Int. J. Pharm.*, 2023, **638**, 122913.
- 26 M. Ridley, *Inf. Technol. Librar.*, 2022, **41**, DOI: [10.6017/ital.v41i2.14683](https://doi.org/10.6017/ital.v41i2.14683).
- 27 A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila and F. Herrera, *Information Fusion*, 2020, **58**, 82–115.
- 28 S. K. Jagatheesaperumal, Q.-V. Pham, R. Ruby, Z. Yang, C. Xu and Z. Zhang, *IEEE Open J. Commun. Soc.*, 2022, **3**, 2106–2136.
- 29 W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl and B. Yu, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**(44), 22071–22080.
- 30 W. J. Murdoch, C. Singh, K. Kumbier and B. Yu, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 22071–22080.
- 31 W. Samek and K.-R. Müller, Towards Explainable Artificial Intelligence, *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, 2019, pp. 5–22.
- 32 W. Samek, T. Wiegand and K.-R. Müller, Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models, *arXiv*, 2017, preprint, arXiv:1708.08296v1, DOI: [10.48550/arXiv.1708.08296](https://doi.org/10.48550/arXiv.1708.08296).
- 33 N. Burkart and M. F. Huber, *J. Artif. Intelligence Res.*, 2021, **70**, 245–317.
- 34 S. Lockey, N. Gillespie, D. Holm and I. A. Someh, *Conference: Hawaii International Conference on System Sciences*, 2021.



- 35 D. Shin, *Int. J. Human-Comput. Stud.*, 2021, **146**, 102551.
- 36 R. Riedl, *Electron. Mark.*, 2022, **32**, 2021–2051.
- 37 D. Leslie, Understanding artificial intelligence ethics and safety, 2019, Zenodo, DOI: [10.5281/zenodo.3240529](https://doi.org/10.5281/zenodo.3240529).
- 38 P. Tagde, S. Tagde, T. Bhattacharya, P. Tagde, H. Chopra, R. Akter, D. Kaushik and H. Rahman, *Environ. Sci. Pollut. Res. Int.*, 2021, **28**, 52810–52831.
- 39 Y. Otsuka, W. Kuwashima, Y. Tanaka, Y. Yamaki, Y. Shimada and S. Goto, *J. Pharm. Sci.*, 2021, **110**, 1142–1147.
- 40 T. Shiratori, S. Goto, T. Sakaguchi, T. Kasai, Y. Otsuka, K. Higashi, K. Makino, H. Takahashi and K. Komatsu, *Biochem. Biophys. Rep.*, 2021, **28**, 101153.
- 41 R. Hiroshige, S. Goto, R. Ichii, S. Shimizu, A. Wada-Hirai, Y.-P. Li, Y. Shimada, Y. Otsuka, K. Makino and H. Takahashi, *J. Inclusion Phenom. Macrocyclic Chem.*, 2022, **102**, 327–338.
- 42 R. Hiroshige, S. Goto, C. Tsunoda, R. Ichii, S. Shimizu, Y. Otsuka, K. Makino, H. Takahashi and H. Yokoyama, *J. Inclusion Phenom. Macrocyclic Chem.*, 2022, **102**, 791–800.
- 43 M. Takatsuka, S. Goto, K. Kobayashi, Y. Otsuka and Y. Shimada, *BBA Adv.*, 2021, **1**, 100030.
- 44 M. Takatsuka, S. Goto, K. Kobayashi, Y. Otsuka and Y. Shimada, *Food Biosci.*, 2022, **48**, 101714.
- 45 Y. Kurosawa, Y. Otsuka and S. Goto, *Colloids Surf., B*, 2022, **212**, 112344.
- 46 Y. Kurosawa, S. Goto, K. Mitsuya, Y. Otsuka and H. Yokoyama, *Phys. Chem. Chem. Phys.*, 2023, **25**, 6203–6213.
- 47 E. R. Henry and J. Hofrichter, *Methods Enzymol.*, 1992, **210**, 129–192.
- 48 R. J. DeSa and I. B. C. Matheson, *Methods Enzymol.*, 2004, **384**, 1–8.
- 49 Y. P. Lee, S. Goto, Y. Shimada, K. Komatsu, Y. Yokoyama, H. Terada and K. Makino, *Phys. Chem. Biophys.*, 2015, **5**, 1000187.
- 50 A. Wada-Hirai, S. Shimizu, R. Ichii, C. Tsunoda, R. Hiroshige, M. Fujita, Y.-P. Li, Y. Shimada, Y. Otsuka and S. Goto, *J. Pharm. Sci.*, 2021, **110**, 3623–3630.
- 51 S. Shimizu, A. Wada-Hirai, Y.-P. Li, Y. Shimada, Y. Otsuka and S. Goto, *J. Pharm. Sci.*, 2020, **109**, 2206–2212.
- 52 E. Flapan, *When Topology Meets Chemistry, A Topological Look at Molecular Chemistry*, Cambridge University Press, New York, 2000.
- 53 P. G. Mezey, Saphe in Chemistry, *An Introduction to Molecular Shape and Topology*, VCH Publishers, Inc., New York, 1993.
- 54 S. Goto, K. Komatsu and H. Terada, *Bull. Chem. Soc. Jpn.*, 2013, **86**, 230–242.
- 55 S. Milgram, *Psychol. Today*, 1967, **2**, 60–67.
- 56 D. J. Watts and S. H. Strogatz, *Nature*, 1998, **393**, 440–442.
- 57 S. Goto and K. Komatsu, *Hiroshima Math. J.*, 2012, **42**, 115–126.
- 58 S. Goto, Y. Hemmi, K. Komatsu and J. Yagi, *Hiroshima Math. J.*, 2012, **42**, 253–266.
- 59 S. Goto, K. Komatsu and J. Yagi, *Hiroshima Math. J.*, 2020, **50**, 185–197.

