

## RESEARCH ARTICLE

## ASBase: The universal database for aggregate science

Junyi Gong<sup>1</sup>  | Wanjun Gong<sup>2</sup>  | Bo Wu<sup>1</sup> | Haoran Wang<sup>1</sup> | Wei He<sup>1</sup> |  
 Zhongyang Dai<sup>3</sup> | Yangzhong Li<sup>3</sup> | Yong Liu<sup>2</sup> | Zhiming Wang<sup>2</sup> | Xiaojun Tuo<sup>2</sup> |  
 Jacky W.Y. Lam<sup>4</sup> | Zijie Qiu<sup>1</sup> | Zheng Zhao<sup>1</sup> | Ben Zhong Tang<sup>1</sup>

<sup>1</sup>Shenzhen Institute of Aggregate Science and Technology, School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Guangdong, China

<sup>2</sup>AIE-Institute, State Key Laboratory of Luminescence Materials and Devices, Center of Aggregation-induced Emission, Key Laboratory of Luminescence from Molecular Aggregates of Guangdong Province, South China University of Technology (SCUT), Guangzhou, China

<sup>3</sup>High Performance Computing Department, National Supercomputing Center in Shenzhen, Shenzhen, China

<sup>4</sup>Department of Chemistry, Hong Kong University of Science and Technology, Kowloon, Hong Kong, China

## Correspondence

Ben Zhong Tang, Shenzhen Institute of Aggregate Science and Technology, School of Science and Engineering, The Chinese University of Hong Kong, 2001 Longxiang Road, Longgang District, Shenzhen, Guangdong, China.  
 Email: tangbenz@cuhk.edu.cn

## Funding information

the National Science Foundation of China, Grant/Award Numbers: 21975077, 52003228, 21788102; Shenzhen Key Laboratory of Functional Aggregate Materials, Grant/Award Number: ZDSYS2021102111400001; the Science and Technology Plan of Shenzhen, Grant/Award Number: JCYJ2021324134613038; Open Fund of Guangdong Provincial Key Laboratory of Luminescence from Molecular Aggregates, Grant/Award Number: 2019B030301003

## Abstract

This paper reports the first universal and versatile database on aggregate materials for the field of aggregate science research. At the current stage, the database (<http://119.91.135.188:8080/>) contains over 1000 entries of organic aggregate materials (mainly luminescent systems at the current stage) with a unique data structure which is designed particularly for aggregate materials and containing the photophysics and physicochemical properties of the compounds in different statuses of aggregation, including dilute solution form, pristine solid-state, stable crystalline, and nanoaggregates formed in solvents. The web-based interface of the database provided functions to index, search, manipulate, fetch and deposit data entries. In addition, a background calculation service optimizes the chemical structure of new entries on different levels of accuracies. The database also provided background API for interactive developments of prediction or regression models based on machine-learning algorithms.

## KEYWORDS

aggregate science, aggregation-induced emission, data-driven model, database

## 1 | INTRODUCTION

The aggregate science, which concentrated not only on the chemical structures but also on the organization of molecules, enlightened the research of chemistry and material science in recent years.<sup>[1,2]</sup> As the representative subarea of aggregate science, research about aggregation-induced emission (AIE) is a breeding ground for thousands of new materials with outstanding optical performance, unique physicochemical behavior, and potential for applications as bio-imaging agents, photovoltaic devices, light-emitting diodes, sensors, and actuators.<sup>[2–9]</sup> The concept of AIE extends from the traditional synthetic organic small molecules to a wide range of systems, including polymers,<sup>[10,11]</sup> natural products,<sup>[12]</sup> biomacromolecules,<sup>[13]</sup> metal-organic coordination compounds,<sup>[14,15]</sup> metal-organic frameworks,<sup>[16,17]</sup>

covalent-organic frameworks,<sup>[18]</sup> and even inorganic compounds,<sup>[19–21]</sup> displaying a rich diversity, complexity, and unlimited possibilities. However, in contrast to the rapid progress in the laboratory, the research methods were stuck in the classical stage of experimental chemistry, highly relying on rules of thumb, experiences in case studies, and trials and error. Moreover, the progress in computer science and data science is still underutilized by aggregate science due to the lack of accumulation and generalization of data.

Databases such as Cambridge structure database (CSD),<sup>[22,23]</sup> Scifinder,<sup>[24]</sup> Reaxys,<sup>[24]</sup> and Materials project<sup>[25]</sup> provide data retrieval, comparison, filtering, and traceability, which are indispensable in modern chemistry and materials research. Specifically for luminescent materials, there were also some good examples.<sup>[26–28]</sup> Nevertheless, all existing databases are somewhat maladaptive

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Aggregate* published by SCUT, AIEI, and John Wiley & Sons Australia, Ltd.

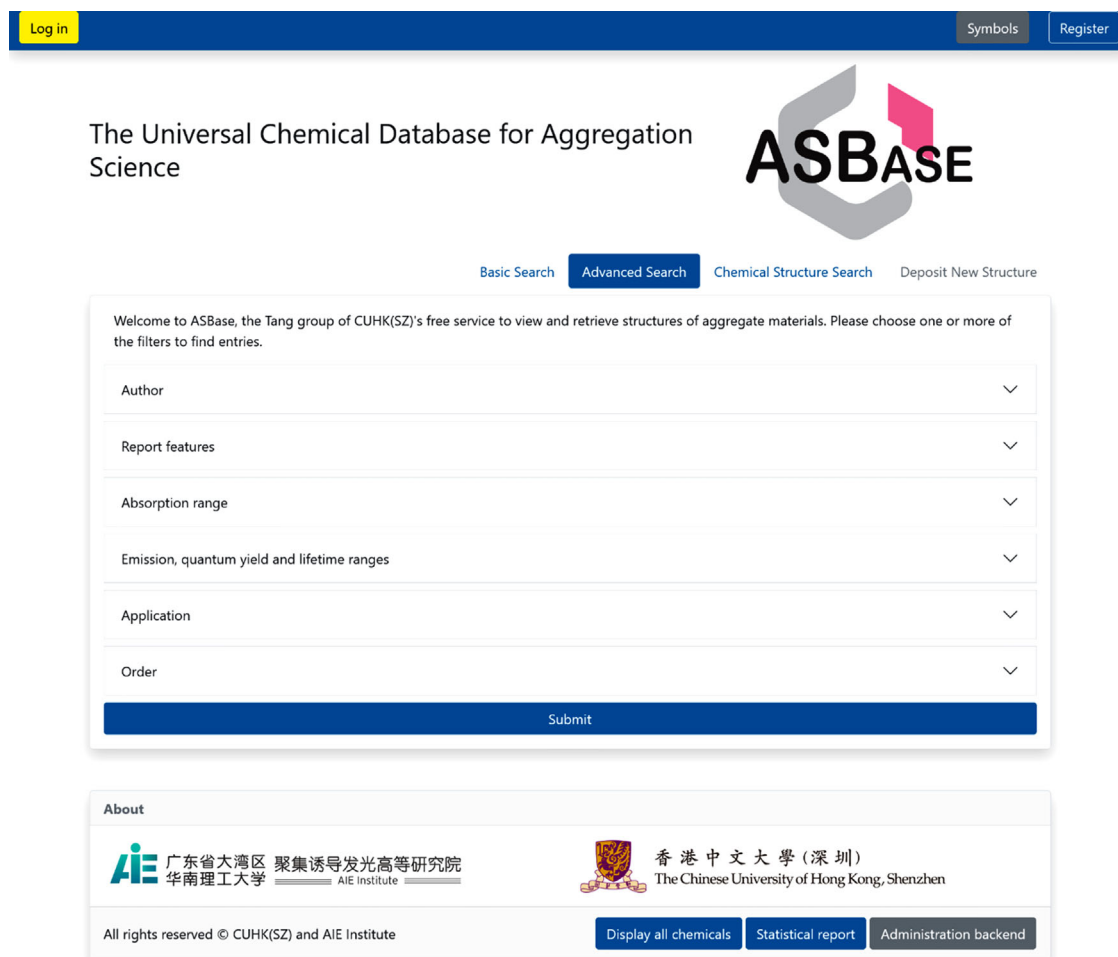


FIGURE 1 Basic web interface of ASBase

in aggregate science. The least fit is in the storage pattern of data, where the aggregate of science requires that the chemical structure and the status of aggregation be regarded as a clue to store macroscopic properties. The researchers of aggregate science may pay more attention to the relationship between properties and the organization forms of molecules. This character facilitated the demand for a specific database design suitable for this unique field of research.

In this work, we reported the first versatile database system, ASBase, designed and developed for aggregate science, with a web-based interface as shown in Figure 1 providing multiple retrieval functions, back-end computation, and statistics services. Comparing to existing databases for luminescent materials, our database did not only include the photophysics about the systems but also many other physicochemical properties which are indispensable for the application research. In addition, the database is a platform with data deposit function, which enabled researchers from all over the world to share their systems on the platform. Currently, the database contains over 1000 entries of aggregate materials with traceable references and reported properties. Each entry contains around 40 fields of uploaded or generated data about chemical structure, literature references, patents, photophysics (absorption, emission, quantum yield and lifetime), physicochemical properties (stabilities and solubilities), and application information. Remarkably, the photophysics of an entry in the database was organized by different aggregation forms, including dilute solution

(solution), pristine solid powder (solid), stable crystalline (crystal), and aggregates formed in solvents (aggregate).

## 2 | MATERIALS AND METHODS

### 2.1 | Database structure and organization

The structure of ASBase was shown in Figure 2. The whole system could be divided into two parts: the server-side and the client-side. The server-side, which was constructed by django 4.0 framework,<sup>[29]</sup> could also be divided into two sub-modules: the data server module and the calculation server module. The data server module was constructed with three sub-tables of the database. An entry should refer to a specific chemical structure and contain different aspects of the information about the compound: (1) the structure information contains the chemical structures, the SMILES, the code name of the compound, and the index in the database; (2) the reference information contains the reported feature (AIE, aggregation-caused quenching (ACQ), room-temperature phosphorescence (RTP),<sup>[30]</sup> thermally activated delayed fluorescence (TADF),<sup>[31]</sup> or aggregation-induced delayed fluorescence (AIDF)<sup>[32]</sup>, the report year, the bibliography, first author and the corresponding author names of the literature reporting the compound, and the patent number if it is available; (3) the photophysics contains the tested solvent and all photophysical properties (the absorption peak

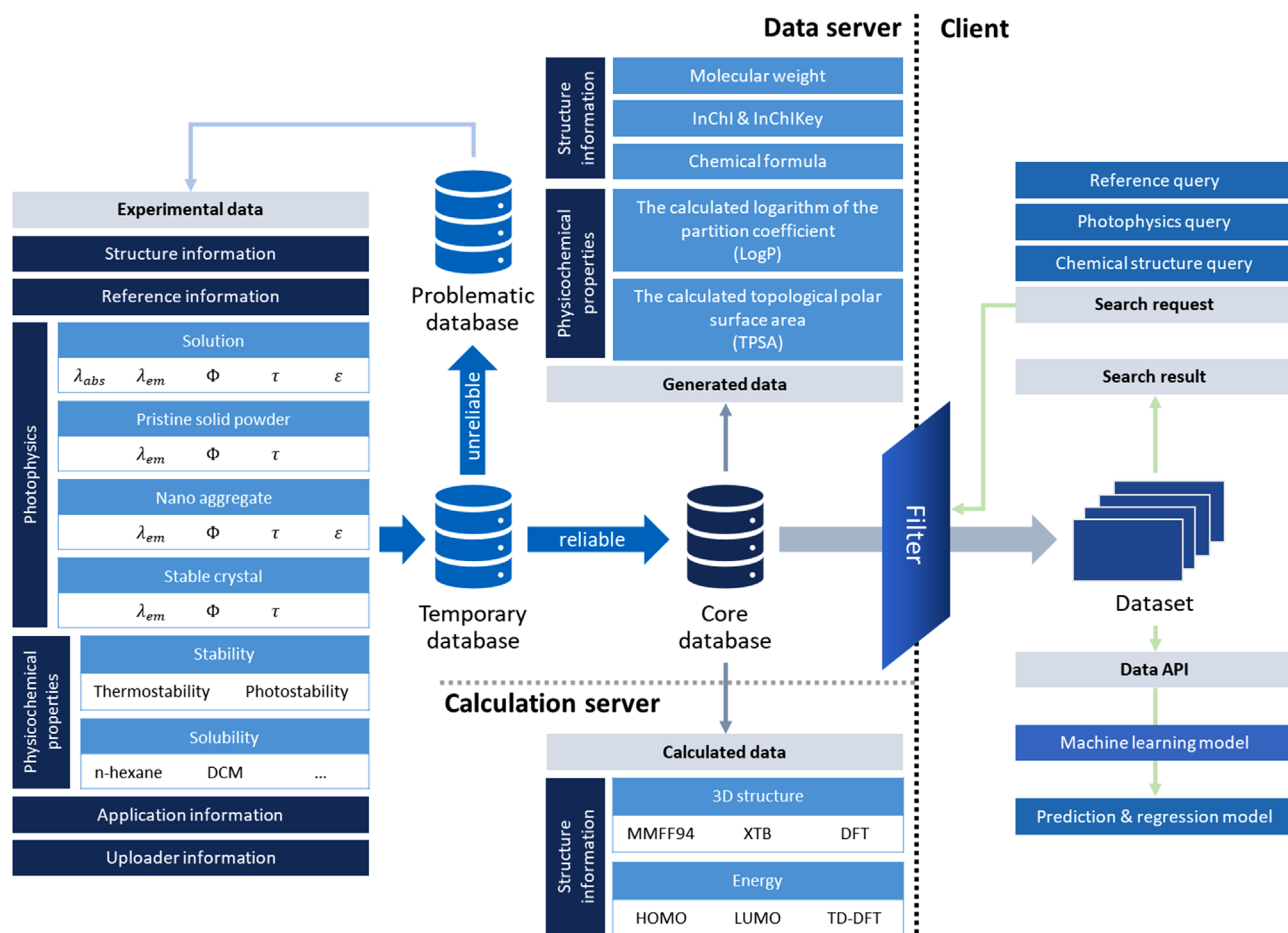


FIGURE 2 The structure of ASBase

wavelength ( $\lambda_{abs}$ ), the emission peak wavelength ( $\lambda_{em}$ ), the molar extinction coefficient ( $\epsilon$ ), the quantum yield ( $\Phi$ ), and lifetime ( $\tau$ ) of photoluminescence) in different statuses of aggregation (in dilute solution of tested solvent, in pristine solid powder form, in the stable crystalline form and in aggregates formed in solvents); (4) the physicochemical properties contain the stability (photostability and thermostability) and the solubilities in different solvent systems; (5) the application information contains four keywords of the potent applications of the compound; (6) the uploader information contains the information about the uploader of the entry.

## 2.2 | Data uploading and verification

A registered user could upload an entry through the data deposit interface in the webpage and the uploaded entry will be stored in the temporary database for verification. At current stage, we will review the data entries in the temporary database periodically to determine their reliability, enter the reliable entries into the core databases, and put the unreliable ones into the problematic database. The unreliable data will be returned to the uploader for double check. Once the reliable data were recorded in the core database, some information including the calculated logarithm of the partition coefficient ( $\log P$ ), the calculated topological polar surface area (TPSA), the structure identifiers (including InChI and InChIKey), the chemical formula and the molecular weight will be generated. The scalable vector graphics (SVG) images

of the chemical structures in two-dimensional (2D) will be also generated in the data-server. Periodically, the entries in the core database will be sent to calculation server to do more complicated computations for generation of the 3D structure in different levels (MMFF94,<sup>[33]</sup> GFN2-xTB,<sup>[34]</sup> and density functional theory (DFT)), the orbital energies of highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) and transition energies through TD-DFT.

## 2.3 | Data retrieval and searching

The client side provided a comprehensive web-interface for user to interactively utilize the database. The most important function in the client side is the data retrieval system. The web-interface provided three different implementations for retrieving the data entries: (1) the basic search approach enabled user to search entries conditionally with a simple filter such as the name of the authors, the photophysics in different statuses of aggregation, the reported feature, and the reported application; (2) the advanced search function helps user to retrieval entries with multiple filters simultaneously; (3) the chemical structure search provided a chemical structure editor and a SMILES input box for searching compounds with inputted chemical sub-structure. All the search results could be sorted by ID, absorption/emission peak wavelengths, and solid-state quantum yields. The detailed information of entries meets the search requirements could be

**TABLE 1** Five couples of properties with highest values of Pearson's correlation coefficient

Property A	Property B	Correlation coefficient
$\tau$ (aggregate)	$\tau$ (solution)	0.9182
$\lambda_{em}$ (aggregate)	$\lambda_{em}$ (solid)	0.9085
$\Phi$ (crystal)	$\Phi$ (solid)	0.9077
MW	$\log P$	0.8966
$\Phi$ (crystal)	$\Phi$ (aggregate)	0.8923

reviewed as shown in Figure 3. Table 1 provided a detailed table of the recorded information and properties of a given entry in the database.

## 2.4 | Statistics

Another critical feature of ASBase is statistical report generation. The server-side can generate real-time statistical reports based on the core database's data entries and render the graphs on the webpage. In the current version, the statistical report contains six graphs, including three histograms about the distribution of absorption peak wavelengths in solution, emission peak wavelengths and fluorescence quantum yields in solid-state; two pie graphs about the distribution of reported features and reported photoluminescence mechanisms. The statistical report also includes a correlation scatter plot of absorption peak wavelengths as the x-axis and emission peak in the solid-state as the y-axis.

## 2.5 | Algorithms

**Pearson's correlation coefficient.** For an ensemble of variable  $x_i$  and another ensemble of variable  $y_i$ , the Pearson's correlation coefficient between  $x$  and  $y$  is defined as:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, -1 \leq R \leq 1 \quad (1)$$

The value of  $R$  is between  $-1$  and  $1$ . When  $R > 0$ , the correlation between  $x$  and  $y$  is positive. When  $R < 0$ , the correlation is negative. The absolute value of  $R$  indicates the linear correlation between  $x$  and  $y$ .

**Spearman's correlation coefficient.** For an ensemble of variable  $x_i$  and another ensemble of variable  $y_i$ , the Spearman's correlation coefficient between  $x$  and  $y$  is defined as:

$$R = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, -1 \leq R \leq 1 \quad (2)$$

where  $d_i$  is the difference in ranks of the elements in  $x_i$  and  $y_i$  ensembles. The value of Spearman's correlation coefficient is also between  $-1$  and  $1$ . When  $R > 0$ , the correlation between  $x$  and  $y$  is positive. When  $R < 0$ , the correlation is negative. The absolute value indicates the monotonicity between  $x$  and  $y$ .

**t-Distributed stochastic neighbor embedding (t-SNE).** t-SNE is a tool to reduce the dimensions and visualize high-dimensional data. The main idea of t-SNE is to convert similarities between data points to joint probabilities and to minimize the Kullback–Leibler divergence between the joint probabilities of the low-dimensional embedding and the high-dimensional data.<sup>[35]</sup>

## 2.6 | Software

The database server-side was held on a  $\times 86$  server at the current version by django ver. 4.0 framework. All background logic is implemented in Python ver. 3.8 or above with RDKit ver. 2021<sup>[36]</sup> or above and OpenBabel ver. 3.1.1.<sup>[37]</sup> The web interface was built by Bootstrap CSS ver. 5.0<sup>[38]</sup> or above with Chart.js<sup>[39]</sup> to render the statistical report. The chemical structure editor is the JavaScript Molecule Editor (JSME) ver. 2022 or above.<sup>[40]</sup>

## 2.7 | Computational methods

The  $\log P$ , TPSA, chemical formula, and molecular weight were generated originally from the SMILES of the entries by algorithms provided by RDKit. The SVG images of 2D chemical structures were also generated from the SMILES by OpenBabel. OpenBabel optimized the 3D structures in the MMFF94 level. Grimme's xTB software optimized the 3D structure in the GFN2-xTB level. The 3D structure in the DFT level, and energies of HOMO and LUMO were optimized by ORCA ver. 5.0.2 under BLYP/def2-SVP level.<sup>[41]</sup> The implementation of the t-SNE algorithm is provided by Scikit-learn package ver. 1.1.<sup>[42]</sup>

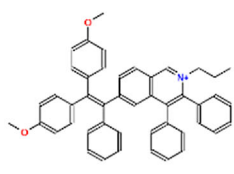
## 3 | RESULTS AND DISCUSSION

### 3.1 | The purpose of creating ASBase

The potential audiences of ASBase were not only academic researchers but also engineers in the industry who are going to develop products for particular specific demands, investors who need some reference about the significance and potential to evaluate a project, and policymakers who are willing to get macroscopic views of the development of the field of aggregate science. As shown in Figure 4, the primary purpose of ASBase is to provide a tool for users mentioned above with functions of searching entries, stating data, exploring new theories and predicting photophysics by data-mining or machine-learning methods. Besides, the prediction function

[Back to homepage](#)
[Symbols](#)

## General Information

Chemical Structure	First Author	SMILES
 <a href="#">3D structure</a>	GUI Chen	<chem>CCC[n+]6cc4ccc/C(=C(c1ccc(OC)cc1))\c2ccc(OC)cc2)c3ccccc3)cc4c(c5ccccc5)c6c7ccccc7</chem>
	Corresponding Author	InChI
	TANG Ben Zhong	InChI= 1S/C46H40NO2 /c1-4-30-47-32-39-21-20-38(31-42(39)45(34-16-10-6-11-17-34)46(47)37-18-12-7-13-19-37)44(33-14-8-5-9-15-33)43(35-22-26-40(48-2)27-23-35)36-24-28-41(49-3)29-25-36 /h5-29,31-32H,4,30H2,1-3H3/q+1
	Reference	InChIKey
	Chem. Sci., 2017, 8, 1822-1830	CVWRQIXEYUPJM-UHFFFAOYSA-N
ID	Patent	Chemical formula
1	TTC.PA. 0871TTC.PA. 0871	C46H40NO2+
Code Name	Report year	Molecular weight
TPE-IQ-20	2017	638.815 Da
AggIndex		
1		

## Photophysics

$\lambda_{\text{abs}}/\text{nm}$	$\lambda_{\text{em}}/\text{nm}$ (solution)	$\Phi$ (solution)	$\tau/\text{ns}$ (solution)
430.00000	530.00000	0.00600	None
Powder color	$\lambda_{\text{em}}/\text{nm}$ (solid)	$\Phi$ (solid)	$\tau/\text{ns}$ (solid)
yellow	620.00000	0.13100	None
$\epsilon/\text{M}^{-1}\text{cm}^{-1}$ (solution)	$\lambda_{\text{em}}/\text{nm}$ (aggregates)	$\Phi$ (aggregates)	$\tau/\text{ns}$ (aggregates)
None	None	None	None
$\epsilon/\text{M}^{-1}\text{cm}^{-1}$ (aggregates)	$\lambda_{\text{em}}/\text{nm}$ (crystal)	$\Phi$ (crystal)	$\tau/\text{ns}$ (crystal)
None	None	None	None

## Physicochemical Properties

Calculated LogP	TPSA
10.8959	22.34 nm <sup>2</sup>
Photostable	Solubility
✓	DCM/Chloroform ✓ THF/EA ? Methanol/Ethanol ? MeCN/Acetone ?
Thermostable	DMF/DMSO ✓ water ✓ Carbohydrate ? Toluene ?
?	

Applications
cell staining      anticancer      None      None

**FIGURE 3** An example page of an entry in the search result



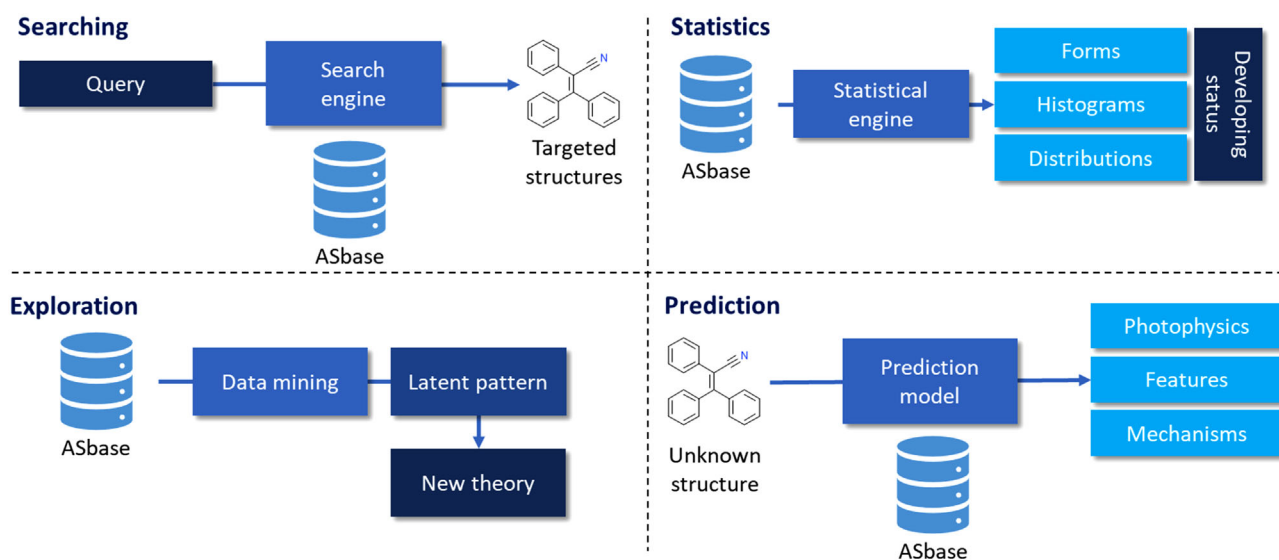


FIGURE 4 The schematic view of the utilization of ASBase

was still under development.<sup>[43]</sup> The main idea for such prediction is to encode the chemical structure with a serial bit string, like molecular fingerprints, and then trained a prediction model with labeled data in the database. Then, we can get prediction from the serially encoded chemical structure of a given unknown compound from this trained model.

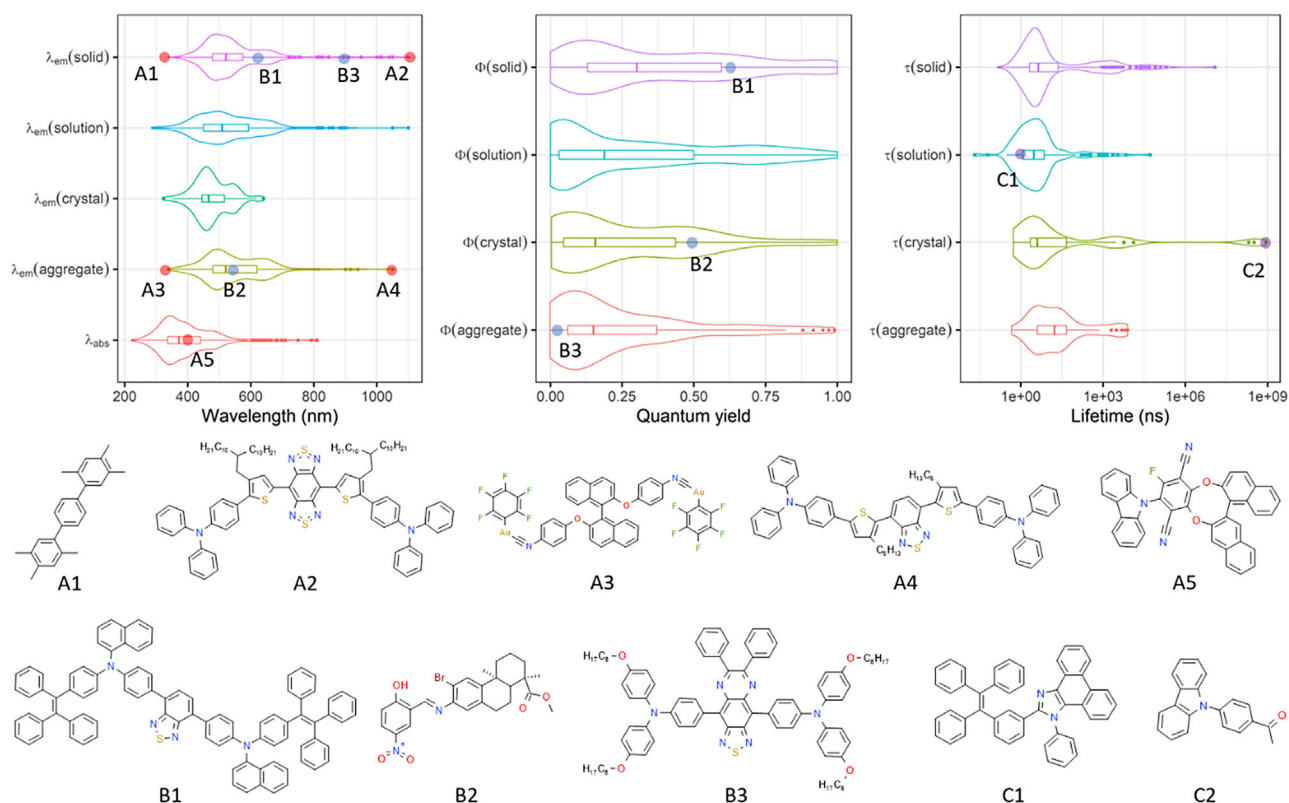
### 3.2 | Current data

At the current stage, the database contains 1100+ entries. Each entry corresponds to a reported chemical in aggregate science, including AIE,<sup>[3]</sup> ACQ,<sup>[44]</sup> RTP,<sup>[30]</sup> TADF,<sup>[31]</sup> and AIDF.<sup>[32]</sup> For their reported mechanism of photoluminescence, all entries could also be categorized into intramolecular charge transfer (ICT), twisted intramolecular charge transfer (TICT),<sup>[45]</sup> and excited-state intramolecular proton transfer (ESIPT),<sup>[46]</sup> neutral aromatic (NA), and clusterluminescence (CL) systems,<sup>[47]</sup> respectively. It is noteworthy that the NA mechanism denoted the system with aromatic rings without obvious donor, acceptor, and ESIPT substructures such as pyrene, tetraphenylethylene, anthracene, and so on. Besides, there were many entries without the specifically reported features and mechanisms in the source literatures and will be categorized into an “other” category. In addition, some systems with features or mechanisms which are too unique to be categorized into a normal ensemble would be also included in the “other” category. Figure 5 displays the distribution of photophysics of the first 988 entries in the database. From the violin plots, we could get some exciting information that cannot be observed easily from isolated papers and case studies. For example, the distributions of wavelengths indicated around 120 nm of the average Stokes’ shift. In addition, the mean value of emission peak wavelengths in the crystal is the lowest among all four statuses of aggregation, which might be caused by the rigidification effects in the crystal that inhibit the excited-state relaxation process.

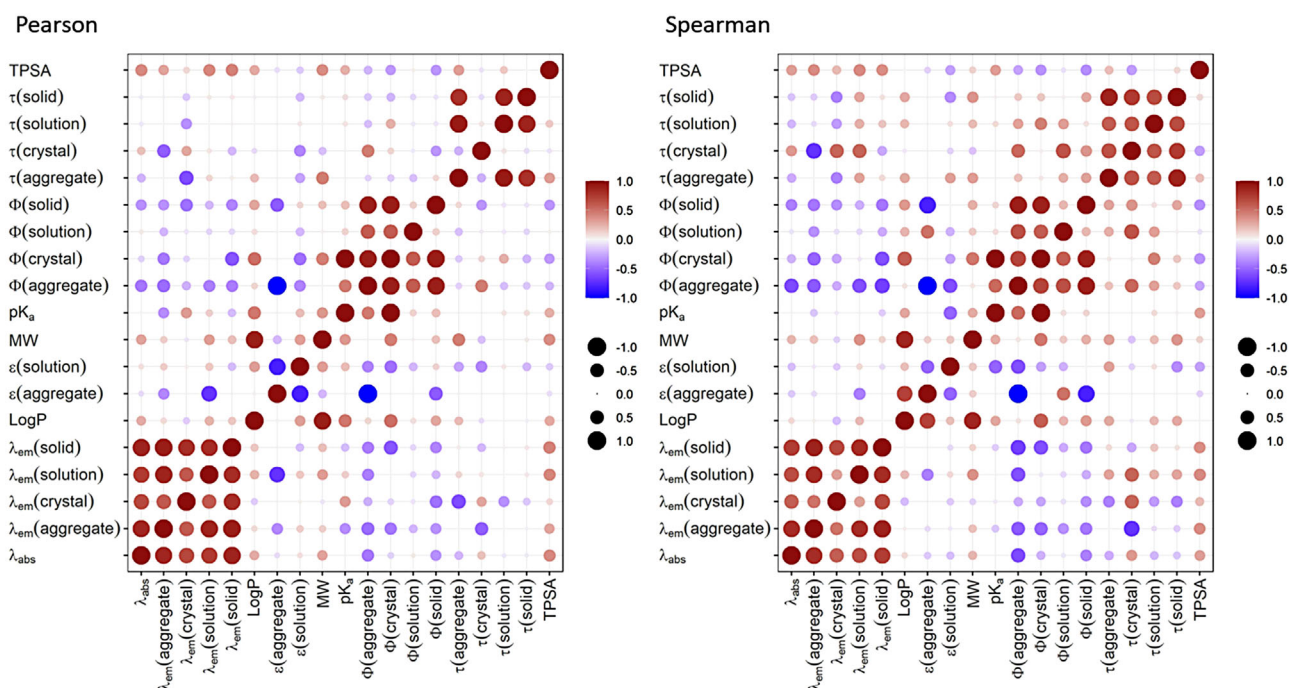
Chemical A1 and A2 are the bluest and reddest emitter in pristine solid powder form in the stated entries. Chemical A1 contains three connected benzene rings with

many methyl groups serving as spatial hindrances, which diminish the conjugation and cause high transition energy. In contrast, chemical A2, which has an emission peak wavelength at around 1100 nm in pristine solid powder, contains a donor- $\pi$ -acceptor- $\pi$ -donor-shaped skeleton, with two triphenylamines as the electron donors, two thiophenes as the conjugation bridges and a 1*H*,5*H*-benzo[1,2-*c*:4,5-*c'*]bis([1,2,5]thiadiazole) as a strong electron acceptor. Near infrared ray (NIR) emitters like chemicals A2 and A4 always have strong donors, acceptors, and large conjugation scale. In the distributions of quantum yields, the quantum yields in aggregates, crystal and pristine solid powder have two prominent density peaks. However, the quantum yields in dilute solution have only one obvious density peak at around 0.05. This result provided empirical evidence that the aggregation effect behaves as an amplifier for photoluminescence for the compounds in ASBase, which is identical to the topic for data collection. It is noteworthy that not every entry in the database has completed data in all the fields of the photophysics. Generally, most source literature reported the quantum yields in solid and solution. But fewer papers reported the quantum yields in crystal and aggregates form due to the more complicated measurement. This unbalanced data collection caused the density of high quantum yields in dilute solution higher than those in aggregates and crystal. In regular cases, the quantum yields negatively correlate to the emission wavelengths due to the fast non-radiative decay process from states with small energy gaps. So, chemical B1 is a precious system with red emission and high fluorescence quantum yield at around 0.63. Chemical B2 (with a quantum yield of 0.49) and B3 (with a quantum yield of 0.03) exhibited typical quantum yields for their emission color in the green–yellow and NIR regions, respectively. The range of lifetime in the crystal is undoubtedly the largest among all four statuses of aggregation because the crystal is the critical requirement for most RTP systems to display an ultra-long lifetime. As a typical RTP compound, chemical C2 in Figure 5 was reported in 2021 with 868 ms in crystal.

One significant database usage is detecting latent patterns in the recorded entries. Figure 6 displays the correlation coefficient matrices of stated data at the current state.



**FIGURE 5** The violin plots of distributions of different experimental photophysics properties in the database at the current stage and the chemical structures of examples noted in the distribution plots



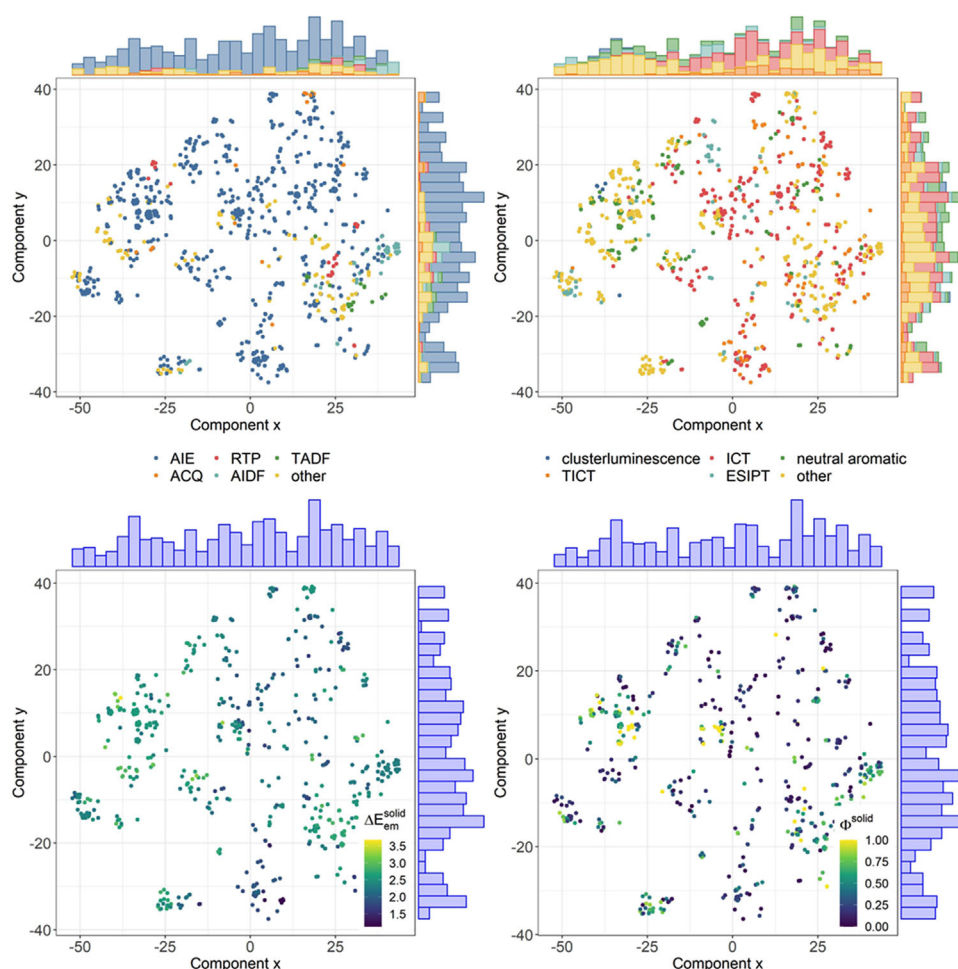
**FIGURE 6** The correlation coefficient matrices of properties in the database at the current stage

Tables 1 and 2 recorded the couples of factors with the highest absolute values of positive and negative correlation coefficients. The Pearson's and Spearman's correlation reveal the linear and monotonic correlation between corresponding parameters, respectively. The positive or negative values of coefficients implied positive or negative correlation

separately. From the matrices, it could be indicated that a strong correlation could be observed between wavelengths, quantum yields, and lifetimes. A negative correlation between molar extinction coefficients and quantum yields in aggregate could be explained by the relative calculation algorithms of quantum yields. The quantum yield is the ratio

**TABLE 2** Five couples of properties with highest absolute values of negative Pearson's correlation coefficient

Property A	Property B	Correlation coefficient
$\varepsilon$ (aggregate)	$\Phi$ (aggregate)	−1.0000
$\varepsilon$ (solution)	$\varepsilon$ (aggregate)	−0.8047
$\varepsilon$ (aggregate)	$\lambda_{\text{em}}$ (solution)	−0.7080
$\tau$ (aggregate)	$\lambda_{\text{em}}$ (crystal)	−0.5970
$\lambda_{\text{em}}$ (solid)	$\Phi$ (crystal)	−0.5609

**FIGURE 7** The scatter plots of t-SNE components of molecular access system (MACCS) keys mapped by features, mechanisms, emission energy ( $\Delta E_{\text{em}}^{\text{solid}}$ ) and quantum yield ( $\Phi^{\text{solid}}$ ) in pristine solid powder of the compounds in the database at current stage

of emitted and absorbed photons. It is easy to understand that a strong molar extinction coefficient could reduce the quantum yield of a given compound if the emission intensity is kept constant. The conclusion above mentioned are almost in line with expectations. However, there are also some strange correlations which may never be considered without the accumulation of data. For example, high molar extinction coefficients are positively but not linearly correlated to the log  $P$  values. These two factors are hardly ever discussed together in case studies. A possible explanation is that high molar extinction coefficients were always achieved by a vast conjugation scale, which may enhance the lipophilicity. The positive correlation between molecular weights and the log  $P$  values could also be explained by such a theory.

To explore the relationship between chemical structures and the properties recorded in the database, we encoded

all chemical structures into the molecular access system (MACCS) keys' fingerprints<sup>[48]</sup> and performed a t-SNE decomposition analysis. As shown in Figure 7, each scatter point represented a specific pattern of MACCS keys' fingerprints, which referred to a particular suite of features in the chemical structure of a given compound. The distances between the points are correlated with the difference in chemical structures. All scatter points were color-mapped by their features, mechanisms, emission transition energy, and quantum yield in solid-state (pristine solid powder). It could be observed that systems with AIE features existed in almost all clusters of scatter points, implying the diversity in chemical structures of AIE systems. Nevertheless, for mechanisms, many categories displayed structural dependences. Some clusters only contained entries with the same photoluminescence mechanisms. Besides, there is a higher structural



correlation in the transition energy than in the quantum yield in pristine solid powder form, indicating that the quantum yield is more unpredictable than emission wavelengths for a given chemical structure.

## 4 | CONCLUSION

Hereby, we present the versatile database for aggregate science ASBase, sourcing mainly from the reported literature. For better adaptability of the database to the requirements of aggregate science, we uniquely designed the database structure by organizing the photophysical properties with the status of aggregation. ASBase is the first database designed specifically for aggregate science research. Currently, the database contains 1100+ unique entries with around 40 different fields of information for each and a comprehensive web interface to retrieve, deposit, and manipulate data. Besides, a statistical report generation function provided a convenient approach to get a macroscopic view of the development of aggregate materials and the progress of aggregate science, which is of great significance for not only academic researchers but also engineers, investors, and policymakers. The database will continue operating and expanding with novel data and new features. We hope that the ASBase could promote the data-driven research paradigm in the field of aggregate science.

## ACKNOWLEDGMENTS

This work is supported by the National Science Foundation of China (21975077, 52003228 and 21788102), Shenzhen Key laboratory of Functional Aggregate Materials (ZDSYS20211021111400001), the Science and Technology Plan of Shenzhen (JCYJ2021324134613038) and the Open Fund of Guangdong Provincial Key Laboratory of Luminescence from Molecular Aggregates (2019B030301003).

## FUNDING

the National Science Foundation of China, Grant Numbers: 21975077, 52003228, 21788102; Shenzhen Key Laboratory of Functional Aggregate Materials, Grant Number: ZDSYS20211021111400001; the Science and Technology Plan of Shenzhen, Grant Number: JCYJ2021324134613038; the Open Fund of Guangdong Provincial Key Laboratory of Luminescence from Molecular Aggregates, Grant Number: 2019B030301003

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

## ORCID

Junyi Gong  <https://orcid.org/0000-0001-7592-7132>

Wanjuan Gong  <https://orcid.org/0000-0002-3099-032X>

## REFERENCES

- H. Zhang, Z. Zhao, A. T. Turley, L. Wang, P. R. McGonigal, Y. Tu, Y. Li, Z. Wang, R. T. K. Kwok, J. W. Y. Lam, B. Z. Tang, *Adv. Mater.* **2020**, *32*, 2001457.
- R. Hu, G. Zhang, A. Qin, B. Z. Tang, *Pure Appl. Chem.* **2021**, *93*, 1383.
- J. Mei, N. L. C. Leung, R. T. K. Kwok, J. W. Y. Lam, B. Z. Tang, *Chem. Rev.* **2015**, *115*, 11718.
- X. Han, F. Ge, J. Xu, X.-H. Bu, *Aggregate* **2021**, *2*, e28.
- V. M. Granchak, T. V. Sakhno, I. V. Korotkova, Y. E. Sakhno, S. Y. Kuchmy, *Theor. Exp. Chem.* **2018**, *54*, 147.
- K. Hatano, K. Matsuoka, D. Terunuma, *Trends Glycosci. Glycotechnol.* **2012**, *24*, 78.
- Z. Li, X. Ji, H. Xie, B. Z. Tang, *Adv. Mater.* **2021**, *33*, 2100021.
- E. Ubba, Y. Tao, Z. Yang, J. Zhao, L. Wang, Z. Chi, *Chem. Asian. J.* **2018**, *13*, 3106.
- H. Wu, W. Huang, X. Zhou, Y. Min, *Front. Immunol.* **2020**, *11*, 575816.
- W. Du, X. Liu, L. Liu, J. W. Y. Lam, B. Z. Tang, *ACS Appl. Polym. Mater.* **2021**, *3*, 2290.
- T. Han, X. Wang, D. Wang, B. Z. Tang, *Top. Curr. Chem.* **2021**, *379*, 7.
- D. Liu, Z. Zhao, B. Z. Tang, *Sci. Sin. Chim.* <https://www.sciengine.com/SSC/doi/10.1360/SSC-2022-0082>
- H. Liu, L.-H. Xiong, R. T. K. Kwok, X. He, J. W. Y. Lam, B. Z. Tang, *Adv. Opt. Mater.* **2020**, *8*, 2000162.
- B. Jiang, C.-W. Zhang, X.-L. Shi, H.-B. Yang, *Chin. J. Polym. Sci.* **2019**, *37*, 372.
- A. Pinto, N. Svahn, J. C. Lima, L. Rodriguez, *Dalton Trans.* **2017**, *46*, 11125.
- T. Mehmood, J. P. Reddy, in *Advances in Aggregation Induced Emission Materials in Biosensing and Imaging for Biomedical Applications*, Pt B (Eds: R. S. Bhosale, V. Singh), Elsevier Academic Press Inc, San Diego **2021**, p. 179.
- L. Zhu, B. Zhu, J. Luo, B. Liu, *ACS Mater. Lett.* **2021**, *3*, 77.
- S. Dalapati, E. Jin, M. Addicoat, T. Heine, D. Jiang, *J. Am. Chem. Soc.* **2016**, *138*, 5797.
- Y. Gao, C. Qin, Y. Nie, W. Liu, T. Li, X. Jiang, *Chin. J. Org. Chem.* **2020**, *40*, 2254.
- L.-C. Mao, X.-Y. Zhang, Y. Wei, *Chin. J. Polym. Sci.* **2019**, *37*, 340.
- Y. Wang, H. Li, D. Wang, B. Z. Tang, *Adv. Funct. Mater.* **2021**, *31*, 2006952.
- F. H. Allen, *Acta. Crystallogr. B.* **2002**, *58*, 380.
- C. R. Groom, F. H. Allen, *Angew. Chem. Int. Ed.* **2014**, *53*, 662.
- J. Silhanek, *Chem. Listy* **2014**, *108*, 83.
- Materials Project Home, <https://materialsproject.org/> (accessed: May 2022).
- M. Taniguchi, J. S. Lindsey, *Photochem. Photobiol.* **2018**, *94*, 290.
- J. F. Joung, M. Han, M. Jeong, S. Park, *Sci. Data* **2020**, *7*, 295.
- C.-W. Ju, H. Bai, B. Li, R. Liu, *J. Chem. Inf. Model.* **2021**, *61*, 1053.
- Django Software Foundation. **2019**, <https://www.djangooproject.com> (accessed: May 2022).
- G. YongYang, T. YeQiang, L. Hong, Z. YiRen, Y. WangZhang, Z. YongMing, S. JingZhi, B. Z. Tang, *Sci. China Chem.* **2013**, *56*, 1183.
- Z. Yang, Z. Mao, Z. Xie, Y. Zhang, S. Liu, J. Zhao, J. Xu, Z. Chi, M. P. Aldred, *Chem. Soc. Rev.* **2017**, *46*, 915.
- H. Liu, J. Guo, Z. Zhao, B. Z. Tang, *Chem. Photochem.* **2019**, *3*, 993.
- T. A. Halgren, *J. Comput. Chem.* **1996**, *17*, 490.
- C. Bannwarth, S. Ehlert, S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 1652.
- L. van der Maaten, G. Hinton, *J. Mach. Learn. Res.* **2008**, *9*, 2579.
- RDKit, <https://www.rdkit.org/> (accessed: May 2022).
- N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *J. Cheminf.* **2011**, *3*, 33.
- M. O. Contributors Jacob Thornton, and Bootstrap, Bootstrap, <https://getbootstrap.com/> (accessed: May 2022).
- Chart.js | Open source HTML5 Charts for your website, <https://www.chartjs.org/> (accessed: May 2022).
- B. Bienfait, P. Ertl, *J. Cheminf.* **2013**, *5*, 24.
- F. Neese, *Rev. Comput. Mol. Sci.* **2012**, *2*, 73.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., *J. Mach. Learn. Res.* **2011**, *12*, 2825.
- J. Gong, Z. Deng, H. Xie, Z. Zhao, B. Z. Tang, **2022**, doi:10.26434/chemrxiv-2022-r4bfq-v4.
- Ein Konzentrationsumschlag der Fluoreszenz. *Z. Phys. Chem.* **1954**, *1*, 275.
- S. Sasaki, G. P. C. Drummen, G. Konishi, *J. Mater. Chem. C.* **2016**, *4*, 2731.
- X.-L. Lu, W. He, *Chin. J. Anal. Chem.* **2021**, *49*, 184.

47. H. Zhang, Z. Zhao, P. R. McGonigal, R. Ye, S. Liu, J. W. Y. Lam, R. T. K. Kwok, W. Z. Yuan, J. Xie, A. L. Rogach, B. Z. Tang, *Mater. Today* **32**, **2020**, 32, 275.
48. J. L. Durant, B. A. Leland, D. R. Henry, J. G. Nourse, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273.

**How to cite this article:** J. Gong, W. Gong, B. Wu, H. Wang, W. He, Z. Dai, Y. Li, Y. Liu, Z. Wang, X. Tuo, J. W. Y. Lam, Z. Qiu, Z. Zhao, B. Z. Tang, *Aggregate* **2023**, *4*, e263.  
<https://doi.org/10.1002/agt2.263>