Towards a Robotic Scientist for Synthesis of Nanocrystals

Haitao Zhao (✉ ht.zhao@siat.ac.cn)
Chinese Academy of Sciences   https://orcid.org/0000-0002-2448-8448

Wei Chen
Chinese Academy of Sciences

Zhehao Sun
Australian National University

Fuming Lai
Chinese Academy of Sciences

Baicheng Zhang
University of Science and Technology of China

Zhuo Wang
The University of Nottingham Ningbo China

Hao Huang
Chinese Academy of Sciences

Oyawale Adetunji Moses
Chinese Academy of Sciences

Mukhtar Lawan Adam
Chinese Academy of Sciences

Zijian Chen
Chinese Academy of Sciences

Cheng Heng Pang
The University of Nottingham Ningbo China

Tao Wu
The University of Nottingham Ningbo China

Paul K. Chu
City University of Hong Kong

Yang Lu
City University of Hong Kong   https://orcid.org/0000-0002-9280-2718

Jun Jiang
University of Science and Technology of China

ZONGYOU YIN
Australian National University

Xue-Feng Yu

Article

Keywords:

Abstract

Data-driven materials synthesis is heralded as a new paradigm to substitute for trial-and-error experiments and labor-intensive tasks by human scientists. Herein, a Robotic Scientist platform that can deliver unprecedented performance for rational design, controllable synthesis, and retrosynthesis of nanocrystals is described. By taking advantage of interdisciplinary fields including artificial intelligence, robotic automation, and big data, the Robotic Scientist platform is trained to synthesize Au nanocrystals. Existing knowledge and machine learning models are integrated into the rational design process. Controllable synthesis is achieved by synergistic coupling of robot-assisted synthesis on the macro-scale and nanocrystal growth on the nano-scale. By means of the Robotic Scientist platform, over 2,300 samples are synthesized in conjunction with *in-situ* characterization to accomplish the complete task of design-synthesis-retrosynthesis. The platform and methodology of Robotic Scientist pave the way for digital synthesis of nanocrystals and facilitate the paradigm shift to data-driven materials synthesis.

Introduction

Data-driven materials synthesis is heralded as a new paradigm to transfer labor-intensive tasks and trial-and-error experiments from human scientists to robotic chemists[1] or chemical synthesis machines[2]. The advanced Human-AI-Robot collaboration system is accelerating the interdisciplinary revolution of materials synthesis towards a Robotic Scientist for automated synthesis. In this emerging field, it is necessary to converge chemical knowledge, theoretical models, purpose-oriented database, programmable cyber systems, as well as robotic physical systems. One of the promising missions is digital synthesis of materials[3] by acquiring knowledge progressively, unveiling data linkages efficiently, and developing solutions constructively over time based on previous iterations.

In the past decade, tremendous efforts have been devoted to developing digital manufacturing/synthesis of materials. In particular, layer-by-layer digital additive manufacturing of three-dimensional materials has been developed on the macro-scale[4]. On the micro-scale, synthetic biology is another benchmark for digital synthesis of biomaterials utilizing cells as the hardware and genes as the software[5]. Recently, there has been rapid development in organic programming language[6] and automated platforms[2,7] for organic synthesis on a small scale. At the same time, a robotic chemist has been reported to search for photocatalysts[1], thus opening up the opportunity for robot-assisted inorganic materials investigation on the nano-scale. However, there are still many limitations hampering automated synthesis, for example, materials search without theoretical models[1], blind materials optimization without scientific methodologies[1], as well as lack of synergy between hardware and software to achieve materials innovation[2]. Herein, we show how these difficulties can be tackled by the Robotic Scientist platform that enables rational design, controllable synthesis, and retrosynthesis of nanocrystals as a proof of concept.

Results

Robotic Scientist platform. The platform towards a Robotic Scientist for digital synthesis nanocrystals involves convergence of the materials databases, cyber systems, and physical systems (Fig. 1).

To accomplish rational design of nanocrystals, the software and AI algorithms are integrated into the cyber system. In addition, process automation by means of a simulated operation system is utilized to pre-examine and monitor the designed synthesis procedures.

In the physical system, crystal growth on the nano-scale is accomplished by automatic synthesis and characterization is performed on the macro-scale to guide controllable synthesis. Concurrently, Robotic Execution Excel (REE) files are designed to provide preliminary instructions for the execution of automatic synthesis using crucial parameters. The database is expanded continuously by the design and controllable synthesis processes. Furthermore, the relationship between the target nanocrystal morphologies (as outputs) and key synthesis parameters in the database (as inputs) are identified to provide constructive guidance to achieve retrosynthesis. Finally, the close loop combining rational design, controllable synthesis, and retrosynthesis provides the unprecedented ability to manipulate the morphologies of nanocrystals. It is expected that the Robotic Scientist can be trained for digital synthesis of customized nanocrystals with the essential capacities similar to those provided by human scientists.

In manual synthesis, the tasks are normally time consuming and error prone and moreover, the raw precursors expire or degrade shortly after preparation in some cases. In order to achieve automatic synthesis in a timely fashion, the Robotic Scientist platform is set up with many desirable features as shown in Fig. 2. The robot, robotic arms, digital pipettors, mobile camera, and microplate reader are connected to a series of modules that are capable of performing robot-assisted high-throughput synthesis and *in-situ* characterization. The photograph, schematic representation, and operation video of the platform are presented in Fig. 2a, Fig. 2b, and SI, respectively. The Robotic Scientist platform is expected to revolutionize traditional synthesis processes that rely on well-trained scientists and technicians.

Rational design. Traditionally, the manual chemical and materials synthesis processes differ slightly from person to person and sometimes introduce inadvertent errors/bias leading to diverse outcome. Moreover, it typically takes several months and even years for a scientist to acquire the required repertoire of synthetic knowledge. Hence, there is a substantial demand to conduct rational design on the Robotic Scientist platform while leveraging the expertise of human scientists. Here, crystal informatics, existing knowledge about synthesis, thermodynamic models, and kinetic models as data-driven scientific hypotheses are integrated into the Robotic Scientist for rational design of nanocrystals (Fig. 3).

Firstly, a crystal database with over 90,000 different crystal facets from seven crystal systems is incorporated into the Robotic Scientist based on our previous research[8]. The typical morphologies in the cubic system are identified in Fig. 3a and Fig. 3b. The morphology information is then digitally converted to the fractional surface area (FSA) and aspect ratio (AR) and the correlations are analyzed as shown in Fig. S1, in which the FSAs of the (001) and (00$\bar{1}$) planes versus AR of the corresponding nanorods are

identified revealing a gradually decreasing trend (Fig. 3a). Afterwards, by exploiting the advantages of the artificial neural network (ANN) model to understand the complex morphology evolution process, the relationship between the crystal equilibrium morphology (FSA and AR related) and surface energy ratio is established using a well-trained ANN model (Fig. S2) based on the crystal informatics database.

To train the Robotic Scientist, Au nanocrystals synthesis knowledge with key parameters is extracted from 1,300 related literatures by data mining with the aid of the Automated Literature Recommendation System[9]. Fig. S3 shows the frequency distribution of the synthesis parameters reported in the literatures and Fig. 3c indicates that L2 is the most frequently used concentrations. Hence, by taking advantage of data mining, the Robotic Scientist is initially trained to capture synthesis parameters and the identified parameters are then adopted by the Robotic Scientist platform to refine predictions. For example, longitudinal surface plasmon resonance (LSPR) can be characterized *in-situ* by the Robotic Scientist platform (Fig. S4 and Table S1) and some of the samples are characterized by *ex-situ* TEM (Fig. S5), XRD, and HR-TEM (Fig. S6) to provide the necessary conditions for the Au nanocrystals[10]. Consequently, the relationship between the customized morphological FSA and AR is established with accurate one-to-one correspondence as shown in Fig. 3b.

To bridge robot-assisted macro-scale operational synthesis and nano-scale crystal growth, the classic thermodynamic model (derived in Method section) and ML-predicted model are explored by taking $[Ag^+]$ as an example (Figs. 3d and 3e). The correlation between LSPR (related to morphology) and $[Ag^+]$ concentration is developed by investigation of the classic model. Furthermore, we have found that the ML-predicted model achieves an extended LSPR range (600-925 nm in Fig. 3e) and accurate prediction (R^2 = 0.99 in Fig. 3e) in comparison with the classical model (666-878 nm with R^2 = 0.98 in Fig. 3d). Therefore, with the assistance of ML and thermodynamic models, the relationship among morphology, surface energy, LSPR, and $[Ag^+]$ concentration is established by the Robotic Scientist platform. Establishment of the thermodynamic model allows the Robotic Scientist platform to realize rational design of desirable nanocrystals using the concentration of synthesis parameters as the input, surface energy and LSPR as the bridge, and nanocrystal morphology as the output.

The kinetics in nanocrystals synthesis is another key model in rational design that can train the Robotic Scientist for tailoring morphology. In this respect, a microplate reader and color-ultra-sensitive camera are employed to monitor the UV-Vis-NIR absorption spectra and color changes during nanocrystal growth. The dynamic-state and steady-state optical absorption spectra are displayed in Figs. S7-S10 together with representative results in Fig. 3f-3h for different C(HCl). The dynamic UV-Vis-NIR absorption spectra with peaks of LSPR and transverse surface plasmon resonance (TSPR) are identified in Fig. 3f. The normalized OD_{LSPR} change with time is shown in Fig. 3g, which indicates the pseudo-first-order kinetics (derived in the Method section and shown in Fig. S7 and Table S2). A similar trend showing the color change (RGB values) with time is presented in Fig. 3h and Fig. S7. These *in-situ* characterization results are employed to establish the nanocrystal kinetic models. Hence, the Robotic Scientist is guided by the

thermodynamic and kinetic models with ML trained models to explore controllable synthesis and retrosynthesis.

Controllable synthesis. The complexity of materials synthesis increases exponentially with the number of variables, thereby stifling full exploration of the materials space. The key to controllable synthesis process is convergence of macro-scale automatic synthesis and nano-scale crystal growth to bridge the synthesis parameters (as input) and corresponding morphologies (as output) on the Robotic Scientist platform. In order to achieve this objective, data-intensive rational design and automated synthesis are integrated. Meanwhile, machine learning and experimental data are utilized to construct models based on the appropriate synthesis variables. As a result, orthogonal, single-, double-, and triple-factor experiments can be conducted systematically in the order of iterations to construct the database for effective training of ML models.

Firstly, orthogonal experiments are conducted by executing materials synthesis with parameters by data mining from 1,300 papers (Fig. 3c). They are designed to address the limitations of blind optimization for all the factors at different levels[1]. The design of experiments with different factors and levels (Table S3), UV-Vis-NIR absorption results (Fig. S11), and multivariate analysis of the variance (Table S4-S5) are presented. Based on the experimental conditions from the high-dimensional experimental space, the initial optimized levels are decided for further single-factor study.

To analyze the potentials in 1D space, 24 levels are studied for each single factor (Table S6) and the models of single factors are presented in Figs. 4a–4c and Fig. S12, respectively. Moreover, 96-level experiments are carried out within the boundaries identified from the 24-level experimental results to provide more training data for the ML models. The ML models, corresponding coefficient, and accuracy of ML prediction are presented in Tables S7-S8 and Fig. S13. All the results can be fitted well with the ML predicted models, which are beyond the capacity of the classical model (merely fitted with the results of $AgNO_3$ factor) in Fig. 3d. Primarily, there is a border range of AR tuned by CTAB, $AgNO_3$, and HCl (compared with Au Seeds, AA, and $HAuCl_4$), which are identified and defined as structure-directing agents (SDAs)[11–14]. The different types of the SDAs can be used as triggers on the macro-scale to control the surface energy during nanocrystal growth on the nano-scale. For example, the factor of $AgNO_3$ can be adjusted by the Robotic Scientist platform to change the AR values of the nanocrystals in Fig. 4a. Therefore, the relationship between the SDAs-based synthesis parameters (inputs) and nanocrystal morphologies (outputs) is identified as the key to achieving controllable synthesis.

To train a sophisticated Robotic Scientist, double-factor experiments are conducted for two identified SDAs from the single-factor experiments. In this way, the chemical space is expanded into the 2D response surface with an 8×8 grid (64 experiments) compared to 1D curves derived from single-factor experiments. Based on 64 preliminary experiments, 96 experimental conditions are generated by a normal distribution mathematical array for active training of the ML model. The design of the double-factor experiments and ML predicted models are presented in Tables S9-S13 and the results are illustrated in Figs. S14-S16. The robust double-factor ML models are then trained with two inputs for morphological

control. It is found that CTAB and $AgNO_3$ play dominant roles and there are noticeable interactions (Fig. 4d), which are consistent with the observation that CTAB and Ag^+ form a face-specific capping agent to achieve cooperated morphological control[15]. Interestingly, the CTAB and HCl factors exhibit similar behavior of the cooperated morphology control (Fig. 4e). However, there is only additive behavior for the $AgNO_3$ and HCl factors (Fig. 4f) and $AgNO_3$ plays a leading role in the two-factor experiment. A complex response profile is created for the three-factor experiments by adjusting three SDAs. The design of triple-factor experiments and ML predicted models are shown in Tables S14-S16. The visualized response of AR to the three factors is presented in Fig. 4g and Fig. S17. Therefore, the function of the SDAs' parameters as inputs and AR features as outputs can be established for controllable synthesis of the nanocrystals in a free 3D space.

At the same time, the color features as potential outputs can be investigated by the Robotic Scientist platform. The results from single-factor, double-factor, and triple-factor experiments are shown in Figs. S18-S20, respectively and the corresponding LSPR and RGB values are listed in Tables S17-S19 for ML training. The trained ML model and comparison between experimental and ML predicted values (Table S20 and Fig. S21), in which a satisfactory ML model with an R^2 of 0.94, are obtained. As shown in Fig. 4h, the color results as another large-sample data-set match the spectra well. In this way, the Robotic Scientist can be trained to digitally recognize colors, thus contributing to the materials genome database with color features.

Finally, with the aid of the Robotic Scientist platform, over 2,300 samples are synthesized together with *in situ* characterization to build up the Au nanocrystals genome (various morphologies with LSPR from 600 to 1,000 nm) (Fig. 4i). It is estimated that this task would have taken a human scientist up to four months (18 samples per day) in comparison with less than one week (384 samples on four 96-well microplates per day) taken by the Robotic Scientist. The Robotic Scientist continues to improve by receiving training with expanding experimental data and ML predicted data to realize the ultimate goal of an intelligent system for digital nanocrystal synthesis and potential of retrosynthesis based on the data sources as described in the next section.

Retrosynthesis. The Robotic Scientist is further developed with the intention of retrosynthesis based on the learned knowledge from controllable synthesis. The Au nanocrystals genome plays a vital role in supporting a closed-loop synthesis process. The genome with typical LSPR from 600 to 1,000 nm displayed in Fig. 5a consists of experimental data, ML predictable data, and TEM validation results (Fig. 5b). Building such a genome within a six-variable experimental space seems like an impossible task with the manual approach due to the experimental complexity that scales exponentially with the number of variables[1]. The relationship between the identified SDAs and morphologies is illustrated as 'Input' and 'Output' in Fig. 5a, respectively. By normalizing different parameters of SDAs to form different nanocrystal morphologies with the trigger of the surface energy on the nano-scale, precise morphological control is accomplished. It is constructed for effective retrosynthesis (Fig. 5c) and efficient scale-up production of Au nanocrystals (Fig. 5d and 5e) to facilitate digital synthesis of Au nanocrystals.

Retrosynthesis and optimization are the Robotic Scientist's creative endeavors. The data of the target Au nanocrystals (such as LSPR as 808 ± 10 nm, 780 ± 10 nm, and 633 ± 10 nm), which are commonly used in biotechnology and information technology (for example, HIV drug delivery[16], surface-enhanced Raman scattering[17], wireless neuromodulation[18], and sensing[19,20]), are extracted from the genome for further retrosynthesis study. Using 808 ± 10 nm as an example, 99 samples are selected from previous single-, double-, and three-factor experiments as shown in Fig. 5c and Table S21. At the same time, by focusing on the best samples, additional samples are predicted by the ML models. Afterwards, optimization experiments are executed by the Robotic Scientist platform. It is generally accepted that a larger OD ratio (OD_{LSPR}/OD_{TSPR}) and narrower FWHM (at fixed LSPR) represent more uniform morphology. Hence, the experiments are designed with a decision plate to optimize the target nanocrystals with higher shape uniformity by evaluating the OD ratio and FWHM of samples from different synthesis routes (Fig. 5c and Table S22). Finally, the best samples in the decision plate with the best quality are recommended for the scale-up experiments.

Three scale-up experiments are conducted sequentially, i.e., high-throughput microplate assay on the Robotic Scientist platform (in Fig. 5d), bench-scale test on a magnetic stirrer, and pilot-scale test in an agitated vessel (in Fig. 5e). Firstly, 2 mL- (on 12-well microplate), 4 mL- (on 6-well microplate), 20 mL- and 40 mL- (on single-well plate) scale experiments are performed on the Robotic Scientist platform for 633, 780, and 808 nm samples synthesis. During the scale-up process, an interesting feature in retrosynthesis is that LSPR gradually red-shifts compared to results in the nanocrystal genome (Fig. 5d), which provides new insights into the scaling law. By taking advantage of kinetics study, SDAs (such as HCl) is identified as the effective input to play the minor modification role in the scale-up process. A slight decrease of c(HCl) adjusts the LSPR according to the established scaling law. Modification by adjusting c(HCl) is proven to be applicable and then a pilot-scale experiment (15 L) is demonstrated in an agitated vessel (Fig. S22). Therefore, this study reveals retrosynthesis and scale-up methodology by taking advantage of the Au nanocrystals genome and kinetics study on the Robotic Scientist platform, which is expected to have broad applications in the production of similar nanocrystals.

Discussion

Training scientists with the required knowledge takes considerable resources and different chemical and materials synthesis routes may lead to diverse outcomes even for trained personnel. Moreover, most of inorganic synthesis involves trial-and-error and laborious tasks with unavoidable unintentional errors/bias. The Robotic Scientist platform described here demonstrates a notable advancement of automation pertaining to the synthesis of nanocrystals and presents an essential step towards data-driven materials synthesis. The sophisticated close loop involving rational design, controllable synthesis, and retrosynthesis is achieved by converging of the Robotic Scientist-assisted synthesis on the macro-scale and nanocrystal growth on the nano-scale. The existing chemical knowledge based on data mining, thermodynamic and kinetic models, as well as ML models are combined to accelerate rational design of the nanocrystal's morphology with initial hypotheses. To avoid blind materials optimization, orthogonal

experiments, and single-, double-, and triple-factor experiments are conducted systematically in iterations and then the database is constructed for effective training of the ML models to enable controllable synthesis of nanocrystals. In these processes, the accessible large data-set (*in-situ* characterized UV-Vis-NIR absorption spectra and RGB color results) and small data-set (*ex-situ* TEM validation) are generated to establish the Au nanocrystals genome and interpretation of the genome plays a vital role in supporting the retrosynthesis process. It is demonstrated that the Robotic Scientist can be trained like a human scientist for retrosynthesis and scale-up synthesis of the targeted Au nanocrystals. This work focuses on establishing the closed-loop (design-synthesis-retrosynthesis) of automation in nanocrystal synthesis using the Robotic Scientist platform. Although the complete Robotic Scientist is an ambitious objective, the prototype is a good start towards a Robotic Scientist with the essential capabilities of scientific hypotheses, experiments by synergizing the hardware and software components, and result interpretation. It is believed that future efforts will close the gap with eventual automation of all aspects of nanocrystals synthesis. Although the Robotic Scientist is only demonstrated for Au nanocrystals in this work, the insights gained reveal the possibility of automation to accelerate data-driven materials innovation on the nano-scale.

Methods

Operation of the Robotic Scientist platform. The operating video of the Robotic Scientist platform with features is provided in the supplementary information. An illustration of the experimental preparation (sample storage and consumable intelligent management, a mobile robot for microplate transport at central line, a synthesis platform for *in-situ* sampling, three automatic pipettors for liquid handling, shake module for integrating operation, and a robotic arm for commercial equipment service in right circle) and experimental characterization (microplate reader for characterizing of UV-Vis-NIR absorption spectra and color-ultra-sensitive mobile camera for *in-situ* color characterization) is presented.

Data mining of synthesis parameters. The parameters involved in Au nanocrystal synthesis are recommended by our recently developed Automated Literature Recommendation System, a software package that can read scientific paper with Chemical Named Entity Recognition[9], expressions and grammatical structures[21], and some special rules in the nanomaterials research field. Using computers to read and digest reported works of many research groups, we were able to found the statistically representative synthesis parameters from 1300 relevant journal papers downloaded from publishers such as Springer Nature, ACS Publicaiton, RSC Publishing, Wiley, Science, and Science Direct Elesvier. From the plotted frequency distribution maps, we extracted the most frequently used parameters for designing experiments on the Robotic Scientist platform.

Raw materials preparation by the Robotic Scientist platform. All the chemicals were used as received. Hexadecyltrimethylammonium bromide (CTAB, ≥99%) and L-ascorbic acid (AA, 99%) were purchased from Sigma Aldrich. Hydrogen tetrachloroaurate trihydrate ($HAuCl_4 \cdot 3H_2O$, ≥99.9%), silver nitride ($AgNO_3$, 99.8%), sodium borohydride ($NaBH_4$, 98%) and hydrochloric acid (HCl, 37%) were purchased from

Aladdin. The seed solution for gold nanocrystals was prepared firstly by mixing $HAuCl_4 \cdot 3H_2O$ (0.05 M, 0.25 mL) with CTAB (0.1 M, 10 mL) of the solution, followed by rapid injection of ice-cold 0.6 mL fresh 0.01 M $NaBH_4$ under vigorous stirring for 2 min. The seed solution was kept at room temperature for 2 h before use. The raw materials solutions including CTAB (0.1 M, 100 mL), $HAuCl_4$ (0.01 M, 5 mL), HCl (1 M, 2 mL), $AgNO_3$ (0.01 M, 1 mL) and AA (0.1 M, 1ml) were prepared manually for one batch of 96 experiments conducted on the platform. Different automatic REE procedures such as orthogonal, single-, double-, and triple-factor experiments were designed.

ML prediction. As a supervised machine learning (ML) algorithm, the sure independence screening and sparsifying operator (SISSO)[22,23] is a compressed sensing-based approach to determine the critical factors necessary to present and predict the interested properties (UV-Vis-NIR absorption spectra and RGB color values). The results are characterized by the microplate reader and ultra-color-sensitive camera on the Robotic Scientist platform. For the construction of experimental feature spaces during ML model training, the set of implemented operators are:

$$opset \equiv \{+, -, \times, \div, log, exp, sin, cos, \char`\^2, \char`\^3, \sqrt{}, \sqrt[3]{}\}$$

The detailed ML models are presented in SI and the predicted results are shown in Fig. 4 to assist controllable synthesis.

Thermodynamic models. To achieve rational design of the morphology, thermodynamic models are derived by integration of Wulff construction[24], Gibbs adsorption isotherm[25,26], and Langmuir adsorption model[27,28] as follows. The relationship of the surface area of equilibrium morphology, surface energy, and concentration of reagent is investigated.

$$\Delta G = \sum_{hkl} O_{hkl} \gamma_{hkl}$$

1

where O_{hkl} is the surface area, and γ_{hkl} is the surface energy (energy required to create a surface of the unit area parallel to the (*hkl*) plane of the crystal). The surface energy of the (*hkl*) surface is proportional to the distance from the crystal's centre to the corresponding surface:

$$\frac{\gamma_{hkl}}{d_{hkl}} = constant$$

2

The equilibrium morphology of Wulff construction relies on the surface energy ratio. However, direct measurement of the surface energy remains challenging. In this work, the FSAs are used as the input parameters and ANN to acquire the surface energy. It should be pointed out that the absolute value of the surface energy cannot be obtained from the equilibrium morphology and only the surface energy ratio can be determined. By selecting a reference surface such as $\{h_0k_0l_0\}$, equation (1) can be written as

$$\Delta G = \gamma_{h_0 k_0 l_0} \sum_{hkl} A_{hkl} \gamma_{hkl} / \gamma_{h_0 k_0 l_0}$$

3

where $\gamma_{h_0 k_0 l_0}$ is the surface energy of $\{h_0 k_0 l_0\}$ facet, A_{hkl} is the fractional surface area, and $A_{hkl} = O_{hkl}/O_{total}$, O_{total} is the total surface area. The thermodynamic relationship between the concentration c and the surface adsorption excess Γ are described by Gibbs adsorption isotherm[19,20]:

$$\Gamma = -\frac{c}{RT} \frac{\partial \sigma}{\partial c},$$

4

where T is the temperature, R is the universal gas constant, and σ is the surface tension. The surface excess is usually evaluated based on the Langmuir adsorption model[12,21]:

$$\Gamma = \Gamma_s \frac{\alpha c}{1 + \alpha c},$$

5

where Γ_s saturated adsorption and α the Langmuir constant. Combining equation (4) and equation (5), we have:

$$\frac{\partial \sigma}{\partial c} = -\Gamma_s RT \frac{\alpha}{1 + \alpha c}$$

6

Integrating equation (6), we have

$$\sigma = -\Gamma_s RT \int \frac{\alpha}{1 + \alpha c} dc = -\Gamma_s RT \ln(1 + \alpha c) + \text{constant}.$$

7

Since the surface energy and surface tension are numerically equal for the liquid and isotropic solid surfaces, equation (7) can be simplified to

$$\gamma = e_0 \ln(1 + c e_1) + e_2,$$

8

where $e_0 = -\Gamma_s RT$, $e_1 = \alpha$ and $e_2 = \text{constant}$.

Kinetics models. The kinetics models are then derived by monitoring the UV-Vis-NIR absorption spectra and color (RGB value) changes during nanocrystal growth. The values of AR (and RGB) change with time and follow the exponential decay function[29]:

$$\mathrm{AR}(t) = \mathrm{AR}_\infty + (\mathrm{AR}_m - \mathrm{AR}_\infty) e^{-(t - t_m)/\tau}$$

9

where, AR_∞ is the final AR, AR_m is the maximum AR in the growth process, t_m is the time corresponding to the maximum AR, and τ is the decay rate of the AR. Based on the results in Fig. 3g and Fig. S7, a first-order reaction[30] can be expressed as:

$$r = -\frac{d\,[P]}{dt} = k\,[P]$$

10

Where the optical concentration of [Au⁺] is extracted from UV-Vis-NIR spectra and expressed in the form of $OD_{[Au+]}$, r is the reaction rate, and k is the reaction rate constant. The integrated rate law for the pseudo-first-order reaction can then be obtained by:

$$\ln\,[P] = -kt + \ln[P]_0,$$

11

where $[P]_0$ is the initial optical concentration of [Au⁺].

Declarations

Acknowledgements

Data availability

All data in support of the findings of this study are available within the Article and its Supplementary Information.

Code availability

Computer code, algorithm, and the related data are available in the Zenodo repository with the digital object identifier (10.5281/zenodo.5893947 with the link: https://zenodo.org/record/5893264#.Ye0sS9pBw2x) to generate results that are reported in the paper and central to its main claims. The algorithms of ANN[8], Chemical Named Entity Recognition[9], expressions and grammatical structures[21], SISSO[22,23] are adaptable as described in detail in ref. 8, ref. 9, ref. 21, ref. 22, and ref. 23, respectively.

References

1. Burger, B. *et al.* A mobile robotic chemist. *Nature* **583**, 237–241, doi:10.1038/s41586-020-2442-2 (2020).

2. Angelone, D. *et al.* Convergence of multiple synthetic paradigms in a universally programmable chemical synthesis machine. *Nature Chemistry* **13**, 63–69, doi:10.1038/s41557-020-00596-9 (2021).

3. Moses, O. A. *et al.* Integration of data-intensive, machine learning and robotic experimental approaches for accelerated discovery of catalysts in renewable energy-related reactions. *Materials Reports: Energy*, 100049, doi:https://doi.org/10.1016/j.matre.2021.100049 (2021).

4. Lin, L., Kollipara, P. S. & Zheng, Y. Digital manufacturing of advanced materials: Challenges and perspective. *Materials Today* **28**, 49–62, doi:https://doi.org/10.1016/j.mattod.2019.05.022 (2019).

5. An, B. *et al.* Programming Living Glue Systems to Perform Autonomous Mechanical Repairs. *Matter* **3**, 2080-2092, doi:https://doi.org/10.1016/j.matt.2020.09.006 (2020).

6. Steiner, S. *et al.* Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, eaav2211, doi:10.1126/science.aav2211 (2019).

7. Perera, D. *et al.* A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science* **359**, 429–434, doi:10.1126/science.aap9112 (2018).

8. Lai, F. *et al.* Machine Learning-Aided Crystal Facet Rational Design with Ionic Liquid Controllable Synthesis. *Small* **n/a**, 2100024, doi:https://doi.org/10.1002/smll.202100024.

9. Jessop, D. M., Adams, S. E., Willighagen, E. L., Hawizy, L. & Murray-Rust, P. OSCAR4: a flexible architecture for chemical text-mining. *Journal of Cheminformatics* **3**, 41, doi:10.1186/1758-2946-3-41 (2011).

10. Katz-Boon, H. *et al.* Stability of Crystal Facets in Gold Nanorods. *Nano Letters* **15**, 1635–1641, doi:10.1021/acs.nanolett.5b00124 (2015).

11. Sun, Y. & Xia, Y. Shape-Controlled Synthesis of Gold and Silver Nanoparticles. *Science* **298**, 2176–2179, doi:10.1126/science.1077229 (2002).

12. Kim, F., Connor, S., Song, H., Kuykendall, T. & Yang, P. Platonic Gold Nanocrystals. *Angewandte Chemie International Edition* **43**, 3673–3677, doi:https://doi.org/10.1002/anie.200454216 (2004).

13. Xia, X., Zeng, J., Oetjen, L. K., Li, Q. & Xia, Y. Quantitative Analysis of the Role Played by Poly(vinylpyrrolidone) in Seed-Mediated Growth of Ag Nanocrystals. *Journal of the American Chemical society* **134**, 1793–1801, doi:10.1021/ja210047e (2012).

14. Wang, Y. *et al.* Synthesis of Silver Octahedra with Controlled Sizes and Optical Properties via Seed-Mediated Growth. *ACS Nano* **7**, 4586–4594, doi:10.1021/nn401363e (2013).

15. Jackson, S. R., McBride, J. R., Rosenthal, S. J. & Wright, D. W. Where's the Silver? Imaging Trace Silver Coverage on the Surface of Gold Nanorods. *Journal of the American Chemical society* **136**, 5261–5263, doi:10.1021/ja501676y (2014).

16. Garrido, C. *et al.* Gold nanoparticles to improve HIV drug delivery. *Future Medicinal Chemistry* **7**, 1097–1107, doi:10.4155/fmc.15.57 (2015).

17. Li, P. *et al.* Evaporative Self-Assembly of Gold Nanorods into Macroscopic 3D Plasmonic Superlattice Arrays. *Advanced Materials* **28**, 2511–2517, doi:https://doi.org/10.1002/adma.201505617 (2016).

18. Li, X. *et al.* Nanotransducers for wireless neuromodulation. *Matter* **4**, 1484–1510, doi:https://doi.org/10.1016/j.matt.2021.02.012 (2021).

19. Xie, H. *et al.* Electrostatic Self-Assembly of Ti3C2Tx MXene and Gold Nanorods as an Efficient Surface-Enhanced Raman Scattering Platform for Reliable and High-Sensitivity Determination of Organic Pollutants. *ACS Sensors* **4**, 2303–2310, doi:10.1021/acssensors.9b00778 (2019).

20. Huang, H. *et al.* Morphological control of gold nanorods via thermally driven bi-surfactant growth and application for detection of heavy metal ions. *Nanotechnology* **29**, 334001, doi:10.1088/1361-6528/aac6b2 (2018).

21. Hawizy, L., Jessop, D. M., Adams, N. & Murray-Rust, P. ChemicalTagger: A tool for semantic text-mining in chemistry. *Journal of Cheminformatics* **3**, doi:10.1186/1758-2946-3-17 (2011).

22. Ouyang, R., Curtarolo, S., Ahmetcik, E., Scheffler, M. & Ghiringhelli, L. M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Physical Review Materials* **2**, 083802, doi:10.1103/PhysRevMaterials.2.083802 (2018).

23. Ouyang, R., Ahmetcik, E., Carbogno, C., Scheffler, M. & Ghiringhelli, L. M. Simultaneous learning of several materials properties from incomplete databases with multi-task SISSO. *Journal of Physics: Materials* **2**, 024002, doi:10.1088/2515-7639/ab077b (2019).

24. Wulff, G. Xxv. zur frage der geschwindigkeit des wachsthums und der auflösung der krystallflächen. *Zeitschrift für Kristallographie-Crystalline Materials* **34**, 449–530, doi:10.1524/zkri.1901.34.1.449 (1901).

25. Martínez-Balbuena, L., Arteaga-Jiménez, A., Hernández-Zapata, E. & Márquez-Beltrán, C. Applicability of the Gibbs Adsorption Isotherm to the analysis of experimental surface-tension data for ionic and nonionic surfactants. *Advances in colloid and interface science* **247**, 178–184, doi:10.1016/j.cis.2017.07.018 (2017).

26. Scatchard, G. The Gibbs adsorption isotherm. *The Journal of Physical Chemistry* **66**, 618–620, doi:10.1021/j100810a011 (1962).

27. Langmuir, I. The adsorption of gases on plane surfaces of glass, mica and platinum. *Journal of the American Chemical society* **40**, 1361–1403 (1918).

28. Wang, X. *et al.* The synergistic elimination of uranium (VI) species from aqueous solution using bi-functional nanocomposite of carbon sphere and layered double hydroxide. *Chemical Engineering Journal* **342**, 321–330, doi:10.1016/j.cej.2018.02.102 (2018).

29. Wang, W. *et al.* Real-time SAXS and ultraviolet-visible spectral studies on size and shape evolution of gold nanoparticles in aqueous solution. *The European Physical Journal B* **76**, 301–307, doi:10.1140/epjb/e2010-00177-3 (2010).

30. Ruditskiy, A., Zhao, M., Gilroy, K. D., Vara, M. & Xia, Y. Toward a quantitative understanding of the sulfate-mediated synthesis of Pd decahedral nanocrystals with high conversion and morphology yields. *Chemistry of Materials* **28**, 8800–8806, doi:10.1021/acs.chemmater.6b04528 (2016).
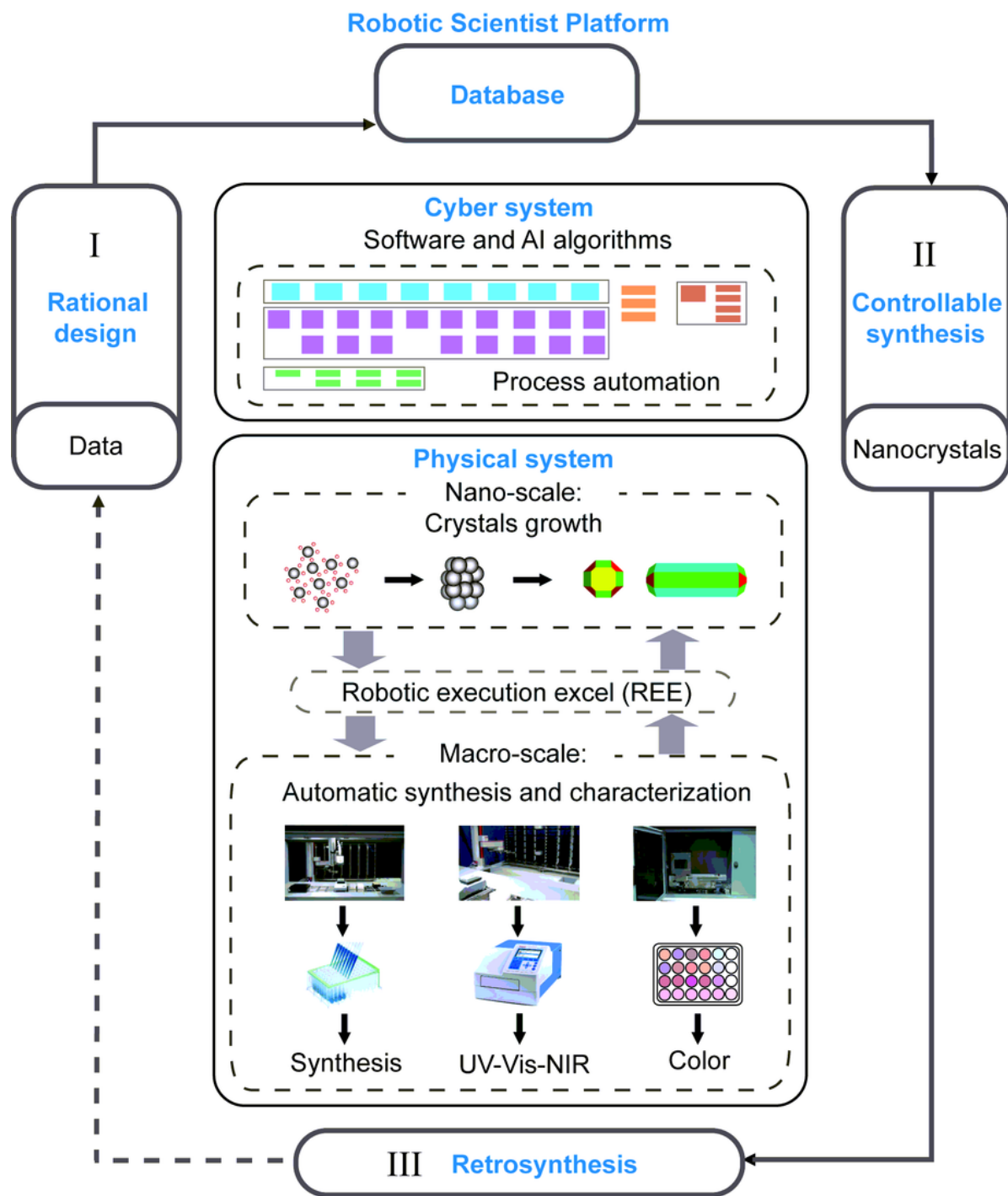
Figures

Figure 1

Robotic Scientist platform. Convergence of the database, cyber system, and physical system and process flow: **I.** Rational design, **II.** Controllable synthesis, and **III.** Retrosynthesis for closed-loop synthesis of nanocrystals based on the Robotic Scientist platform.
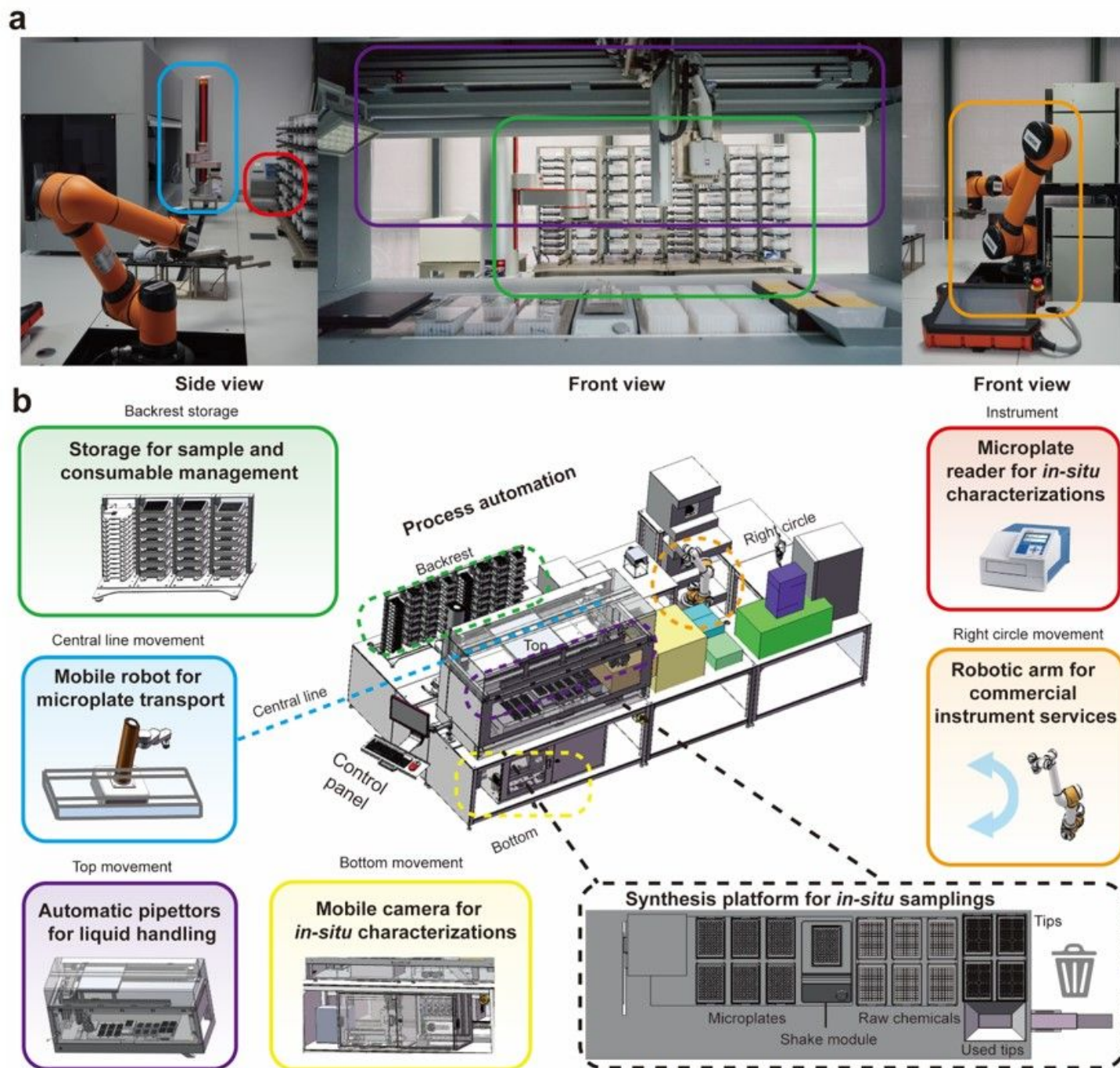
Figure 2

Illustration of the Robotic Scientist platform. a, Photograph. **b,** Schematic representation. The color frames in the photograph and schematic representation match each other. Backrest: Storage for the sample, microplates and pipette tips; Central line: Mobile robot for microplate transport; Top: Three automatic pipettors for liquid handling; Bottom: Mobile color-ultra-sensitive camera for *in situ* color characterization; Platform: Synthesis platform for *in situ* sampling; Instrument: Microplate reader for *in situ* UV-Vis-NIR absorption spectrophometry; Right circle: Robotic arm for instrument services.
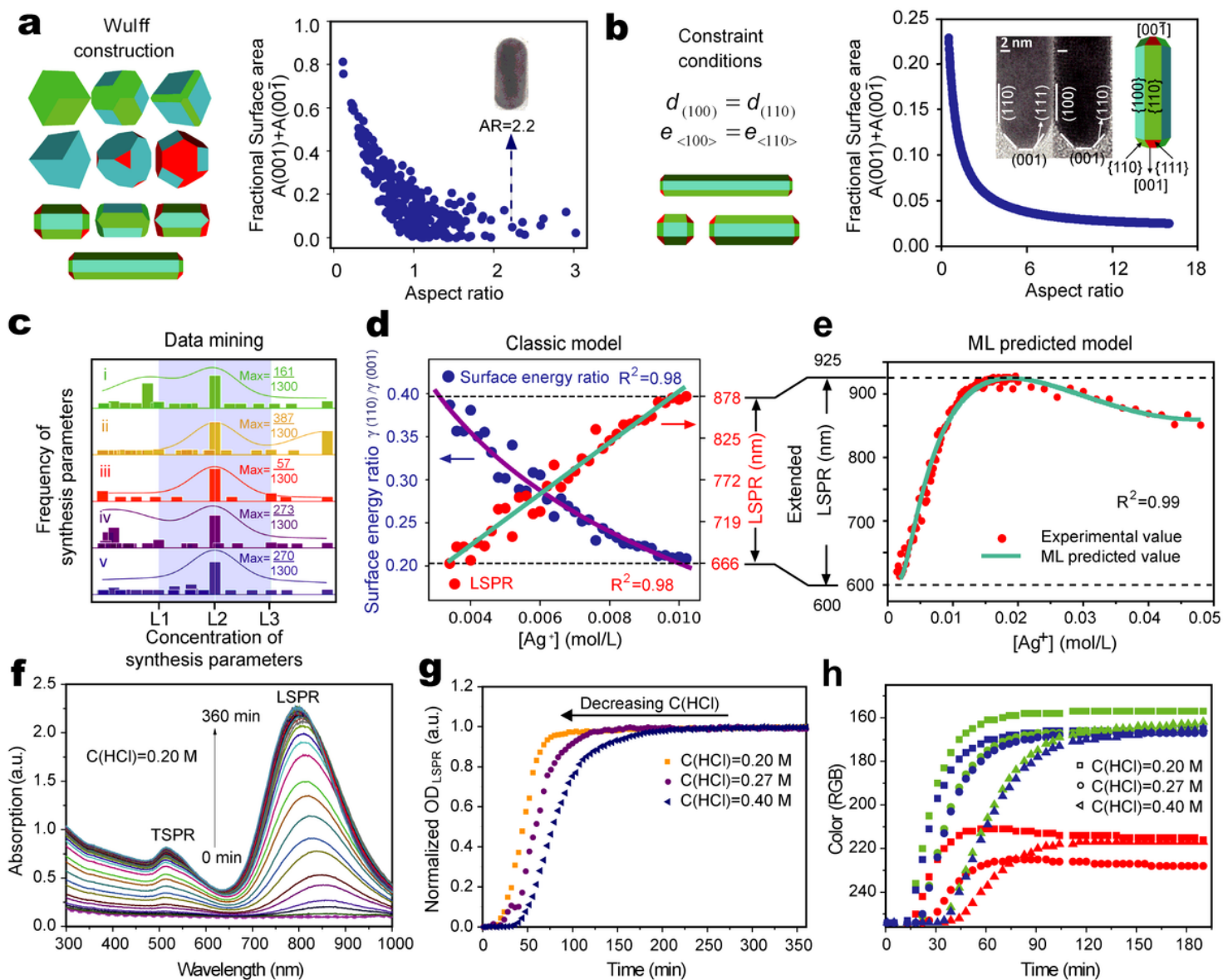
Figure 3

Data-driven rational design. a, Database of cubic Au nanocrystals: Left: Morphologies of Wulff construction; Right: Correlations of fractional surface area and aspect ratio (more figs in Fig S1). **b,** Database of Au nanorods: Left: Constraint conditions of the morphologies; Right: Refined correlations of the fractional surface area and aspect ratio of specific nanorods. **c,** Data mining of Au nanocrystals synthesis parameters (i, ii, iii, iv and v refer to the parameters of $AgNO_3$, hexadecyltrimethylammonium bromide (CTAB), HCl, $HAuCl_4$ and L-ascorbic acid (AA), respectively. L1, L2, and L3 are low, middle, and high levels). **d,** Classic thermodynamic model. **e,** Machine learning (ML) trained model. **f,** Kinetic analysis of UV-Vis-NIR absorption spectra. **g,** Normalized OD_{LSPR} results of nanocrystals growth under varied C(HCl). **h,** Kinetic analysis of color changes with time. The color (red, green and blue)-time relationships with different C(HCl) is presented by the symbols of square, circle and triangle, respectively.
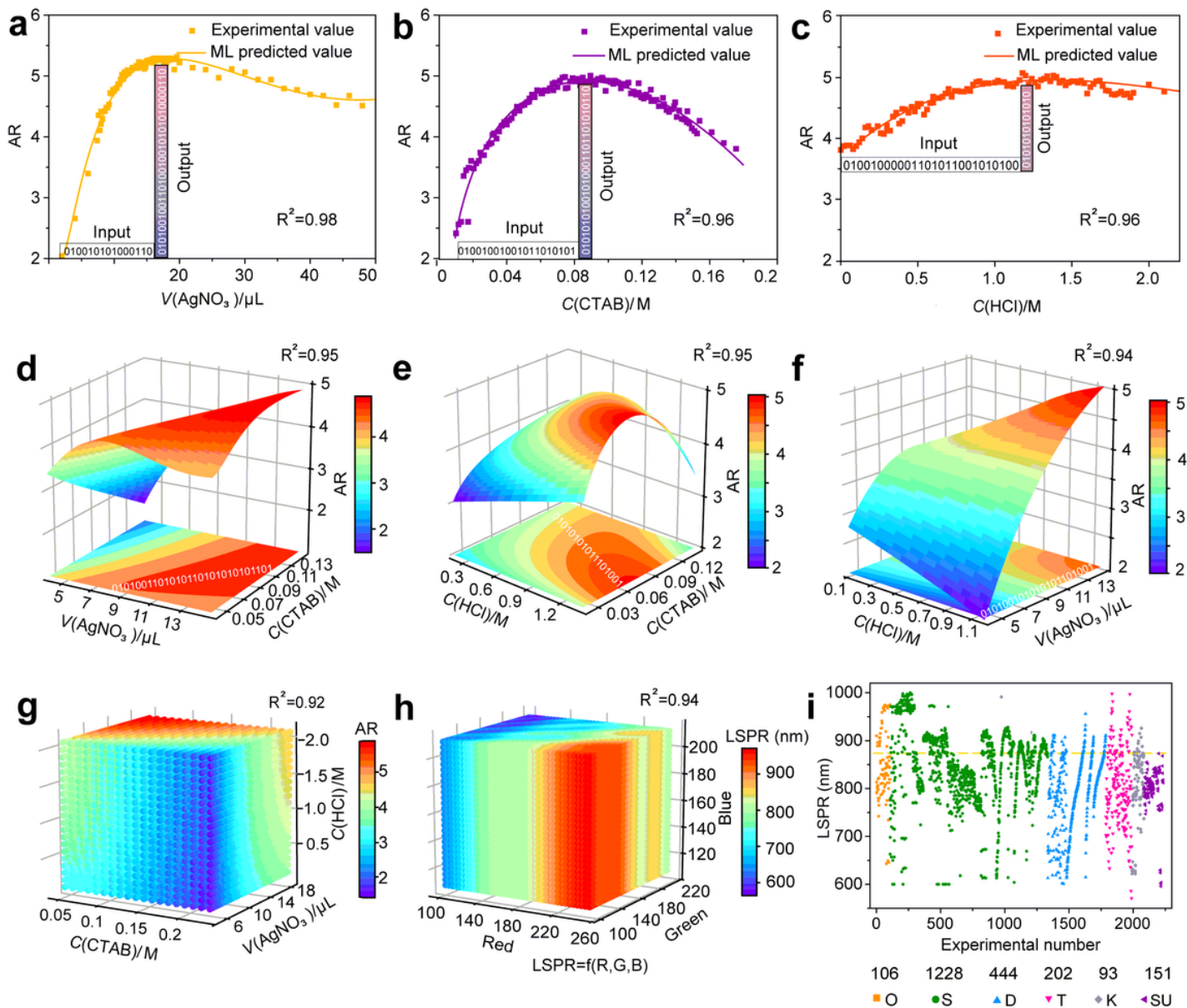
Figure 4

Controllable synthesis, ML prediction, and database construction. a-c, Single-factor ML predicted models. **d-f,** Double-factor ML predicted models. **g,** Triple-factor ML predicted models. **h,** LSPR-color model. **i,** Overview of the number of experiments: O, S, D, T, K, and SU represent the orthogonal, single-, double-, triple-factor, kinetics, and scale-up experiments, respectively. The relationship between the experimental factors (as inputs) and AR (as outputs) is identified, and '01010101' is the schematic diagram of the controllable range.
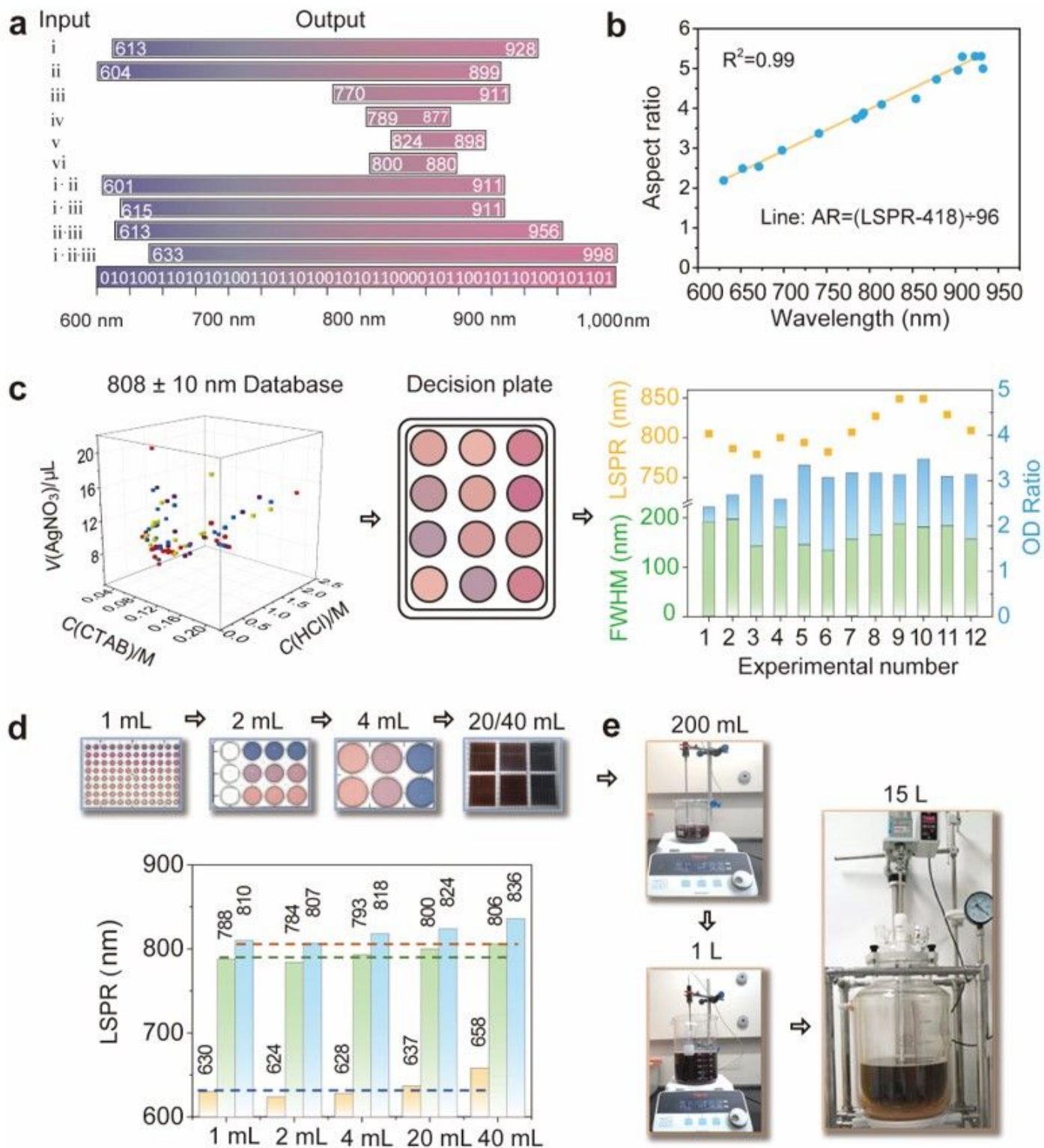
Figure 5

Robotic Scientist platform facilitated retrosynthesis. a, Au nanocrystals genome: Experimental data and ML predictable range, and **b,** Linear relationship between LSPR and AR. (validation by TEM images in Fig. S5) **c,** Retrosynthesis and optimization of the target Au nanocrystals (LSPR as 808 ± 10 nm, for example), showing the experimental database, decision plate and the evaluation criteria for optimization. **d,** Sequential scale-up of representative nanocrystals with parameters for minor modification study (LSPR,

located within 633 ± 10 nm, 780 ± 10 nm, and 808 ± 10 nm, 1 mL on a 96-well microplate, 2 mL on a 12-well microplate, 4 mL on a 6-well microplate, 20 and 40 mL on one plate). **e,** Bench-scale experiments on a magnetic stirrer (200 and 1,000 mL), and pilot-scale experiment in an agitated vessel (15 L).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MOVIES1.mp4
- SupportingInformation.docx