Chinese Chemical Letters 34 (2023) 107829

Contents lists available at ScienceDirect

Chinese Chemical Letters

journal homepage: www.elsevier.com/locate/cclet

Diagnosis and subtype classification on serum peptide fingerprints by mesoporous polydopamine with built-in metal-organic framework

Zixing Xu[a], Haolin Chen[a], Huimin Chu[a], Xizhong Shen[b], Chunhui Deng[a,b,**], Nianrong Sun[b,*], Hao Wu[b,*]

[a] Department of Gastroenterology and Hepatology, Zhongshan Hospital, Department of Chemistry, Institute of Metabolism & Integrate Biology (IMIB), Fudan University, Shanghai 200433, China
[b] Department of Gastroenterology and Hepatology, Zhongshan Hospital, Fudan University, Shanghai 200032, China

ARTICLE INFO

Article history:
Received 19 June 2022
Revised 29 August 2022
Accepted 14 September 2022
Available online 17 September 2022

Keywords:
Mesoporous dopamine
Metal-organic frameworks
Serum peptides
Mass spectrometry
Crohn's disease
Disease diagnosis

ABSTRACT

Crohn's disease (CD) as a big issue to public health needs an accurate diagnosis urgently that is the common challenge among internal diseases. Herein, we design a mesoporous polydopamine with built-in metal-organic frameworks (dubbed MMP-*b*-MOFs) to combine with high-throughput mass spectrometry to extract serum peptide fingerprints from CD and healthy controls (HC). Benefitting by the size-exclusion and strong hydrophilicity of MMP-*b*-MOFs, the extracted peptide fingerprints present extremely high quality. CD and HC are explicitly discriminated with orthogonal partial least squares discriminant analysis (OPLS-DA), the corresponding area under the curve (AUC) value is 1.000. Moreover, eight peptides with clear identity are screened out and achieve the accurate diagnosis and subtype classification of CD, with all AUC values up to 1.000. Moreover, the unsupervised model is also established to precisely classify HC and CD based on these eight clearly identified peptides. This work brings great benefits for clinical detection especially internal medicine.

Crohn's disease (CD) is a progressive and destructive chronic inflammatory bowel disorder of the gastrointestinal tract with an increasing incidence worldwide and unclear etiology [1], easily causing abdominal pain, diarrhea, and gastrointestinal bleeding [2,3]. In addition to clinical symptoms, the integration of radiology, endoscopy and histological biopsy is also demanded for CD detection, which is accompanied by long-time consumption, high cost and low acceptance among patients, leading to serious complications [4]. Up to 50% of patients have undergone intestinal resection after CD is diagnosed [5]. Thus, an early detection of CD is paramount to guide therapy.

Peptidomics refers to determinate all endogenous peptides expressed in a biological sample, which is a rather heterogeneous entity including small-size proteins, fragments of protein degradation, peptide hormones and neuropeptides. Their expression levels tend to vary from one moment to another, and can reflect valuable information of the particular physiological and pathological status. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), as a rapid and high-throughput technique with high resolution and high sensitivity, has been a common method for the detection of peptides, metabolites, proteins, DNA or other molecules [6–9]. Up to date, much efforts have been devoted to combine beads with MALDI-TOF MS to reveal differential expression patterns by making a comparison on peptidome between diseased samples and healthy samples, aiming at discovery of biomarkers for diseases or indicators for pathological events. For example, Zhang *et al.* revealed that the simultaneous increase of inter-α-trypsin inhibitor (*m/z* 3272), apolipoprotein A-II (28–94, *m/z* 7800) and apolipoprotein A-I (*m/z* 28,043) may indicate the early ovarian cancer [10]. Villanueva *et al.* reported that fibrinopeptide A, bradykinin and their several fragments, as well as other fragments from fibrinogen α-chain, C3, C4a, inter-α-trypsin inhibitor, apolipoprotein A-I, apolipoprotein A-IV, apolipoprotein E, clusterin, high–molecular-mass kininogen, factor XIII and transthyretin may serve the prostate, bladder, or breast cancers [11]. In addition, des-Ala-fibrinopeptide A (*m/z* 1466) and fibrinogen α-chain (175–200, *m/z* 2664) were reported for liver and oral cancer, respectively [12,13]. Interestingly, we found that these peptide fragments prefer to be hydrophilic. As a

* Corresponding authors.
** Corresponding author at: Department of Gastroenterology and Hepatology, Zhongshan Hospital, Department of Chemistry, Institute of Metabolism & Integrate Biology (IMIB), Fudan University, Shanghai 200433, China.
E-mail addresses: chdeng@fudan.edu.cn (C. Deng), sunnianrong@fudan.edu.cn (N. Sun), wu.hao@zs-hospital.sh.cn (H. Wu).

https://doi.org/10.1016/j.cclet.2022.107829
1001-8417/

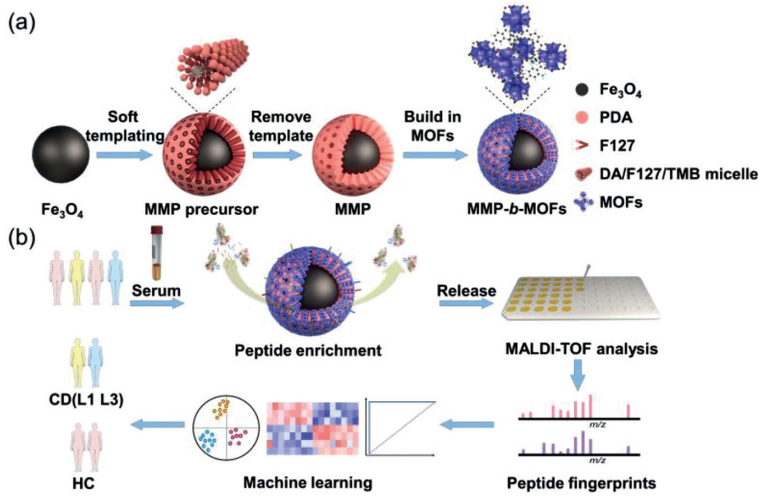

Fig. 1. (a) The synthesis scheme of MMP-*b*-MOFs. (b) Scheme of establishing a diagnostic model based on serum peptidome fingerprints for diagnosis and classification of Crohn's disease patients.

consequence, hydrophilic endogenous peptides may hold greater potential in serving as biomarkers.

Mesoporous materials, which feature with large surface area and ordered porous channels, have experienced flourishing development in peptidomics research over the past decades [14,15]. Generally, mesoporous silica was intensively employed because of its strong modifiability. However, an appropriate linker is required to connect the functional groups and silica wall. Compared with conventional silica, polydopamine (PDA), another intensively applied linker, is a better candidate for fabricating mesoporous materials, for it possesses abundant catechol and amine groups that can endow itself with great hydrophilicity and facilitate modification [16]. Moreover, it has been well known that PDA deriving from the self-polymerization reaction of dopamine under a mild condition can deposit on nearly all kinds of surfaces, which is a huge advantage that can dramatically simplify the surface modification process of materials. In recent years, mesoporous PDA has been considerably explored and applied in numerous fields [17], including electrocatalysis [18], adsorption [19], photothermal therapy [20], sensors [21], drug delivery and tumor therapy [22]. More recently, we tentatively fabricated the mesoporous PDA on magnetic nanoparticles for immobilization of metal ions and achieved high-efficiency enrichment of phosphopeptides in human saliva [23]. Inspired by the above, mesoporous PDA can be both an outstanding structural scaffold and linker to contribute a great to peptidomics research.

Metal organic frameworks (MOFs), as an emerging class of porous materials, consist of metal ions and organic ligands, based on which MOFs exhibit promising functional tunability and have been widely used as adsorbents in peptidomics research [24–26]. Among various MOFs, UiO-66-NH_2 stands out by the exceptional merits such as simple preparation, strong hydrophilicity and stability [27–29]. In this study, we fabricated large mesoporous PDA on magnetic nanoparticles to serve as structural scaffold and linker to build in hydrophilic UiO-66-NH_2 (denoted as MMP-*b*-MOFs). The UiO-66-NH_2 could be firmly coated on the pore wall, endowing MMP-*b*-MOFs with excellent hydrophilicity and size-exclusion effect. By combining MMP-*b*-MOFs and MALDI-TOF MS, high-quality serum peptide fingerprints were extracted from 50 CD and 50 healthy controls (HC). Based on these serum peptide fingerprints, CD was distinguished from HC with area under the curve (AUC) value of 1.000. Moreover, eight potential peptides were screened out and realized the discrimination and subtype classification of CD successfully, all of AUC values were up to 1.000. This exciting result disclosed the wide prospective of peptides as molecular markers for disease diagnosis and subtype classification, even the prognosis monitoring.

The synthesis procedure and application of core-shell MMP-*b*-MOFs in serum samples are illustrated in Figs. 1a and b. Briefly, a vesicular structure was firstly constructed on the surface of magnetic nanoparticles by using triblock copolymer Pluronic F127 and 1,3,5-trimethyl benzene (TMB) as structure-directing agents, and dopamine as functional monomer for *in situ* polymerization. Then an acetone refluxing process was adopted to remove the structure-directing agents to obtain MMP which was subsequently applied to support the growth of hydrophilic MOFs through metal-oxygen units. Afterwards, the obtained MMP-*b*-MOFs material was used to pre-treat serum samples for disease detection. Transmission electron microscope (TEM) and field emission scanning electron microscope (FESEM) are initially employed to characterize the morphology and structure of MMP-*b*-MOFs and MMP. From images in Figs. 2a-d, it can be observed that both MMP-*b*-MOFs and MMP present uniformly spherical morphology as well as the core-shell structure. Compared to images of MMP, the mesoporous channels seem to be narrower, which are further tested by recording nitrogen adsorption–desorption isotherms. As shown in Fig. S1a (Supporting information), there are two main pore sizes in MMP-*b*-MOFs, 3.80 nm and 6.79 nm, both of which are smaller than MMP (12.84 nm) [23]. This means the successful imbedding of hydrophilic MOFs in the relatively large mesoporous channels of MMP, and the narrow pore size will be beneficial to peptide adsorption and large protein exclusion. In addition, the Brunauer-Emmett-Teller surface area of MMP-*b*-MOFs is evaluated as 230.65 m^2/g, which is larger than MMP (182.76 m^2/g) [23]. This should be attributed to the introduction of hydrophilic MOFs.

Furthermore, a series of characterizations are employed to depict the successful introduction of hydrophilic MOFs. At first, energy dispersive X-ray spectrum in Fig. S1b (Supporting information) manifests the existence of Fe, Zr, O, C and N, which is consistent with the results of element mappings in Fig. 2e. Moreover, from Fig. 2e it can be deduced that MOFs evenly display on MMP. Also, Fourier transform infrared (FT-IR) spectra are utilized to confirm and compare the composition of MMP and MMP-*b*-MOFs (Fig. S1c in Supporting information). The absorption bands at 576 cm^{-1} is attributed to the Fe−O stretching vibration [30]. Compared to the FTIR spectrum of Fe_3O_4, new peaks at 1352 and 1496 cm^{-1} are respectively ascribed to C–N bending vibration and aromatic C=C stretching vibration, which are derived

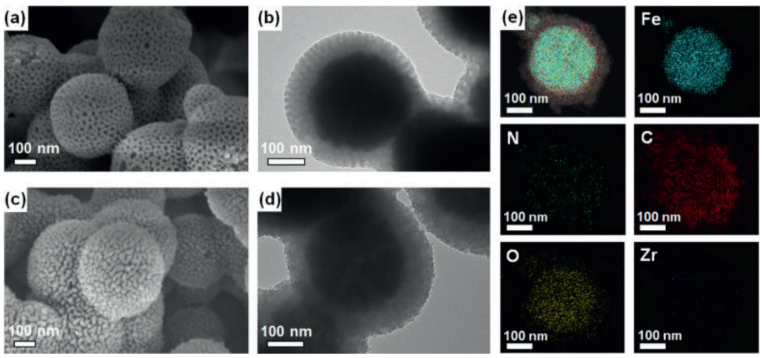

Fig. 2. FESEM images of MMP (a) and MMP-*b*-MOFs (c), TEM images of MMP (b) and MMP-*b*-MOFs (d). Element mappings of MMP-*b*-MOFs (e).

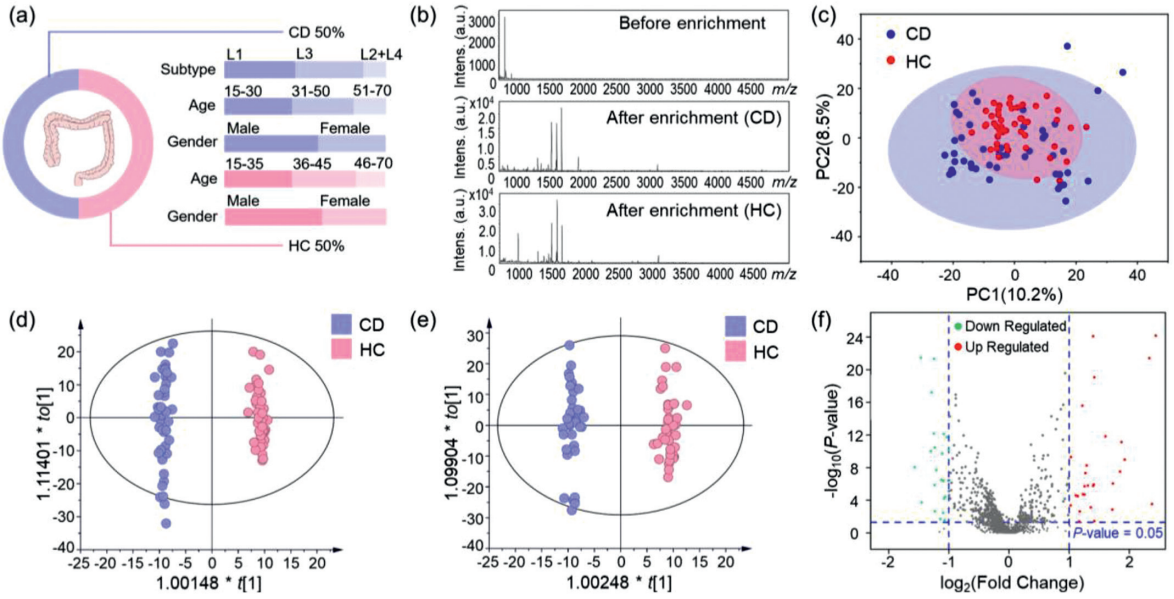

Fig. 3. (a) Gender and age distribution of 50 CD and 50 HC. (b) The representative serum peptide fingerprinting. (c) PCA analysis to distinguish CD from HC in all samples based on 1526 peaks. OPLS-DA model between (d) all the samples and (e) discovery cohort. (f) Volcano plot of discovery cohort.

from polydopamine [31]. Compared to FTIR spectrum of Fe_3O_4 and MMP, peaks in the range of 1500–1700 cm^{-1} are originated from asymmetrical and symmetrical stretching vibrations of the carboxylate groups of the organic ligand in MOFs, and the peaks at 600–800 cm^{-1} are assigned to Zr–O bond [32]. All the above results imply the successful modification of MMP-*b*-MOFs. Water contact angles of Fe_3O_4, MMP precursor, MMP and MMP-*b*-MOFs are respectively measured as 39.96°, 30.08°, 28.36° and 18.75° (Fig. S1d in Supporting information), demonstrating the build-in of MOFs enhances hydrophilicity. Finally, magnetic hysteresis loop in Fig. S1e (Supporting information) reveals the magnetization saturation value of MMP-*b*-MOFs is 47.12 emu/g, enabling rapid separation with the help of exterior magnetic field.

In this work, 100 serum samples are collected including 50 CD and 50 HC, the gender and age distribution of which are displayed in Fig. 3a. Non-parametric Mann-Whitney U tests between CD and HC prove there is no significant difference in gender and age (gender: $Z = -0.202$, $P = 0.840$; age: $Z = -1.749$, $P = 0.080$, Tables S1 and S2 in Supporting information). The representative spectra are presented in Fig. 3b.

Before enrichment, peptide signals are strongly suppressed, while after enrichment by MMP-*b*-MOFs, clear serum peptide signals from CD or HC can be observed. Then specific peptide features will be screened out by combining all serum peptide fingerprints and machine learning algorithm for establishing robustly discrimination model towards CD including diagnosis and subtype classification. The detailed workflow of producing specific peptide features is displayed in Fig. S2 (Supporting information). Herein, after processing by MALDIquant, a total of 1526 features are determined. Based on all these features, principal component analysis (PCA) that an unsupervised machine learning algorithm is implemented to tentatively distinguish CD from HC. As viewed in Fig. 3c, these two groups cannot be separated. Subsequently, OPLS-DA that a supervised machine learning algorithm is employed. Delightfully, two groups cluster respectively ($R^2Y(cum) = 0.992$, $Q^2(cum) = 0.938$, Fig. 3d), indicating the feasibility of OPLS-DA. The permutation test using 200 iterations was proceeded as shown in Fig. S3a (Supporting information), R^2 and Q^2 points corresponding to analog values on the right are higher than points ascribed to actual values on the left, and the intercept of blue regression line (Q^2) is less than zero, proving the OPLS-DA model based on peptide features is reliable.

Inspired by the above results, all of 100 samples are randomly divided into discovery cohort containing 40 CD and 40 HC, and validation cohort containing 10 CD and 10 HC, for further analysis. We apply OPLS-DA model to all features from the discovery cohort. As expected, CD and HC can be distinctly separated ($R^2Y(cum) = 0.988$, $Q^2(cum) = 0.926$, Fig. 3e, the permutations plot is shown in Fig. S3b (Supporting information), and the corresponding receiver operating characteristic (ROC) curve is shown in Fig. S3c (Supporting

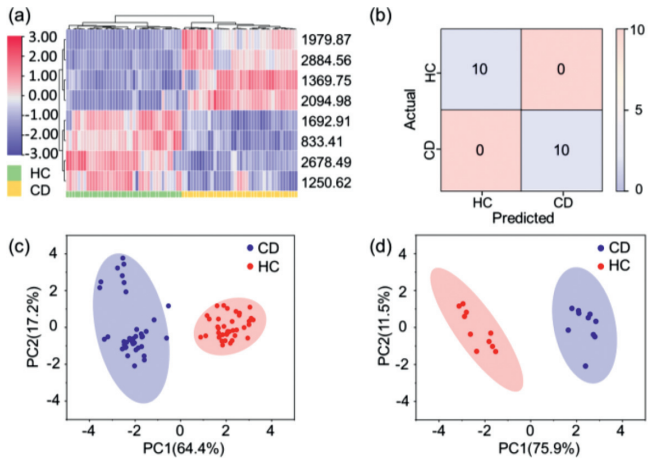

Fig. 4. (a) Heat map of specific 8 peptide features in discovery cohort. (b) Confusion matrix of the classification results of the validation cohort in the OPLS-DA model using 8 selected features. PCA analysis based on 8 features to distinguish CD and HC in (c) discovery cohort and (d) validation cohort.

information), from which the AUC value achieves up to 1.000, indicating the perfect discrimination effect of peptide features towards CD from HC.

Afterwards, all these features from discovery cohort are subjected to multiple feature selection algorithms to be sorted, including T-test of the expression level analysis, fold change analysis, variable importance on projection (VIP) analysis. In detail, volcano plot on fold change and discovery cohort and expression level shows the up- or down- regulation of these features (Fig. 3f), further combining with VIP values greater than 2, 8 specific features are finally selected. As shown in Fig. 4a, the heat map of these 8 specific features in discovery cohort shows their great contribution to distinguish CD from HC. The diagnosis efficiency of the 8 specific features was then tested using the validation cohort. Surprisingly, the confusion matrix of the classification results confirm CD and HC can be correctly distinguished (Fig. 4b), suggesting the 8 specific features may serve as biomarkers in diagnosing CD clinically. More surprisingly, based on the 8 specific features, CD and HC can be clearly distinguished using PCA (Figs. 4c and d), for which the first two principal components respectively explain 81.6% and 87.4% of variation in discovery cohort and validation cohort, with 95% confidence interval (CI), proving their high responsibility towards difference between CD and HC. In short, all the above results demonstrated remarkable diagnosis ability of the selected features.

Furthermore, we combine LC-MS/MS with Uniprot to determine these specific peptide features using several CD and HC samples. The detailed information is listed in Table S3 (Supporting information). All these specific features are hydrophilic peptides, with grand average of hydropathicity (GRAVY) value less than zero, indicating the effective extraction of MMP-*b*-MOFS. Moreover, the results show that these specific peptide features are respectively identified as protein fragments of alpha-1-antichymotrypsin, clusterin, complement C3, alpha-1-antitrypsin, alpha-2-macroglobulin, transthyretin, indicating the reasonability of these specific peptide features in CD detection combining with previous reports [33,34]. For example, Derer *et al.* reported that the inappropriate activation of the complement system is involved in pathogenetic mechanisms of CD [35]. Complement C3 is a crucial protein in activating antibody production and secreting cytokines, and exerts crucial effects in pathogen clearance. It presents obviously upregulated in this work, which is corresponding to that complement C3 may serve as a biomarker for directly evaluating the severity of CD [33]. Also, researches pointed out that the imbalance between the production and elimination of reactive oxygen species, namely oxidative stress, is a hallmark of inflammatory bowel disease [36]. The elevated clusterin in CD was to protect tissues from oxidative stress during pathological status and inflammatory response [34]. Moreover, alpha-2-macroglobulin, alpha-1-antichymotrypsin and alpha-1-antitrypsin are general protease inhibitors [37], and proteases were reported to take the main responsibility to maintain gastrointestinal homeostasis, its dysregulation normally advanced the inflammatory process [38].

Clinical manifestations of CD are varied because of the complex pathogenic factors, precise classification will be of great significance for picking appropriate therapy protocols. CD is clinically categorized into four subtypes containing terminal ileum (L1), colon (L2), ileocolon (L3), and upper gastrointestinal (L4) on the grounds of the different disease sites. In this work, we attempt to conduct subtype classification of CD by using those specific peptide features. Given the sample number of L2 subtype and L4 subtype is not enough, only 22 CD with L1 subtype and 21 CD with L3 subtype are applied to explore possibility of classifying CD subtypes. Based on the 8 specific peptide features, the OPLS-DA models are built among HC, L1 subtype and L3 subtype. As viewed in Figs. S4a and b (Supporting information), L1 subtype and HC (R^2Y(cum) = 0.936, Q^2(cum) = 0.926), L3 subtype and HC (R^2Y(cum) = 0.934, Q^2(cum) = 0.923) can be effectively distinguished in the OPLS-DA models. The corresponding permutations plots are shown in Figs. S3d and e (Supporting information), manifesting the feasibility of classifying CD with different subtypes. All of the classification models got AUC values as 1.000 (Figs. S4c and d in Supporting information). Moreover, as shown in Fig. S4e (Supporting information), the L1 subtype, L3 subtype, and HC occupy different intervals in the OPLS-DA model (R^2Y(cum) = 0.873, Q^2(cum) = 0.850). Similarly, the 3D scatter in Fig. S4f (Supporting information) also expressed that the three groups can be accurately classified. The 200 permutations in Fig. S3f (Supporting information) indicate this method is reliable for subtype classification. These results confirmed the rationality of the selected features and veracity of the established models.

In conclusion, we successfully built the MOFs in mesoporous polydopamine to obtain hydrophilic MMP-*b*-MOFs. Benefitting from the ordered mesopores and great hydrophilicity, as well as high-throughput MALDI-TOF MS, the serum peptide fingerprints were rapidly acquired from CD and HC samples. Based on these peptide finger-prints, OPLS-DA model achieved great diagnosis efficiency for CD from HC, with AUC value up to 1.000 (R^2Y(cum) = 0.992, Q^2(cum) = 0.938). With multiple feature selection algorithms, 8 feature peaks (p-value $\langle$ 0.05, VIP $\rangle$ 2, fold change values $>$ 2 or $<$ 0.5) were selected as potential biomarkers for CD detection and subtype classification. Based on the selected 8 features, CD serum samples can be well distinguished from HC. Furthermore, subtype classification of CD was realized based on these 8 peptide features (R^2Y(cum) = 0.873, Q^2(cum) = 0.850). All the results expressed the 8 features could be good indicators of CD, elucidating that our approach has great potential in high-throughput disease screening.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was financially supported by National Key R&D Program of China (No. 2018YFA0507501) and the National Natural Science Foundation of China (Nos. 22074019, 21425518, 22004017), and Shanghai Sailing Program (No. 20YF1405300).

References

[1] J. Lloyd-Price, C. Arze, A.N. Ananthakrishnan, et al., Nature 569 (2019) 655–662.
[2] B. Huang, Z. Chen, L. Geng, et al., Cell 179 (2019) 1160–1176.
[3] J. Torres, S. Mehandru, J.F. Colombel, L. Peyrin-Biroulet, Lancet 389 (2017) 1741–1755.
[4] S. Danese, W.J. Sandborn, J.F. Colombel, et al., Gastroenterology 157 (2019) 1007–1018.
[5] G. Roda, S.Chien Ng, P.G. Kotze, et al., Nat. Rev. Dis. Primers 6 (2020) 22.
[6] S. Lin, S. Dikler, W.D. Blincoe, et al., Science 361 (2018) 6402.
[7] Q. Zhang, H. Xiao, J. Zhan, et al., Chin. Chem. Lett. 33 (2022) 4746–4749.
[8] M. Chen, Z. Gui, K. Chen, et al., Chin. Chem. Lett. 33 (2022) 2086–2090.
[9] M. Chen, C. Qi, X. Tang, et al., Chin. Chem. Lett. 33 (2022) 3772–3776.
[10] Z. Zhang, R.C. Bast, Y. Yu Jr., et al., Cancer Res. 64 (2004) 5882–5890.
[11] J. Villanueva, D.R. Shaffer, J. Philip, et al., J. Clin. Investig. 116 (2006) 271–284.
[12] E. Orvisky, S.K. Drake, B.M. Martin, et al., Proteomics 6 (2006) 2895–2902.
[13] A.J. Cheng, L.C. Chen, K.Y. Chien, et al., Clin. Chem. 51 (2005) 2236–2244.
[14] J. Wang, Q. Ma, Y. Wang, et al., Anal. Chem. 93 (2021) 9486–9494.
[15] H. Xu, J. Guo, L. Yang, et al., Anal. Chem. 93 (2021) 9486–9494.
[16] X. Hu, Y. Lu, X. Shi, et al., Chem. Commun. 55 (2019) 14785–14788.
[17] T. Zhao, L. Chen, R. Lin, et al., Acc. Mater. Res. 1 (2020) 100–114.
[18] B.Y. Guan, L. Yu, X.W. Lou, J. Am. Chem. Soc. 138 (2016) 11306–11311.
[19] P. Pan, T. Zhang, Q. Yue, et al., Adv. Sci. 7 (2020) 2000443.
[20] L. Zhang, P. Yang, R. Guo, et al., Int. J. Nanomed. 14 (2019) 8647–8663.
[21] P. Zhang, Y. Zhang, X. Xiong, et al., Sens. Actuators B: Chem. 321 (2020) 128626.
[22] Y. Xing, J. Zhang, F. Chen, et al., Nanoscale 9 (2017) 8781–8790.
[23] Z. Xu, Y. Wu, H. Wu, et al., Anal. Chim. Acta 1146 (2021) 53–60.
[24] H. Qi, Z. Li, J. Ma, Q. Jia, J. Mater. Chem. B 10 (2022) 3560–3566.
[25] T. Zhu, Q. Gu, Q. Liu, et al., Talanta 240 (2022) 123193.
[26] H. Qi, G. Chen, Q. Jia, Talanta 247 (2022) 123563.
[27] J. Chun, S. Kang, N. Park, et al., J. Am. Chem. Soc. 136 (2014) 6786–6789.
[28] G.W. Peterson, J.J. Mahle, J.B. DeCoste, et al., Angew. Chem. Int. Ed. 128 (2016) 6343–6346.
[29] Y. Wang, W. Zhang, D. Li, et al., Adv. Sci. 8 (2021) 2004456.
[30] M. Zhao, C. Deng, X. Zhang, Chem. Commun. 50 (2014) 6228–6231.
[31] Z. Yang, L. Zhu, L. Chen, J. Colloid Interface Sci. 539 (2019) 76–86.
[32] J. Yang, Y. Dai, X. Zhu, et al., J. Mater. Chem. A 3 (2015) 7445–7452.
[33] K. Okada, H. Itoh, M. Ikemoto, Heliyon 7 (2021) e06554.
[34] A. Rostagno, M. Calero, J.L. Holton, et al., Neurobiol. Dis. 158 (2021) 105452.
[35] C. Sina, C. Kemper, S. Derer, Semin. Immunol. 37 (2018) 66–73.
[36] E.A. Franzosa, A. Sirota-Madi, J. Avila-Pacheco, et al., Nat. Microbiol. 4 (2019) 293–305.
[37] J. Vandooren, Y. Itoh, Front. Immunol. 12 (2021) 803244.
[38] N. Solà-Tapias, N. Vergnolle, A. Denadai-Souz, F. Barreau, J. Crohns Colitis 14 (2020) 1149–1161.