Artificial intelligence algorithms drive the deciphering of traditional Chinese medicine by analyzing the chemicalome, targetome, and bioactivome

Huipeng Song[1#], Zeyuan Liang[2#], Xinru Zhang[1#], Fengyao Yang[2], Mingyue Zheng[3*] and Guizhong Xin[2*]

[1] *Liaoning University of Traditional Chinese Medicine, Dalian, Liaoning 116600, China*
[2] *Department of Chinese Medicines Analysis, China Pharmaceutical University, Nanjing, Jiangsu 211198, China*
[3] *Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China*
[#] Authors contributed equally: Huipeng Song, Zeyuan Liang, Xinru Zhang
[*] Correspondence: myzheng@simm.ac.cn (Zheng M); xinguizhong1984@126.com (Xin G)

Abstract

Artificial intelligence (AI) is reshaping the research paradigm of traditional Chinese medicine (TCM) in a profound way. This review offers a systematic account of how AI algorithms propel the modernization of TCM through the integrated analysis of three core concepts: the chemicalome, referring to the collection of *in vitro* and *in vivo* chemical constituents derived from TCM; the targetome, defined as the set of biological macromolecules that engage in interactions with TCM components; and the bioactivome, signifying the range of integrated biological activities and phenotypic outcomes induced by TCM interventions. First, it illustrates how AI enables comprehensive characterization of complex *in vitro* and *in vivo* chemicalomes by revolutionizing mass spectrometry analysis and metabolite identification techniques. Next, it examines how AI works in conjunction with experimental technologies to systematically predict the targetome and validate these predictions. Furthermore, the review clarifies how AI deciphers the bioactivome arising from TCM interventions and uncovers mechanisms through the integration of multi-omics datasets. Finally, it explores methodologies for establishing comprehensive interconnections among the chemicalome, targetome, and bioactivome. This analytical framework demonstrates that AI functions not just as a tool to enhance research efficiency, but also as a foundational methodology capable of systematically decoding TCM and linking traditional wisdom with modern science.

Introduction

In recent years, the rapid progress of artificial intelligence (AI) algorithms has been fundamentally reshaping the mode of drug research and development. Traditional experimental methods are often time-intensive and resource-heavy, yet AI-based techniques can efficiently pinpoint potential patterns from vast volumes of chemical and biological datasets, thus greatly boosting the efficiency and success rate of drug development[1]. Within medicinal chemistry research, AI algorithms enable precise prediction of the absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of compounds. Such predictive power offers critical insights into lead compound optimization and helps effectively mitigate the risk of failures in the late stages of drug development[2]. For drug target identification, graph neural networks (GNNs) can accurately forecast drug-target interactions (DTIs) through direct learning of molecular topological architectures, outperforming traditional descriptor-driven methods in this aspect[3]. Notably, generative AI models including deep reinforcement learning frameworks have proven capable of designing novel molecular structures with tailored properties, which shifts AI's role from passive data prediction to active molecular creation[4]. Furthermore, AI excels at processing high-dimensional cell imaging data, mining actionable insights from scientific literature, integrating multi-omics datasets, and even offering novel perspectives for unraveling the complexities of biological systems[5]. Overall, these progressive technological advances have solidified AI's position as a core driver of paradigm transformation in drug research and development, and delivered a powerful impetus for translating basic scientific research into clinical practice.

Natural products, particularly the bioactive ingredients isolated from traditional Chinese medicine (TCM), have long been a vital source for developing innovative drugs, owing to their rich structural diversity and well-characterized biological activities[6]. However, the synergistic mechanisms of action of TCM involve a variety of active components and their corresponding molecular targets, which together form a highly intricate regulatory network. This characteristic makes the traditional single-target research strategy insufficient for the systematic investigation of TCM's pharmacological effects[7]. A core challenge in contemporary TCM research is to systematically decipher the complex interconnections between TCM's chemical components, molecular targets and its overall biological activities, and traditional research methods have an inherent inability to tackle such systemic problems effectively[8]. Notably, breakthroughs in metabolomics, proteomics, and transcriptomics have established a robust database for interpreting the mechanisms of TCM action from a systems biology perspective[9]. At the same time, newly developed experimental technologies such as chemical proteomics and thermal shift analysis are approaching technical maturity, which allows for the direct identification of drug targets throughout the entire proteome[10]. Presently, the combination of these state-of-the-art technologies with AI algorithms has led to the construction of a rigorous methodological system, thereby offering an unprecedented chance to unravel the systemic complexity underlying TCM's pharmacological effects.

This review aims to systematically elaborate on the AI algorithms driving the modern elucidation of TCM from 2012 to 2025, by analyzing the chemicalome, targetome, and bioactivome of TCM systems. The chemicalome is defined as the full set of chemical constituents derived from TCM, including prototype compounds

and their metabolites[11]. The targetome encompasses the complete range of biological macromolecules, including primarily receptors, enzymes, and ion channels, with which the TCM chemicalome interacts[12]. The bioactivome denotes the full spectrum of integrated biological activities and phenotypic effects arising from the systemic modulation of the targetome by the TCM chemicalome[13]. Based on the above-defined concepts of the chemicalome, targetome, and bioactivome, this review is structured as follows. First, we demonstrate how AI technology, by leveraging technological progress in mass spectrometry analysis and metabolite identification, enables the comprehensive characterization of the complex chemicalome of TCM. Next, we explore how various algorithms, such as deep learning and GNNs, facilitate the identification of the TCM targetome. In addition, we expound on how AI integrates multi-omics datasets, systematically interprets the bioactivome induced by TCM intervention, and uncovers the multi-pathway synergistic mechanisms of action of TCM. Finally, we explore the potential of AI algorithms in elucidating the holistic mechanisms of TCM, which is achieved by mapping the interactive relationships among the chemicalome, targetome, and bioactivome. This analytical framework highlights that AI acts not only as a tool to improve research efficiency, but also as a foundational method for systematically dissecting TCM systems.
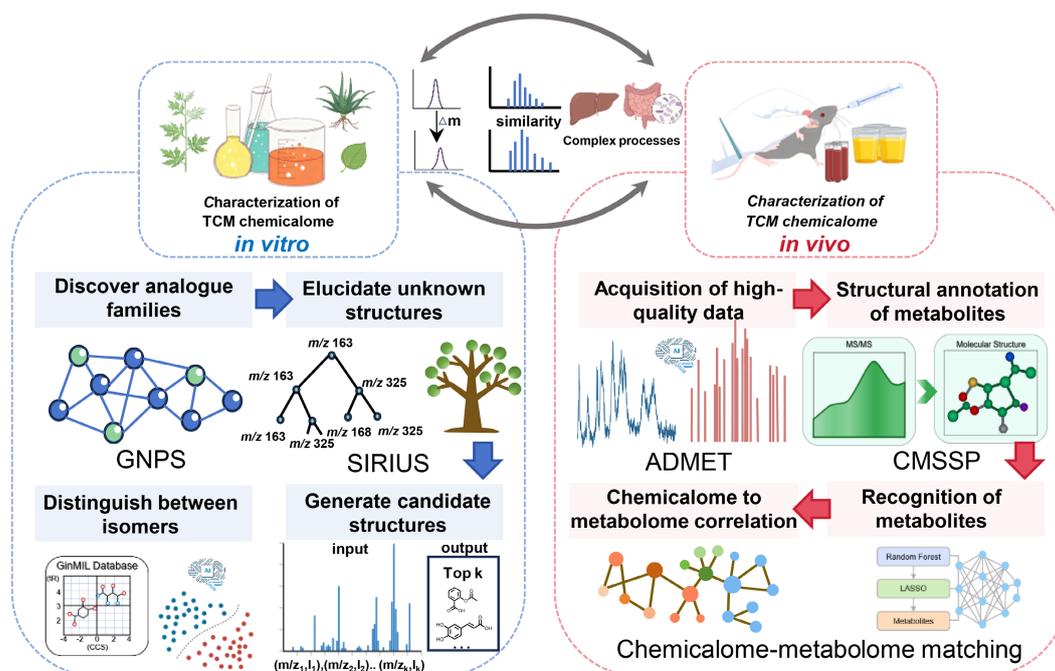
AI-driven revelation of the chemicalome in TCM

AI-driven analysis of the *in vitro* TCM chemicalome

Analyzing the *in vitro* TCM chemicalome acts as a foundational step for identifying its pharmacologically active components[14]. Traditional analytical chemistry, however, encounters considerable challenges in tackling the complex characteristics of herbal medicines. By contrast, AI algorithms possess distinct advantages: they can extract the intrinsic "fragment-structure" correlations from large-scale mass spectrometry datasets, enabling efficient resolution of these challenges[15]. This advancement has spurred technological evolution and transformed research from the traditional, experience-driven manual database querying to proactive prediction and generation.

In terms of applications within this field, AI technology has advanced steadily (Fig. 1, left panel). In the early phases, molecular network algorithms such as Global Natural Products Social Molecular Networking (GNPS) could compute and visualize the similarity of clustered mass spectrum fragments, revealing numerous structural homologies. This addressed the arduous task of manual substance tracing and notably accelerated the discovery of known compound families[16]. Later, integrated computing tools like SIRIUS, designed to identify unknown structures from tandem mass spectrometry, leveraged deep learning to conduct fragmentation tree analysis and structural searches, markedly enhancing the precision of de novo structural analysis for unknown compounds[17].

Along with the development of neural networks, the reverse design concept of "structure-based mass spectrometry prediction" has been adopted. It has become possible to predict mass spectrometry behavior with high accuracy, which has revolutionized the method of compound verification[18]. Recently, there have been some new breakthroughs at the forefront of technology. For example, tools like mass spectrometry-driven de novo molecular structure generation (MSNovelist) can now deduce possible molecular structures directly from a single MS/MS spectrum. This means that the role of AI has shifted from primarily relying on database comparison to actively generating structures. Importantly, this method no longer depends entirely on physical reference substances[19]. As a result, even in the absence of known compound samples, researchers now have a completely new pathway to identify novel molecules in natural products.

Fig. 1 Schematic diagram for AI-driven revelation of the TCM chemicalome. The diagram delineates comprehensive characterization strategies for the TCM chemicalome, categorized into *in vitro* (left panel) and *in vivo* (right panel) approaches. The *in vitro* workflow integrates GNPS for analog discovery via molecular networking, SIRIUS for structural elucidation using fragmentation tree analysis, and GinMIL for isomer differentiation. The *in vivo* workflow employs mass spectrometry for high-quality ADMET data acquisition, CMSSP for deep learning-based metabolite annotation, and a chemicalome-metabolome matching approach for the construction of metabolic networks.

In practical applications, AI has proven capable of addressing complex classification tasks. For instance, researchers integrated deep neural networks (DNNs) with pseudo-targeted metabolomics to differentiate between various types of easily confused ginseng varieties, achieving classification accuracy that outperformed traditional statistical methods by a significant margin. The key strength of DNNs lies in its capacity to automatically discover hidden patterns within large volumes of complex data. Beyond validating the model's high reliability, this study also precisely identified the core chemicalome responsible for the differences among these closely related ginseng types[20]. In another research effort, scientists combined machine learning with ion mobility separation technologies to develop a multi-dimensional information database (GinMIL), which encompasses the full range of ginsenosides. By effectively leveraging retention time and collision cross section data, this technological system has successfully overcome the major challenge faced by traditional mass spectrometry, that is its difficulty in distinguishing ginsenoside isomers[21].

AI-driven analysis of the *in vivo* TCM chemicalome

Generally, AI has become a core tool for systematically analyzing the complex *in vivo* chemicalome of TCM (Fig. 1, right panel). Its role has evolved from an auxiliary tool to a core driving force that can identify and discover new structures. It allows for a more thorough understanding of the full range of components actually contained in TCM. In fact, what really matters about TCM is how it affects our bodies after administration. Therefore, many researchers have focused on elucidating the *in vivo* chemicalome of TCM[22]. The *in vivo* TCM chemicalome is the collection of parent compounds and their metabolites that exist in the blood and tissues after administration, which constitutes the direct material basis of the pharmacological action. Compared with the *in vitro* analysis, the *in vivo* analysis of the TCM chemicalome is extremely difficult because of the low concentration, the unknown metabolic processes, and numerous interferences brought by the complex biological matrix[23].

The application of AI algorithms, online platforms, and related tools has greatly broken through the limitations of traditional analysis[24]. They have thus become the core support for efficient research on the *in vivo* TCM chemicalome. Moreover, innovations in AI algorithms have significantly improved the accuracy of predicting relevant properties of the *in vivo* TCM chemicalome. Specifically, among these algorithms, those based on multitask deep featurization use molecular graph convolution technology. They actively learn the core molecular features linked to ADMET properties. This kind of algorithm has raised ADMET prediction accuracy by a breakthrough margin. It has also effectively solved a key shortcoming of traditional methods: they cannot accurately predict novel chemical substances[25]. In addition, researchers have developed a combined classification model for blood-brain barrier permeability with AI technology. Tailored specifically for natural product research, this model has been successfully applied to predict the blood-brain barrier permeability of TCM components and their *in vivo* metabolites. Comprehensive validation through these experiments confirmed the model's reliability, with its prediction accuracy hitting 81%. Additionally, the model clearly pinpointed the key molecular factors that influence the blood-brain barrier permeability of TCM components[26]. At the same time, a variety of AI-assisted online platforms and tools provide convenient and practical support for analyzing the *in vivo* TCM chemicalome. SwissADME, in particular, is a free online tool designed for this research field. It can rapidly assess key indicators of small TCM molecules, such as pharmacokinetic properties and drug-likeness[27]. Another notable tool is

ADMETlab 2.0, an integrated online prediction platform that can comprehensively and accurately forecast over 70 ADMET-related detection indicators for TCM components and their metabolites[28].

Beyond the ADMET prediction approaches discussed above, combining mass spectrometry with AI algorithms also offers an effective means of analyzing the *in vivo* TCM chemicalome. Leveraging AI technologies, researchers have developed the Contrastive Mass Spectra-Structure Pretraining (CMSSP) model. This model is capable of building a unified representation space for mass spectra and molecular structures, which significantly boosts the accuracy of metabolite structural annotation. When applied to the analysis of *Glycyrrhiza glabra* metabolites, the CMSSP model achieved exceptional performance, with its identification accuracy surpassing that of existing methods by a considerable margin. As such, it has become an effective tool for the precise identification of low-abundance metabolites within the *in vivo* TCM chemicalome[29].

AI can also be integrated with matching techniques for *in vitro* and *in vivo* chemicalomes to support the completion of a range of key procedures, such as raw data processing, endogenous background elimination, and distinction between prototype components and their metabolites. Through these procedures, metabolic networks of complex systems can be effectively established. For example, following the administration of Mai-Luo-Ning injection to rats, the metabolites in the animals were successfully linked to their corresponding parent compounds using the chemicalome-metabolome matching approach[11]. For TCM formulas, AI-supported integrated strategies allow for systematic analysis of *in vitro* and *in vivo* chemicalomes. These strategies incorporate techniques including mass defect filtering and characteristic product ion filtering, which have been used to successfully identify and match 37 prototypes and 39 metabolites in Huanghou antidiarrhea dropping pills[30], as well as 21 prototypes and 70 metabolites in Chai-Gui decoction[31]. Additionally, these strategies have clarified the tissue distribution patterns of these prototypes and metabolites. Furthermore, AI serves as the foundation for a time-series-dependent global data filtering strategy, which can quickly screen out xenobiotic metabolites in dynamically complex matrices. In a co-culture system of *Ginkgo biloba* extract and gut microbiota, this strategy identified 107 flavonoid-derived metabolites and clearly outlined their temporal metabolic profiles[32]. Overall, AI algorithms fulfill their roles through diverse approaches: they optimize the analytical workflows of the *in vivo* TCM chemicalome, integrate multi-dimensional analytical data, and clarify dynamic metabolic characteristics, thereby providing efficient and viable technical support for the systematic analysis of the *in vivo* TCM chemicalome. The core technologies, key functionalities, advantages, and limitations of the above AI tools are summarized in Table 1.

AI-facilitated discovery of the targetome in TCM

Traditional drug discovery has predominantly followed a single-target paradigm, attributing the therapeutic effect of drugs to their high-affinity binding to a key biological target in order to regulate physiological activities[33]. However, biological systems are essentially constructed by complex networks. Many drugs with clinical value are not effective by acting on a single target, but by simultaneously regulating multiple targets. To systematically describe this multi-target mode of action, the concept of targetome has been introduced[12]. It refers to the entire set of biological targets that a molecule or a complex drug system acts on and regulates *in vivo*.

Table 1. Comparative analysis of AI tools for TCM *in vitro* and *in vivo* chemicalome characterization.

| Tool name	Core technologies	Key functionalities	Advantages	Limitations
GNPS	Molecular network algorithm, MS/MS fragment clustering, unsupervised learning	1. MS/MS fragment similarity visualization; 2. Structural homology mining; 3. Known compound family screening	1. Intuitive MS data visualization; 2. Enables rapid natural product tracing; 3. Accelerates discovery of known TCM compound families	1. Reliance on existing MS databases; 2. Low structural annotation accuracy for rare TCM components with unique fragments; 3. Clustering affected by low-quality MS data
SIRIUS	Deep learning, fragmentation tree analysis, tandem MS database matching, de novo structural analysis	1. Tandem MS-based unknown structure identification; 2. MS fragment decomposition; 3. Structural similarity search	1. Enables computational de novo prediction for unknown TCM components; 2. Integrated strategies enable higher annotation accuracy than single-algorithm tools; 3. Supports batch processing of large-scale TCM MS data	1. Reduced de novo accuracy for TCM components with multiple chiral centers; 2. Relies on high-quality tandem MS data; 3. Slow analysis for high-diversity TCM samples
MSNovelist	MS-driven de novo molecular structure generation, deep neural network, single MS/MS spectrum interpretation	1. Direct molecular structure deduction from single MS/MS spectrum; 2. Reference-free unknown structure generation	1. Breaks dependence on physical references; 2. Requires only a single MS/MS spectrum; 3. Improves novel TCM compound discovery efficiency	1. Diverse structural candidates require manual verification; 2. Low accuracy for TCM macromolecules; 3. Poor adaptability for rare TCM components
GinMIL	Machine learning, ion mobility separation, multi-dimensional database construction	1. Comprehensive ginsenoside database construction; 2. Ginsenoside isomer differentiation	1. Solves the traditional MS bottleneck in TCM isomer distinction (ginsenosides); 2. Multi-dimensional data fusion enables high component identification accuracy; 3. Specialized for saponin-rich TCM targeted analysis	1. Ginsenoside-focused; 2. Relies on ion mobility separation equipment; 3. Unable to analyze unknown saponin isomers
SwissADME	Machine learning, QSAR model, molecular descriptor calculation	1. Rapid ADME indicator assessment for small molecules	1. Fast prediction for high-throughput TCM active component screening; 2. Predicted indicators align with TCM small-molecule *in vivo* metabolic characteristics; 3. Visualizable results	1. Applicable only to small molecules; 2. Low prediction accuracy for rare TCM components with unique structures; 3. Lacks prediction of *in vivo* metabolic pathways for TCM components
ADMETlab 2.0	Integrated machine learning, multi-task feature learning, large-scale ADMET database	1. Comprehensive prediction of ADMET indicators; 2. Toxicity/metabolic stability prediction; 3. Drug-drug interaction prediction	1. Provides extensive ADMET indicators relevant to TCM research; 2. High accuracy; 3. Supports batch/customized prediction; 4. Predicts TCM compound prescription drug-drug interactions	1. Basic molecular descriptor knowledge required; 2. TCM metabolite ADMET prediction needs metabolic pathway tools; 3. Lacks specific guidance for lead optimization of TCM components
CMSSP	Contrastive learning model, molecular graph convolution, MS-structure unified representation space construction	1. Metabolite structural annotation; 2. Low-abundance metabolite identification in complex biological matrices; 3. *In vivo* MS data interpretation	1. The MS-structure unified representation space significantly enhances annotation accuracy; 2. Strong anti-interference ability for complex biological matrices; 3. Reduces dependence on *in vivo* metabolite databases	1. Requires large-scale natural product MS-structure paired training data; 2. Slow analysis for TCM *in vivo* samples with massive metabolites; 3. Sensitive to high-abundance endogenous biological matrix interferences
Chemicalome-Metabolome Matching Platform	Machine learning, metabolic network construction, chemicalome-metabolome correlation	1. *In vivo* MS raw data processing; 2. Endogenous background elimination; 3. Parent compound and metabolite distinction; 4. Screening of xenobiotic metabolites	1. Integrates *in vitro-in vivo* matching; 2. Time-series filtering outlines TCM component temporal metabolic characteristics; 3. Constructs complex TCM metabolic networks; 4. Adapts to dynamic biological matrix processing	1. High *in vitro* and *in vivo* MS data quality requirements; 2. Large-scale metabolic network construction needs high computing resources; 3. Weak automatic identification for rare TCM metabolic pathways

These targets mainly include enzymes, receptors, ion channels, and proteins involved in key biological processes such as signal transduction, metabolic regulation, transcriptional regulation, and epigenetic modification[34]. The schematic diagram for AI-facilitated discovery of the TCM targetome is shown in Fig. 2.

AI-facilitated algorithmic strategies for the TCM targetome

As a powerful tool for exploring the TCM targetome, AI-enabled algorithms support the systematic analysis of this complex system, which in turn helps clarify the multi-component and multi-target mechanisms of TCM through advancements in compound-target interaction prediction, affinity evaluation, and target deconvolution. GNNs and convolutional neural networks (CNNs) serve as the fundamental basis for these algorithmic strategies: GraphDTA employs molecular graphs to predict compound-target affinity[35],

whereas DeepDTA utilizes CNNs to achieve 1D sequence encoding[36]. To further improve the interpretability and generalization ability of these models, attention mechanisms and transformer architectures are integrated: DrugBAN combines bilinear attention with domain adaptation techniques to model local drug-target interactions[37]; AttentionMGT-DTA integrates drug molecular graphs and protein pocket characteristics[38]; TransformerCPI realizes compound-protein interaction prediction using only sequence information for targets without known 3D structures[39]. EviDTI adopts evidential deep learning to conduct uncertainty quantification, thus prioritizing compound-target pairs with high confidence[40].

A series of innovative algorithmic tools further extend the scope of TCM targetome exploration: CATNIP is capable of predicting the compatibility between enzymes and their substrates, which provides insights into the enzyme-mediated mechanisms of TCM[41]; DeepDTAGen adopts a multitask learning approach to
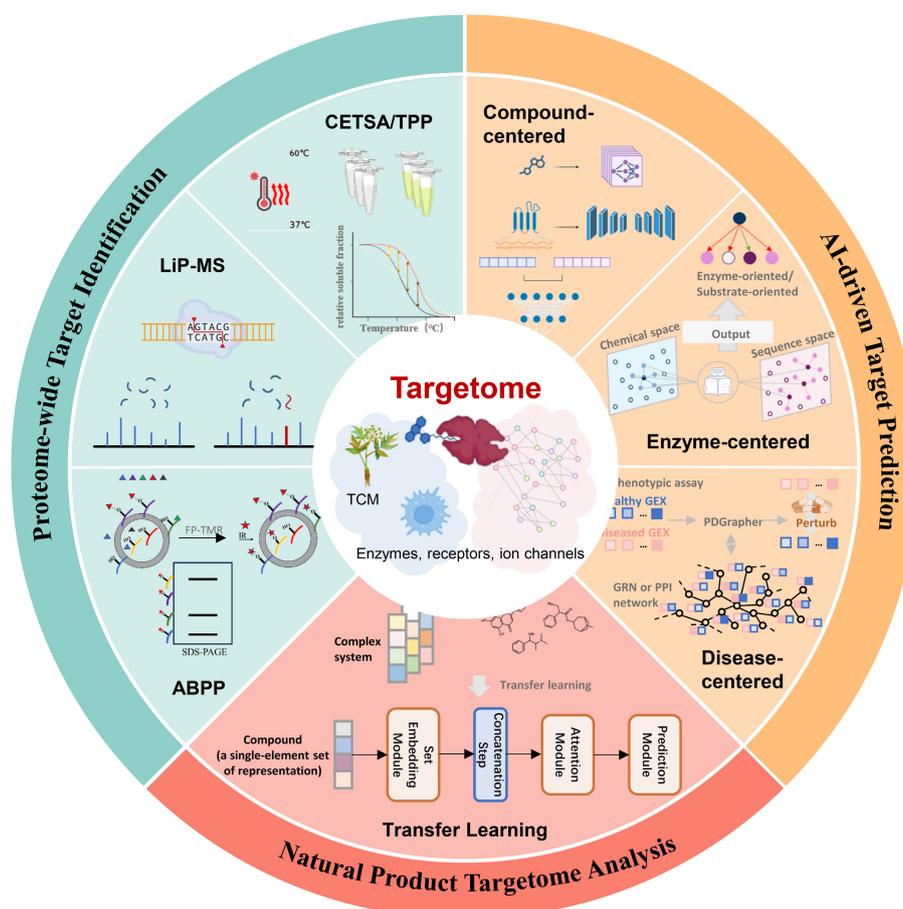
Fig. 2 Schematic diagram for AI-facilitated discovery of the TCM targetome. The diagram integrates three methodological components: AI-driven target prediction by compound-centered, enzyme-centered, and disease-centered methods; proteome-wide target identification by CETSA, TPP, LiP-MS, and ABPP; and natural product targetome analysis by transfer learning models.

simultaneously perform drug-target affinity prediction and target-aware molecule generation[42]; SurfDock applies diffusion models to achieve accurate prediction of protein-ligand complexes[43]. In terms of disease-associated targets, TRESOR incorporates omics data to screen and identify therapeutic targets[44]; PDGrapher is designed to predict combinatorial perturbagens that can reverse disease phenotypes[45]; NetFlow3D elucidates the impacts of mutation-driven target networks[46]. In summary, these AI-enabled algorithms show considerable potential in deciphering the complex targetome of TCM. By enabling precise target identification and detailed interaction profiling, they can provide strong support for interpreting the therapeutic mechanisms of TCM from a target-oriented perspective.

AI-integrated experimental technologies for TCM targetome validation

The gap between computational inference and real biological processes still needs to be bridged by AI predictions that rely on high-quality data and experimental verification[47]. Thermal stability detection methods, such as thermal proteome profiling (TPP) and cellular thermal shift assay (CETSA), have become the gold standard for high-throughput target verification. By monitoring ligand-induced changes in protein thermal stability in native cellular environments, these label-free methods can resolve complex proteoform-specific interactions and verify target engagement at the proteome level[48–50]. In order to attain a higher degree of structural resolution, analytical methods based on proteolysis and accessibility do offer complementary research perspectives. Techniques such

as limited proteolysis-mass spectrometry (LiP-MS) and LiP-Quant can detect subtle conformational changes induced by ligand binding, as well as changes in the accessibility of residues on the protein surface[51,52]. Additionally, based on activity-based protein profiling (ABPP) and emerging label-free platforms, covalent and high-affinity non-covalent interactions across the entire proteome are mapped via reactive probes or competitive labeling[53,54]. Unlike thermal analysis that focuses on detecting physical changes, these chemical proteomic techniques can directly produce chemical evidence of target occupancy.

The combination of experimental techniques and computational models leads to a "prediction-verification-optimization" closed loop. AI methods can screen and lock high-confidence candidate targets within a huge search space, and experimental methods can effectively eliminate false positives through empirical verification[55]. The high-quality interaction data thus obtained can be used as high-quality ground truth to retrain and optimize AI models, enabling more effective enhancement of model performance[56].

Representative applications of AI-facilitated TCM targetome discovery

Natural products have complex structures and the characteristics of multi-target binding, which indeed pose challenges to traditional target identification. In early computational processes, researchers mostly used structural similarity-based inference methods to predict potential targets[57]. In recent years, with the development of AI technology, target prediction has been shifting to a

phenotype-driven and task-specific research model. To address the insufficient adaptability of structural descriptors in new scaffold systems, phenotype-driven reverse inference models have begun to mine high-dimensional biological activity data, thereby establishing a correlation between the transcriptome features induced by compounds and protein interaction networks[58].

The integration of deep learning and transfer learning has made it feasible to carry out systematic modeling of multi-component TCM systems, and related frameworks have shown significant value in practical applications. The Set Embedding and Transfer learning model for Complex systems (SETComp), for example, incorporates permutation invariance and treats TCM formulas as unordered molecular sets, rather than as a single entity for isolated processing. This model can efficiently transfer knowledge from single-drug pharmacology to multi-drug combination systems by pre-training on large-scale single-drug transcriptomic data and fine-tuning on natural product datasets. In the rigorous evaluation of complex TCM systems, the model attained a high prediction accuracy, demonstrating AI's capability to translate the disordered characteristics of TCM systems into molecular hypotheses with practical significance[59]. This strategy marks a transformation in the research paradigm, shifting from predicting the correspondence between isolated compounds and targets to analyzing the overall and systematic regulatory network.

Beyond computational models, key research has illustrated the applicability of chemical proteomic approaches in elucidating the action mechanisms of natural products and facilitating TCM targetome discovery. By combining activity-based probes with quantitative mass spectrometry, researchers have systematically delineated the covalent target profiles of terpenoid compounds (e.g., eriocalyxin B), pinpointed their specific binding interactions in the ubiquitination pathway with high precision, and thus clarified the molecular underpinnings of their anti-proliferative activity[60]. As a valuable complement to covalent capture strategies, emerging label-free analytical platforms now allow for the concurrent characterization of non-covalent binding interactions and target occupancy statuses in living cells. For instance, this analytical method has been successfully applied to identify core RNA metabolism-associated targets of withaferin A in triple-negative breast cancer, which also validates the platform's capacity to resolve high-occupancy binding interactions in complex biological milieus[61].

Furthermore, the integration of fragment-based target research (FBTR) and phenotype-target coupled drug screening (PTDS) methodologies can facilitate a more in-depth understanding of the multi-target binding characteristics of natural products[62,63]. Another representative framework is Herb-CMap, which is specifically tailored for TCM research. It leverages multimodal data to combine gene expression features with protein-protein interaction networks, enabling the identification of active ingredients in complex TCM systems and their corresponding therapeutic targets, thus serving as an effective tool for clarifying the molecular mechanisms underlying TCM systems[64].

Numerous practical studies have validated the value of AI technologies in the targetome discovery of TCM formulas. For the treatment of metabolic-associated fatty liver disease (MAFLD) with Qigui Jiangzhi Formula, AlphaFold3 predicted the structures of key targets AMPK and SIRT1 with high confidence. Combined with molecular docking, these predictions clarified the formula's component binding and its regulatory mechanism through the AMPK/SIRT1-TFEB axis[65]. For XiaoErFuPi granules against functional dyspepsia, an unsupervised machine learning strategy classified its compounds into functional modules, clarifying its multi-target mechanism and identifying potential active components[66]. Additionally, AI-assisted target analysis revealed that Fuzheng Jiedu Decoction inhibits coronaviruses by blocking SARS-CoV-2 Spike protein-ACE2 binding and regulating inflammatory pathways, with caffeic acid and octyl gallate as core active components[67]. Overall, these integrated efforts demonstrate that AI technologies are effective in elucidating the TCM targetome.
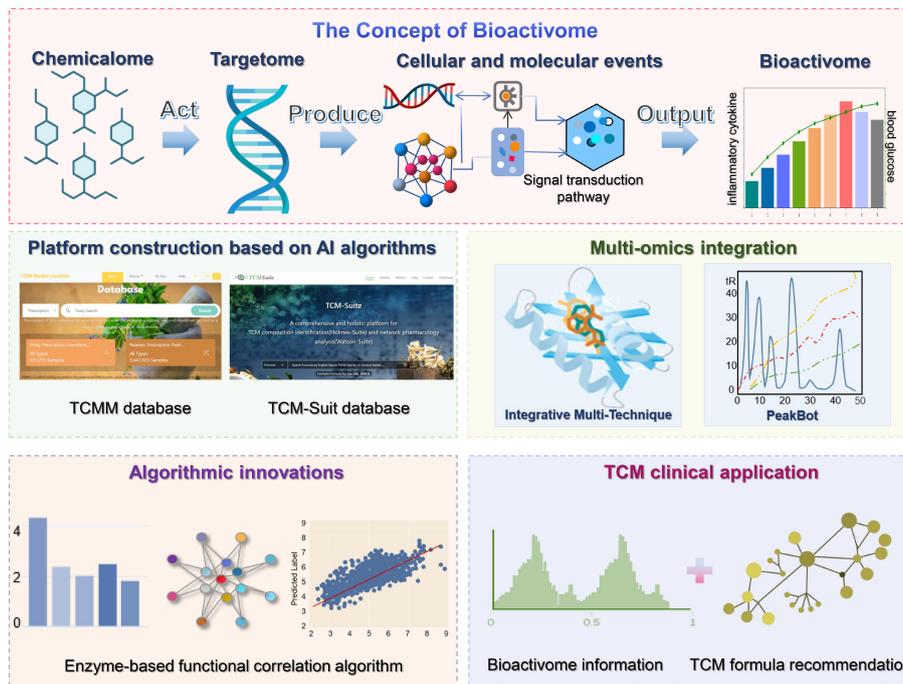
AI-accelerated revelation of the bioactivome in TCM

The exploration of the functional mechanisms of TCM is undergoing a profound transformation driven by AI. By systematically analyzing the causal links between the chemical components of TCM, their corresponding action targets, and the resulting biological activity responses, AI has established a novel methodological system for research into TCM modernization[68]. Within this newly developed system, the bioactivome plays a vital bridging role. It connects the upstream targetome and further converts the interactions between the chemicalome and targetome into dynamic functional changes at both the cellular and molecular levels[69]. When these microscopic alterations are comprehensively integrated, they eventually give rise to the observable pharmacological effects of TCM. As a result, the bioactivome has become the core focus for understanding TCM[70]. AI technologies combine multi-omics data with advanced computational models, aiming to systematically decode this complex network of functional responses. This decoding process enables us to understand the mechanisms underlying the synergistic action of multiple components in TCM[24].

Conceptual analysis of the bioactivome

To clarify this concept, we define the bioactivome in this review as the set of biological activities arising from the molecular and cellular events triggered by the action of the TCM chemicalome on the targetome (Fig. 3). These responses are typically multifaceted, combining to form a comprehensive profile of functional activity. This feature can be characterized by multidimensional dynamic indicators: those mainly at the molecular, pathway, and network levels[71]. More specifically, at the molecular level, we can observe quantitative variations in biomolecules, including gene expression levels, proteins and their modification states, as well as metabolite concentrations[72]. At the pathway level, the degree of activation or inhibition of signaling pathways can be assessed through aggregated analyses and special algorithms[73]. At the network level, the topology and dynamic changes of functional interaction networks such as gene coexpression networks and protein-protein interaction networks are described using parameters such as node centrality and module stability[74]. The integration of these multifaceted indicators forms the quantitative basis of the bioactivome. Therefore, a key focus of current research driven by AI is to elucidate the complex connections among various levels within the bioactivome as well as the relationship between the functional level and the final macroscopic therapeutic outcomes[75].

The targetome research, which is exploring the question of "which biological macromolecules the TCM components will bind to", depicts a static picture of potential interactions. By contrast, the bioactivome research focuses on a distinct question: what cellular and molecular functional changes do these interactions trigger? It uncovers dynamic functional responses within biological systems. For example, a single TCM component of andrographolide may bind to multiple target proteins, in turn inhibiting the activity of key kinases, a molecular-level event that falls within the scope of the

Fig. 3 Schematic diagram for AI-accelerated revelation of the TCM bioactivome. The diagram introduces the concept of the bioactivome, which refers to the set of biological activities derived from the molecular and cellular events induced by the action of the TCM chemicalome on the targetome. Research on the bioactivome mainly focuses on platform construction based on AI algorithms, multi-omics integration, algorithmic innovations, and TCM clinical application.

bioactivome. This inhibition can then block tumor cell proliferation pathways and activate senescence, which represent pathway- and network-level events of the bioactivome, ultimately resulting in specific pharmacological effects in organisms[76]. In this way, while the targetome defines the potential for TCM components to act, the bioactivome reveals the practical functionality of these actions. As such, it serves as a critical bridge connecting intermolecular interactions to the overall pharmacological phenomena observed in TCM treatments.

AI-driven methodologies for the bioactivome research in TCM

Through multi-level and multi-dimensional research methods, the decoding of the TCM bioactivome is being systematically carried out by AI technology[77]. A comprehensive research framework encompassing data infrastructure, algorithm innovation, and clinical application has been constructed. At the data infrastructure level, those knowledge platforms like the TCM modernization (TCMM) and linking traditional Chinese medicine with modern medicine platform (LTM-TCM) have actually laid a solid foundation for the bioactivome research. The TCMM database, which integrates 20 modern Chinese medicine concepts and 46 kinds of biological relationships, contains more than 3.4 million records[78]. Meanwhile, the LTM-TCM platform utilizes biomedical natural language processing technology and carefully organizes the multi-level interactions from symptoms, prescriptions, components to targets, thus also creating a huge network with millions of relations[79]. These platforms effectively connect the TCM theories and the modern medical insights, thus creating an indispensable data foundation for systematically decoding the multi-level functional outputs following TCM intervention.

In the area of algorithms and multi-omics integration, significant progress has been made with the enzyme-based functional correlation (EBFC) algorithm. This algorithm has advanced the traditional statistical correlation analysis to the functional association level.

In investigating BX (a drug derived from *Psoralea corylifolia* L.) for osteoporosis, researchers adopted the approach of function-oriented multi-omics analysis to integrate bone mineral density parameters, metabolic mechanism data and gut microbiome data. This method identified 59 distinct metabolites along with 9 key metabolic pathways, while the EBFC algorithm uncovered how major bacterial species regulate enzymes involved in purine and tryptophan metabolism. Through these findings, researchers have clarified and established a functional link between gut microbiota and host metabolites[80]. Similarly, in research focused on treating chronic kidney disease with the Jian-Pi-Yi-Shen formula, this algorithm has successfully unraveled the metabolite-microbiota-enzyme functional network, enabling TCM to reconstruct the interactions between host metabolism and gut microbiota[81]. It is important to note that the core technical challenge in building a reliable bioactivity graph lies in integrating heterogeneous multi-omics data, which is characterized by high noise levels, multiple origins, and batch effects. Resolving this challenge requires a rigorous methodological framework anchored in robust AI and computational strategies, including standardized data preprocessing, integrated noise-reduction modeling, rigorous validation, and biological interpretation. More importantly, it is essential to ensure that these interpretations are both biologically meaningful and reproducible[82,83].

In TCM clinical practice, AI algorithms have been extensively utilized to overcome the drawbacks of formula application relying on subjective experience, with their primary focus on intelligent prescription recommendation and prediction. FordNet, a system based on deep learning, incorporates bioactivome information for TCM formula recommendation and attains remarkable performance by leveraging large volumes of electronic health records for model training[84]. For personalized prescription recommendation, TCMPR adopts a subnetwork-oriented symptom term mapping approach, which can effectively characterize the features of unrecorded symptoms and thereby improve recommendation accuracy[85]. As for

TCMFP, it integrates TCM clinical experience, AI technologies and network science algorithms to screen optimal herbal formulas for specific diseases such as Alzheimer's disease, offering a novel approach to formula optimization[86]. Collectively, these AI-driven methods facilitate the precise treatment of TCM, and accelerate the transformation of TCM research models.

Mapping interactions among the chemicalome, targetome, and bioactivome

While previous discussions have focused on individual research on the chemicalome, targetome, and bioactivome, the next critical step is to explore the cross-links among these three omes. This integration is key to understanding systemic perturbations and demonstrating the efficacy of TCM. These three types of information allow researchers to describe the effects of TCM in organisms in a more systematic manner.

The hierarchical framework of the chemicalome, targetome, and bioactivome

The framework that integrates the chemicalome, targetome, and bioactivome can serve as the primary paradigm for characterizing the overall mechanism of TCM. AI stands as a key technology, permeating every stage of this multi-level analytical cascade. The chemicalome constitutes the fundamental material foundation of TCM. Instrumental analytical techniques such as mass spectrometry are combined with AI-powered structural prediction to systematically characterize both the *in vitro* and *in vivo* chemicalome of TCM[87,88]. This approach helps tackle the chemical complexity of TCM and identify the biologically relevant substances that engage with living systems. The intermediate layer of these interactions is referred to as the targetome. GNNs and Transformers are AI models that can combine molecular structures with protein sequences to predict high-confidence DTIs[89,90]. These approaches in conjunction with experimental validation methods, such as CETSA and chemical proteomics, are used to construct and validate the polypharmacological networks in which TCM components interact with enzymes, receptors, and signaling proteins[91]. Finally, the bioactivome corresponds to the functional output layer. AI integrates multi-omics data to elucidate the dynamic cellular and molecular responses triggered by targetome engagement. These responses can be changes in biomarkers, regulation of pathways and even creation of synergistic functional modules[92].

Research approaches for chemicalome-targetome-bioactivome interactions of TCM

Depending on different research goals, the starting points of AI analysis will vary, and relevant interactions can be roughly categorized into three research directions (Fig. 4). The chemicalome-oriented research starts from the chemical structure of natural products and investigates the effect brought about by the interactions between compounds and biological targets[93]. The targetome-oriented research focuses on targets like receptors and enzymes and seeks to understand how target perturbations spread across molecular networks and translate into changes in system-level functions[94]. AI has made this process much more efficient by
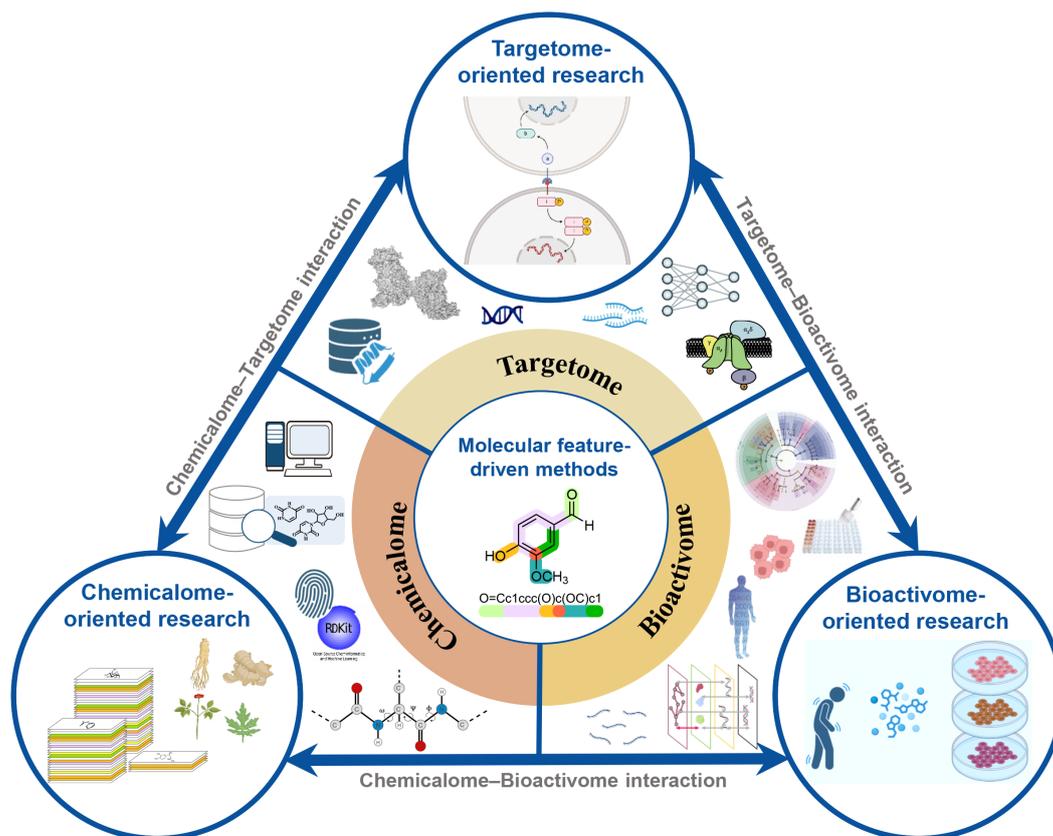
Fig. 4 Schematic diagram for interactions among the chemicalome, targetome, and bioactivome. The framework includes chemicalome-oriented, targetome-oriented, and bioactivome-oriented research to construct their interactions. Molecular feature-driven methods play an important role in this process.

providing an interpretable output from target perturbations to system-level output. The bioactivome-oriented research employs a reverse-reasoning strategy. It uses observed system-level phenotypes as clues to retrospectively trace and infer how TCM components collectively shape complex biological responses[95]. In order to achieve the integration of the aforementioned research directions, it is important to use the AI methods. By encoding multi-dimensional features of compounds, targets, and biological activities, researchers can discover new perturbation-response patterns, thereby realizing the integration of the information from the three scales at a computational level[96].

TCM formulas are inherently composed of multiple components, and their therapeutic effects are primarily exerted through mechanisms involving multiple targets and pathways[97]. This inherent complexity makes TCM formulas an ideal system for studying the entire chain of "chemical composition-target action-biological activity", as they perfectly align with our newly proposed concepts of the chemicalome, targetome, and bioactivome. Therefore, via multi-omics, metabolic perturbation mapping, and deep learning, we can better integrate TCM chemicalome, targetome, and bioactivome data into interaction networks. These data-driven approaches will not take the place of traditional pharmacology. Instead, they serve as a valuable complement to it. The two together drive progress in the field by speeding up hypothesis generation, identifying effective component combinations, and enabling the explanation of TCM mechanisms from a network perspective[98,99].

Limitations of the present AI models in integrating the chemicalome, targetome, and bioactivome

Although AI demonstrates great capabilities in processing large-scale data, numerous fundamental limitations that need to be addressed still remain when AI is applied to complex biological systems. Additionally, effective methods for dataset partitioning and model performance evaluation are currently lacking. Specifically, traditional model training relies on random splitting approaches, which can lead to structural similarities between molecules in the training and testing sets. The similarities prevent rigorous assessment of the model's capability to handle completely novel chemical structures[100]. Some researchers contend that UMAP (Uniform Manifold Approximation and Projection) clustering, which is based on chemical similarity, better mimics the structural diversity of molecules. However, the dimensionality reduction process inherent to UMAP may lead to the loss of critical structural features. Furthermore, UMAP is particularly effective at preserving local structures rather than global ones, meaning it may fail to accurately reflect overall dataset trends or inter-category relationships in certain scenarios[101,102]. If AI models are unable to achieve effective skeleton hopping, their capacity to explore the unknown chemical space of TCM will be limited.

Additionally, we must confront the long-standing criticisms surrounding AI applications in TCM research: the lack of model interpretability acts as a barrier to elucidating underlying mechanisms. While complex AI models deliver strong predictive capabilities, their decision-making logic remains unclear, as they generate outputs based solely on input data without transparent reasoning. In life sciences research, explaining the rationale behind why therapies work is just as important as predicting their efficacy. We can use AI to pinpoint potential pathways for mechanism validation, but we currently lack logical explanations for how or why these specific pathways are identified. This gap in reasoning creates a disconnect between computational predictions and experimental validation

processes, undermining trust in the reliability of AI-driven findings[103].

Conclusions and future prospects

This review clarifies how AI technology is incorporated into TCM research to systematically decode its action pathways within biological organisms. By integrating the chemicalome, targetome, and bioactivome, AI offers significant potential for uncovering the intricate mechanisms that underlie the synergistic effects of TCM, which are characterized by its multi-component, multi-target, and multi-pathway nature. We highlight AI's influence across various domains of TCM research: it transforms the way complex TCM constituents are characterized, supports the prediction of multiple targets, processes large-scale bioactivity data efficiently, and aids in the reverse engineering of TCM mechanisms. For thousands of years, TCM has depended mainly on empirical guidance for medication use, and the complexity of how its components act has long posed a stubborn challenge for modern scientific research. The incorporation of AI technology has further advanced contemporary TCM studies, signifying a major paradigm shift in the field.

Looking ahead, the integration of AI and TCM research points to several critical frontiers for future exploration. Research efforts moving forward should focus on developing explainable and causal AI models, which will allow the field to move beyond mere correlation analysis and uncover authentic therapeutic pathways underlying TCM's therapeutic effects. The next key challenge lies in establishing dynamic and multi-scale models to capture the interactions within and across the three core omics layers. This modeling task encompasses a broad range of biological processes, from the ADME processes of TCM components to the phenotypic changes observed at the whole-organism level. Additionally, generative AI offers tremendous potential for designing new and network-based therapeutic agents that draw on TCM's accumulated wisdom. Fulfilling this potential depends heavily on two key factors: building high-quality TCM data ecosystems and fostering in-depth interdisciplinary collaboration. Additionally, combining multi-omics profiles with the analysis of TCM's three core domains will pave the way for personalized TCM practice. This advancement will usher in an era of precision herbal medicine, where ancient TCM wisdom is systematically translated into modern therapeutic approaches.

Ethical statements

Not applicable.

Author contributions

The authors confirm contributions to the review as follows: conception and design: Xin G, Song H; draft manuscript preparation: Song H, Liang Z, Zhang X, Yang F; analysis and interpretation: Zheng M, Xin G. All authors reviewed and approved the final version of the manuscript.

Data availability

Data sharing is not applicable to this review as no datasets were generated or analyzed.

Acknowledgments

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

References

[1] Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, et al. 2019. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery* 18:463−477

[2] Yang H, Sun L, Wang Z, Li W, Liu G, et al. 2018. ADMETopt: a web server for ADMET optimization in drug design via scaffold hopping. *Journal of Chemical Information and Modeling* 58:2051−2056

[3] Torng W, Altman RB. 2019. Graph convolutional neural networks for predicting drug-target interactions. *Journal of Chemical Information and Modeling* 59:4131−4149

[4] Olivecrona M, Blaschke T, Engkvist O, Chen H. 2017. Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics* 9:48

[5] Cui H, Wang C, Maan H, Pang K, Luo F, et al. 2024. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods* 21:1470−1480

[6] Atanasov AG, Waltenberger B, Pferschy-Wenzig EM, Linder T, Wawrosch C, et al. 2015. Discovery and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnology Advances* 33:1582−1614

[7] Li S, Xiao W. 2025. General expert consensus on the application of network pharmacology in the research and development of new traditional Chinese medicine drugs. *Chinese Journal of Natural Medicines* 23:129−142

[8] Chen YH, Bi JH, Xie M, Zhang H, Shi ZQ, et al. 2021. Classification-based strategies to simplify complex traditional Chinese medicine (TCM) researches through liquid chromatography-mass spectrometry in the last decade (2011-2020): Theory, technical route and difficulty. *Journal of Chromatography A* 1651:462307

[9] Patti GJ, Yanes O, Siuzdak G. 2012. Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology* 13:263−269

[10] Wang Y, Zhao J, Xu Y, Tao C, Tong J, et al. 2023. Uncovering SOD3 and GPX4 as new targets of Benzo[α]pyrene-induced hepatotoxicity through Metabolomics and Chemical Proteomics. *Redox Biology* 67:102930

[11] Gong P, Cui N, Wu L, Liang Y, Hao K, et al. 2012. Chemicalome and metabolome matching approach to elucidating biological metabolic networks of complex mixtures. *Analytical Chemistry* 84:2995−3002

[12] Zheng X, Cai X, Hao H. 2022. Emerging targetome and signalome landscape of gut microbial metabolites. *Cell Metabolism* 34:35−58

[13] Zhao MY, Chen YH, Wang WY, Sun WP, Xiao HH, et al. 2022. A strategy to comprehensively analyze the bioactivity of complex herbal prescriptions via peak-by-peak cutting and knock-out chromatography: Qiliqiangxin capsule as an example. *Journal of Separation Science* 45:2446−2457

[14] Liu FT, Chen YH, Wang WY, Zhang JN, Sun RB, et al. 2025. A comprehensive analysis of volatile and nonvolatile components in berberis fortunei and its inhibition against HT29 colorectal cancer cells through GC-IMS, LC-QTOF-MS, and docking-based network analysis. *Phytochemical Analysis* 36:2349−2361

[15] Yang F, Liang Z, Zhao H, Zheng J, Liu L, et al. 2025. Mass spectral database-based methodologies for the annotation and discovery of natural products. *Chinese Journal of Natural Medicines* 23:410−420

[16] Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, et al. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* 34:828−837

[17] Yan Z, Ning J, Luo Z, Li D, Wang H, et al. 2025. Systematic analysis of the chemical components of gentiana urnula harry sm using SIRIUS and liquid chromatography high-resolution mass spectrometry. *Rapid Communications in Mass Spectrometry* 39:e10005

[18] Wei JN, Belanger D, Adams RP, Sculley D. 2019. Rapid prediction of electron-ionization mass spectrometry using neural networks. *ACS Central Science* 5:700−708

[19] Stravs MA, Dührkop K, Böcker S, Zamboni N. 2022. MSNovelist: de novo structure generation from mass spectra. *Nature Methods* 19:865−870

[20] Jiang M, Sha Y, Zou Y, Xu X, Ding M, et al. 2025. Integration of deep neural network modeling and LC-MS-based pseudo-targeted metabolomics to discriminate easily confused ginseng species. *Journal of Pharmaceutical Analysis* 15:101116

[21] Wang H, Zhang L, Li X, Sun M, Jiang M, et al. 2024. Machine learning prediction for constructing a universal multidimensional information library of Panax saponins (ginsenosides). *Food Chemistry* 439:138106

[22] Mi N, Cheng T, Li H, Yang P, Mu X, et al. 2019. Metabolite profiling of traditional Chinese medicine formula Dan Zhi Tablet: An integrated strategy based on UPLC-QTOF/MS combined with multivariate statistical analysis. *Journal of Pharmaceutical and Biomedical Analysis* 164:70−85

[23] Chen Z, Lin Z, Wan H, Li C, Jin W, et al. 2025. Rapid and accurate metabolite identification of traditional Chinese medicine based on UPLC-Q-TOF-MS coupled with UNIFI analysis platform and quantitative structure-retention relationship: Danshen-Honghua herbal pair as an example. *Journal of Pharmaceutical and Biomedical Analysis* 257:116696

[24] Li Y, Liu X, Zhou J, Li F, Wang Y, et al. 2025. Artificial intelligence in traditional Chinese medicine: advances in multi-metabolite multi-target interaction modeling. *Frontiers in Pharmacology* 16:1541509

[25] Feinberg EN, Joshi E, Pande VS, Cheng AC. 2020. Improvement in ADMET prediction with multitask deep featurization. *Journal of Medicinal Chemistry* 63:8835−8848

[26] Zhang X, Liu T, Fan X, Ai N. 2017. *In silico* modeling on ADME properties of natural products: Classification models for blood-brain barrier permeability, its application to traditional Chinese medicine and *in vitro* experimental validation. *Journal of Molecular Graphics and Modelling* 75:347−354

[27] Daina A, Michielin O, Zoete V. 2017. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports* 7:42717

[28] Xiong G, Wu Z, Yi J, Fu L, Yang Z, et al. 2021. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Research* 49:W5−W14

[29] Chen L, Xia B, Wang Y, Huang X, Gu Y, et al. 2024. CMSSP: a contrastive mass spectra-structure pretraining model for metabolite identification. *Analytical Chemistry* 96:16871−16881

[30] Li L, Wang X, Ma R, Hou M. 2024. An integrating strategy for systematic profiling of Chinese patent drug's chemicalome and associated metabolome: Huanghou antidiarrhea dropping pills as a case study. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* 1234:124029

[31] Zhu H, Chang W, Zhou C, Xu C, Su W, et al. 2021. Chemicalome and metabolome profiling of Chai-Gui Decoction using an integrated strategy based on UHPLC-Q-TOF-MS/MS analysis. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* 1185:122979

[32] Wang HY, Qu C, Li MN, Li CR, Liu RZ, et al. 2022. Time-series-dependent global data filtering strategy for mining and profiling of xenobiotic metabolites in a dynamic complex matrix: application to biotransformation of flavonoids in the extract of *Ginkgo biloba* by gut microbiota. *Journal of Agricultural and Food Chemistry* 70:14386−14394

[33] Song HP, Chen J, Hong JY, Hao H, Qi LW, et al. 2015. A strategy for screening of high-quality enzyme inhibitors from herbal medicines based on ultrafiltration LC-MS and in silico molecular docking. *Chemical Communications* 51:1494−1497

[34] Hao H, Kirchhoff F. 2025. Navigating drug discovery with targetome. *Targetome* 1:e001

[35] Nguyen T, Le H, Quinn TP, Nguyen T, Le TD, et al. 2021. GraphDTA: predicting drug-target binding affinity with graph neural networks. *Bioinformatics* 37:1140−1147

[36] Öztürk H, Özgür A, Ozkirimli E. 2018. DeepDTA: deep drug-target binding affinity prediction. *Bioinformatics* 34:i821−i829

[37] Bai P, Miljković F, John B, Lu H. 2023. Interpretable bilinear attention network with domain adaptation improves drug–target prediction. *Nature Machine Intelligence* 5:126−136

[38] Wu H, Liu J, Jiang T, Zou Q, Qi S, et al. 2024. AttentionMGT-DTA: A multi-modal drug-target affinity prediction using graph transformer and attention mechanism. *Neural Networks* 169:623−636

[39] Chen L, Tan X, Wang D, Zhong F, Liu X, et al. 2020. TransformerCPI: improving compound-protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics* 36:4406−4414

[40] Zhao Y, Xing Y, Zhang Y, Wang Y, Wan M, et al. 2025. Evidential deep learning-based drug-target interaction prediction. *Nature Communications* 16:6915

[41] Paton AE, Boiko DA, Perkins JC, Cemalovic NI, Reschützegger T, et al. 2025. Connecting chemical and protein sequence space to predict biocatalytic reactions. *Nature* 646:108−116

[42] Shah PM, Zhu H, Lu Z, Wang K, Tang J, et al. 2025. DeepDTAGen: a multitask deep learning framework for drug-target affinity prediction and target-aware drugs generation. *Nature Communications* 16:5021

[43] Cao D, Chen M, Zhang R, Wang Z, Huang M, et al. 2025. SurfDock is a surface-informed diffusion generative model for reliable and accurate protein-ligand complex prediction. *Nature Methods* 22:310−322

[44] Namba S, Iwata M, Nureki SI, Yuyama Otani N, Yamanishi Y. 2025. Therapeutic target prediction for orphan diseases integrating genome-wide and transcriptome-wide association studies. *Nature Communications* 16:3355

[45] Gonzalez G, Lin X, Herath I, Veselkov K, Bronstein M, et al. 2025. Combinatorial prediction of therapeutic perturbations using causally inspired neural networks. *Nature Biomedical Engineering* (*Epub ahead of print*)

[46] Zhang J, Leung AK, Kang JJ, Sun Y, Wu G, et al. 2025. A multiscale functional map of somatic mutations in cancer integrating protein structure and network topology. *Nature Communications* 16:975

[47] Li J, Dan W, Zhang C, Liu N, Wang Y, et al. 2024. Exploration of berberine against ulcerative colitis via TLR4/NF-κB/HIF-1α pathway by bioinformatics and experimental validation. *Drug Design, Development and Therapy* 18:2847−2868

[48] Kurzawa N, Leo IR, Stahl M, Kunold E, Becher I, et al. 2023. Deep thermal profiling for detection of functional proteoform groups. *Nature Chemical Biology* 19:962−971

[49] Leo IR, Kunold E, Audrey A, Tampere M, Eirich J, et al. 2025. Functional proteoform group deconvolution reveals a broader spectrum of ibrutinib off-targets. *Nature Communications* 16:1948

[50] Jafari R, Almqvist H, Axelsson H, Ignatushchenko M, Lundbäck T, et al. 2014. The cellular thermal shift assay for evaluating drug target interactions in cells. *Nature Protocols* 9:2100−2122

[51] Malinovska L, Cappelletti V, Kohler D, Piazza I, Tsai TH, et al. 2023. Proteome-wide structural changes measured with limited proteolysis-mass spectrometry: an advanced protocol for high-throughput applications. *Nature Protocols* 18:659−682

[52] Piazza I, Beaton N, Bruderer R, Knobloch T, Barbisan C, et al. 2020. A machine learning-based chemoproteomic approach to identify drug targets and binding sites in complex proteomes. *Nature Communications* 11:4200

[53] Luo P, Liu D, Zhang Q, Yang F, Wong YK, et al. 2022. Celastrol induces ferroptosis in activated HSCs to ameliorate hepatic fibrosis via targeting peroxiredoxins and HO-1. *Acta Pharmaceutica Sinica B* 12:2300−2314

[54] Biggs GS, Cawood EE, Vuorinen A, McCarthy WJ, Wilders H, et al. 2025. Robust proteome profiling of cysteine-reactive fragments using label-free chemoproteomics. *Nature Communications* 16:73

[55] Peng W, Yang Y, Lu H, Shi H, Jiang L, et al. 2024. Network pharmacology combines machine learning, molecular simulation dynamics and experimental validation to explore the mechanism of acetylbinankadsurin A in the treatment of liver fibrosis. *Journal of Ethnopharmacology* 323:117682

[56] Edwards AM, Owen DR. 2025. Protein-ligand data at scale to support machine learning. *Nature Reviews Chemistry* 9:634−645

[57] Mayr F, Möller G, Garscha U, Fischer J, Rodríguez Castaño P, et al. 2020. Finding new molecular targets of familiar natural products using in silico target prediction. *International Journal of Molecular Sciences* 21:7102

[58] Zhong F, Wu X, Yang R, Li X, Wang D, et al. 2022. Drug target inference by mining transcriptional data using a novel graph convolutional network framework. *Protein and Cell* 13:281−301

[59] Wang B, Pan B, Zhang T, Liu Q, Li S. 2025. Transfer learning and permutation-invariance improving predicting genome-wide, cell-specific and directional interventions effects of complex systems. *Advanced Science* 12:e09456

[60] Zhao WS, Chen KF, Liu M, Jia XL, Huang YQ, et al. 2023. Investigation of targets and anticancer mechanisms of covalently acting natural products by functional proteomics. *Acta Pharmacologica Sinica* 44:1701−1711

[61] Nie HJ, Fu YJ, Long S, Wang JY, Zhao WS, et al. 2025. Chemoproteomics reveals proteome-wide covalent and non-covalent targets of withaferin A. *Acta Pharmacologica Sinica* 46:1782−1793

[62] Wang C, Shi J, Rao Q, Shen B, Su C, et al. 2025. Obtain substance of anti-glioblastoma from Erigeron breviscapus through fragment-based target research (FBTR): An efficient strategy for pharmacology investigation and optimization of natural products. *Journal of Pharmaceutical Analysis* 15:101366

[63] Zhou W, Gao Y. 2025. Phenotype–target coupled drug screening: a high-efficiency framework for innovative drug discovery from CHMs. *Engineering* 57:10−13

[64] Wang Y, Sui Y, Yao J, Jiang H, Tian Q, et al. 2024. Herb-CMap: a multimodal fusion framework for deciphering the mechanisms of action in traditional Chinese medicine using Suhuang antitussive capsule as a case study. *Briefings in Bioinformatics* 25:bbae362

[65] Zhang L, Zheng Y, Shao M, Chen A, Liu M, et al. 2025. AlphaFold-based AI docking reveals AMPK/SIRT1-TFEB pathway modulation by traditional Chinese medicine in metabolic-associated fatty liver disease. *Pharmacological Research* 212:107617

[66] Guo F, Tang X, Zhang W, Wei J, Tang S, et al. 2020. Exploration of the mechanism of traditional Chinese medicine by AI approach using unsupervised machine learning for cellular functional similarity of compounds in heterogeneous networks, XiaoErFuPi granules as an example. *Pharmacological Research* 160:105077

[67] Liu K, Hong B, He ST, Du S, Ke J, et al. 2025. The potential mechanisms and material basis of Fuzheng Jiedu decoction broad-spectrum inhibiting coronaviruses. *Virologica Sinica* 40:125−136

[68] Yan D, Zheng Q, Chang K, Hua R, Liu Y, et al. 2025. Artificial intelligence in traditional Chinese medicine: from systems biological mechanism discovery, real-world clinical evidence inference to personalized clinical decision support. *Chinese Journal of Natural Medicines* 23:1310−1328

[69] Zhang JN, Pei ZD, Wang WY, Zhao MY, Pei WH, et al. 2024. Integration of high-resolution LC-Q-TOF mass spectrometry and multidimensional chemical-biological analysis to detect nanomolar-level acetylcholinesterase inhibitors from different Parts of *Zanthoxylum nitidum*. *Journal of Agricultural and Food Chemistry* 72:17328−17342

[70] Wang S, Fu JL, Hao HF, Jiao YN, Li PP, et al. 2021. Metabolic reprogramming by traditional Chinese medicine and its role in effective cancer therapy. *Pharmacological Research* 170:105728

[71] Zhang R, Zhu X, Bai H, Ning K. 2019. Network pharmacology databases for traditional Chinese medicine: review and assessment. *Frontiers in Pharmacology* 10:123

[72] Vandereyken K, Sifrim A, Thienpont B, Voet T. 2023. Methods and applications for single-cell and spatial multi-omics. *Nature Reviews Genetics* 24:494−515

[73] Liu J, Chen Y, Cao Z, Guan B, Peng J, et al. 2020. Babao Dan inhibits the migration and invasion of gastric cancer cells by suppressing epithelial-mesenchymal transition through the TGF-β/Smad pathway. *Journal of International Medical Research* 48:1−13

[74] Wang S, Zhao Y, Hu X. 2023. Exploring the mechanism of Suanzaoren decoction in treatment of insomnia based on network pharmacology and molecular docking. *Frontiers in Pharmacology* 14:1145532

[75] Song HP, Wu SQ, Hao H, Chen J, Lu J, et al. 2016. A chemical family-based strategy for uncovering hidden bioactive molecules and multi-component interactions in herbal medicines. *Scientific Reports* 6:23840

[76] Zhang J, Li C, Zhang L, Heng Y, Wang S, et al. 2022. Andrographolide, a diterpene lactone from the Traditional Chinese Medicine Andrographis paniculate, induces senescence in human lung adenocarcinoma via p53/p21 and Skp2/p27. *Phytomedicine* 98:153933

[77] Yang L, Wang H, Zhu Z, Yang Y, Xiong Y, et al. 2025. Network pharmacology-driven sustainability: AI and multi-omics synergy for drug discovery in traditional Chinese medicine. *Pharmaceuticals* 18:1074

[78] Ren Z, Ren Y, Li Z, Xu H. 2024. TCMM: A unified database for traditional Chinese medicine modernization and therapeutic innovations. *Computational and Structural Biotechnology Journal* 23:1619−1630

[79] Li X, Ren J, Zhang W, Zhang Z, Yu J, et al. 2022. LTM-TCM: A comprehensive database for the linking of Traditional Chinese Medicine with modern medicine at molecular and phenotypic levels. *Pharmacological Research* 178:106185

[80] Wei Q, Zhou Y, Hu Z, Shi Y, Ning Q, et al. 2024. Function-oriented mechanism discovery of coumarins from *Psoralea corylifolia* L. in the treatment of ovariectomy-induced osteoporosis based on multi-omics analysis. *Journal of Ethnopharmacology* 329:118130

[81] Yin YH, Li CH, Huang HP, Zhang C, Zhang SB, et al. 2024. A novel enzyme-based functional correlation algorithm for multi-omics reveals the potential mechanisms of traditional Chinese medicines: taking Jian-Pi-Yi-Shen formula as an example. *Journal of Pharmaceutical and Biomedical Analysis* 241:115973

[82] Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, et al. 2020. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology* 21:111

[83] Patel-Murray NL, Adam M, Huynh N, Wassie BT, Milani P, et al. 2020. A multi-omics interpretable machine learning model reveals modes of action of small molecules. *Scientific Reports* 10:954

[84] Zhou W, Yang K, Zeng J, Lai X, Wang X, et al. 2021. FordNet: Recommending traditional Chinese medicine formula via deep neural network integrating phenotype and molecule. *Pharmacological Research* 173:105752

[85] Dong X, Zheng Y, Shu Z, Chang K, Xia J, et al. 2022. TCMPR: TCM prescription recommendation based on subnetwork term mapping and deep learning. *BioMed Research International* 2022:4845726

[86] Niu Q, Li H, Tong L, Liu S, Zong W, et al. 2023. TCMFP: a novel herbal formula prediction method based on network target's score integrated with semi-supervised learning genetic algorithms. *Briefings in Bioinformatics* 24:bbad102

[87] Xue Z, Shang Y, Yang L, Li T, Yang B. 2025. A tailored database combining reference compound-derived metabolite, metabolism platform and chemical characteristic of Chinese herb followed by activity screening: application to magnoliae officinalis cortex. *Journal of Pharmaceutical Analysis* 15:101066

[88] Zhang D, Zhang J, Chen S, Zhang H, Yang Y, et al. 2024. An integrated approach for studying exposure, metabolism, and disposition of traditional Chinese medicine using PATBS and MDRB tools: a case study of semen *Armeniacae amarum*. *Chinese Medicine* 19:158

[89] Zeng J, Jia X. 2025. Quantifying compatibility mechanisms in traditional Chinese medicine with interpretable graph neural networks. *Journal of Pharmaceutical Analysis* 15:101342

[90] Qiao Y, Hu L, Zhang J, Hu P, Luo X. 2025. Identifying novel therapeutic targets of natural compounds in traditional Chinese medicine herbs with hypergraph representation learning. *Briefings in Bioinformatics* 26:bbaf399

[91] Wei W, Zeng Q, Wang Y, Guo X, Fan T, et al. 2023. Discovery and identification of EIF2AK2 as a direct key target of berberine for anti-inflammatory effects. *Acta Pharmaceutica Sinica B* 13:2138−2151

[92] Khanal S, Kumar A, Kumar P, Thakur P, Chander AM, et al. 2025. Unraveling bioactive potential and production in Ganoderma lucidum through omics and machine learning modeling. *Chinese Herbal Medicines* 17:414−427

[93] Liu KX, Li N, Yin YH, Zhong ZJ, Li P, et al. 2022. An in-house database-driven untargeted identification strategy for deep profiling of chemicalome in Chinese medicinal formula. *Journal of Chromatography A* 1666:462862

[94] Tian Y, Wan N, Zhang H, Shao C, Ding M, et al. 2023. Chemoproteomic mapping of the glycolytic targetome in cancer cells. *Nature Chemical Biology* 19:1480−1491

[95] Zhou M, Wu B. 2025. Chinese medicine in regulating immune balance for the treatment of autoimmune diseases. *American Journal of Chinese Medicine* 53:1615−1640

[96] Liu M, Wang Y, Deng W, Xie J, He Y, et al. 2024. Combining network pharmacology, machine learning, molecular docking and molecular dynamic to explore the mechanism of Chufeng Qingpi decoction in treating schistosomiasis. *Frontiers in Cellular and Infection Microbiology* 14:1453529

[97] Chen J, Wei X, Zhang Q, Wu Y, Xia G, et al. 2023. The traditional Chinese medicines treat chronic heart failure and their main bioactive constituents and mechanisms. *Acta Pharmaceutica Sinica B* 13:1919−1955

[98] Shang L, Wang Y, Li J, Zhou F, Xiao K, et al. 2023. Mechanism of Sijunzi Decoction in the treatment of colorectal cancer based on network pharmacology and experimental validation. *Journal of Ethnopharmacology* 302:115876

[99] Ma S, Liu J, Li W, Liu Y, Hui X, et al. 2023. Machine learning in TCM with natural products and molecules: current status and future perspectives. *Chinese Medicine* 18:43

[100] Gangwal A, Ansari A, Ahmad I, Azad AK, Kumarasamy V, et al. 2024. Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities. *Frontiers in Pharmacology* 15:1331062

[101] Guo Q, Hernandez-Hernandez S, Ballester PJ. 2025. UMAP-based clustering split for rigorous evaluation of AI models for virtual screening on cancer cell lines. *Journal of Cheminformatics* 17:94

[102] Vigil-Vásquez C, Schüller A. 2022. De novo prediction of drug targets and candidates by chemical similarity-guided network-based inference. *International Journal of Molecular Sciences* 23:9666

[103] Qadri YA, Shaikh S, Ahmad K, Choi I, Kim SW, et al. 2025. Explainable artificial intelligence: a perspective on drug discovery. *Pharmaceutics* 17:1119