ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202508.00231

Central Intelligent Drug: Design of a Multi-Agent-Based End-to-End System for Drug Design and Delivery

Authors: Tian Zihan, Tian Zihan

Date: 2025-08-15T03:11:56+00:00

Abstract

[Objective] To address the "technical silo" problem in the drug discovery pipeline, where target screening, molecular design, efficacy evaluation, and clinical translation rely on isolated models and lack a lightweight end-to-end integrated framework. [Methods] We constructed a multi-agent system based on the AutoGen+Ollama framework, integrating six agents: input classification, disease-target mapping, gene translation, ligand generation, ADMET screening, and nanocarrier design, forming a closed-loop pipeline. [Results] Validated on the Alzheimer's disease BACE1 target: 404 ligands were generated, and after six-dimensional ADMET screening, 6 drug candidates were obtained, with customized blood-brain barrier delivery solutions output; the system can be deployed and run on consumer-grade GPUs. [Limitations] The drug delivery design phase primarily relies on large model outputs and empirical rules, lacking physical mechanism constraints such as molecular dynamics simulations; workflow flexibility needs further improvement. [Conclusion] This study pioneers the multi-agent system "中枢智药" that integrates the full process of drug design and delivery, achieving an end-to-end automated closed loop from target identification to delivery solution design, providing a new paradigm for AI-driven efficient drug discovery.

Full Text

CentraPharma: An End-to-End Multi-Agent System for Drug Design and Delivery

Tian Zihan
[1]School of Computer Science and Technology, Xinjiang University, Xinjiang 830017, China

Abstract:
[**Objective**] To address the "technology silo" problem in drug research and development, where isolated models across target screening, molecular design, efficacy evaluation, and clinical translation lack a lightweight, end-to-end integrated framework. [**Methods**] We constructed a multi-agent system based on the AutoGen+Ollama framework, integrating six specialized agents: input classification, disease-target mapping, gene translation, ligand generation, ADMET screening, and nanocarrier design, forming a closed-loop pipeline. [**Results**] Validation using the Alzheimer's BACE1 target generated 404 ligands, yielding 6 druggable molecules after six-dimensional ADMET screening, along with customized blood-brain barrier delivery solutions; the system operates on consumer-grade GPUs. [**Limitations**] The drug delivery design primarily relies on large model outputs and empirical rules, lacking constraints from physical mechanisms such as molecular dynamics simulations; workflow flexibility requires further enhancement. [**Conclusions**] This study introduces "CentraPharma," the first multi-agent system integrating the complete drug design and delivery workflow, achieving end-to-end automated closed-loop from target identification to delivery design, providing a new paradigm for AI-driven efficient drug research and development.

Keywords: Multi-Agent System; Drug Design; Drug Delivery; AI Drug Discovery

---

The rapid advancement of artificial intelligence technology is reshaping the paradigm of drug research and development. From targeted molecular design to multi-modal data integration, AI-driven innovative tools have significantly enhanced the efficiency and precision of drug discovery. Transformer-based DrugGPT achieved 99.9% validity in molecular generation, while reinforcement learning-optimized DrugGen reduced binding energy by 24% [**?**]; lightweight deployment platforms like Ollama enable medical edge computing on consumer-grade GPUs through dynamic quantization; and large models such as BioGPT and BioMedGPT provide new perspectives for target mechanism analysis and drug-target relationship prediction through cross-modal knowledge fusion [**?**, **?**]. However, despite remarkable breakthroughs in individual stages, end-to-end drug R&D still faces systemic challenges—target screening, molecular design, efficacy evaluation, and clinical translation rely on isolated models or toolchains, lacking a lightweight multi-agent collaboration framework for integrated workflow. This "technology silo" phenomenon restricts data flow, creates inefficient task handoffs, and prevents the formation of a closed-loop innovation chain from basic research to clinical application. How to construct a scalable, highly secure multi-agent system that coordinates the entire drug R&D workflow has become a critical bottleneck urgently requiring breakthrough in AI-driven pharmaceutical research.

2 Related Work

The deep integration of artificial intelligence technology in drug research and development has significantly accelerated the transformation of innovation paradigms. Autoregressive generation models such as DrugGPT, representing the forefront of molecular design, transform protein sequences and ligand SMILES into finite vocabularies through Byte Pair Encoding (BPE) algorithms, learning protein-ligand binding data via Transformer architecture to generate high-affinity molecules for specific targets [**?**]. Experimental results demonstrate 99.9% validity of generated molecules, with support for ligand prompt input and atom count control, providing efficient tools for drug repositioning and novel pharmacophore design. Furthermore, the Iranian team-developed DrugGen enhances the DrugGPT architecture through reinforcement learning (PPO) and integrates a binding affinity prediction module, improving generated molecule binding energy from 7.22 to 5.81 and significantly enhancing clinical translation potential [**?**]. Such breakthroughs mark that AI's precision in molecular design has approached experimental validation levels.

At the technical deployment level, the lightweight and local deployment demands of large models have driven computational paradigm innovation. The Ollama platform reduces hardware barriers through quantization optimization—for instance, the Qwen3-8B dynamic quantization 2.0 version achieves 45 tokens/second inference speed on 12GB VRAM devices (e.g., RTX 4070s 12G) while supporting 128K long context and multilingual generation, providing feasible solutions for medical edge computing scenarios. Its simplified configuration strategies (e.g., Temperature=0.6 to avoid circular generation) further enhance deployment efficiency, enabling real-time inference in resource-constrained environments. This progress provides flexible computational infrastructure support for AI-driven drug discovery.

Biomedical large model research has also achieved significant breakthroughs. Microsoft's BioGPT, pre-trained on large-scale biological literature, supports drug-target interaction extraction, document classification, and question-answering tasks, with its Hugging Face version achieving 98.6% accuracy in drug interaction prediction [**?**]. Meanwhile, multi-modal models like BioMedGPT integrate text, sequence, and structural data, promoting cross-innovation in target mechanism research and drug design [**?**]; OpenAI's GPT-5 provides assisted diagnosis and treatment recommendations through certified medical knowledge bases (20 million documents) and has obtained HIPAA compliance certification, demonstrating AI's compliance and practicality in clinical scenarios [**?**].

At the system architecture level, multi-agent collaboration frameworks (e.g., AutoGen) can construct closed-loop drug R&D chains of "target screening-molecular generation-efficacy evaluation" through task decomposition and cyclic feedback mechanisms. Their core advantages include: first, flexible networking capability, where agents collaborate through natural language instructions without hard-coded logic; second, tool integration capability, compatible with

Python toolchains (e.g., RDKit, PLAPT) and supporting pharmaceutical tool calls; third, local security, ensuring data remains on-premises to meet medical privacy requirements. Such frameworks provide standardized solutions for collaborative division of labor in complex drug R&D tasks [**?**].

Specialized agents for complex biomedical tasks (e.g., Biomni) further expand AI application boundaries. Biomni integrates retrieval-augmented planning with code execution capabilities, supporting CRISPR experimental design, single-cell sequencing analysis (scRNA-seq), and ADMET prediction [**?**]. Its modular tool library integrates FDA drug databases and clinical guideline parsers, connecting external tool servers through the Model Context Protocol (MCP) to achieve functional extension. Notably, this agent has demonstrated autonomous research capabilities, such as generating verifiable hypotheses from natural language instructions (e.g., "predict 32 genes regulating T-cell exhaustion"), providing new approaches for hypothesis-driven drug discovery.

Despite significant progress in specific stages, current research still faces critical challenges: existing systems focus on single-task optimization, lacking lightweight multi-agent collaboration frameworks for end-to-end targeted drug R&D. Specifically, target screening, molecular generation, efficacy evaluation, and clinical trials still depend on independent models or toolchains, making it difficult to form closed-loop systems. Moreover, agent deployment efficiency in resource-constrained scenarios and multi-modal data collaboration capabilities remain insufficiently validated. Therefore, constructing a lightweight, highly secure multi-agent architecture to integrate the entire drug R&D workflow remains a key breakthrough for advancing AI pharmaceuticals from laboratory to industrialization.

3.1 Therapeutic Target Database (TTD)

The Therapeutic Target Database (TTD) is a global, freely accessible online database developed and maintained by the National University of Singapore, with the core objective of systematically collecting and providing target information related to drug research and development [**?**, **?**]. The database currently contains over 3,500 known and exploratory therapeutic protein and nucleic acid targets, along with detailed information on nearly 40,000 associated drug molecules. TTD not only provides basic target information but also integrates corresponding targeted diseases, involved pathways, and relevant drugs (including approved and in-development) for each target, clearly demonstrating the multi-layered association network of "drug-target-disease." To more precisely guide research and evaluate target value, TTD proposes an efficacy-based target identification strategy that specifically distinguishes between "efficacious" and "non-efficacious" targets, further categorizing "efficacious" targets with therapeutic potential into four tiers: successful targets (corresponding to approved drugs), clinical trial targets (corresponding to drugs in clinical trials), patent-recorded targets (documented in patent literature), and literature-reported targets (proposed in scientific publications). TTD emphasizes data

traceability and reliability, with all included information supported by clear, verifiable references. Users can conveniently retrieve specific disease-related potential therapeutic drugs and their targets through multiple search methods including disease name, drug name, or target name. Additionally, TTD provides rich cross-links to internationally renowned bioinformatics databases such as UniProtKB, PDB, KEGG, OMIM, and BRENDA for in-depth target background knowledge. Since its launch in 2002, the database has undergone continuous content updates and functional optimization, with the latest data updated to January 10, 2024.

3.2 UniProt Knowledgebase

UniProt (Universal Protein) is the most comprehensive and extensive protein database internationally. It was constructed by integrating data from three previously independent databases—Swiss-Prot, TrEMBL, and PIR-PSD. Its core data primarily originates from subsequent protein sequence analysis of completed genome sequencing projects and aggregates extensive protein biological function information reported in literature. The database's core is the UniProt Knowledgebase (UniProtKB) [**?**]. UniProtKB itself consists of two main components: UniProtKB/Swiss-Prot is a high-quality, non-redundant dataset with strict manual curation and annotation by experts, with entries primarily based on published research and quality-controlled through computational analysis (e.g., E-value verification) to ensure reliability (e.g., over 560,000 entries in the 2023 release); UniProtKB/TrEMBL mainly contains protein sequences annotated through automated pipelines, designed to effectively address the processing pressure from massive data generated by genome projects. It automatically collects and annotates coding sequence translations from three major nucleic acid databases (EMBL-Bank/GenBank/DDBJ) and gene prediction sequences from protein structure databases (PDB), Ensembl, RefSeq, and CCDS (with a huge number of entries, e.g., over 220 million in 2023). Additionally, UniProt includes UniProt Archive (UniParc), a comprehensive non-redundant protein sequence archive aiming to collect sequences from all major public protein databases. UniParc addresses redundancy issues between different databases and within different versions of the same database by assigning a stable unique identifier (UPI) to each unique protein sequence. Notably, UniParc only stores sequence information itself without any annotation content. With its comprehensive sequence information, rich functional annotations (especially in the Swiss-Prot section), and ability to resolve sequence redundancy, the UniProt database has become an indispensable foundational resource for protein-related research.

3.3 MyGene

MyGene.info is a RESTful API web interface service for gene annotation data, funded by the National Institute of General Medical Sciences (NIGMS) under the National Institutes of Health (NIH) [**?**]. Its core objective is to provide

users with a simple and easy-to-use method to query and retrieve integrated gene annotation information from multiple authoritative sources. MyGene.info synchronizes and updates data weekly from over 20 bioinformatics databases (including NCBI Entrez Gene, Ensembl, UniProt, and UCSC), aiming to provide a relatively comprehensive and up-to-date gene annotation view. Although some of its integrated original data sources may have specific usage restrictions, the MyGene.info service itself is free, with its source code hosted on GitHub. The service provides two core functions: gene query service and annotation retrieval service, both returning structured results in JSON format. The MyGene.info API continues to be updated and optimized—for example, its v3 version (as of the description) includes improvements and bug fixes in data representation (e.g., including RefSeq accession version information), relationship mapping (e.g., enhanced RNA-protein associations, improved Ensembl to Entrez Gene ID mapping), and data structure. To facilitate users in different programming environments, the MyGene.info community has developed and open-sourced official client libraries, including the MyGene R Client and MyGene Python Client (mygene module), enabling users to conveniently access MyGene.info web services in their R or Python analysis pipelines. For example, in Python environments, using the mygene module can easily implement batch conversion between gene identifiers (e.g., Entrez ID, Ensembl ID, Symbol) or query detailed annotation information for specific genes, significantly simplifying gene data integration steps.

This study constructs a multi-agent collaborative system based on the AutoGen+Ollama framework, using the Qwen3:8B large model as the core reasoning engine for agents and integrating professional toolchains to achieve automation in biomedical R&D. The system implements end-to-end automated workflow from target identification to molecular screening through the collaborative operation of six specialized agents: the input classification agent first parses user request types (disease/gene/protein/nanocarrier); the disease-target mapping agent establishes disease-target-protein sequence mapping relationships by integrating the TTD database and UniProt services; the gene translation agent performs bioinformatics conversion from gene identifiers to protein sequences using the MyGene API; the ligand generation agent drives the DrugGPT model for small molecule ligand generation; the ADMET screening agent integrates QED models and five-dimensional pharmacokinetic prediction modules for joint molecular property screening; and the nanocarrier design agent focuses on multi-parameter optimization of blood-brain barrier-penetrating carriers. Agents exchange data through strictly defined JSON protocols, manage output files using hierarchical storage structures, and form a closed-loop automated pipeline for biomedical R&D. The overall workflow architecture is shown in Figure 1 [Figure 1: see original paper].

4.1 Data Processing Agents

Data processing agents constitute the foundational support layer of the system, comprising three core units: input classification, disease-target mapping, and gene translation. These agents collectively accomplish standardized conversion of biomedical data, transforming unstructured biological entity information into machine-processable protein sequence data to provide structured input foundations for downstream molecular generation. Their design objective is to eliminate error risks from manual data conversion through automated processes, enhancing the reliability of the overall R&D pipeline.

(1) Input Classification Agent
The input classification agent employs a Qwen3:8B large model-based semantic parsing engine for initial request classification. This agent processes user input text through a technical route combining regular expression pattern matching and keyword feature extraction. The core processing workflow first applies the re.sub() function for text normalization to eliminate special character interference; then uses a predefined regular expression pattern library for entity recognition, including disease term patterns, UniProt ID patterns, and FASTA sequence feature patterns; finally, the contextual understanding capability of Qwen3:8B resolves semantic ambiguity. Classification results strictly follow JSON Schema specifications, containing two required fields—category and content—to ensure downstream agents receive standardized, unambiguous task instructions.

(2) Disease-Target Mapping Agent
The disease-target mapping agent constructs a cross-domain conversion channel based on biomedical databases. This agent implements core functionality through the disease_{{{to}}*{{protein}}}*{sequences}() function, with technical implementation comprising three key stages: first, loading the TTD disease-target database text file, using the csv.reader() module to parse the INDICATI field line-by-line, applying regular expressions to validate and extract target IDs in standard format; then calling the UniProt service interface from the bioservices library, using the UniProt.retrieve() method to batch obtain corresponding protein sequences; finally, using BioPython's SeqIO.write() function to integrate multi-sequence FASTA files. The processing pipeline includes a term mapping fault-tolerance mechanism that automatically activates Levenshtein distance algorithms for similar term expansion retrieval when exact matching fails, ensuring effective mapping of disease term variants (e.g., "Alzheimer" vs. "Alzheimers"). Output files follow the internationally accepted FASTA format standard, with each record containing complete target identification and amino acid sequence information.

(3) Gene Translation Agent
The gene translation agent implements bioinformatics conversion from gene identifiers to protein sequences through the gene_{{{target}}*{{to}}*{{{protein}}_{{sequence}}}() function. This agent integrates myGene public services and UniProt sequence retrieval services to form a dual-layer parsing architecture: the first stage

calls the mygene.MyGeneInfo.query() interface, setting species="human" and fields="uniprot.Swiss-Prot" parameters to obtain UniProt ID sets corresponding to gene symbols, using a priority ranking algorithm to prefer reviewed entries; the second stage retrieves FASTA-format sequence data through the UniProt.retrieve() method from the bioservices library, and applies RDKit's Chem.MolFromSequence() for amino acid composition validity verification. To address the complexity challenge of gene nomenclature, a dynamic routing strategy is designed: when standard gene symbol parsing fails, the input is automatically attempted as a UniProt ID for direct sequence retrieval; when retrieval results are abnormal, a homologous gene expansion retrieval mechanism is activated. The final output file adopts standardized FASTA format, with headers containing complete gene identifier and UniProt ID mapping information.

This class of agents constructs the core infrastructure for biomedical data conversion, achieving reliable transformation from unstructured biological entities to machine-readable protein sequences through strictly defined function interfaces and standardized data processing workflows. Each agent adopts modular design principles, supporting functional extension through parameterized configuration, and providing high-quality structured input data for downstream molecular generation stages.

4.2 Drug Design (Generation) Agent

The drug design agent constitutes the core computational unit of the system, responsible for transforming protein target sequences into small molecule ligands with potential biological activity. This agent establishes an automated design workflow from target structure to candidate drugs by integrating the deep learning-driven molecular generation model DrugGPT, providing a data foundation for subsequent ADMET property screening. Its primary function is converting standardized protein sequences from upstream data processing into candidate molecular structures in chemical space.

The drug design (generation) agent implements core functionality through the generate_{ligands}() function, constructing a molecular generation computational pipeline based on the DrugGPT framework. The function first specifies GPU computing resources through environment configuration parameters to ensure necessary hardware acceleration for molecular generation; then performs input preprocessing by writing protein sequences into temporary files compliant with FASTA 2.0 standards, with headers containing metadata such as target identifiers and generation timestamps; finally, calls the DrugGPT command-line interface to execute molecular generation tasks, with key runtime parameters including -p for specifying input protein sequence strings, -n for controlling the number of generated molecules (default 50), and -o for defining hierarchical output directory structures. The molecular generation model employs a conditional variational autoencoder architecture, extracting spatial geometric features and physicochemical property distributions of protein binding pockets

through three-dimensional convolutional neural networks. These features serve as conditional vectors to guide molecular structure sampling in latent space, ensuring generated ligands match target binding sites in spatial complementarity and interaction patterns. Generation results are stored in structured CSV files, with each molecular record containing four key data fields: molecular hash identifier (128-bit MD5 encrypted string), canonical SMILES representation (IUPAC-compliant), generation probability score (0-1 confidence metric), and estimated binding free energy (ΔG approximation based on molecular mechanics optimization, in kcal/mol).

The drug design (generation) agent forms a tightly collaborative mechanism with upstream data processing agents. When receiving multi-target FASTA files output from the disease-target mapping agent, it automatically activates the batch_{{{generate}}*{{ligands}}}() *batch processing workflow, parsing file content through bioinformatics tools and creating independent computation tasks for each target, with parallel processing quantity controlled by the max*{targets} parameter (default limit 3 targets). When collaborating with the gene translation agent, it directly accepts single-sequence FASTA file input, validates amino acid composition effectiveness through cheminformatics tools, and automatically extracts gene identifiers as target naming bases. The collaboration process follows strict data contract specifications: inputs must comply with FASTA 2.0 format requirements, sequence lines must not contain spaces or special characters; outputs adopt hierarchical directory structures; an exception status code mechanism is established, returning "invalid_{sequence}" codes when sequence validation fails and "resource_{unavailable}" status when GPU resources are insufficient; all error logs are automatically recorded in dedicated report text files in the output directory. This agent constructs an automated design bridge from biological targets to candidate drugs, with its deep learning model integrating structure generation and preliminary activity evaluation. Modular design ensures seamless connection with upstream data processing stages, standardized output structures provide reliable input for downstream ADMET screening, and strictly defined data contracts and exception handling mechanisms significantly enhance the efficiency and reproducibility of early-stage drug discovery workflows.

4.3 ADMET Drug Screening Agent

The ADMET drug screening agent constitutes the core of the system's pharmacokinetic evaluation, responsible for comprehensive assessment of absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of candidate molecules output from the ligand generation stage. This agent establishes quantitative associations between molecular properties and druggability through integrated multi-parameter prediction models, providing critical decision-making basis for subsequent nanocarrier design and experimental validation. Its primary function is systematic screening of candidate molecules in chemical space to identify potential drug candidates with favorable pharmacoki-

netic characteristics.

The ADMET drug screening agent implements core functionality through the admet_{{{filter}}*{{tool}}}() function, constructing a comprehensive evaluation pipeline based on RDKit and ADMET prediction models. The agent first locates molecular data files output from the ligand generation stage (hash{{{ligand}}*{{mapping}}}.csv) through the find{{{csv}}*{{directories}}}() function; then calls the ADMET prediction model to perform multi-dimensional property assessment for each molecule, with key computational metrics including blood-brain barrier permeability (BBB{Martins}), drug-likeness score (QED), lipophilicity (logP), polar surface area (tpsa), cardiac toxicity risk (hERG), and mutagenicity (AMES). The screening process employs a composite conditional decision tree: first applying the QED.qed() function to calculate drug-likeness scores, ensuring molecules meet basic Lipinski's rule requirements; then constructing a druggability screening funnel through multi-parameter joint constraints (BBB$ 0.70, QED 0.50, 1.0 $logP$ 4.5, $tpsa$ 90, $hERG$ 0.5, $AMES$ $0.5). Molecules passing screening are stored in CSV format, with each record retaining original hash identifiers and SMILES structures to ensure data traceability. Technical implementation adopts modular design, encapsulating prediction algorithms through the ADMETModel() class to support batch processing and incremental updates.

The ADMET drug screening agent forms an end-to-end collaborative workflow with the ligand generation agent. Upon completion of ligand generation computation, it automatically triggers screening tasks, precisely locating input files through hierarchical directory structures (output root directory/target ID/seq_{min}). The collaboration follows standardized data contracts: input files must contain hash_{{{ligand}}*{{mapping}}}.csv and comply with hash identifier-SMILES binary structure; output files adopt the {filtered}.csv suffix naming convention, retaining original directory structures. Exception handling mechanisms include: automatically marking invalid SMILES structures as "invalid_{smiles}," recording molecules with prediction failures as "prediction_{error}." Screening results feed back to the system hub through JSON-format status reports, containing successfully processed file counts, output path lists, and detailed screening statistics, providing decision support for subsequent nanocarrier design.

This agent constructs a pharmacokinetic evaluation system for candidate molecules, significantly enhancing druggability potential through multi-parameter joint screening mechanisms. Its modular design and standardized interfaces ensure seamless connection with upstream ligand generation stages, while quantitative screening standards provide high-quality input for nanocarrier design and in vitro/in vivo experiments, substantially reducing late-stage R&D failure risks and optimizing resource allocation efficiency in drug discovery.

4.4 Drug Delivery Agent Design

The drug delivery design agent addresses critical technical challenges in central nervous system delivery of candidate drug molecules. This module focuses on blood-brain barrier penetration challenges, designing targeted nanocarrier systems through computation-driven material selection and structural optimization. Its core objective is constructing safe and efficient delivery solutions for active molecules optimized by the ADMET screening stage.

The drug delivery design agent employs a rational design approach based on biomaterial databases. The design process comprises three sequential stages: first, material compatibility analysis based on physicochemical parameters such as drug molecule logP values, molecular weight, and polar surface area to screen optimal carrier matrices from pre-validated material libraries (including 12 pharmaceutical polymers such as PLGA, PEG-PLGA, and chitosan); second, carrier structure optimization to determine key parameters including particle size range (80-120 nm), surface charge (-5 to +5 mV), and drug loading capacity through hydrodynamic simulation; third, targeted modification design, selecting appropriate targeting ligands (e.g., transferrin, Angiopep-2 peptide) based on target receptor expression characteristics (e.g., transferrin receptor, low-density lipoprotein receptor) and calculating optimal modification density (2-4 ligand molecules per square nanometer). Design solution outputs comprise three parts: carrier composition, preparation process, and critical quality attributes, forming complete formulations ready for experimental validation.

This agent establishes a transformation pathway from active molecules to delivery systems, with its empirical data-based decision-making mechanism providing reliable guidance for nanoformulation development. Through systematic integration of material property databases, physiological barrier models, and target expression data, it significantly enhances design efficiency for central nervous system delivery solutions, laying a technical foundation for subsequent experimental research.

4.5 Web Interface Design

The system employs the Streamlit framework to construct a lightweight web interface serving as an intuitive interaction channel between users and the agent cluster. This design enables complete visualization of the drug R&D workflow, allowing researchers to access system functions directly through browsers without programming knowledge or command-line operations. The interface adopts a conversational interaction pattern, simulating natural research collaboration processes while providing real-time workflow status feedback. The actual web interface is shown in Figure 2 [Figure 2: see original paper].

The interaction system architecture comprises three core modules: 1) a dialogue management module that maintains interaction history between users and agents through st.session_{state} for state persistence; 2) a task distribution module that automatically routes to corresponding agents based on user input

types; and 3) a result presentation module that transforms complex technical results into highly readable natural language descriptions. Security mechanisms employ session isolation design to ensure data from different users remains separate.

In implementation, the interface initialization sets titles and layouts through st.set_{{{page}}*{{config}}}(), creating a central control panel. The user input area implements real-time dialogue interaction through st.chat*{input}(), with agent responses displayed in chat bubble format. The core processing workflow is implemented through asynchronous task mechanisms: user input is first passed to the classification agent for task type identification, which parses text semantics based on the Qwen3:8B model to accurately distinguish between disease-target mapping, gene translation, ligand generation, or nanocarrier design tasks; based on classification results, the system automatically calls corresponding agent workflows: disease names trigger the disease-target mapping agent to access the TTD database, gene identifiers activate the gene translation agent to query the MyGene API, and protein sequences are passed to the ligand generation agent to drive DrugGPT computation. All processing results are encapsulated into InputInfo objects, whose data structures transform technical parameters into natural language descriptions through overridden ___{str}_ methods, with these structured results ultimately displayed in the interaction interface's history area through Streamlit's st.chat_{message}() component in chat bubble format, enabling researchers to intuitively view complete workflow progress.

The user operation process is intuitive and straightforward: researchers input natural language instructions in the chat box (e.g., "Convert disease Alzheimer to protein target sequences and develop related drugs"), the system automatically parses requirements and triggers corresponding agent workflows. Real-time status prompts are displayed during processing, with final results containing key information (e.g., target count, file paths) and subsequent operation suggestions. For multi-step tasks (e.g., disease→target→ligand generation→ADMET screening), the system automatically chains workflows, reducing manual user intervention.

This design achieves visual access to complex agent systems through the Streamlit framework, significantly lowering usage barriers. The conversational interaction pattern aligns with researcher work habits, real-time status feedback enhances system transparency, provides an efficient collaboration platform for cross-disciplinary teams, and effectively promotes the engineering application of computational drug discovery.

5 End-to-End Example: Alzheimer's Disease Targeted Drug Development

This chapter demonstrates the complete system workflow from target to delivery solution using the key Alzheimer's disease target β-secretase 1 (BACE1) as

an example. The target FASTA sequence has a length of 509 amino acids (sequence identifier: MAQALPWLLL…QHDDFADDISLLK), serving as the input foundation for the workflow.

5.1 From Target to Ligand

Using the key Alzheimer's disease target β-secretase 1 (BACE1) as the research object, we input its 509-amino-acid FASTA sequence. This sequence is processed by the ligand generation agent driving the DrugGPT model, which employs three-dimensional convolutional neural networks to extract spatial topological features of the BACE1 binding pocket and performs directed sampling in chemical space to generate ligand structures. The generation process strictly follows preset parameters: inputting the complete target sequence and setting the generation quantity to 404 candidate molecules. Output results are stored in hash identifier-SMILES key-value pair format, with each molecule assigned a 128-bit MD5 hash value as a unique identifier (e.g., 751c49fa3f906b94948f4ae22bea329004b24bae) and corresponding canonical SMILES structure recorded. All generated molecules are completely preserved in csv files, achieving automated transformation from biological target structures to chemical ligands.

5.2 ADMET Screening

The 404 candidate molecules enter the ADMET screening workflow for multi-dimensional pharmacokinetic joint assessment. Screening criteria require blood-brain barrier permeability (BBB_{Martins}) $ $0.70, drug-likeness score (QED) $ $0.50, lipophilicity (logP) between 1.0-4.5, polar surface area (tpsa) $ 90^{2}$, cardiac toxicity risk (hERG) $ $0.5, and mutagenicity (AMES) $ 0.5.$The screening process is implemented through the admet_{filter_{tool}} function : first loading the pre-trained ADMET prediction model, then batch-calculating physicochemical properties and
001 to BACE1-006) and their complete ADMET parameters are recorded in the filtered_{preds}.csv file, where typica
001 exhibit characteristic parameters including $logP = 3.66, BBB = 0.59,$ and $tpsa =$
77.15^{2}$.

5.3 Corresponding Delivery Solutions

Based on the ADMET characteristics of the six candidate molecules, the system generates customized delivery solutions. For high lipophilicity molecules BACE1-001 and BACE1-002, an LRP1-targeted PLGA-PEG polymeric micelle system is designed, using PLGA(50:50) as the core material, surface-modified with Angiopep-2 peptide as the targeting ligand, and co-loaded with P-gp inhibitor Tariquidar to overcome efflux effects, with particle size strictly controlled in the 80-120nm range. For low blood-brain barrier permeability molecules BACE1-003 and BACE1-004, a dual-targeted responsive nanoparticle is constructed, using DPPC/cholesterol lipid-polymer hybrid structures as carriers, surface-modified with transferrin and RVG29 peptide (1:1 ratio)

simultaneously, achieving reactive oxygen species-responsive drug release through thioether bonds, with particle size range of 70-100nm. For balanced molecules BACE1-005 and BACE1-006, an enzyme-pH dual-responsive mesoporous silica carrier is developed, using mesoporous silica as the matrix (pore size 4nm), achieving drug loading through pH-sensitive hydrazone bonds, employing MMP-9 cleavable peptides as gate systems, and surface-coated with chitosan-g-PEG to improve biocompatibility. All solutions detail material composition, modification density, and process parameters, providing clear guidance for experimental translation.

This case completely demonstrates the system's end-to-end R&D capability: starting from target sequence input, through ligand generation and multi-dimensional ADMET screening, to final customized delivery solution output, forming a computation-driven drug discovery closed loop. Each stage achieves seamless connection through standardized file formats, providing an efficient research paradigm for neurodegenerative disease drug development.

6 Conclusion and Outlook

This study constructs a multi-agent drug R&D system based on the AutoGen+Ollama framework, achieving the first automated closed-loop from target identification to delivery solution design. The system innovatively integrates six functional modules with Qwen3:8B large model as the agent reasoning core: input classification agent for natural language task parsing; disease-target mapping agent establishing disease-target association networks through the TTD database; gene translation agent completing bioinformatics conversion from gene identifiers to protein sequences; ligand generation agent driving DrugGPT model for directed generation from target structures to ligands; ADMET screening agent constructing six-dimensional pharmacokinetic joint evaluation funnel; and nanocarrier design agent outputting customized solutions for central nervous system delivery challenges. In the Alzheimer's BACE1 target validation case, the system successfully transformed a 509-amino-acid protein target sequence into 404 candidate ligands, obtaining 6 druggable molecules after strict screening and generating matching delivery solutions. Core innovations include: pioneering the agent-based integration architecture for the complete drug R&D workflow, achieving consumer-grade GPU deployment through lightweight design; deeply integrating professional toolchains (TTD/UniProt/DrugGPT) with large model capabilities to form a digital R&D chain; and transforming complex technical workflows into natural language operations through Streamlit-based conversational interfaces. Current limitations primarily manifest in delivery design relying on large model outputs without physical knowledge constraints such as molecular dynamics simulations, and workflow flexibility requiring improvement. Future research will focus on embedding physics-based validation mechanisms such as free energy perturbation calculations in the delivery stage, developing customizable workflow engines, and expanding to emerging paradigms such as antibody drugs, accelerating AI pharmaceuticals from computational

design to clinical translation.

[1]Sheikholeslami Mahsa,Mazrouei Navid,Gheisari Yousof, et al. DrugGen enhances drug discovery with large language models and reinforcement learning.[J]. Scientific reports, 2025, (1): 13445.
[2]Luo Renqian,Sun Liai,Xia Yingce, et al. BioGPT: generative pre-trained transformer biomedical text generation and mining[J]. Oxford University Press (OUP), (6).
[3]Kai Zhang,Rong Zhou,Eashan Adhikarla, et al. A generalist vision–language foundation model for diverse biomedical tasks[J]. Nature Medicine, 2024, 30: 3129-3141.
[4]Li Yuesen,Gao Chengyi,Song Xin, et al. DrugGPT: A GPT-based Strategy for Designing Potential Ligands Targeting Specific Proteins[Z]: Cold Spring Harbor Laboratory.
[5]Singhal Rahul-K.Arora-Jason-Wei-Rebecca-Soskin-Hicks-Preston-BowmanJoaquin-Quin˜onero-Candela- Foivos-Tsimpourlas-Michael-SharmanMeghan-Shah-Andrea-Vallone-Alex-Beutel-Johannes-Heid eckeKaran. HealthBench:Evaluating Large Language ModelsTowards Improved Human Health[J],
[6]Chin Shuang-Ying,Why Dr.-Ng-Kok. Comparative of Multi-Agent System Frameworks: Crewai, Langchain, and Autogen[Z]: Elsevier BV. huang
[7]Kexin Serena-Zhang-Hanchen-Wang-Yuanhao-Qu-Yingzhou-Lu-Yusuf-Roohani-Ryan-Li-Lin-Qiu-Gavi n-Li-Junze-Zhang-Di-Yin-Shruti-Marwaha-Jennefer-N.-Carter-Xin-Zhou-Matthew-Wheeler-Jonat han-A.-Bernstein-Mengdi-Wang-Peng-He-Jingtian-Zhou-Michael-Snyder-Le-Cong-Aviv-Regev-J ure-Leskovec. Biomni: A General-Purpose Biomedical AI Agent[EB/OL]: Cold Spring Harbor Laboratory, 2025. https://doi.org/10.1101/2025.05.30.656746. 10.1101/2025.05.30.656746.
[8]Zhu Feng,Han BuCong,Kumar Pankaj, et al. Update of TTD: Therapeutic Target Database[J]. Oxford University Press (OUP), (suppl_1): D787-D791.
[9]Chen X.. TTD: Therapeutic Target Database[J]. Oxford University Press (OUP), (1): 412-415.
[10]Uniprot consortium The. UniProt: the universal protein knowledgebase[J]. Oxford University Press (OUP), (5): 2699-2699.
[11]Welcome to MyGene.py's documentation! — MyGene.py v3.1.0 documentation[EB/OL]. https://docs.mygene.info/projects/mygene-py/en/latest/.

(Corresponding author: Tian Zihan E-mail: 3208566786@qq.com)

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.