MMP9 in Pan-Cancer and Computational Study to Screen MMP9 Inhibitors

Bo Wu
Department of orthopaedic trauma, Center of Orthopaedics and Traumatology, the First Hospital of Jilin University, Street Xinmin 71, Changchun, China.

Xinyu Wang
Department of orthopaedic trauma, Center of Orthopaedics and Traumatology, the First Hospital of Jilin University, Street Xinmin 71, Changchun, China.

Chunzheng Chen
Department of orthopaedic trauma, Center of Orthopaedics and Traumatology, the First Hospital of Jilin University, Street Xinmin 71, Changchun, China.

Qifan Yang
Department of orthopaedic trauma, Center of Orthopaedics and Traumatology, the First Hospital of Jilin University, Street Xinmin 71, Changchun, China.

Pengcheng Zhou
Department of orthopaedic trauma, Center of Orthopaedics and Traumatology, the First Hospital of Jilin University, Street Xinmin 71, Changchun, China.

Shiyuan Lu
Department of orthopaedic trauma, Center of Orthopaedics and Traumatology, the First Hospital of Jilin University, Street Xinmin 71, Changchun, China.

Liangxin Lin
Department of orthopaedic trauma, Center of Orthopaedics and Traumatology, the First Hospital of Jilin University, Street Xinmin 71, Changchun, China.

Yanwei Liu
Department of orthopaedic trauma, Center of Orthopaedics and Traumatology, the First Hospital of Jilin University, Street Xinmin 71, Changchun, China.

Ming Li
Honghui Hospital, Xi'an Jiaotong University College of Medicine, Xi'an 710054, P. R. China.

Dong Zhu (✉ zhu_dong@mail.jlu.edu.cn)
Department of orthopaedic trauma, Center of Orthopaedics and Traumatology, the First Hospital of Jilin University, Street Xinmin 71, Changchun, China.

Article

Keywords: MMP9, Pan-cancer, Ligand-based pharmacophore model, Machine learning model, Virtual screening

Posted Date: February 28th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2589286/v1

Additional Declarations: No competing interests reported.

Abstract

Purpose Stromal metalloproteinase 9 (MMP9) is a stromal cell protein associated with extracellular matrix (ECM) degradation and remodeling, which promotes tumor invasion and metastasis and regulates the activity of cell adhesion molecules and cytokines. This study aims to evaluate the MMP9 in pan-cancer and to screen out new major compounds and drug candidates that can inhibit MMP9.

Methods The pan-cancer dataset was downloaded from UCSC database. The MMP9 expression in pan-cancer was evaluated and the correlation of MMP9 with tumor microenvironment (TME), RNA Modification Genes and tumor mutation burden (TMB) in pan-cancer. Then, MMP9 crystal structures was download and a ligand-based pharmacophore model was constructed. machine learning model was also constructed for further screen. The identify compounds were pooled into Discovery Studio 4.5 for ADME (absorption, distribution, metabolism, excretion) and toxicity prediction. Molecular docking was performed using the Glide module of the Schrödinger kit to demonstrate the binding affinity and mechanism between compounds and MMP9. To assess whether the ligand-receptor complex is stable, Molecular Dynamics Simulation of Discovery Studio 4.5 was used.

Results The results of our pan-cancer analysis showed difference expression of MMP9 in tumor tissues. Prognostic analysis showed that higher MMP9 expression means both worse survival rate and tumor progression in GMBLGG, KIPAN, UVM, LGG, ACC, and LIHC. MMP9 expression in GMBLGG, KIPAN, UVM, LGG, ACC, and LIHC was significantly positively correlated with TME. After screened by ligand-based pharmacophore model and machine learning model, 49 small molecules were identified. ADME and toxicity prediction results indicated that CEMBL82047 and CEMBL381163 were identified as potential MMP9 inhibitor. CEMBL82047 and CEMBL381163 showed robust binding affinity with MMP9, and the complex can exist stably in the natural environment.

Conclusion CHEMBL82047 and CHEMBL381163 are ideal potential leading compounds for inhibiting MMP9. The findings of this study and these selected drug candidates have greatly contributed to the design and improvement of novel MMP9-targeted drugs.

Introduction

It is well known that tumor is formed by proliferation and abnormal differentiation of normal cells under the action of various initiating and promoting factors. It destroys normal tissues and organs, and can cause gradual organ dysfunction until failure or death due to compression, consumption or destruction, especially malignant tumor[1]. As one of the leading causes of death worldwide[2], malignant tumor has extremely low cure rates in both developed and developing countries[3, 4]. Although great progress has been made in tumor therapy, a large number of patients still have a poor prognosis and low survival rates. Thus, novel therapeutic methods and drugs are still urgent need to develop.

MMP-9 (Matrix Metalloproteinase-9), a member of the zinc-dependent endopeptidase family, is one of gelatinases. It is involved in a variety of biological processes such as proteolytic Extracellular matrixc

degradation, cell–ECM or cell–cell interactions and cell surface cleavage activities, in addition to degrading and regulating extracellular matrix proteins and releasing bioactive proteins including cytokines, chemokines and growth factors[5, 6]. MMP-9 degrades type IV collagen and disrupts basement membranes associated with tumor invasion and metastasis. The expression level of MMP-9 mRNA was significantly higher in Nasopharyngeal Carcinoma tissues than in nasopharyngeal tissues, and overexpression of MMP-9 accelerated tumor growth by inducing angiogenesis and enhanced local cell invasion and metastasis by degrading the extracellular matrix[7]. For oesophageal cancer, MMP-9 overexpression correlated significantly with the depth of tumor infiltration, lymphatic infiltration, lymph node metastasis and degree of pathological differentiation[7]. The ECM is a key component of local tumor microenvironment (TME) and undergos extensive remodelling during breast cancer (BC) evolution. MMP-9 is reported as a key player in ECM remodelling during cancer initiation and progression through a variety of mechanisms[8].

At present, several chemotherapeutic agents targeting MMP-9 exist. MMP-9-IN-1, a highly selective MMP-9 inhibitor with oral efficacy[9, 10], could selectively inhibit MMP-9 to control the development, progression, invasion and metastasis of nasopharyngeal carcinoma, but also affects the function of the human respiratory system and reduces the activity of other proteases and cytokines due to its strong and effective inhibitory effect[7, 11]. JNJ0966 is another highly selective MMP-9 inhibitor that blocks the conversion of MMP-9 zymogen to a catalytically active enzyme[12]. However, it is currently only used for scientific research. Besides, other MMP9 inhibitor exist with extensive effect targets, which means more side effects. Therefore, novel drugs targeting MMP-9 are needed and worthy to be identified.

In this study, we combined pharmacophore model and machine learning model to screen novel inhibitors of MMP-9. Pharmacophores are combinations of characterized three-dimensional structural elements[13, 14], and it has been used to design and screen new drugs based on specific ligand structures[15, 16]. Machine learning has been used for prediction or classification based on data analysis[17]. It has been reported helpful in many fields such as clinical data procession[18, 19]. We explored the role of MMP-9 in pan-cancer and assessed the relevance of MMP-9 to the tumor immune microenvironment and mRNA modifications. Then, pharmacophore model and machine learning model were constructed to screen inhibitors for MMP9, following ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) analysis, protein-ligands docking and MD (Molecular Dynamics) simulation. This research provided a novel investigation strategy and a group of therapeutic candidates for MMP9, which might serve as a strong foundation for further agonist research.

Methods

Analysis of the Expression Level of MMP9 in Pan-Cancer Datasets

We downloaded the unified and standardized pan-cancer dataset from the UCSC (https://xenabrowser.net/) database: TCGA Pan-Cancer (PANCAN, N = 10535, G = 60499), we extracted

the expression data of ENSG00000100985 (MMP9) gene in each sample, further screened the samples from Solid Tissue Normal, Primary Blood Derived Cancer-Peripheral Blood, and Primary Tumor, and performed log2(x + 0.001) transformation on each expression value, we excluded cancers with fewer than three samples in a single cancer. We calculated the difference in expression between normal and tumor samples in each tumor using R software (version 3.6.4), performed a significance analysis using unpaired Wilcoxon Rank Sum and Signed Rank Tests, and finally completed the plot of the difference in expression of MMP9.

Identification of the Correlation between MMP9 Expression levels and Survival in Pan-Cancer

We extracted several metrics (overall survival [OS] and progression-free survival [PFS]) from TCGA samples to investigate the association between MMP9 expression and patient outcomes. We obtained a high-quality prognostic dataset of TCGA from a previously published TCGA prognosis study published in Cell and excluded cancers with fewer than 10 samples in a single cancer and samples with a follow-up time of less than 30 days. We used the R software package "survival" to obtain a "forest map" for Cox to analyze the relationship between MMP9 gene expression and survival in each tumor. The patients in each tumor type in the TCGA dataset were divided into two groups according to the best cut-off value of MMP9 to compare the prognosis differences, and the prognostic differences between the two groups were further analyzed by the "survfit" function of the R software package "survival", and finally the "logrank test" method was used to evaluate the significant prognosis difference between the samples of different groups.

Association between MMP9 Expression and TME in Pan-Cancer

We extracted the gene expression profiles of each tumor separately, mapped the expression profiles to "Gene Symbol", and further used the R software package "ESTIMATE" to calculate the stromal, immune, and ESTIMATE scores of each patient in each tumor according to gene expression. We used the "corr.test" function of the R software package "psych" to calculate the Pearson's correlation coefficient between genes and immune invasion and immune cell invasion scores in each tumor to determine a significant correlation of immune invasion scores.

Correlation between MMP9 Expression and mRNA-modifying genes in Pan-Cancer

We extracted the expression data of the marker genes of MMP9 gene and three types of RNA modification genes (m1A, m5C, m6A) in each sample, we then screened samples from Primary Blood Derived Cancer - Peripheral Blood and Primary Tumor, and calculated the pearson correlation between MMP9 and the marker genes of the five types of immune pathways by filtering all normal samples and transforming each expression value. This data was used to estimate the important role of RNA

modifications in cancer using the gene expression dataset, and to further summarize their therapeutic potential for abnormal deposition in cancer.

Association between MMP9 Expression and Tumor Mutation Burden (TMB) in Pan-Cancer

We downloaded simple nucleotide variation data from the database and processed the data. A simple nucleotide variation dataset was used to plot the mutational landscape of MMP9 in 4 tumor types. TMB scores were calculated based on mutation data from four tumor samples, the Cancer Genome Atlas (TCGA), and patients were divided into low TMB and high TMB groups according to the TMB score quartile. Differentially expressed genes (DEGs) were identified in the low and high TMB groups.

Construction and verification of pharmacodynamic mass models

Pharmacophore models have proven useful for screening ideal compounds, and we know of two pharmacophores: structure-based pharmacological models derived directly from the X-ray structure of protein-ligand complexes; and ligand-based pharmacological models derived from the structure of known active compounds. Analyze the crystal structure of human MMP9 receptors with different ligands (PDB IDs: 2OW0, 2OW1, 4H3X & 4WZV) by LigandScout v4.3, which provides automated construction of three-dimensional pharmacophores. LigandScout identifies 3D chemical features, ligand options containing hydrogen bond donors (HBD) and acceptors (HBA) are shown as concentrated vectors along with negative and positive ignitable spheres moreover lipophilic regions are indicated by spheres. In addition, to expand selectivity, the LigandScout indicator incorporates spatial data about regions into each promising inhibitor. By using the web server Pharmit (http://pharmit.csb.pitt.edu/), we enter pharmacophore signatures and search for and identify small molecules that bind to the target molecule (MMP9 receptor) based on structural and chemical similarities between small molecules. By joining the code from the PDB, 1752844 possible small molecules are obtained. Then, the deep learning model was built by DeepScreening (http://deepscreening.xielab.net/) for further screen and the performance of the model was evaluated based on test loss, accuracy, recall, precision, F1, and MCC.

ADME (Absorption, Distribution, Metabolism, and Excretion) and Toxicity Prediction

The ADME module of Discovery Studio 4.5 is used to calculate the absorption, distribution, metabolism and excretion (ADME) of selected compounds, including their water solubility, blood-brain barrier permeability, cytochrome P-450 2D6 (CYP2D6) inhibition, hepatotoxicity, human enteric absorption, and plasma protein binding levels. Use the topcat module of Discovery Studio 4.5 to calculate toxicity and other properties of all potential compounds, such as National Toxicology Program rodent carcinogenicity, Ames mutagenic city, developmental toxicity potential, and the median oral lethal dose (LD50) and chronic oral minimum observed adverse reaction level (LOAEL) in rats. These pharmacological properties were fully considered when selecting the appropriate drug candidates for MMP9.

Protein Molecule Docking

Molecular docking using the Glide module of the Schrödinger kit to collect the active conformation of small molecules interacting with the MMP9 receptor. Prepare top-level compounds from pharmacophore screening in Maestro using the LigPrep module to obtain the starting structure for docking. Ligand-acceptor interactions include hydrogen bond interactions, van der Waals interactions, π-π stacked interactions, and ionic interactions. The results of molecular docking were analyzed based on the binding energy (kcal/mol) between small molecules and amino residues and the number of binding interactions.

Molecular Dynamics Simulation

The best binding conformations of the ligand-MMP9 complexes among the poses predicted by the molecule docking program were submitted to MD simulation using Discovery Studio 4.5. Place the ligand-acceptor complex into an orthogonal box and solvate with an explicit periodic boundary solvated water model. To simulate the physiological environment, sodium chloride is added to a system with an ionic strength of 0.145. The system is then subjected to a CHARMM force field for analogy-based ligand parameterization. For this system, the following simulation protocols are applied: 1000 minimization steps for the fastest descent and conjugate gradient; 5ps equilibrium simulation at 300 K temperature (slow drive 2ps from 50 K initial temperature) and atmospheric pressure set; 25ps-MD simulation (production mode) at NPT (atmospheric pressure and temperature). The Particle Grid Ewald (PME) algorithm is used to calculate remote electrostatic and the Linear Constraint Solver (LINCS) algorithm is suitable for fixing all bonds involving hydrogen. With the initial complexity setting as a reference, the trajectories of the root mean square deviation (RMSD), potential energy, and structural features were determined by the Discovery Studio 4.5 analysis trajectory protocol in Discovery Studio 4.5.

Result

MMP9 Expression in Pan-cancer

The whole process of data analysis is depicted in Fig. 1.We analyzed the expression data of 26 cancer types and found that MMP9 was highly expressed in the vast majority of tumor samples, and there was a statistical difference in most tumors, including GBM CESC LUAD COAD COADREAD BRCA ESCA STES KIRP KIPAN STAD PRAD UCEC HNSC KIRC LUSC LIHC READ PCPG BLCA KICH CHOL ($P < 0.05$). Although MMP9 is highly expressed in LGG, CESC, and PAAD, due to the small sample size of the control group (Normal), no significances were uncovered. Besides, for THCA, there was no difference in the expression of MMP9 compared with the normal samples (Fig. 2A) .

Pan-cancer Prognostic Analysis of MMP9

To further explore the association between MMP9 and the prognosis of pan-cancer, we performed prognostic analysis on 39 cancer types. The overall survival (OS)(Supplementary Fig. 1A) results show that for GMBLGG, KIPAN, UVM, LGG, ACC, KIRC, LIHC, BLCA, and TGCT, the higher the expression of

MMP9, the lower the survival rate of the tumor ($P < 0.05$); for SKCM, and SKCM-M, the higher the expression of MMP9, the higher the survival rate of the tumor, which means MMP9 is a beneficial factor for this two tumors ($P < 0.05$); for the other 28 tumors, there was no significances ($P > 0.05$). We also plot the survival curves of GMBLGG, KIPAN, UVM, LGG, ACC, KIRC, LIHC, BLCA, and TGCT (Fig. 2B, Supplementary Figs. 1B-C). In addition, we analyzed the progression-free survival (PFS) of pan-cancer (Supplementary Fig. 1D), we found that for GMBLGG, KIPAN, KIRC, UVM, LGG, ACC, THCA, GBM, and KICH, the higher the expression of MMP9, the faster the tumor progression ($P < 0.05$); for DLBC and OV, the higher the expression of MMP9, the slower the tumor progression was, which means the MMP9 was a suppressor gene of tumor development ($P < 0.05$); for the other 28 tumors, there was no significances ($P > 0.05$). To sum up, higher MMP9 expression means both worse survival rate and tumor progression in GMBLGG, KIPAN, UVM, LGG, ACC, and LIHC.

Correlation of MMP9 Expression with TME and Immune Infiltration

The tumor microenvironment (TME) is composed of various components such as immune cells, non-immune stromal cells, and extracellular matrix proteins, including innate immune cells, adaptive immune cells, extracellular immune factors, and cell surface molecules. TME has unique internal interactions and important roles in tumor biology and is known as the tumor immune microenvironment (TIME)[20, 21]. To further explore the correlation between MMP9 and tumor immune infiltration, we performed immune analysis on 6 tumors with MMP9 expression. We found that MMP9 expression in GMBLGG, KIPAN, UVM, LGG, ACC, and LIHC was significantly positively correlated with Immune score, ESTIMATE score, and Stromal score (Fig. 2C, Supplementary Fig. 2).

In addition, we also analyzed the correlation of MMP9 expression with immune cells in each tumor (Fig. 3A). We found that macrophages were significantly associated with MMP9 expression, among which M0 macrophages were significantly positively correlated with all 6 tumors; classically activated M1 macrophages were positively correlated in GMBLGG, KIPAN, UVM, LGG, and ACC; alternative activated M2 macrophages were positively correlated in GMBLGG, and LGG. The high expression of macrophages releases more cytokines (such as EGF), which promotes the metastasis and invasion of cancer cells[22, 23], so this our results provided a new explain of high correlation between MMP9 expression and metastasis. Monocytes were significantly negatively correlated with MMP9 expression in 5 tumors except ACC, which means the ability to recognize and kill tumor cells was inhibited[24]. The activated natural killer cells were negatively correlated with MMP9 expression in GMBLGG, KIPAN, KIRC, and ACC, that means their killing power to tumor cells was decreased when tumors express more MMP9. Besides, MMP9 expression was positively correlated with Tregs (T cell regulatory) in GMBLGG, KIPAN, UVM, LGG, and KIRC, which could suppression of the immune system[25].

Correlation of MMP9 Expression with RNA Modification Genes

RNA chemical modifications have important roles in fundamental cellular processes such as cell differentiation, protein production, cell signaling, and maintenance of circadian rhythms[26, 27], and these modifications can play critical roles as tumor suppressors or tumor promoter effect. We found that GBMLGG was positively correlated with most of the genes in m1A modification, and there was a significant statistical difference; The gene ALKBH3 was significantly and positively associated with MMP9expression in four tumors, GBMLGG, KIPAN, ACC and LGG, with statistically significant differences (Fig. 3B). ALKBH3 can promote the proliferation, migration and invasion of cancer cells[28]. In m6A modification (Fig. 3C), MMP9 expression was positively correlated with most genes in GBMLGG, and there was a significant statistical difference; TRMT61A was significantly and positively correlated with MMP9 expression in four tumors, GBMLGG, KIPAN, ACC and LGG, with statistically significant differences (Fig. 3C). A58 in m1tRNA is composed of RNA-binding component TRMT6 and catalytic component TRMT61A, which is very important for maintaining its stability and affecting translation initiation, and has profound effects on various biological processes[29]. In m5C modification (Fig. 3D), MMP9 expression was positively correlated with most genes in GBMLGG, and there was a significant statistical difference; DNMT3B was significantly and positively correlated with MMP9 in four tumors, GBMLGG, KIPAN, ACC and LGG, with statistically significant differences (Fig. 3D). DNMT3B is considered to be involved in de novo DNA methylation in embryonic stem cells and early embryos. It is overexpressed in several human tumors and is an indicator of early tumor recurrence and poor prognosis in hepatocellular carcinoma[30].

Correlation of MMP9 Expression with TMB

We furtherly performed single nucleotide polymorphism (SNP) analysis by dividing patients into two groups: high MMP9 expression group and low MMP9 expression group. In LGG (Supplementary Fig. 3A), the genes IDH1, TP53, and ATRX have high mutation frequencies (> 20%), and EGFR, MYH13, EPPK1, MYO15A, SI, KIAA1109, CDH17, SLCO1B1, SYNE2, CFAP47, SSPO and ZFHX4 also have more mutation rates and mutation types in high MMP9 expression group. In KIRC (Supplementary Fig. 3B), the genes VHL and PBRM1 have high mutation frequencies (> 20%), and the mutation types are mostly missense mutation, frameshift deletion mutation, nonsense mutation, splice site and in-frame insertion; THSD7B, ADGRV1, XPO7, LAMC2 and UBR4 also have more mutation rates and mutation types in the MMP9 high expression group. TP53, CTNNB1 and MUC16 show high mutation frequencies (> 20%) in ACC (Supplementary Fig. 3C), as well as higher mutation rates in high MMP9 expression group; DST, FAT4, ASXL3, CNTNAP5 and NF1 also have more mutation rates and mutation types in high MMP9 expression group. In UVM (Supplementary Fig. 3D), the genes GNAQ, GNA11, BAP1, and SF3B1 have high mutation frequencies (> 20%), while BAP1 is higher in high MMP9 expression group; SF3B1 and EIF1AX in patients with high MMP9 expression show higher mutation rates.

Construction and validation of pharmacophore model

To further screen novel inhibitors of MMP9, we constructed a ligand-based pharmacophore model. We first considered evaluating the major residues, obtained by analyzing the crystal structures (PDB IDs: 2OW0, 2OW1, 4H3X, and 4WZV) to obtain the major residues of the MMP9 receptor (Fig. 4A, 4B, 4C, and

4D), identifying small active molecules and target proteins Physicochemical interaction patterns between them, mapping them to 3D array features (eg, hydrogen bonds, lipophilic contacts, ionic or aromatic interactions).

As shown in Fig. 4A, the crystal structure 2OWO exhibited two hydrophobic interactions binding with residues TYR423, LEU397, LEU418, VAL398 and ZN444. Two hydrogen bond acceptors with the ALA189, GLN402, HOH503, HOH608 and LEU188. In addition, a positively ionized region was also detected. The crystal structure 2OW1(Fig. 4B) exhibited two hydrophobic interactions binding with residues VAL398, LEU418, TYR423, LEU397 and ZN444. Five hydrogen bond acceptors with the LEU188, HOH593, HOH557, ALA189 and GLN402, as well as three hydrogen bond donors were also observed and a positively ionized region. The crystal structure 4H3X (Fig. 4C) exhibited two hydrophobic interactions binding with residues LEU243, TYR248, VAL223 and ZN301. Two hydrogen bond acceptors with the LEU188 and ALA189, as well as three hydrogen bond donors with ALA189, HIS226 and HOH415 were also observed and a positively ionized region. The crystal structure 4WZV (Fig. 4D) exhibited two hydrophobic interactions binding with residues TYR245, MET247, ZN302, VAL223 and TYR248. Four hydrogen bond acceptors with the ALA191, HOH401, LEU188 and ALA189, as well as three with ALA189, HIS230 and GLU227 hydrogen bond donors were also observed and a positively ionized region. As shown in Supplementary Fig. 4A-D, these drugs all exert the most important effect with amino acid residue H401.

Virtual Screening

We performed a prospective virtual screening (VS) of a database of compounds of natural origin and synthetic drugs, in which we used fitted values as pharmacology-based screening criteria. After removing duplicates, we screened 230 small molecules with the same pharmacophore from 1,752,844 small molecules. Then, we built a deep learning model of MMP9 with 3479 small molecules and validated it. the Accuracy, Precision, and AUC of the model we constructed gradually stabilized with the increase of Epoch, and finally stabilized at around 0.9 (Fig. 4E); Recall and F1 also gradually stabilized around 0.9, and Loss and MCC gradually stabilized around 0.45 and 0.7 respectively (Supplementary Fig. 4E-H). After screened by the machine learning model, 49 small molecules (score = 1) from the 230 small molecules were identified.

ADME and toxicity prediction

Pharmacokinetics is an important analytical method for detecting effective compounds in the process of drug discovery, and the analysis of its properties plays an important role in drug design (Supplementary Table 1). Water solubility predictions (defined in water at 25°C) indicated that 33 compounds were soluble in water. For human intestinal absorption, 21 compounds have good absorption levels. There are 40 compounds were highly bound to plasma proteins, and the rest were the opposite. CYP2D6 is one of the important enzymes involved in drug metabolism, and all 49 compounds were predicted to be non-inhibitors of cytochrome P450 2D6 (CYP2D6). For hepatotoxicity, seven compounds were predicted to be nontoxic. CHEMBL82047 and CHEMBL381163 have good water solubility, intestinal absorption and protein binding, and can act as non-inhibitors of CYP2D6 without hepatotoxicity. In addition to this, we

conducted a comprehensive investigation of the safety of these small molecules, the results show that two small molecules, CEMBL82047 and CEMBL381163, are non-mutagenic and are predicted to have less Ames mutagenic, rodent carcinogenic, and developmental toxicity potential than other compounds.

Protein molecular docking

To further study the binding properties of small molecules to proteins, we carried out molecular docking experiments (Fig. 5, Fig. 6A-B, Supplementary Fig. 5A-D). As shown in Table 1, CEMBL82047 and CEMBL381163 have higher binding affinity to the protein compared to the drugs JNJ0966 and MMP-9-IN-1. Supplementary Fig. 5E-F shows the π-dependent interactions and hydrogen bonds performed by the structural calculations. The results of structural calculation studies showed that CEMBL82047 formed 4 pairs of hydrogen bonds with the MMP9 residue acceptor, and in addition, the complex itself formed 4 pairs of π-related interactions with the MMP9 residue acceptor. CHEMBL381163 forms 4 pairs of hydrogen bonds and 7 pairs of π-related interactions with the MMP9 residue acceptor (Tables 2 and 3).

Table 1
COCKER potential energy of compounds.

	COCKER potential energy
CEMBL82047	-12.164
CEMBL381163	-11.623
JNJ0966	-6.629
MMP9IN1	-8.618

Table 2
Hydrogen Bond Interaction Parameters for Each Compound with MMP9 Residues.

Receptor	Compound	Donor Atom	Receptor Atom	Distances (Å)
2OW1	CEMBL82047	LEU188:H	UNK900:O2	1.84939
		ALA189:H	UNK900:O2	2.5637
		GLN402:HE22	UNK900:O5	2.04371
		UNK900:H29	ALA189:O	2.21685
	CEMBL381163	LEU188:H	UNK900:O3	2.73564
		GLN402:HE22	UNK900:O7	1.84541
		HIS411:HD1	UNK900:O4	2.63353
		UNK900:H22	ALA189:O	2.04562
	JNJ0966	LEU188:H	UNK900:N3	2.7175
		UNK900:H1	MET422:O	2.01943
	MMP9IN1	GLN227:HE21	UNK900:N2	2.6066
		UNK900:H11	TYR245:O	1.90625

Table 3
π-Related Interaction Parameters for Each Compound with MMP9.

Receptor	Compound	Donor Atom	Receptor Atom	Distances(Å)
2OW1	CEMBL82047	HIS401	UNK900	4.2886
		UNK900:C15	LEU187	4.82839
		UNK900	LEU188	5.32252
		UNK900	VAL398	4.8719
	CEMBL381163	UNK900:H4	HIS401	2.77631
		UNK900:C1	LEU397	4.16409
		UNK900:C14	LEU187	4.4138
		PHE110	UNK900:C15	4.14676
		HIS411	UNK900	5.04727
		TYR423	UNK900:C1	4.81758
		UNK900	LEU188	4.80899
	JNJ0966	HIS401	UNK900	4.00463
		UNK900	TYR423	5.5835
		ALA189	UNK900:C9	3.89591
		UNK900:C9	LEU188	4.59029
		UNK900:C9	VAL398	4.42465
		UNK900	LEU188	4.21123
		UNK900	VAL398	4.71999
		UNK900	LEU397	5.13481
		UNK900	LEU418	5.36355
	MMP9IN1	UNK900	LEU243	5.32463

Molecular dynamics simulation

Molecular dynamics simulation is a method for simulating the physical motion trajectories and states of atoms and molecules based on Newtonian mechanics. We build a molecular dynamics simulation module to evaluate the stability of small molecule-protein complexes under natural environment conditions. Figure 6C-D shows the potential energy and RMSD plots for each complex. The trajectories of each complex reached equilibrium, and the potential energy and RMSD of complexes CEMBL82047-

MMP9 and CEMBL381163-MMP9 reached a steady state over time. That means the complexes can exist stably in the natural environment.

Discussion

Tumors are one of the leading causes of death worldwide[2], and MMP9 has been previously reported as a cancer biomarker[6] as one of the initiating factors that promote tumor invasion and metastasis, greatly promote the occurrence and development of tumors[5, 6]. Although great progress has been made in the design and development of drugs targeting MMP9, these drugs have many shortcomings. This study aimed to systematically determine the expression pattern and prognostic value of MMP9 in pan-cancer, and to screen out specific targeting drugs against MMP9.

In this study, we explored that the expression level of MMP9 was significant different between tumor samples and normal samples in most of 26 cancers. Higher MMP9 expression implies poorer survival and tumor progression in GMBLGG, KIPAN, UVM, LGG, ACC and LIHC. The above findings are consistent with previous reports such as elevated MMP9 expression in breast cancer as a predictor of shortened patient survival[31–33]; it also acts as a prognostic biomarker for thyroid cancer[34]. In order to further confirm the correlation between MMP9 expression and tumors, we performed immune infiltration analysis on above 6 tumors with high MMP9 expression in the TCGA database. we found that MMP9 expression in tumor patients was significantly correlated with Stromal score, Immune score and ESTIMATE score. We further examined the relationship between MMP9 expression and infiltration of 22 immune cell subtypes, and our findings showed that the level of immune cell infiltration was significantly correlated with MMP9 expression in most cancer types. This also proves that there is immune escape in tumor patients with high expression of MMP9, and on the other hand, it also illustrates the mechanism of MMP9 in tumors. For example, macrophages were significantly positively associated with all 6 tumors, and high expression of macrophages promotes cancer initiation and malignant progression. During tumorigenesis, they create a mutagenic and growth-promoting inflammatory environment; as tumors progress to malignant tumors, macrophages stimulate angiogenesis, enhance tumor cell migration, invasion, and suppress antitumor immunity[34]. Monocytes were significantly negatively associated with 6 tumors, and their ability to generate antitumor effectors and activate antigen-presenting cells was suppressed[24]. NK cells were also significantly inhibited in these 6 tumors, and their functions to directly kill tumor cells and release soluble factors affecting innate and adaptive immune responses were significantly inhibited. In the tumor microenvironment (TME), Tregs can be induced and differentiated by traditional T cells, have strong immunosuppressive functions, inhibit anti-tumor immunity, and promote the occurrence and development of tumors, which also explains why Tregs in these types of tumors positively correlated with them[35]. Activated CD4 memory T cells can suppress anticancer immunity, thereby hindering protective immune surveillance of tumors and hindering effective antitumor immune responses of tumor hosts, thereby promoting tumor development and progression. This is consistent with the finding that expression of activated CD4 memory T cells is positively correlated with tumors[36].

To further explore the mutational landscape of MMP9 in different cancers, we found that UVM, KIRC, ACC, and LGG, four types of tumors with high MMP9 expression, had much higher mutation numbers and types than normal people, this also verified that MMP9 promotes tumorigenesis and development. Tumor mutational burden (TMB) reflects the number of cancer mutations, and higher TMB can be used as a prognostic factor for better outcomes. Mutations are processed as neoantigens and presented to T cells by major histocompatibility complex (MHC) proteins, while higher TMB results in more neoantigens, increasing the chances of T cell recognition and better immunotherapy efficacy Good[37].

Although MMP9 is highly expressed in most tumors and is closely related to tumor metastasis, there are only a few drugs targeting MMP9, and there are few specific targeted drugs, and there are many limitations. JNJ0966 is a specific inhibitor of MMP9 and it has been shown to be involved in the progression and development of various diseases, and can regulate a series of physiological response processes in the body by regulating the expression of MMP9 in the body, but as mentioned above, JNJ0966 is currently only used for scientific research[12]. Similarly, MMP-9-IN-1, as a specific targeted drug against MMP9, has not been put into clinical use on a large scale due to a series of defects such as respiratory system inhibition[7, 11]. Although the mechanism of action of MMP9 in tumor progression is relatively clear, the application of existing drugs is not satisfactory. Therefore, it is necessary to use various cell biology experiments and other methods to further screen and develop new drugs targeting MMP-9 in the future.

We virtually screened 1,752,844 small-molecule compounds in the natural source compound and synthetic drug database. By constructing a pharmacophore model, we screened 230 small-molecule compounds with the same pharmacophore, and then constructed a pharmacophore model. The machine learning model further screened 49 small molecule compounds with high binding affinity to MMP9 and pooled them for further study.

ADME and toxicity prediction results indicated that CEMBL82047 and CEMBL381163 were identified as having good water solubility, absorption level and plasma protein binding properties, no hepatotoxicity and toxicity, less Ames mutagenicity, rodent carcinogenicity and developmental toxicity potential ideal lead compound. Then, we further performed docking analysis and the results showed that CEMBL82047 and CEMBL381163 had higher binding affinity to MMP9 than JNJ0966 and MMP-9-IN-1. Since these two compounds have more chemical bonds with MMP9 than JNJ0966 and MMP-9-IN-1, they have higher interaction force and more stable binding with MMP9, which may enhance their inhibition of MMP9, thereby improving tumor killing effect. Finally, we conducted molecular dynamics simulation and results showed that the potential energy and RMSD of these complexes reached a steady state over time, indicating that the two complexes can remain stable in natural environments.

Although this study was carefully designed and calculated with precision, we still acknowledge that there are still some limitations of this study. Since the research on the two small-molecule compounds we finally screened is not available at present, we cannot buy them for further tests. Therefore, we have not

conducted cell, and animal experiments. Besides, drug safety indicators such as maximum tolerated dose (MTD) and aerobic biodegradability (AB) have not been evaluated.

In conclusion, MMP9 is highly expressed in most cancers. Higher MMP9 expression in GMBLGG, KIPAN, UVM, LGG, ACC, and LIHC means poorer survival and tumor progression. For GMBLGG, KIPAN, UVM, LGG, ACC, and LIHC, higher MMP9 expression means more infiltration of immune cells such as macrophages and regulatory T cells, and more RNA modifications. For UVM, LGG, ACC, and LIHC, higher MMP9 expression means that the tumor has a higher tumor mutation burden. There are 39 candidate inhibitors against MMP9 were screened by ligand-based pharmacophore model and machine learning model. CHEMBL82047 and CHEMBL381163 have good water solubility, absorption levels, and plasma protein binding properties, as well as low Ames mutagenicity, rodent carcinogenicity, and developmental toxicity potential, without hepatotoxicity and toxicity. They have higher binding affinity to proteins and are stable in the natural environment. CEMBL82047 and CEMBL381163 are promising MMP9 inhibitor potential drugs.

Declarations

Author Contribution

Bo Wu, Dong Zhu designed experiments; Ming Li, Xinyu Wang and Chunzheng Chen wrote manuscript; Qifan Yang, Pengcheng Zhou, and Shiyuan Lu carried out data analysis; Liangxin Lin and Yanwei Liu drew plots.

Funding

This research was supported by the National Natural Science Foundation of China (Grant No. 12072129) and the Talents Reserve Program of First Hospital of Jilin University (Grant No. 04028190001).

Data availability statement

The unified and standardized pan-cancer dataset was downloaded from the UCSC (https://xenabrowser.net/) database: TCGA Pan-Cancer (PANCAN, N=10535, G=60499). The human MMP9 receptors with different ligands can be deposited in the Worldwide Protein Data Bank (PDB DOI: 10.2210/pdb2OW0/pdb, PDB DOI: 10.2210/pdb2OW1/pdb, PDB DOI: 10.2210/pdb4H3X/pdb, PDB DOI: 10.2210/pdb4WZV/pdb). The datasets used and analyzed during the current study available from the corresponding author on reasonable request.

Conflict of Interest

There is no conflict of interests.

References

1. Hossain SMM, Khatun L, Ray S, Mukhopadhyay A. Pan-cancer classification by regularized multi-task learning. Sci Rep. 2021;11(1):24252. doi:10.1038/s41598-021-03554-8.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015;65(2):87-108. doi:10.3322/caac.21262.
3. Hu J, Xu J, Feng X, Li Y, Hua F, Xu G. Differential Expression of the TLR4 Gene in Pan-Cancer and Its Related Mechanism. Front Cell Dev Biol. 2021;9:700661. doi:10.3389/fcell.2021.700661.
4. Roche L, Danieli C, Belot A, Grosclaude P, Bouvier AM, Velten M et al. Cancer net survival on registry data: use of the new unbiased Pohar-Perme estimator and magnitude of the bias with the classical methods. Int J Cancer. 2013;132(10):2359-69. doi:10.1002/ijc.27830.
5. Liu N, Wang X, Wu H, Lv X, Xie H, Guo Z et al. Computational study of effective matrix metalloproteinase 9 (MMP9) targeting natural inhibitors. Aging (Albany NY). 2021;13(19):22867-82. doi:10.18632/aging.203581.
6. Huang H. Matrix Metalloproteinase-9 (MMP-9) as a Cancer Biomarker and MMP-9 Biosensors: Recent Advances. Sensors (Basel). 2018;18(10). doi:10.3390/s18103249.
7. Liu Z, Li L, Yang Z, Luo W, Li X, Yang H et al. Increased expression of MMP9 is correlated with poor prognosis of nasopharyngeal carcinoma. BMC Cancer. 2010;10:270. doi:10.1186/1471-2407-10-270.
8. Owyong M, Chou J, van den Bijgaart RJ, Kong N, Efe G, Maynard C et al. MMP9 modulates the metastatic cascade and immune landscape for breast cancer anti-metastatic therapy. Life Sci Alliance. 2019;2(6). doi:10.26508/lsa.201800226.
9. Tamura Y, Watanabe F, Nakatani T, Yasui K, Fuji M, Komurasaki T et al. Highly selective and orally active inhibitors of type IV collagenase (MMP-9 and MMP-2): N-sulfonylamino acid derivatives. J Med Chem. 1998;41(4):640-9. doi:10.1021/jm9707582.
10. Dufour A, Sampson NS, Li J, Kuscu C, Rizzo RC, Deleon JL et al. Small-molecule anticancer compounds selectively target the hemopexin domain of matrix metalloproteinase-9. Cancer Res. 2011;71(14):4977-88. doi:10.1158/0008-5472.Can-10-4552.
11. Song Z, Wang J, Su Q, Luan M, Chen X, Xu X. The role of MMP-2 and MMP-9 in the metastasis and development of hypopharyngeal carcinoma. Braz J Otorhinolaryngol. 2021;87(5):521-8. doi:10.1016/j.bjorl.2019.10.009.
12. Scannevin RH, Alexander R, Haarlander TM, Burke SL, Singer M, Huo C et al. Discovery of a highly selective chemical inhibitor of matrix metalloproteinase-9 (MMP-9) that allosterically inhibits zymogen activation. J Biol Chem. 2017;292(43):17963-74. doi:10.1074/jbc.M117.806075.
13. Ravichandran S, Singh N, Donnelly D, Migliore M, Johnson P, Fishwick C et al. Pharmacophore model of the quercetin binding site of the SIRT6 protein. J Mol Graph Model. 2014;49:38-46. doi:10.1016/j.jmgm.2014.01.004.
14. Vaidyanathan J, Vaidyanathan TK, Ravichandran S. Computer simulated screening of dentin bonding primer monomers through analysis of their chemical functions and their spatial 3D alignment. J Biomed Mater Res B Appl Biomater. 2009;88(2):447-57. doi:10.1002/jbm.b.31134.

15. Pascual R, Almansa C, Plata-Salamán C, Vela JM. A New Pharmacophore Model for the Design of Sigma-1 Ligands Validated on a Large Experimental Dataset. Front Pharmacol. 2019;10:519. doi:10.3389/fphar.2019.00519.
16. Lindvall M, McBride C, McKenna M, Gesner TG, Yabannavar A, Wong K et al. 3D Pharmacophore Model-Assisted Discovery of Novel CDC7 Inhibitors. ACS Med Chem Lett. 2011;2(10):720-3. doi:10.1021/ml200029w.
17. Auslander N, Gussow AB, Koonin EV. Incorporating Machine Learning into Established Bioinformatics Frameworks. Int J Mol Sci. 2021;22(6). doi:10.3390/ijms22062903.
18. Taniguchi H, Sato H, Shirakawa T. A machine learning model with human cognitive biases capable of learning from small and biased datasets. Sci Rep. 2018;8(1):7397. doi:10.1038/s41598-018-25679-z.
19. Deo RC. Machine Learning in Medicine. Circulation. 2015;132(20):1920-30. doi:10.1161/circulationaha.115.001593.
20. Fu T, Dai LJ, Wu SY, Xiao Y, Ma D, Jiang YZ et al. Spatial architecture of the immune microenvironment orchestrates tumor immunity and therapeutic response. J Hematol Oncol. 2021;14(1):98. doi:10.1186/s13045-021-01103-4.
21. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018;24(5):541-50. doi:10.1038/s41591-018-0014-x.
22. Siveen KS, Kuttan G. Role of macrophages in tumour progression. Immunol Lett. 2009;123(2):97-102. doi:10.1016/j.imlet.2009.02.011.
23. Lewis CE, Pollard JW. Distinct role of macrophages in different tumor microenvironments. Cancer Res. 2006;66(2):605-12. doi:10.1158/0008-5472.Can-05-4005.
24. Ugel S, Canè S, De Sanctis F, Bronte V. Monocytes in the Tumor Microenvironment. Annu Rev Pathol. 2021;16:93-122. doi:10.1146/annurev-pathmechdis-012418-013058.
25. Bazewicz CG, Dinavahi SS, Schell TD, Robertson GP. Aldehyde dehydrogenase in regulatory T-cell development, immunity and cancer. Immunology. 2019;156(1):47-55. doi:10.1111/imm.13016.
26. Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS, Novoa EM. The RNA modification landscape in human disease. Rna. 2017;23(12):1754-69. doi:10.1261/rna.063503.117.
27. Gao L, Chen R, Sugimoto M, Mizuta M, Kishimoto Y, Omori K. The Impact of m1A Methylation Modification Patterns on Tumor Immune Microenvironment and Prognosis in Oral Squamous Cell Carcinoma. Int J Mol Sci. 2021;22(19). doi:10.3390/ijms221910302.
28. Chen Z, Qi M, Shen B, Luo G, Wu Y, Li J et al. Transfer RNA demethylase ALKBH3 promotes cancer progression via induction of tRNA-derived small RNAs. Nucleic Acids Res. 2019;47(5):2533-45. doi:10.1093/nar/gky1250.
29. Wang Y, Wang J, Li X, Xiong X, Wang J, Zhou Z et al. N(1)-methyladenosine methylation in tRNA drives liver tumourigenesis by regulating cholesterol metabolism. Nat Commun. 2021;12(1):6314. doi:10.1038/s41467-021-26718-6.

30. Lai SC, Su YT, Chi CC, Kuo YC, Lee KF, Wu YC et al. DNMT3b/OCT4 expression confers sorafenib resistance and poor prognosis of hepatocellular carcinoma through IL-6/STAT3 regulation. J Exp Clin Cancer Res. 2019;38(1):474. doi:10.1186/s13046-019-1442-2.
31. Joseph C, Alsaleem M, Orah N, Narasimha PL, Miligy IM, Kurozumi S et al. Elevated MMP9 expression in breast cancer is a predictor of shorter patient survival. Breast Cancer Res Treat. 2020;182(2):267-82. doi:10.1007/s10549-020-05670-x.
32. Xue Q, Cao L, Chen XY, Zhao J, Gao L, Li SZ et al. High expression of MMP9 in glioma affects cell proliferation and is associated with patient survival rates. Oncol Lett. 2017;13(3):1325-30. doi:10.3892/ol.2017.5567.
33. Niu H, Li F, Wang Q, Ye Z, Chen Q, Lin Y. High expression level of MMP9 is associated with poor prognosis in patients with clear cell renal carcinoma. PeerJ. 2018;6:e5050. doi:10.7717/peerj.5050.
34. Zarkesh M, Zadeh-Vakili A, Akbarzadeh M, Fanaei SA, Hedayati M, Azizi F. The role of matrix metalloproteinase-9 as a prognostic biomarker in papillary thyroid cancer. BMC Cancer. 2018;18(1):1199. doi:10.1186/s12885-018-5112-0.
35. Li C, Jiang P, Wei S, Xu X, Wang J. Regulatory T cells in tumor microenvironment: new mechanisms, potential therapeutic strategies and future prospects. Mol Cancer. 2020;19(1):116. doi:10.1186/s12943-020-01234-1.
36. Togashi Y, Shitara K, Nishikawa H. Regulatory T cells in cancer immunosuppression - implications for anticancer therapy. Nat Rev Clin Oncol. 2019;16(6):356-71. doi:10.1038/s41571-019-0175-7.
37. Jardim DL, Goodman A, de Melo Gagliato D, Kurzrock R. The Challenges of Tumor Mutational Burden as an Immunotherapy Biomarker. Cancer Cell. 2021;39(2):154-73. doi:10.1016/j.ccell.2020.10.001.

Figures

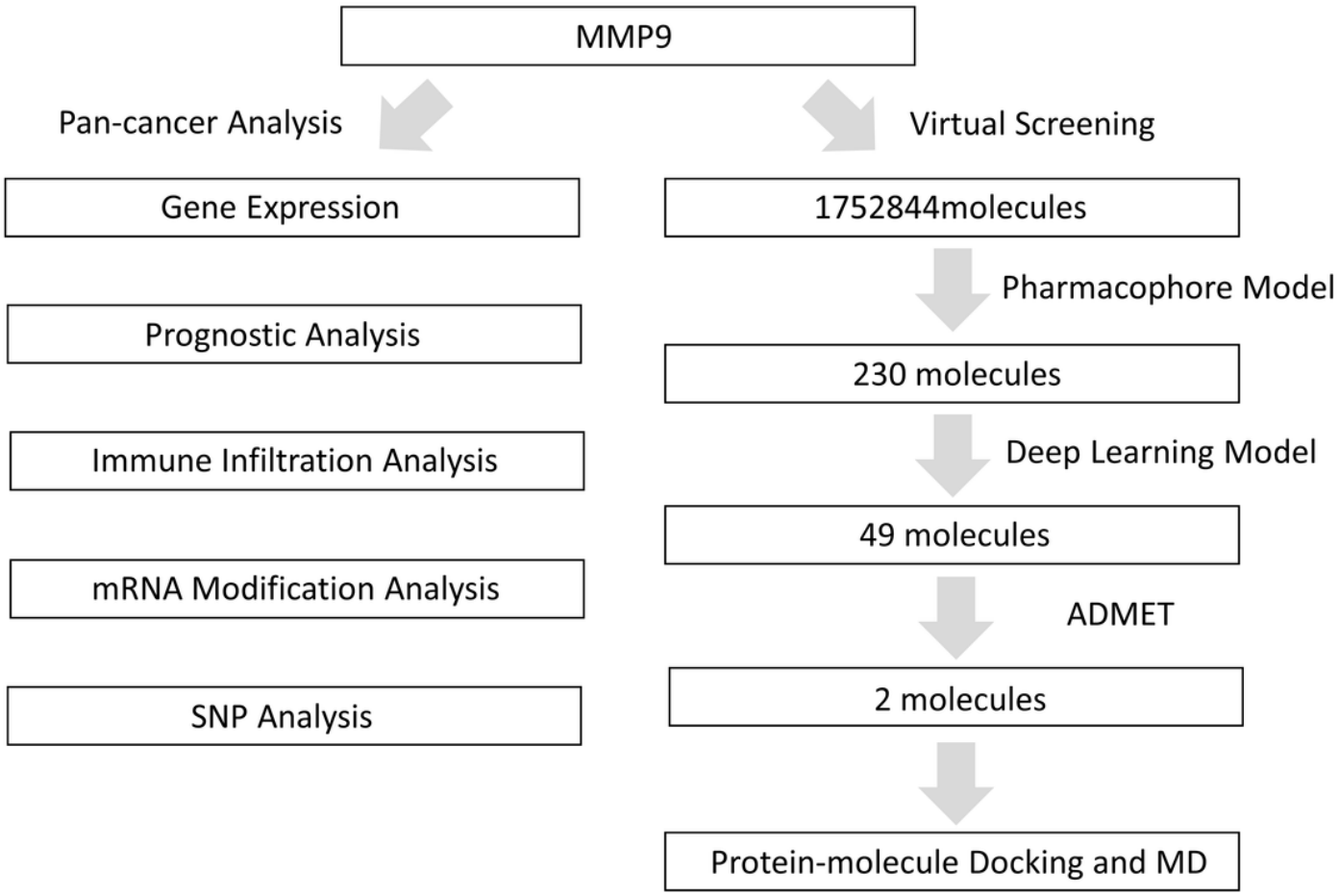

Figure 1

Mind map of this study.

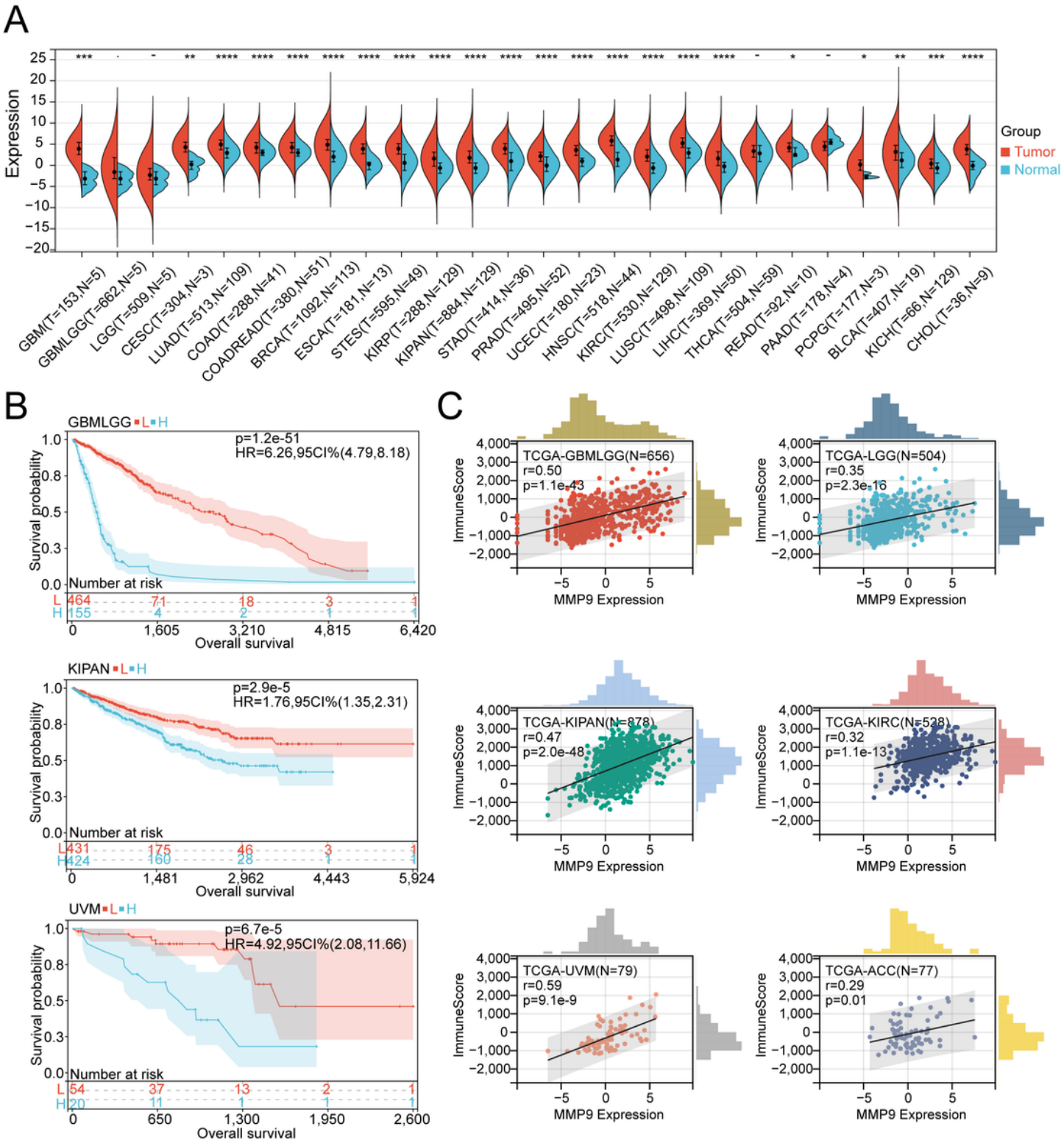

Figure 2

Pan-cancer analysis of MMP9 expression. (A) Differential expression of MMP9 between tumor and normal tissues in pan-cancer analysis. MMP9 expression correlates with overall survival time (OS). (B) Survival curves of MMP9 expression in GBMLGG, KIPAN and UVM. L represents low expression of MMP9 group, H represents high expression of MMP9 group. (C) Pan-cancer cohort (GBMLGG, KICH, KIRC, KIRP, KIPAN and UVM). Correlation between MMP9 expression and immune scores.

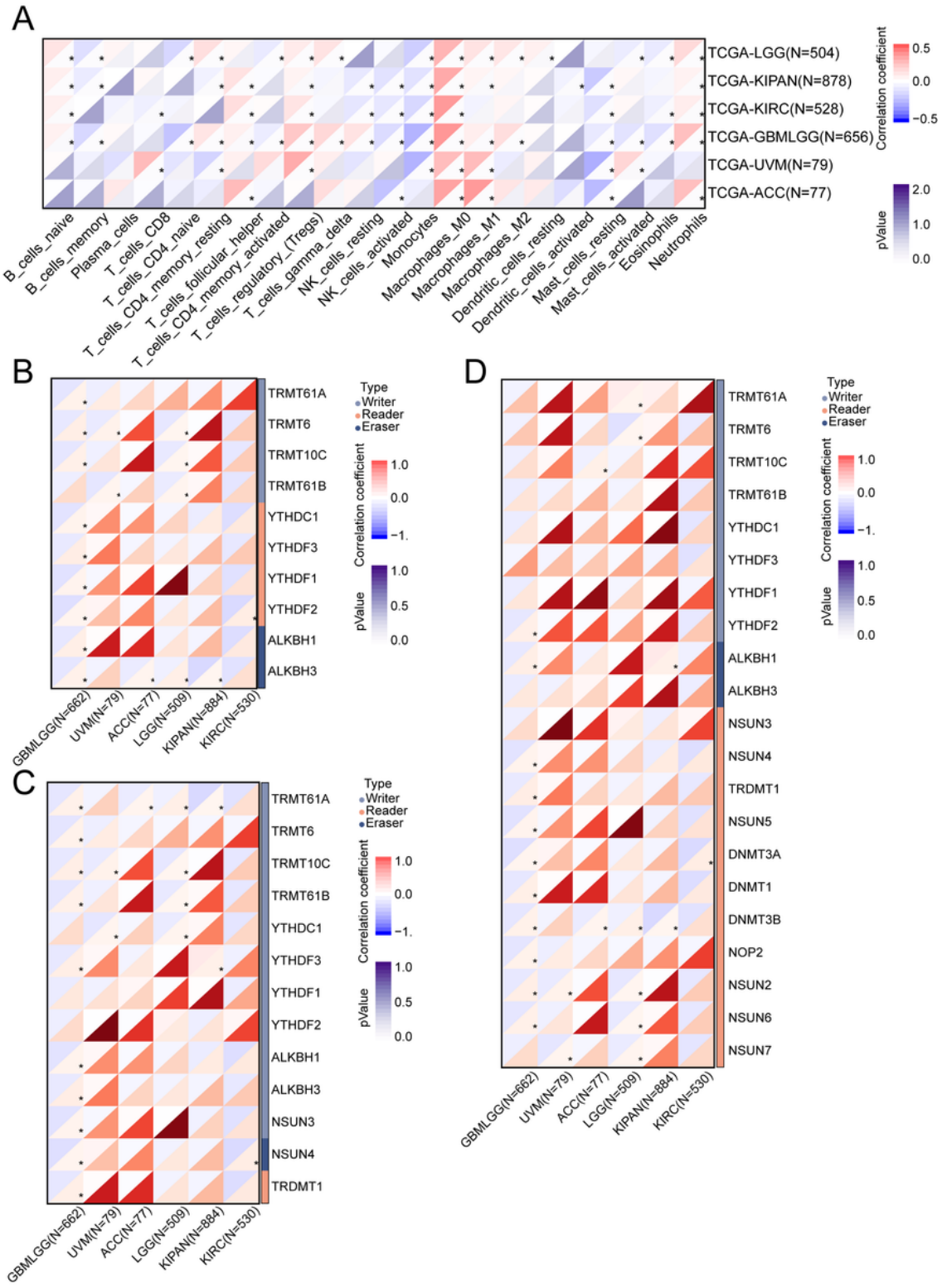

Figure 3

(A) Pan-cancer cohort (GBMLGG, KICH, KIRC, KIRP, KIPAN and UVM). Analysis of the relationship between MMP9 expression and immune cell infiltration. (B-D) Correlation between m1A, m5C and m6A mRNA modification genes and the expression of MMP9 in Pan-cancer cohort (GBMLGG, KICH, KIRC, KIRP, KIPAN and UVM).

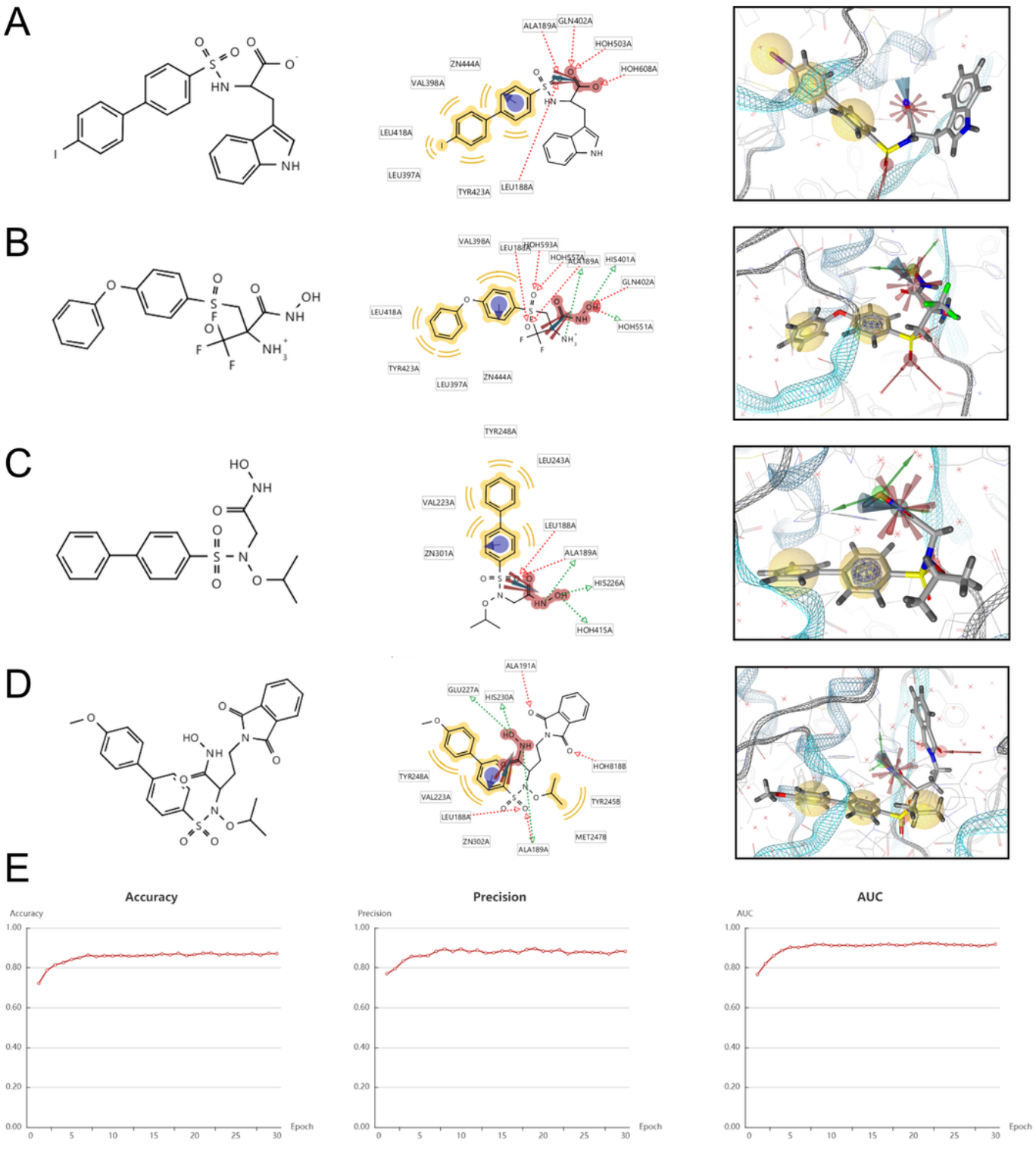

Figure 4

Chemical structure formula and pharmacophore analysis of (A) 2OW0 (B) 2OW1 (C) 4H3X and (D) 4WZV. Chemical features of the co-crystal structures were analyzed for summarizing common features. Red arrows indicate hydrogen bond acceptors, green arrows indicate hydrogen bond donors and yellow spheres indicate hydrophobes. (E) Evaluation index of deep learning model.

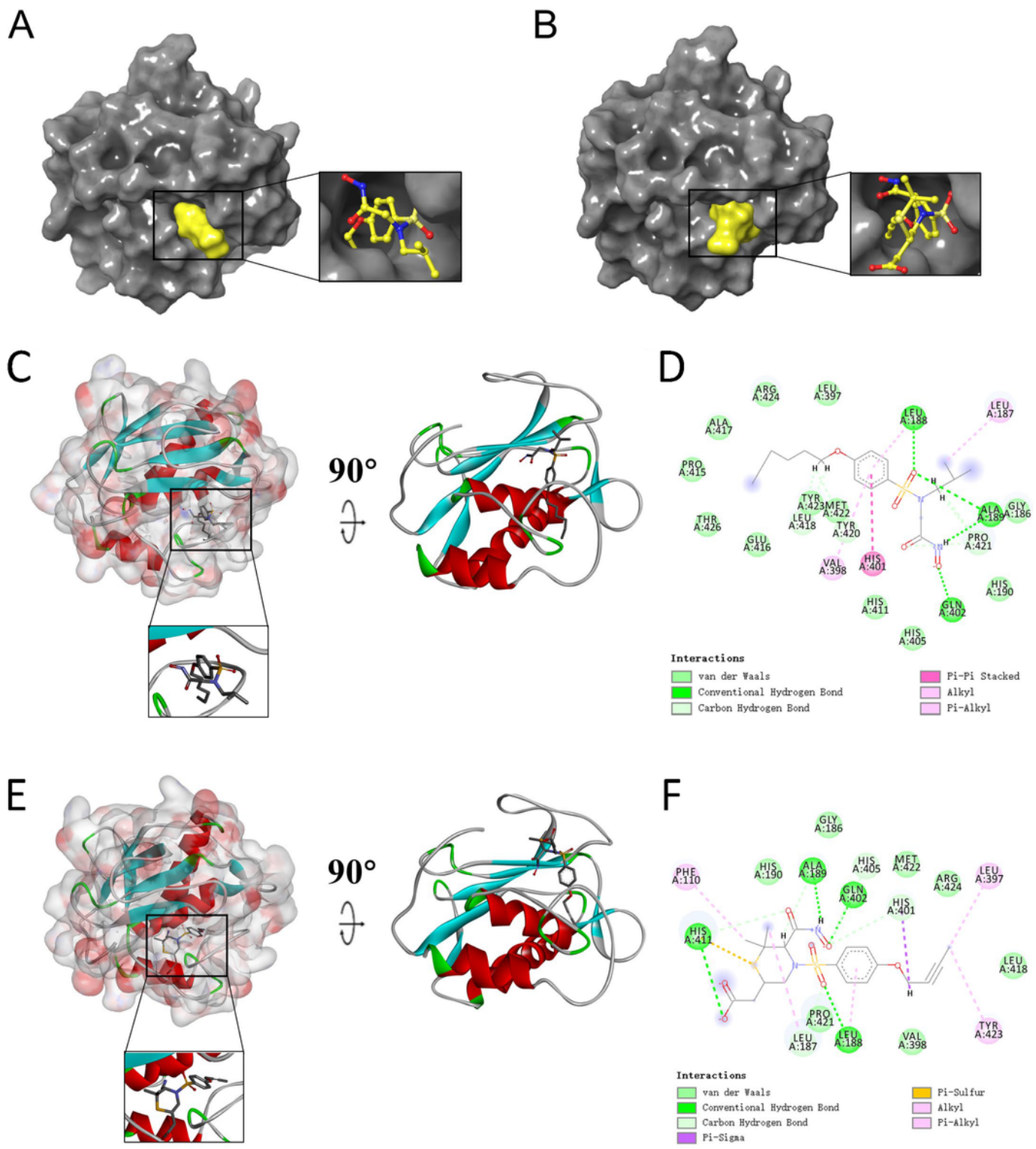

Figure 5

Schematic drawing of interactions between ligands and MMP9. (A) Ligand interaction diagram of CHEMBL82047-MMP9 complex. (B) Ligand interaction diagram of CHEMBL381163-MMP9 complex. (C) CHEMBL82047-MMP9 complex. (D) CHEMBL381163-MMP9 complex. Schematic of intermolecular interaction of the predicted binding modes of (E) CHEMBL82047 with MMP9, (F) CHEMBL381163 with MMP9.

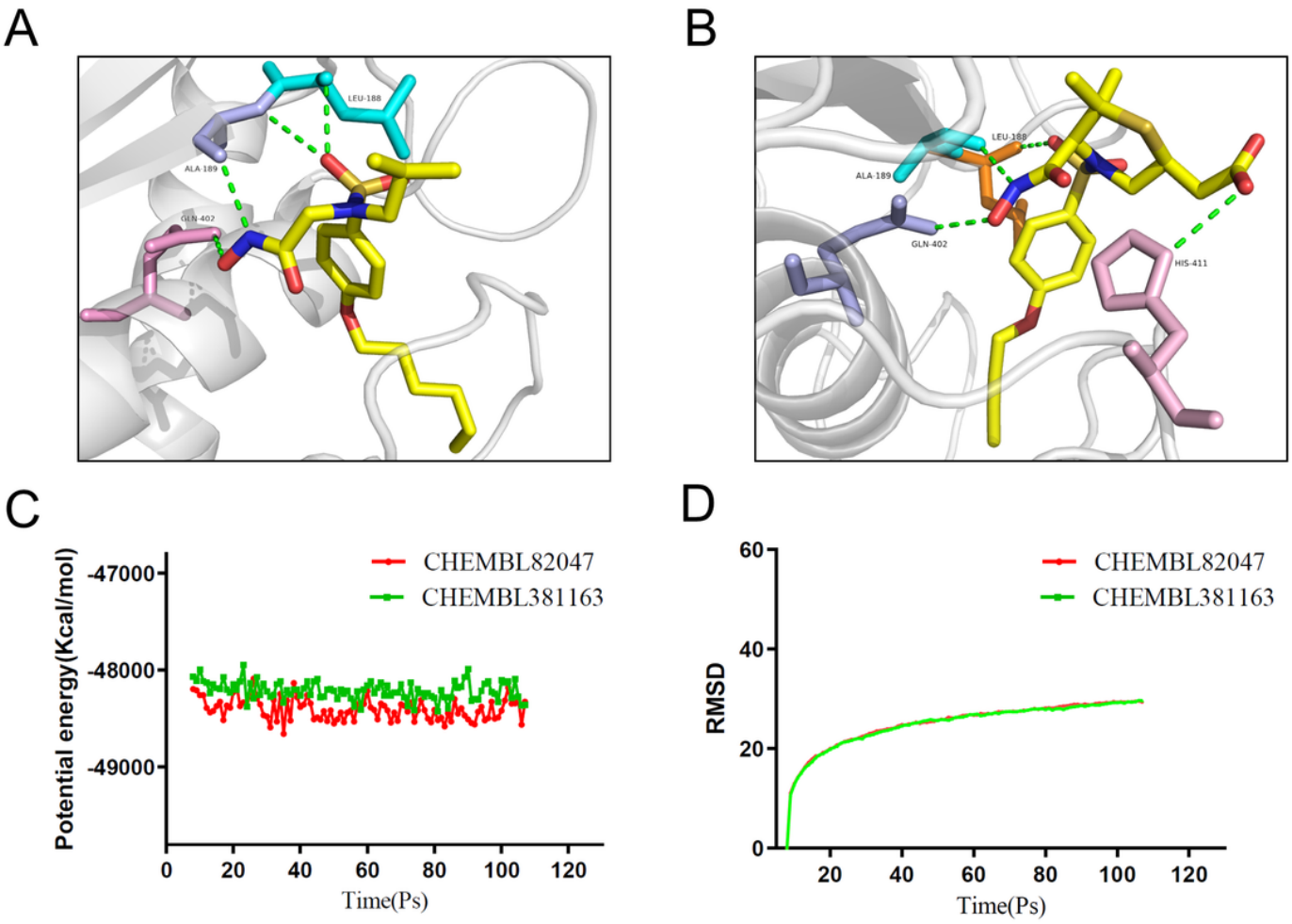

Figure 6

Schematic of intermolecular interaction of the predicted binding modes of (A) CHEMBL82047 with MMP9, and (B) CHEMBL381163 with MMP9. Results of molecular dynamics simulation of the compounds CHEMBL82047-MMP9 complex and CHEMBL381163-MMP9 complex. (C) Potential energy, Average backbone root-mean-square deviation. (D) RMSD, root-mean-square deviation.

Supplementary Files

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

- SupplementaryFigure1.tif
- SupplementaryFigure2.tif
- SupplementaryFigure3.tif
- SupplementaryFigure4.tif
- SupplementaryFigure5.tif
- SupplementaryTable1.docx

- SupplementaryTable2.docx
- SupplementaryFigureandTableLegends.docx