AI-Guided Design of Self-Assembled Flavonoid-Cisplatin Nanoparticles Enhances Triple-Negative Breast Cancer Therapy via Three Cell-Death Pathways

Marmar Tuohan [1, 2], Geng Hu [3], Di Zhu [4], Bo Hu [1, 2], Mengyue Ni [1, 2], Liujia Chan [1, 2], Yu Lu [1,2,*] and Yuji Wang [1,2,*]

[1]Department of Medicinal Chemistry, College of Pharmaceutical Sciences of Capital Medical University, Beijing 100069, P. R. China;

[2]Beijing Key Laboratory of Intelligent Clinical Data and AI Model-driven Innovative Drug Development; Beijing Key Laboratory of Drug Innovation for Neuro-Oncology, Beijing Engineering Research Center of Targeted Drugs and Cell Therapy for CNS Tumors, Laboratory for Clinical Medicine, Capital Medical University, Beijing 100069, P. R. China;

[3]School of Engineering Medicine of Beihang University and National Industry Education Integration Platform (Medical Engineering Cross), Beihang University, Beijing 100191.

[4]Department of Pharmacy, Xuanwu Hospital of Capital Medical University, Beijing, 100053, P. R. China;

*Correspondence:

Yu Lu (luyu@mail.ccmu.edu.cn);

Yuji Wang (wangyuji@ccmu.edu.cn)

Materials and Methods

Methods

Training dataset, data processing, and model design of LMCDPS

The training dataset for LMCDPS was established using more than 70,000 natural compounds collected from public databases including TCMBank and DrugBank. A multidimensional knowledge base was constructed, covering 2D/3D molecular structures, natural sources, and reported bioactivities of these compounds. Molecular fingerprints with 2048 dimensions were generated using RDKit, an open-source cheminformatics toolkit, to characterize molecular properties including atomic type, charge, chirality, hybridization, aromaticity, bond type, and conjugation. Compounds associated with triple-negative breast cancer (TNBC), immune regulation, or drug self-assembly potential were included, while those without complete molecular structure annotations or available bioactivity data were excluded. A 5-fold cross-validation strategy was applied to split the training and validation sets, and an independent experimental set composed of flavonoid-cisplatin combinations was used as the test set. Since LMCDPS was designed as a function-oriented screening platform rather than a conventional binary classification model, the generalizability of the model was mainly evaluated based on experimental consistency.

Pathway-protein contextual encoding

Pathway annotations were obtained from DrugBank, KEGG and Gene Ontology (GO) databases, and only pathway–protein interactions supported by experimental evidence were retained. Pathways relevant to triple-negative breast cancer (TNBC) were selected

based on their association with TNBC biology, molecular self-assembly and cisplatin synergy, as well as direct protein-pathway regulatory relationships. These annotations were incorporated into the model as contextual biological features and processed jointly with molecular fingerprint representations through the Transformer attention mechanism.

LMCDPS-based compound screening

To identify natural compounds capable of self-assembling with cisplatin and exhibiting synergistic antitumor activity in TNBC, we developed a multi-objective scoring framework within the LMCDPS platform. The scoring system incorporated four biological endpoints: breast cancer target relevance, apoptosis regulation, immune modulation, and drug-assembly compatibility. These endpoints were further summarized into two functional dimensions, namely a self-assembly score and a synergistic activity score.

All endpoints were normalized to a 0-1 scale based on molecular structural features and pathway-protein contextual relevance. Candidate compounds showing high scores in both dimensions were retained for further analysis. For these candidates, scores were aggregated using a weighted strategy emphasizing TNBC-related pathways and self-assembly potential. Molecular similarity calculated from 2048-dimensional RDKit fingerprints was incorporated into the weighting scheme to refine ranking. Based on the final composite score, the top 12 candidate compounds were selected for further evaluation.

Molecular Dynamics Simulation Module (LMCDPS-MD)

To validate whether the predicted drug combinations exhibit potential conformational synergy or self-assembly behavior, we developed the LMCDPS-MD module. Candidate molecules are subjected to molecular dynamics (MD) simulations using AmberTools or GROMACS. Then, ten copies of the molecule were randomly placed in a 6 × 6 × 6 nm^3 periodic water box. Periodic boundary conditions were applied throughout the simulation. The cutoff distance for non-bonded interactions was 12 Å, and long-range electrostatic interactions were treated using the particle mesh Ewald (PME) method. Hydrogen bonds were constrained using the LINCS algorithm, with an integration timestep of 2 fs. Temperature was maintained at 310 K using a V-rescale thermostat (coupling constant 1.0 ps), and pressure was controlled at 1 bar using the Berendsen barostat. After system construction, 50,000 steps of energy minimization were performed, followed by 100 ps NVT equilibration. Subsequently, a 100 ns production simulation under the NPT ensemble was conducted, and the resulting trajectories were used for structural analysis. Conformational dynamics and stability are assessed through metrics such as root-mean-square deviation (RMSD) and hydrogen bonding analysis:

$$\mathrm{RMSD}(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\mathbf{r}_i(t) - \mathbf{r}_i(0)\right|^2}$$

The simulations additionally yield bonding energy profiles, hydrophobic/hydrophilic contact matrices, and assembly propensity maps, providing a comprehensive assessment of the structural stability and potential synergistic mechanisms within the drug combination system. A single simulation trajectory was analyzed for each system,

and structural stability was evaluated over the full 100 ns simulation timescale.

Preparation of Persimmon Leaf Flavonoids (PLF)

Take 600 g of persimmon leaves, wash and cut them (size ≤ 4 × 4 cm), soak in distilled water at 80-95 °C for 60 min, collect light green hot water, freeze-dry after reducing pressure and obtained 115 g of freeze-dried powder. The concentration of persimmon leaf extract (PLE) before lyophilization was calculated to be 2.0 mg/mL. PLE was extracted by petroleum ether to remove non-polar impurities, further purified by HP-20 microporous resin, and collected 70% ethanol eluate and lyophilized it to obtain total flavonoid extract (PLF). The total flavonoids content was determined by $NaNO_2$-$Al(NO_3)_3$-NaOH colorimetric method. The flavonoid content is 835.28 μg/mg (83.5%).

LC–ESI–MS/MS analyses

LC-MS analysis was performed on Waters SYNAPT G2-Si HDMS (Waters Corporation, Milford, USA) mass spectrometer equipped with an electrospray ionization source, with an analytical column (ACQUITY UPLC® HSS T3 2.1 × 100, 1.8 μm). The MS analysis was performed in negative ion mode under the following conditions: cone gas (N2), 50 L/h; desolvation gas, 800 L/h; source temperature, 120 °C; desolvation temperature, 450 °C; atomizer gas pressure (N2), 6.0 Bar; capillary voltage, 2.0 kV; sampling cone, 40 V; and source offset, 80 V. The sample was analyzed in MSE continuum mode and the scanning range was 100-2000 Da. A collision energy ramp was set at 20-60 V in the function 2-high energy tunnel. For accurate mass acquisition,

an external reference (0.2 ng/μL Leucine-Enkephalin) was used with a lock spray interface at a flow rate of 10 μL/min, generating reference ions at 554.2615 Da in negative ion mode.

Using the mobile phase consisting of acetonitrile (B) and 0.1% formic acid in water (A), gradient elution was performed as follows: 0.00-3.00 min, 2%-2% B; 3.00-15.00 min, 2%-34% B; 15.00-25.00 min, 34%-34% B; 25.00-50.00 min, 34%-60% B; 50.00-55.00 min, 60%-100% B; 55.00-59.00 min, 100%-100% B; The column temperature, injection volume, and flow rate were 40 °C, 2 μL, and 0.3 ml/min, respectively. The MS data were processed through MassLynx v 4.1 (Waters Corporation, Milford, MA, USA) software, which provided a list of possible elemental formulae. The accuracy for confirmation of the compounds was established on the basis of their error less than 10 ppm and MSn fragment matching.

Self-assembled properties

The particle size, zeta potential, and polydispersity index (PDI) of PLF-Cis NPs were measured using a Malvern Zetasizer Nano ZS particle analyzer. All measurements were performed in deionized water at room temperature; the pH and ionic strength were not adjusted. The morphology of the nanoparticles was observed using a JEM-2100 transmission electron microscope (TEM, Tokyo, Japan).

Cell cultures

All cell lines used in this study, including mouse mammary carcinoma cells (4T1), were obtained from a certified cell bank (Nanjing KeyGen Biotech Co., Ltd.) and routinely tested for mycoplasma contamination using a standard detection assay, with negative

results prior to use. Cells were cultured in DMEM supplemented with 10% fetal bovine serum (FBS, GIBCO, USA) and 1% penicillin–streptomycin in a humidified incubator at 37 °C with 5% CO_2.

MTT Assay

After 24 h of seeding 4T1 cells into 96-well plates, gradient concentrations of Cisplatin, PLF or PLF-Cis NPs were added to the medium to treat the cells for 24 h. To ensure comparability among groups, the administered doses of PLF-Cis NPs were normalized according to their equivalent cisplatin concentration. Subsequently, the MTT reagent was added for approximately 4 h, followed by absorbance measurement using a microplate reader. For inhibitor experiments, CQ (autophagy inhibitor), Fer-1 (ferroptosis inhibitor), Nec-1 (necroptosis inhibitor), and v-ZAD-fmk (apoptosis inhibitor) were added to the cells 1 h prior treatment. The concentrations of inhibitors were as previously described. Each group had 6 replicate wells, and each experiment was repeated three times.

Bliss Independence Model:

Based on the cell inhibition rates detected by the MTT assay, the measured inhibition rates for each group (E_A, E_B) were calculated. According to the Bliss independence model, the theoretical expected inhibition rate was calculated as $E_{expected}= E_A+E_B-(E_A \times E_B)$. Drug interaction was evaluated by calculating the synergy difference value $\Delta E = E_{observed} - E_{expected}$: $\Delta E > 0$ indicates a synergistic effect, $\Delta E = 0$ indicates an additive effect, and $\Delta E < 0$ indicates an antagonistic effect.

Colony formation assay

Colony formation assay was conducted as described previously. Briefly, 600 4T1 cells were seeded in each well of 6-well plates. Following the specified treatments, cells were cultured at 37 ℃ for 10–14 days to facilitate colony formation. The colonies were fixed with 4 % paraformaldehyde, stained with crystal violet, and then counted.

Apoptosis detection

After seeding 4T1 for 24 h, Cisplatin, PLF and PLF-Cis NPs were added to the medium for 24 h or 48 h incubation. Subsequently, all cells were collected and washed twice with PBS, then resuspended in Annexin-V Binding Buffer and stained with Annexin V-FITC/PI for 10 min. Stained cells were analyzed using flow cytometry.

Cellular uptake

We chose 4T1 cells to study the cellular uptake. To elucidate the internalization mechanism, cells were pretreated with specific endocytosis inhibitors (amiloride, chlorpromazine, and methyl-β-cyclodextrin) prior to incubation with PLF-Cis NPs or free cisplatin. The intracellular platinum content was then measured by ICP-MS according to the manufacturer's instructions. Prior to analysis, cell samples were digested with nitric acid to ensure complete dissolution. Rhodium (Rh) was employed as an internal standard to correct for instrumental drift and matrix effects, ensuring accurate quantification. The platinum content was normalized to cell number and expressed as ng per 10^6 cells.

Cell cycle analysis

After seeding 4T1 for 24 h, Cisplatin, PLF and PLF-Cis NPs were added to the medium for 24 h incubation. Following the treatment period, the cells were harvested and

washed with PBS. They were then fixed with 70% (v/v) ethanol at -20 ℃ for at least 24 h. the fixed cells were centrifuged and washed twice with PBS before being stained with propidium iodide dye, along with RNase, for 30 min in the dark at 37 ℃. Flow cytometry was employed to analyze the resulting data, which were independently measured three times.

Wound-healing assay

Cells were seeded in 6-well plates, and once they reached full confluency, uniform straight lines were drawn in the wells using a 200 μL pipette tip. Subsequently, Cisplatin, PLF and PLF-Cis NPs were added to the medium without FBS. Photographs were taken at 0 h and 24 h at the same locations *via* Zen Blue 3.1 Imaging System.

ROS generation assay

After incubating cells with Cisplatin, PLF and PLF-Cis NPs for 24 h or 48 h. After rinsing with PBS, the cells were labeled with DCFH-DA and subsequently detected by flow cytometry (BD LSRFortessa SORP, USA) and normalized to the number of cells in each sample to ensure accurate comparisons between treatment groups.

Western blot analysis

Following cell treatment, the cells were collected and lysed with RIPA buffer. The lysates were quantified using a BCA protein assay kit and normalized by adding PBS to adjust the protein concentrations. Protein samples were separated by 12.5% SDS-PAGE gel electrophoresis and transferred to PVDF membranes. The membranes were blocked with 5% skim milk for 1 h, washed three times with TBST, and then incubated overnight at 4 ℃ with antibodies. Subsequently, appropriate secondary antibodies were

added, and protein bands were detected using enhanced chemiluminescence (ECL).

Results

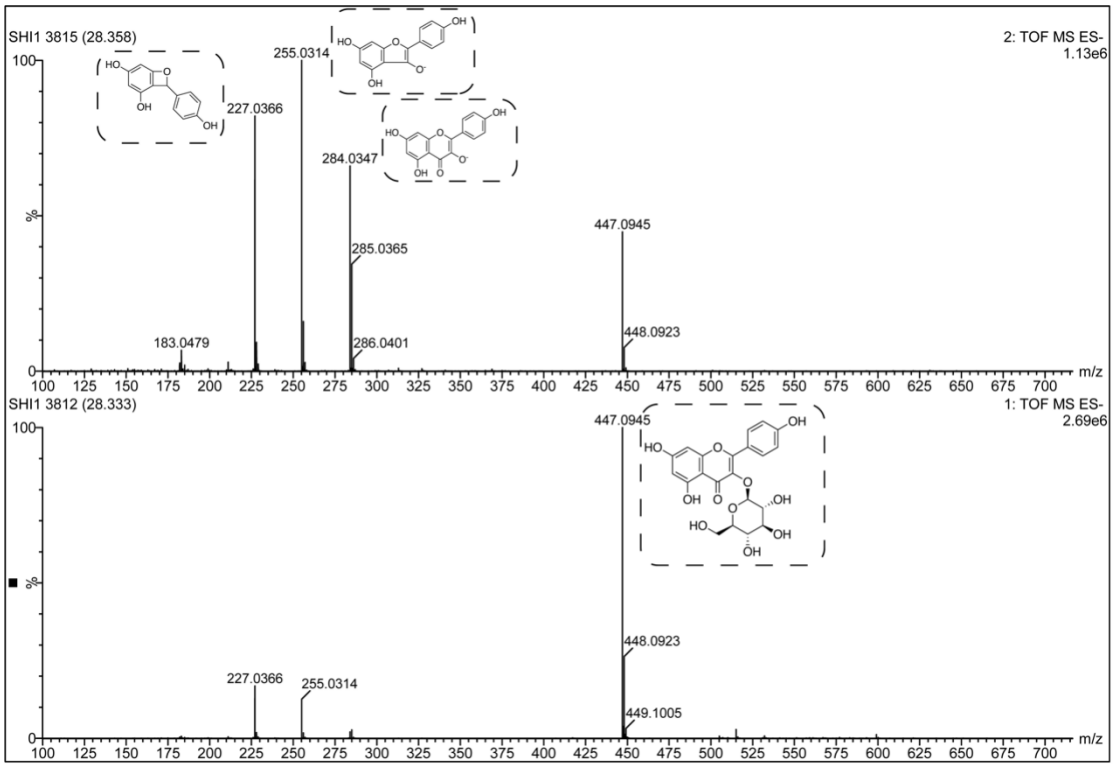


Figure S1. An example-based analysis of fragmentation peaks in LC-MS/MS.

Table S1. UPLC-ESI-MS/MS information of PLF

Peak	RT	MS	MS^n	Formula	ppm	Identify	Integrate
1	6.236	897.1915	711.1351/593.1284/ 407.0729/289.0687	$C_{45}H_{38}O_{20}$	4.1	Prodelphinidin C2; GC-C-C	98331
2	6.699	609.1236	423.0684/305.0686/ 177.0190	$C_{30}H_{26}O_{14}$	-1.3	Helichrysoside	101501
3	7.339	881.1881	407.0729/177.0190/ 289.0687/593.1284	$C_{45}H_{38}O_{19}$	-5.4	Catechin-(4alpha->8)-gallocatechin-(4alpha->8)-catechin	153500
4	7.616	609.1236	423.0684/305.0686/ 177.0190	$C_{30}H_{26}O_{14}$	-1.3	Helichrysoside	43344
5	7.815	865.2003	695.1442/407.0729/ 289.0687/243.0279	$C_{45}H_{38}O_{18}$	2.7	Procyanidin C1	18527
6	7.965	331.0673	125.0254/168.0091/ 177.0190/243.0279	$C_{13}H_{16}O_{10}$	2.4	beta-Glucogallin	75282
7	8.612	593.1284	407.0729/289.0687/ 177.0190	$C_{30}H_{26}O_{13}$	-1.9	Procyanidin	296009

8	9.067	305.0686	289.0761/177.0190/125.0254	$C_{15}H_{14}O_7$	8.2	Epigallocatechin	221726
9	9.255	593.1284	423.0684/305.0686/177.0190/125.0254	$C_{30}H_{26}O_{13}$	-1.9	Epigallocatechin-(4beta->8)-catechin	122356
10	9.7	897.1915	711.1351/593.1284/423.0684/289.0687/177.0190	$C_{45}H_{38}O_{20}$	4.1	Prodelphinidin C2	35610
11	10.183	881.1881	695.1328/407.0729/377.0150/289.0687/243.0279/177.0190	$C_{45}H_{38}O_{19}$	-5.4	Gallocatechin-(4alpha->8)-catechin-(4alpha->8)-catechin	37292
12	10.994	577.1362	407.0729/289.0687/125.0254	$C_{30}H_{26}O_{12}$	2.8	Procyanidin B1	302654
13	11.2	865.2003	695.1442/407.0729/289.0687/243.0279	$C_{45}H_{38}O_{18}$	2.7	Procyanidin C1	34565
14	11.762	401.1825	437.1597(+Cl)/447.1869(+$COOH^-$)	$C_{19}H_{30}O_9$	3.2	Spionoside A	83727
15	12.161	289.0687	245.0813/203.0701	$C_{15}H_{14}O_6$	-8.6	Epicatechin	366423
16	12.666	431.1545	204.0295/153.0201/	$C_{19}H_{28}O_{11}$	-1.9	Benzyl gentiobioside	34516
17	13.405	729.1462	577.1362/407.0729/289.0687/125.0254	$C_{37}H_{30}O_{16}$	0.8	Procyanidin B2 3'-O-gallate	25013
18	14.209	457.1317	323.1381/163.0427/119.0489	$C_{20}H_{26}O_{12}$	-6.3	cis-p-Coumaric acid 4-[apiosyl-(1->2)-glucoside]	99282
19	14.678	401.1476	167.0370/269.1050/437.1597(+Cl)/447.1869(+COOH-)	$C_{18}H_{26}O_{10}$	7	Benzyl O-[arabinofuranosyl-(1->6)-glucoside]	106561
20	15.403	517.2281	205.1225/153.0957/125.0254/553.2084/563.2368			Unknown	116415
21	16.833	415.1567	451.1387/461.1646			Unknown	103541
22	18.412	479.0873	316.0264/287.0232/271.0283	$C_{21}H_{20}O_{13}$	9.8	Myricetin 3-galactoside	551516
23	18.945	631.0887	479.0777/317.0287/271.0283/151.0065	$C_{28}H_{24}O_{17}$	-7.6	Myricetin 3-O-beta-D-galactoside 6''-O-gallate	21412
24	20.02	449.0727	316.0264/271.0283/137.0238	$C_{20}H_{18}O_{12}$	1.6	Myricetin 3-arabinoside	19953
25	21.71	463.089	300.0252/271.0283/255.0314/243.0279	$C_{21}H_{20}O_{12}$	2.8	Hyperoside	1126258

26	22.366	463.089	300.0252/271.0283/255.0314/243.0279	$C_{21}H_{20}O_{12}$	2.8	Isoquercetin	2091298
27	23.443	615.0981	301.0396/313.0602/271.0283/178.9999/151.0065	$C_{28}H_{24}O_{16}$	-0.8	2''-Galloylhyperin	2292035
28	24.3	493.0953	433.0789/315.0181/300.0252/271.0283/255.0314/243.0279	$C_{22}H_{22}O_{13}$	-5.9	Quercetagetin 3'-methylether 7-glucoside	24356
29	26.185	447.0945	284.0347/255.0314/227.0366	$C_{21}H_{20}O_{11}$	4	Trifolin	2167683
30	28.34	447.0945	284.0347/255.0314/227.0366	$C_{21}H_{20}O_{11}$	4	Astragalin	3506801
31	28.888	599.099	313.0602/285.0439/257.0438/229.0479	$C_{28}H_{24}O_{15}$	-7.8	Trifolin 6''-gallate	1727354
32	30.004	599.099	313.0602/285.0439/257.0438/229.0479	$C_{28}H_{24}O_{15}$	-7.8	Astragalin 6''-gallate	3316537
33	32.586	569.097	417.0808/285.0439/	$C_{27}H_{22}O_{14}$	6.9	Kaempferol 3-(2''-galloyl-alpha-L-arabinopyranoside)	43232
34	33.944	569.097	417.0808/285.0439/	$C_{27}H_{22}O_{14}$	6.9	Kaempferol 3-(2''-galloyl-alpha-L-arabinopyranoside)	36372
35	22.64	615.0981	301.0396/313.0602/271.0283/178.9999/	$C_{28}H_{24}O_{16}$	-0.8	2''-Galloylhyperin	646310

207

208

Table S2. Concentrations of PLFs

Compound	Linear Equation	R^2	Concentration(μg/mL)
Catechin	y = 206025x + 49308	0.997	18.8
Kaempferol	y = 280075x + 23183	0.995	-
Quercitrin	y = 31209x + 7286	0.998	0.33
Isoquercitrin	y = 62056x + 18910	0.997	**86.7**
Hyperoside	y = 1,125,348 x + 239,278	0.998	21.4
Astragalin	y = 1,143,091x + 256,623	0.998	**68.6**
Kaempferol-3-O-glucoside	y = 3,184,922 x + 132,405	0.999	16.8
Myricetin-3-O-glucoside	y = 1,387,059x + 3,903	0.999	13.6

Myricetin 3-arabinoside	y = 136487x + 27052	0.998	9.19
Procyanidin B1	y = 154443x + 3065	0.999	7.44
2''-Galloylhyperin	y = 1,290,731 x + 62,282	0.998	**53.4**

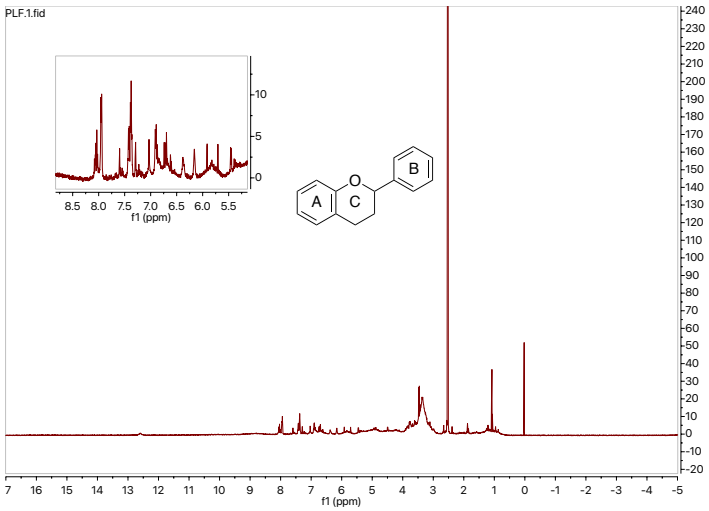


Figure S2. The ^{1}H NMR spectrum of PLF

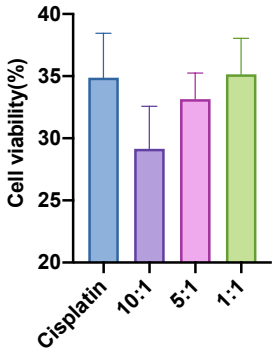


Fig S3. Cell viability of PLF and cisplatin at different ratios

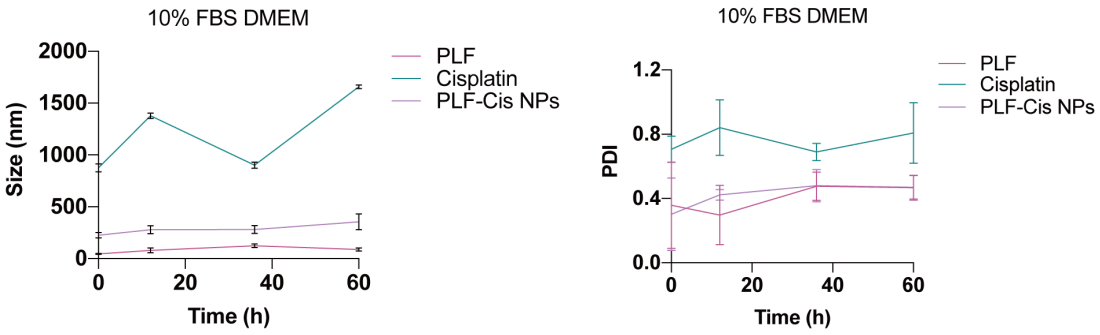


Figure S4. Nanoparticle Stability in DMEM culture medium containing 10% FBS over 60 h.

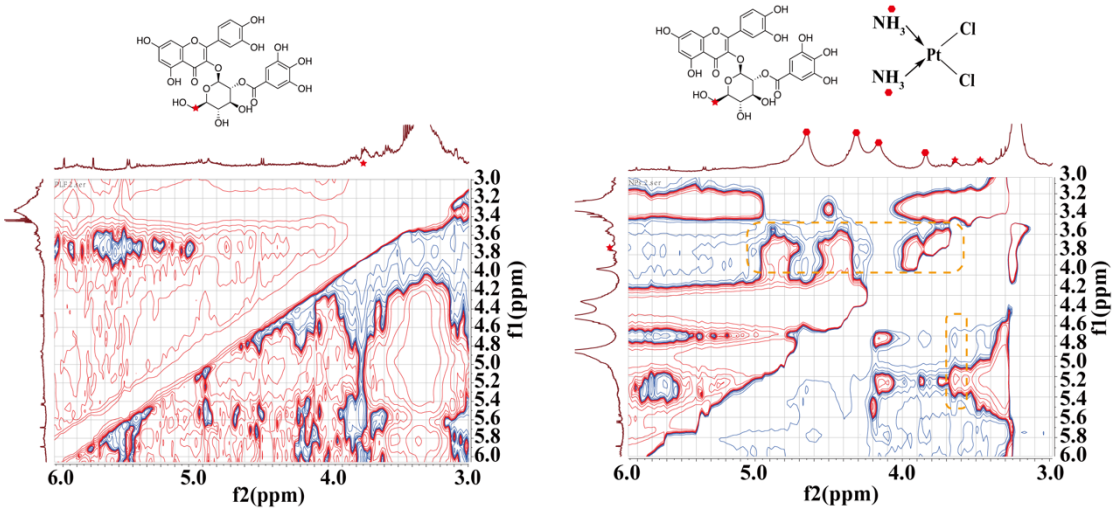


Figure S5. 2D-NOESY-NMR Results

Table S3 Particle Size, Polydispersity Index (PDI), and Zeta Potential of Different Formulations

	Cisplatin	PLF	Cisplatin+PLFs	Cisplatin+Astragalin	PLF-Cis NPs
Size(nm)	950.6±240.4	140.1±7.2	620.0±271.5	680.6±272.5	160.2±30.9
Zeta(mV)	+13.5±5.3	-17.3±3.0	-0.3±0.1	-13.0±1.1	-20.4±3.6
PDI	0.75±0.2	0.35±0.1	0.86±0.3	0.75±0.3	0.24±0.1

Table S4. Cytotoxicity of Cisplatin, PLF, PLF-Cis NPs by MTT assay (IC_{50}, μg mL^{-1}/48h).

Compound	Cell (IC_{50}/μg mL^{-1})		
	4T1	MCF-7	MDA-MB-231
Cisplatin	1.84±0.14	3.46±0.47	>30
PLF	>300	>300	>300
PLF-Cis NPs	0.64±0.023	1.71±0.08	17.46±2.28

Table S5. Cytotoxicity of Cisplatin, PLF-Cis NPs, PLF+Cisplatin by MTT assay (24 h).

Compound	Cell (IC_{50}/μM)
Cisplatin	12.7±1.03
PLF-Cis NPs	3.3±0.81
PLF+ Cisplatin	4.8±0.45

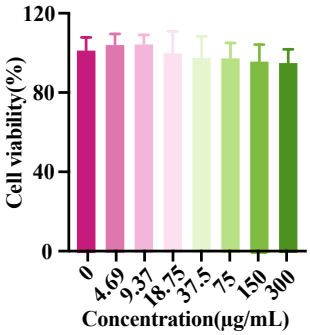


Figure S6. The cytotoxicity of PLF.

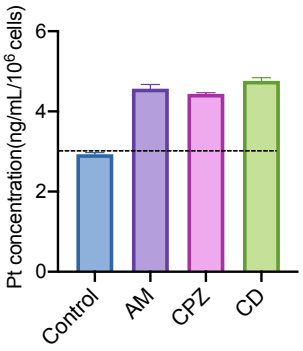


Fig. S7 Investigation of the internalization mechanism of PLF-Cis NPs.

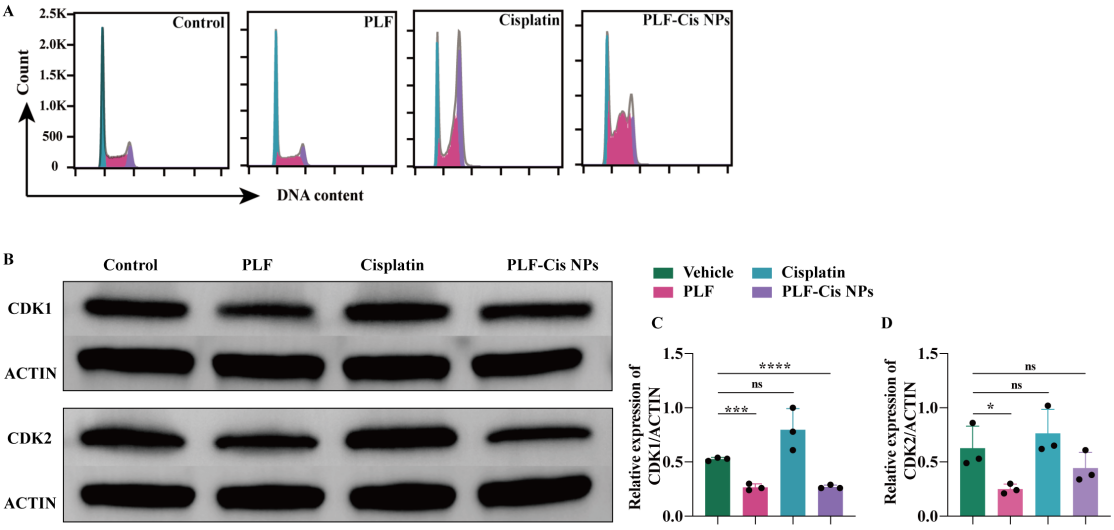


Figure S8. (A) Effect of PLF-Cis NPs on the cell cycle. (B-F) CDK1 and CDK2 expression determined by western blot.

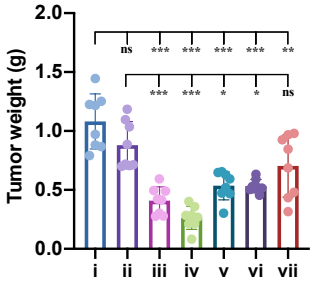


Fig. S9 Tumor weight of 4T1 orthotopic tumor-bearing mice

(i: vehicle; ii: cisplatin 1 mg/kg; iii: cisplatin 2 mg/kg; iv: PLF-Cis NPs-high; v: PLF-Cis NPs-medium; vi: PLF-Cis NPs-low; vii: PLF)

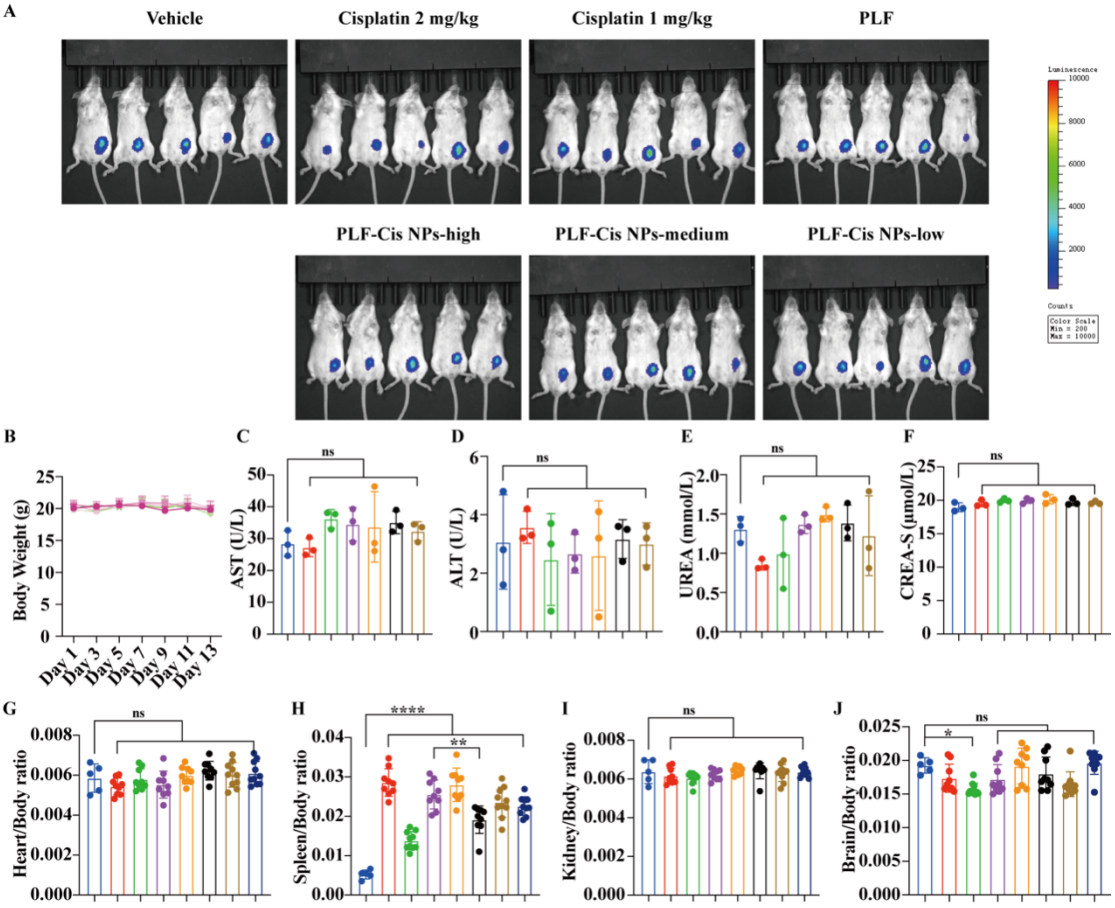
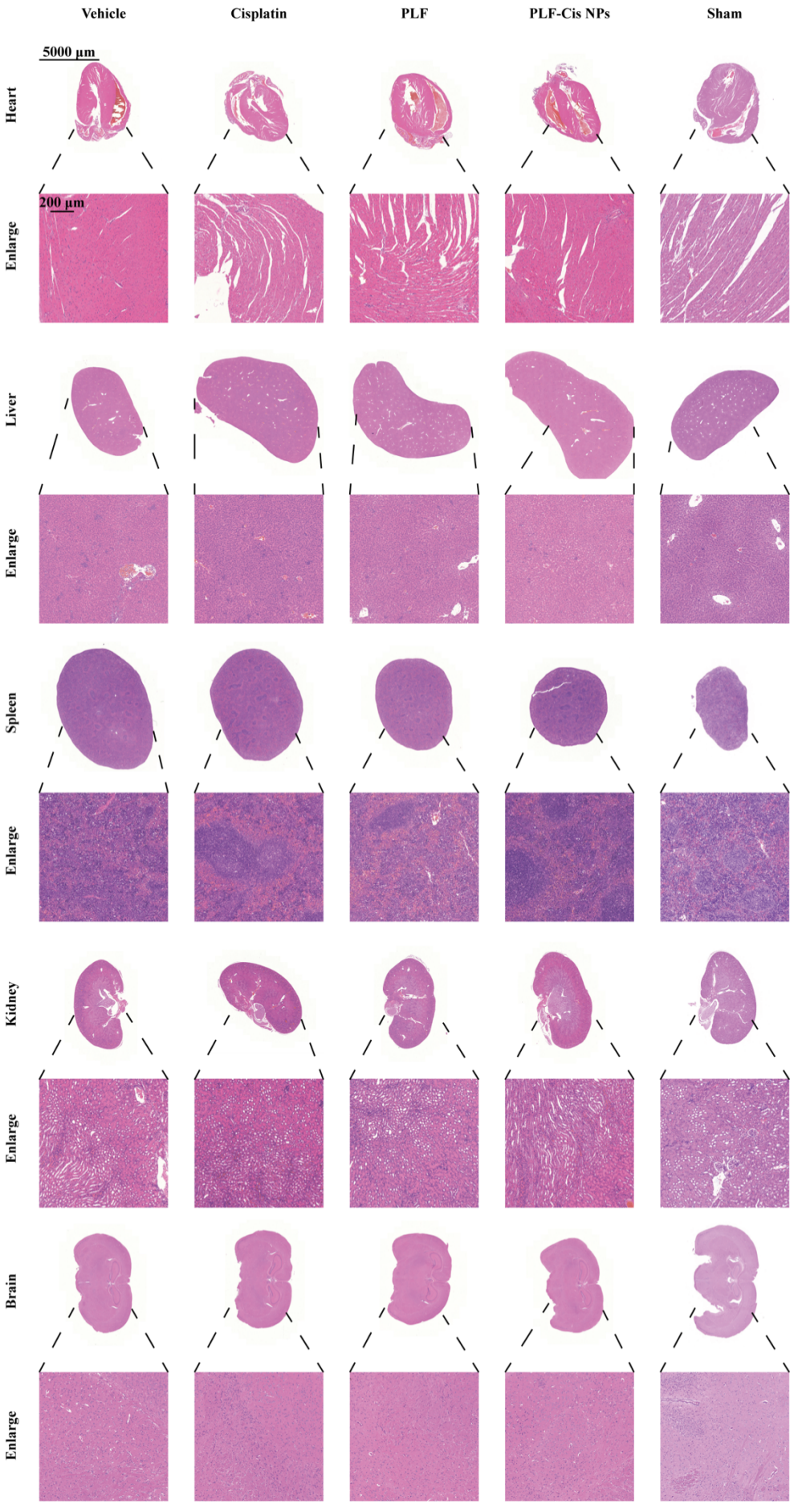


Figure S10. (A) Living images of orthotopic 4T1 tumor-bearing mice on 1st day of treatments. (B) Body weights of orthotopic 4T1 tumor-bearing mice during treatment. (C-F) The changes of serum biochemical indicators (ALT, AST, UREA, CREA-S) after different treatments. (G-J) Viscerosomatic ratio of major organs after different treatments.


Vehicle
Cisplatin
PLF
PLF-Cis NPs
Sham
5000 μm
200 μm
Heart
Enlarge
Liver
Enlarge
Spleen
Enlarge
Kidney
Enlarge
Brain
Enlarge

244

Figure S11. The H&E staining of major organ sections (heart, liver, spleen, kidney and brain) after different treatments.

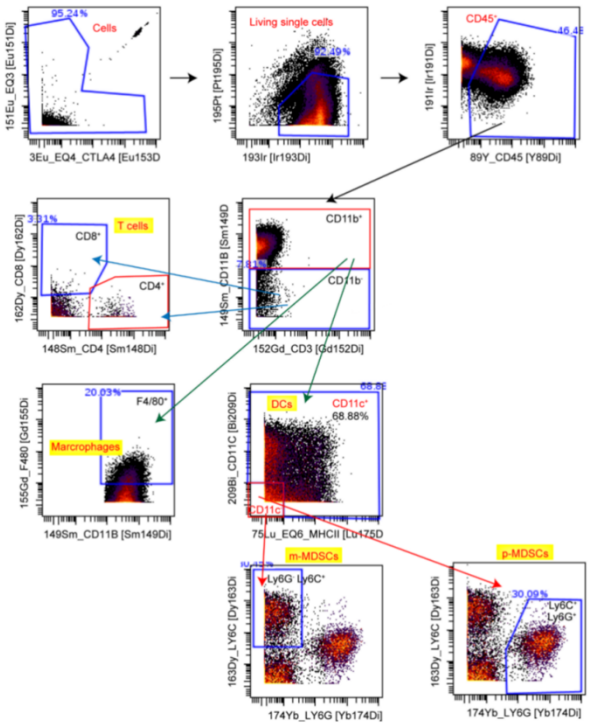


Fig. S12 Gating strategy of CyTOF analysis[1]

Reference:

1. D. Zhu, Y. Lu, L. Gui, W. Wang, X. Hu, S. Chen, Y. Wang and Y. Wang, *Acta Pharm Sin B*, 2022, **12**, 2592-2608.