Supplementary Information

Computationally guided high-throughput design of self-assembling drug nanoparticles

Daniel Reker,[1,2,9] Yulia Rybakova,[1] Ameya R. Kirtane,[1,2] Ruonan Cao,[1,3] Jee Won Yang,[1,2,4] Natsuda Navamajiti,[1,5] Apolonia Gardner,[1,6] Rosanna M. Zhang,[1,6] Tina Esfandiary,[1] Johanna L'Heureux,[1] Thomas von Erlach,[1] Elena M. Smekalova,[1] Dominique Leboeuf,[7] Kaitlyn Hess,[1] Aaron Lopes,[1] Jaimie Rogner,[1] Joy Collins,[1] Siddartha M. Tamang,[1] Keiko Ishida,[1] Paul Chamberlain,[1] DongSoo Yun,[1] Abigail Lytoon-Jean,[1] Christian K. Soule,[1] Jaime H. Cheah,[1] Alison M. Hayward,[1,2] Robert Langer,[1,8] Giovanni Traverso[2,8,*]

Supplementary Notes

Supplementary Note 1 – Number of self-aggregating drugs

Our pipeline identified 788 approved drugs as likely to self-aggregate. The total number of investigated drugs was 1857, meaning that 40% of approved drugs were anticipated as potential candidate structures. This is likely to be an overestimate at least at usually low testing concentrations. Seidler et al J. Med. Chem. 2003 reported that 13% of approved drugs aggregate as measured by unspecific inhibition of three enzymes. However, it is known that colloidal aggregates not always adsorb enzymes and therefore will not always show up in these type of screens (*cf* Feng et al Nat Chem Bio 2005). Other estimates have suggested that about 20% of molecular material might be self-aggregating (*cf* Reker et al Nat Chem 2019). While we recognize that our model overestimates self-aggregation, we believe this to be a useful property to span a large space of possible material for future investigations. Furthermore, most of the previously published estimates have been based on drug concentrations up to 30 µM. We are escalating these concentrations further in the context of nanoparticle formation, which suggests that potentially more drugs will aggregate at such higher concentrations.

Supplementary Note 2 – Choice of solvent and anti-solvent

We chose dimethyl sulfoxide (DMSO) as our primary solvent for several reasons: (i) it enabled dissolution of all our drugs and excipients at the stock concentration of 10mM, (ii) it is easily miscible with water, and (iii) it is routinely used in biological experiments and therefore enables rapid translation without additional purification. Phosphate-buffered saline (PBS) was selected as our anti-solvent since it is commonly used in biological research. Through this choice of solvent and anti-solvent, our particles were created in fixed buffer conditions of 1% DMSO in PBS. From an orthogonal perspective, our protocol directly augments established protocols for the detection of self-aggregation in high-throughput screening campaigns at 50 µM in 1% DMSO PBS to interrogate the ability of our excipients to prevent the colloidal self-aggregation of such drugs at these conditions.

Supplementary Note 3 – Bias in random forest feature importance measure

We observed that the chemical fingerprint was considered slightly less important compared to the physicochemical properties and interaction descriptors, but this could be in part due to the random

forest model's preference for continuous, versus binary descriptors (see Strobl *et al*. BMC Bioinformatics 8, 25, 2007).

Supplementary Note 4 – Challenges in measuring tumour accumulation

While we were unable to measure tumour accumulation given the multifocal, heterogenous and diffused growth of tumours in our genetic model, the accumulation of our particles in the liver combined with the active endocytosis-driven active drug uptake of the nanoparticles into hepatocellular carcinoma cells (Figure 4E) provides good evidence that our nanoparticles are able to actively target the tumours.

Supplementary Note 5 – Effect of route of administration

Given lower tumour targeting of intravenous injected sorafenib dissolved in cremophor-ethanol (Figure 4I) and the lack of AFP reduction through this treatment (Supplementary Figure 15), we can refute that the injection route alone is responsible for the improved efficacy of our sorafenib-glycyrrhizin nanoparticles.

Supplementary Note 6 – Alternative experimental and computational techniques

We used DLS to analyse the relative sizes of our nanoparticles. Multiple other aggregation characterization techniques are available, often with higher accuracy but lower throughput. These other techniques can inform future co-aggregation pipelines. For example, a clear limitation of DLS is its sensitivity to larger aggregates, indicating that our platform might be more prone to false negative readouts in the presence of multiple particle populations. Similarly, a 50% size reduction leading to particles smaller than 1 µm was an appropriate threshold for our study but might not be applicable for all types of particles and applications. Other readouts, experimental protocols and data processing might become helpful as orthogonal assessments of particle populations. In addition to altering experimental readouts, future studies can explore the application of different machine learning models, such as deep neural networks, for co-aggregation prediction. Such studies could potentially improve the accuracy of nanoparticle prediction and detection and would enable the customization of such platforms in laboratories with distinct computational or experimental preferences and capabilities.

Supplementary Note 7 – Alternative nanoformulation approaches

Multiple more matured nanoformulation techniques exist that can provide these capabilities. For example, liposomes will continue to be some of the most important tools in nanoformulations with clear advantages over our current protocol, for example by using established protocols for ligand-based targeting or for the delivery of hydrophilic drugs such as doxorubicin. Our approach is not meant as a replacement for other delivery techniques but instead an exciting new option with clear benefits in terms of drug loading and ease of synthesis. Ultimately, the decision to pursue a certain formulation technique is driven by a host of contributing factors and our study validates self-assembled drug nanoparticles as a competitive and orthogonal technique that expands the drug formulation armamentarium. To the best of our knowledge, there has been no direct head-to-head comparison of self-assembling drug nanoparticles with other nanomedicine approaches. Such a comparison would enable a clearer understanding of the advantages and disadvantages of applying different nanoformulation approaches.

Supplementary Methods

Identification of candidate drugs A random forest machine learning model was trained on publicly available aggregation data provided by the Brian Shoichet laboratory (bkslab.org). Briefly, data was downloaded in Excel format (bkslab.org/static/files/aggregators/aggregator_hts.xls) and all SMILES structures were encoded through Morgan fingerprints (radius 3, 2048 bits) and physicochemical descriptors (rdkit.Chem.Descriptors._descList) as surrogates for the original ECFP6 and MOE descriptors used in the original publication. "Ambiguous" annotations were discarded. This resulted in a binary classification dataset of 263 self-aggregating drugs (28.7%) and 653 non-self-aggregating drugs (71.3%) described through 2248-dimensional descriptors of chemical substructures and physicochemical properties. To generalize this dataset, a random forest classifier (scikit-learn, 1000 trees, no maximum features) was trained and evaluated using 10x ten-fold cross validation. To predict the self-aggregation behaviour of approved drugs with this model, all 1857 small molecular structures in the "approved" category were extracted in SMILES format from DrugBank 5.0. To enable aggregation prediction, these structures were also encoded through the same Morgan fingerprints (radius 3, 2048 bits) and physicochemical descriptors (RDKit). Missing values in the physicochemical descriptions were estimated through mean values using a scikit-learn imputation module. Aggregation likelihood was estimated through the number of trees in the random forest model that consider a DrugBank compound to be an aggregator (maximum predictive confidence) and a permissive threshold of 0.2 that reflects the class imbalance in the training set. 788 drugs were identified as potential candidate compounds for our workflow. This number is likely to be an overestimation of the number of self-aggregating drugs at low testing concentrations, which has been previously estimated at 13%-20%, but thereby provides a larger set of reasonable candidate structures for further study at elevated concentrations with higher aggregation propensity. From these 788 candidates, we identified 20 drugs for testing manually according to commercial availability, covering different indications, spanning the space of FDA-approved drugs according to chemical diversity (cf Supplementary Figure 1). Code and data for diversity assessment can be found at https://git.io/JJySb. We also added compounds we had previously investigated in our model drug libraries of compounds relevant for drug delivery research.

Transmission electron microscopy Nanoparticles were created at 50 µM drug concentration in 1% DMSO PBS. Samples were observed using TEM with or without negative staining. In either case, 7 µl of nanoparticle solution was transferred onto a 200 mesh copper grid coated with a continuous carbon film, and excess solution was wicked away after 60 seconds. In case of negative staining, 10 µl of staining solution (phosphotungstic acid 1% in water) was added and excess solution was wicked away immediately. Subsequently another round of staining was performed using the same volume of staining solution and excess solution was wicked away after 40 seconds from the edges of the samples. Finally, all samples were then dried at room temperature before imaging. Afterwards, the grid was mounted on a JEOL single tilt holder. Imaging was done on a JEOL 2100 FEG microscope. The microscope was operated at 200 kV and with a magnification in the range of 3,000 to 60,000 for assessing particle shape and size and atomic arrangement. All images were recorded on a Gatan 2kx2k UltraScan CCD camera. STEM imaging was done using a HAADF (high-angle annular dark field) detector with 0.5 nm probe size and 12 cm camera length. A X-Max 80mm^2 EDX (Oxford Instrument, UK) was used to determine the chemical information of the samples.

Analytics Sorafenib (250µM), nelfinavir (50 µM), and celecoxib (500 µM) formulations were analysed using an Agilent 4.6 x 50 mm EC C-18 Poroshell column with 2.7 µm particles, maintained at 50 °C. The

optimized mobile phase consisted of 0.1 % formic acid in water (A) and acetonitrile (B) using a flow rate of 1 mL/min. Gradient separation was achieved over a 13 minute run time at the following parameters: 0 min 95% A and 5% B; 4.4 min 50% A and 50% B; 6.4 min 50% A and 50% B; 9 min 20% A and 80% B; 9.5 min 20% A and 80% B; and 10 min 95% A and 5% B. The injection volume was 10 µl, and the selected ultraviolet (UV) detection wavelength was 248 nm with no reference parameters at an acquisition rate of 40 Hz. Atovaquone formulations (250 µM) were analysed using a Phenomenex Sphereclone 4.6 x 250 mm ODS (I) column with 5 µm particles, maintained at 50 °C. The optimized mobile phase consisted of 5% water, 15% methanol, and 80% acetonitrile using a flow rate of 1 mL/min over a 15 minute run time. The injection volume was 10 µl, and the selected ultraviolet (UV) detection wavelength was 295 nm with no reference parameters at an acquisition rate of 40 Hz. Danazol (250 µM) formulations were analysed using an Agilent Zorbax Eclipse XDB C-18 4.6 x 150 mm column with 5 µm particles, maintained at 35 °C. The optimized mobile phase consisted of 15%, 0.1% formic acid in water and 85% methanol using a flow rate of 1 mL/min over a 6 minute run time. The injection volume was 5 µl, and the selected ultraviolet (UV) detection wavelength was 285 nm with no reference parameters at an acquisition rate of 10 Hz. Terbinafine (250 µM) formulations were analysed using an Agilent Zorbax Eclipse XDB C-18 4.6 x 150 mm column with 5 µm particles, maintained at 40 °C. The optimized mobile phase consisted of 0.1 % formic acid in water (A) and methanol (B) using a flow rate of 1 mL/min. Gradient separation was achieved over a 10 minute run time (3 min post run) at the following parameters: 0 min 70% A and 30% B; 5 min 30% A and 70% B. The injection volume was 10 µl, and the selected ultraviolet (UV) detection wavelength was 242 nm with no reference parameters at an acquisition rate of 10 Hz. For all experiments, signals were converted to drug concentration through standard curves that were measured together with the samples. Measured concentrations were subsequently normalized against concentrations observed in control samples prepared in acetonitrile or methanol rather than aqueous solutions to ensure complete solubility of all tested drugs and account for inaccuracies in the stock solutions. All experiments were repeated at least three times.

For homogenized terbinafine skin samples, analysis was performed on a Waters ACQUITY UPLC®-I-Class System aligned with a Waters Xevo® TQ-S mass spectrometer (Waters Corporation, Milford MA). Liquid chromatographic separation was performed on an Acquity UPLC® BEH C18 (50mm × 2.1mm, 1.7 µm particle size) column at 50 °C. The mobile phase consisted of aqueous 0.1% formic acid, 10mM ammonium formate solution (Mobile Phase A) and acetonitrile: 10 mM ammonium formate, 0.1% formic acid solution (95:5 v/v) (Mobile Phase B). The mobile phase had a continuous flow rate of 0.6 mL/min using a time and solvent gradient composition. The initial composition, 95% Mobile Phase A, was held for 0.25 minutes. Afterwards, the composition was changed linearly to 5% Mobile Phase A and 95% Mobile Phase B until 1.00 minutes. The composition was held constant at 95% Mobile Phase B until 2.75 minutes. At 3.00 minutes the composition returned to 95% Mobile Phase A, where it remained for column equilibration for the duration of the run, ending at 4.00 minutes. The mass to charge transitions (m/z) used to quantitate terbinafine were 292.294>141.167 and 292.294>93.181 for quantitation and confirmation respectively. For internal standard, naftitine, 288.26>117.17 and 288.26>141.17 m/z transitions were used for quantitation and confirmation respectively. Sample introduction and ionization was by electrospray ionization (ESI) in the positive ionization mode. Waters MassLynx 4.1 software was used for data acquisition and analysis. Stock solutions were prepared in methanol at a concentration of 500 µg/mL. A twelve-point calibration curve was prepared in analyte-free, blank skin homogenate ranging from 1.25-5000 ng/mL. 100 µl of each sample was spiked with 200 µl of 250 ng/mL internal standard in acetonitrile to elicit protein precipitation. Samples were vortexed, sonicated for 10 minutes,

and centrifuged for 10 minutes at 13,000 rpm. 200 µl of supernatant was pipetted into a 96-well plate containing 200 µl of water.

Concentration escalation experiments Nanoparticles were created by mixing 5 µl of drug stock solution in DMSO with 5 µl of excipient stock solution in DMSO; unformulated drug solutions were created by mixing 5 µl drug stock solution with 5 µl DMSO. Solvent exchange was performed by adding 990 µl sterile filtered and degassed PBS. All samples were generated in 4ml glass vials and stored sealed at room temperature. These experiments were repeated at increasing stock solution concentrations until either 5 mM final drug concentrations were reached or the unformulated drug visibly precipitated immediately after solvent exchange. Images were taken at 1 mM for atovaquone, 500 µM for celecoxib, 5 mM for terbinafine, 1 mM for sorafenib, 500 µM for danazol and 500 µM for nelfinavir.

MEK phosphorylation analysis Infrared secondary antibodies, IRdye 680RD goat anti-mouse IgG and IRdye 800CW anti-rabbit (both 1:1000) were purchased from LiCor. Phospho-MEK1/2 (Ser217/221) Antibody (9121) and MEK1/2 Antibody (9122) were purchased from Cell Signaling Technology (both 1:100). For In-Cell Western, cellular proteins were quantitated *in situ* based on infrared intensity. Samples were immunolabelled with an infrared conjugated IgG secondary antibody using standard immunofluorescence protocol. Briefly, cells were fixed with 4 % (v/v) formalin in dH2O (Sigma) for 15 minutes at room temperature, washed with PBS, permeabilized with 0.25 % (v/v) TritonX-100/PBS for 2 minutes, washed with PBS and then blocked with 4 % (w/v) bovine serum albumin in PBS. Primary and secondary antibodies were incubated in blocking buffer for 1 h at room temperature. Samples were subsequently imaged using an Odyssey Fc Infrared Imaging System (LiCor). The resulting signal intensity was subsequently quantified by using the Odyssey CLx Image Studio Analysis Software.

Sorafenib-glycyrrhizin toxicity assessment Two healthy, 6 week old, female FVB/N mice were treated with elevated dosages of 60 mg/kg sorafenib-glycyrrhizin nanoparticles for one week (three administrations). Blood was sampled after the final administration for serum chemistry assessment. We screened seven markers associated with liver damage (alanine aminotransferase ALT, aspartate aminotransferase AST, alkaline phosphatase ALP, albumin, gamma-glutamyl transferase GGT, total protein, direct bilirubin) by the Massachusetts Institute of Technology Division of Comparative Medicine.

AFP analysis Mouse blood was collected into BD Vacutainer® EDTA tubes, incubated on ice for 10 min and then centrifuged at 2000g for 20 min at 4°C. 3 µl of serum (the supernatant) was mixed with 47 µl PBS (Gibco) and 10 µl 6X reducing SDS buffer (Boston Bio Products), the mixture was heated at 68°C for 30 min, and 10 µl of the mixture were subjected to SDS/PAGE. Proteins were transferred to Nitrocellulose membranes (BioRad) and blocked with Li-COR Odyssey® Blocking Buffer (TBS) for 1h at RT. Biotinylated anti-mouse AFP antibody (R&D Systems, BAF5369) was reconstituted at 0.2 mg/mL in sterile PBS and used at 1:2000 dilution. The binding was done for either 2 h at RT or overnight at 4°C. Protein bands were visualized by incubation with secondary antibodies labelled with infrared fluorophores (Li-COR IRDye® 680LT Streptavidin, 926-68031, dilution 1:10,000) for 30 min at RT. Membranes were scanned on a Li-COR Odyssey Scanner and fluorescence intensity was subsequently quantified using the ImageJ software. Band density was normalized to background per gel. The normalized band density for each sample was divided by the average normalized band density for the “no treatment” controls to assess relative band density, reflecting the difference in AFP expression levels in the treated mice *vs* control mice. The relative change of band density level before (week 5), during (week 7), and after treatment (week 10) was used to assess tumour progression.

Supplementary Tables

Supplementary Table 1: Retrospective ten times 10-fold cross validation of self-aggregation model for candidate drug selection. Mean performance and one standard deviation are reported.

Metric	Mean ± standard deviation
MCC	0.56 ± 0.02
Accuracy	0.83 ± 0.01
Precision	0.77 ± 0.02
F1	0.66 ± 0.02
ROC AUC	0.88 ± 0.002
AUPR	0.76 ± 0.004

Supplementary Table 2: Drugs that failed to show aggregation at fixed testing concentration and were therefore not amenable to our high-throughput formulation approach.

Drug	DLS radius [nm]
dyphylline	65
simvastatin	80
budesonide	60
minoxidil	70

Supplementary Table 3: Retrospective machine learning performance using 10x "leave one drug out" evaluation of various tested descriptors for classification. The full model, which relied on chemical descriptors (FP) and molecular dynamics (MD), performed significantly better than any sub-model, although the "only FP" model (using radial chemical substructure descriptors and physicochemical descriptors but no MD simulations) was competitive with a similar retrospective performance. *p* corresponds to two-sample, two-sided student T-test of the ten repeats comparing the full model with the respective model in the row.

	10-fold cross validation											
	MCC		**F1**		**Precision**		**Accuracy**		**AUROC**		**AUPRC**	
	mean	***p***	**mean**	***p***	**mean**	***p***	**mean**	***p***	**mean**	***p***	**mean**	***p***
FP & MD	0.32	--	0.28	--	0.62	--	0.94	--	0.86	--	0.37	--
only FP	0.30	1e-3	0.27	2e-3	0.60	3e-2	0.94	4e-1	0.85	3e-2	0.36	2e-3
only MD	0.07	4e-14	0.14	6e-11	0.13	3e-17	0.88	1e-21	0.66	6e-20	0.11	4e-20
	Leave one drug out validation											
	MCC		**F1**		**Precision**		**Accuracy**		**AUROC**		**AUPRC**	
	mean	***p***	**mean**	***p***	**mean**	***p***	**mean**	***p***	**mean**	***p***	**mean**	***p***

FP & MD	0.28	--	0.26	--	0.55	--	0.94	--	0.71	--	0.23	--
only FP	0.26	1e-3	0.26	7e-1	0.55	8e-1	0.89	4e-5	0.70	6e-2	0.22	4e-1
only MD	0.09	8e-20	0.15	2e-15	0.14	6e-25	0.88	5e-25	0.63	9e-16	0.1	3e-23

Supplementary Table 4: The 50 most important features of the random forest model. Feature name includes the source of the feature ("drug", "excipient", "simulation") followed by the feature definition. Names of features are derived from rdkit.Chem.Descriptors._desclist, numbered Morgan fingerprints ("fp_X") or manually labelled for the simulation descriptors. For the latter, "contact_av" is the average number of contacts between drugs and excipients during the simulation, "AvDistance "is the average distance between one of the four possible pairings of two drugs and two excipients, "Variance" is the variance of their distance. "d(distance)" is the first derivative of the distance (i.e. the speed at which excipients and drugs move away or towards each other, where "max" indicates the maximum and "avg" indicates the average. "kineticE" and "potentialE" are the kinetic and potential energy derived from the last frame of the simulation through the force field.

Rank	Feature name	Feature importance
1	simulation_d(distance)_avg_4	0.0053
2	simulation_d(distance)_max_2	0.0052
3	excipient_MolMR	0.0051
4	excipient_EState_VSA2	0.0048
5	simulation_contact_av	0.0048
6	simulation_d(distance)_max_3	0.0047
7	excipient_Chi0	0.0047
8	simulation_AvDist_3	0.0046
9	excipient_qed	0.0046
10	simulation_kineticE	0.0045
11	simulation_d(distance)_avg_3	0.0045
12	excipient_Chi0n	0.0045
13	simulation_d(distance)_max_4	0.0044
14	excipient_Chi2n	0.0044
15	excipient_Chi1v	0.0044
16	simulation_AvDist_1	0.0044
17	excipient_Ipc	0.0043
18	simulation_Variance_2	0.0043
19	simulation_potentialE	0.0043
20	simulation_d(distance)_avg_2	0.0042
21	simulation_d(distance)_avg_1	0.0042
22	simulation_d(distance)_max_1	0.0041
23	excipient_NumAromaticCarbocycles	0.0041
24	excipient_RingCount	0.004
25	excipient_BertzCT	0.004
26	excipient_Chi0v	0.0039
27	simulation_AvDist_4	0.0039

28	simulation_Variance_3	0.0038
29	simulation_Variance_1	0.0037
30	excipient_Chi1	0.0037
31	excipient_Chi2v	0.0036
32	excipient_SlogP_VSA5	0.0036
33	excipient_NumValenceElectrons	0.0036
34	excipient_Chi1n	0.0036
35	excipient_HeavyAtomCount	0.0036
36	excipient_MinAbsPartialCharge	0.0035
37	excipient_Chi3n	0.0035
38	excipient_SMR_VSA10	0.0035
39	excipient_Kappa1	0.0034
40	simulation_AvDist_2	0.0034
41	excipient_Chi4n	0.0034
42	excipient_NumHeteroatoms	0.0033
43	excipient_LabuteASA	0.0033
44	drug_PEOE_VSA9	0.0033
45	excipient_PEOE_VSA7	0.0032
46	excipient_ExactMolWt	0.0031
47	excipient_SlogP_VSA6	0.0031
48	excipient_HeavyAtomMolWt	0.0031
49	excipient_fp_1160	0.0031
50	excipient_VSA_EState8	0.0031

Supplementary Table 5: Retrospective machine learning model comparison. Retrospective performance was evaluated using 10x ten-fold cross validation (top) or 10x "leave one drug out" evaluation (bottom). Random forest (RF), Naïve Bayes (NB), nearest neighbor (kNN), decision trees (DT), Neural Networks (NN) and Support-Vector machines (SVM) were implemented from scikit-learn with default parameters. *p* corresponds to two-sample, two-sided student T-test of the ten repeats comparing the RF model with the respective model in the row. Highest performance value is indicated by underscore.

	10-fold cross validation											
	MCC		**F1**		**Precision**		**Accuracy**		**AUROC**		**AUPRC**	
	mean	**p**	**mean**	**p**	**mean**	**p**	**mean**	**p**	**mean**	**p**	**mean**	**p**
RF	0.32	--	0.28	--	0.62	--	0.94	--	0.86	--	0.37	--
NB	0.05	1e-16	0.05	8e-16	0.19	4e-18	0.93	2e-14	0.48	2e-18	0.08	8e-22
kNN	0.2	1e-08	0.23	4e-05	0.34	6e-13	0.92	7e-13	0.66	9e-22	0.18	1e-16
DT	0.27	1e-03	0.32	4e-03	0.32	5e-14	0.91	2e-13	0.64	2e-22	0.34	9e-04
NN	0.05	2e-14	0.12	1e-11	0.1	2e-18	0.85	4e-08	0.52	2e-23	0.15	3e-17
SVM	0.1	1e-12	0.17	3e-09	0.1	5e-19	0.69	2e-11	0.57	2e-15	0.09	2e-21
	Leave one drug out validation											
	MCC		**F1**		**Precision**		**Accuracy**		**AUROC**		**AUPRC**	
	mean	**p**	**mean**	**p**	**mean**	**p**	**mean**	**p**	**mean**	**p**	**mean**	**p**
RF	0.28	--	0.26	--	0.55	--	0.94	--	0.71	--	0.23	--
NB	0.05	1e-16	0.05	2e-14	0.19	4e-18	0.93	8e-16	0.36	2e-34	0.06	1e-23
kNN	0.14	1e-13	0.17	4e-19	0.25	6e-17	0.92	1e-10	0.61	5e-29	0.13	8e-22

DT	0.19	3e-08	0.25	6e-11	0.23	2e-15	0.89	1e-02	0.6	5e-21	0.28	3e-09
NN	0.05	2e-12	0.12	2e-08	0.11	9e-17	0.87	3e-10	0.52	2e-20	0.14	6e-15
SVM	0.09	3e-10	0.16	6e-16	0.1	8e-19	0.68	2e-08	0.56	8e-11	0.09	2e-19

Supplementary Table 6: Adversarial control model's using y shuffling ("y") or x randomization ("x"), both applied in 10x ten-fold cross validation or 10x leave one drug out validation. *p* corresponds to two-sample, two-sided student T-test of the ten repeats comparing the model with the respective adversarial control model.

	10-fold cross validation											
	MCC		**F1**		**Precision**		**Accuracy**		**AUROC**		**AUPRC**	
	mean	**p**	**mean**	**p**	**mean**	**p**	**mean**	**p**	**mean**	**p**	**mean**	**p**
RF	0.32	--	0.28	--	0.62	--	0.94	--	0.86	--	0.37	--
y	0.01	8e-16	0.00	2e-24	0.10	3e-04	0.93	3e-11	0.46	1e-13	0.06	4e-21
x	0.00	5e-28	0.00	1e-26	0.00	7e-33	0.93	9e-14	0.49	1e-15	0.07	1e-22
	Leave one drug out validation											
	MCC		**F1**		**Precision**		**Accuracy**		**AUROC**		**AUPRC**	
	mean	**p**	**mean**	**p**	**mean**	**p**	**mean**	**p**	**mean**	**p**	**mean**	**p**
RF	0.28	--	0.26	--	0.55	--	0.94	--	0.71	--	0.23	--
y	0.00	2e-15	0.00	5e-16	0.03	1e-11	0.93	9e-06	0.50	4e-17	0.07	1e-20
x	0.00	3e-16	0.00	3e-16	0.00	3e-18	0.93	1e-05	0.56	6e-15	0.08	2e-19

Supplementary Table 7: DLS results of predicted self-assembling nanoparticles. Reported are the mean particle radius, the reduction factor of aggregate sizes between drugs and co-aggregates, the polydispersity, and the normalized signal intensity. Furthermore, we report the mean size and weight percentage of particles detected at the different acquisition ranges (R1-R4) as well as the percentage of autonomously flagged DLS acquisitions by the Wyatt Dynamics software.

Drug	**Sample**	**Radius [nm]**	***X***	**PDI [%]**	**Inten -sity**	**R1** 1-10	**R2** 10-100	**R3** 100-1000	**R4** >1000	**% flag**
Sorafenib	drug alone	846.0	*19*	Multi-modal	2e8	--	--	478 [31]	3344 [69]	90
+ glycyrrhizin		45.3		17.0	6e7	--	44.2 [100]	--	--	0
Atovaquone	drug alone	1079.3	11	Multi-modal	6e10	6.9 [3]	--	--	1163 [97]	90
+ a tocopherol		95.7		26.1	1e11	--	92.1 [100]	--	--	0
Nelfinavir	drug alone	1102.0	9.4	Multi-modal	2e10	--	30.3 [0]	--	1398 [100]	60
+ riboflavin		117.5		22.1	8e10	--	--	115.4 [100]	--	0
Danazol	drug alone	558.2	9.4	Multi-modal	3e8	---	13.1 [1]	778 [99]	--	90
+ tannic acid		59.4		18.1	2e8	--	58	--	--	0

							[100]			
Celecoxib	drug alone	646.1	11	Multi-modal	4e8	--	--	377.92 [31]	2661 [69]	60
+ indomethacin		58.1		16.2	3e8	--	54.95 [100]	--	--	0
Terbinafine	drug alone	596.3	4.4	Multi-modal	1e10	--	--	179 [39]	2832 [61]	80
+ taurocholic acid		134.2		23.1	3e10	--	--	110 [100]	--	0

Supplementary Table 8: Interactions found in nanoparticles according to last frame of MD simulation. "HBonds" were quantified as Baker-Hubbard Hydrogen Bonds using MDTraj. All other interactions were identified using the MolBridge webserver. The largest number per row is highlighted in bold.

	Sorafenib-glycyrrhizin	**Terbinafine-taurocholic acid**	**Celecoxib-Indomethacin**	**Atovaquone-alpha tocopherol**	**Danazol-tannic acid**	**Nelfinavir-riboflavin**
HBonds	**84**	44	72	9	57	28
C-Cl..Cl	0	0	**2**	**2**	0	0
C-Cl..F	0	0	**5**	0	0	0
C-Cl..N	0	0	**1**	0	0	0
C-Cl..O	**11**	0	2	2	0	0
C-F..Cl	3	0	**4**	0	0	0
C-F..F	5	0	**8**	0	0	0
C-F..N	**8**	0	2	0	0	0
C-F..O	**7**	0	3	0	0	0
C-H..N	33	40	37	0	5	**147**
C-H..O	477	257	114	192	**836**	317
C-H..S	0	**1**	**1**	0	0	0
N-H..S	0	1	**23**	0	0	16
N-H..N	22	1	**36**	0	0	8
N-H..O	16	3	**46**	0	0	19
O-H..N	3	0	2	0	0	**30**
O-H..O	174	49	3	21	**530**	45
O-H..S	0	**16**	1	0	0	1
O-H..π	0	1	3	1	**33**	6
S..π	0	**8**	**8**	0	0	0
π..π	8	1	14	3	**32**	11
C-H..π	100	97	121	120	164	**180**
N-H..π	**7**	6	2	0	0	8
C-Cl..π	2	0	**3**	0	0	0

C-F..π	**7**	0	6	0	0	0

Supplementary Table 9: Drug loading and encapsulation efficiency data for terbinafine-taurocholic acid particles. Concentrations are in ug/ml and were determined through high-performance liquid chromatography on two independent samples.

Sample	**Conc 1**	**Conc 2**	**Average**			
Total terbinafine	369.2	364.5	366.9	**Encapsulation 1** 280 / 369.2 = 75.8 %	**Encapsulation 2** 274.3 / 364.5 = 75.3 %	**Encapsulation** 75.6 ± 0.4 %
Particle terbinafine	280.0	274.3	277.2			
Particle taur. acid	22.3	19.5	20.9	**Loading 1** 280 / 302.3 = 92.6 %	**Loading 2** 274.3 / 293.8 = 93.3 %	**Loading** 93.0 ± 0.5 %

Supplementary Table 10: DLS results of further characterized sorafenib nanoparticles. Reported are the mean particle radius, the reduction factor of aggregate sizes between drugs and co-aggregates, the polydispersity, and the normalized signal intensity. Furthermore, we report the mean size and weight percentage of particles detected at the different acquisition ranges (R1-R4) as well as the percentage of autonomously flagged DLS acquisitions by the Wyatt Dynamics software.

Drug	**Sample**	**Radius [nm]**	**X**	**PDI [%]**	**Inten-sity**	**R1** 1-10	**R2** 10-100	**R3** 100-1000	**R4** >1000	**% flag**
Sorafenib	drug alone	920.0	**-----**	M	1e8	--	--	129 [34]	3718 [66]	90
	+ glycyrrhizin	42.1	**21.8**	17.0	7e7	--	39.3 [100]	--	--	0
	+ candesartan cilexitel	38.1	**24.1**	19.4	1e7	--	38.8 [100]	--	--	0
	+ tannic acid	56.2	**16.3**	14.5	3e8	--	52.3 [100]	--	--	0
	+ indomethacin	65.2	**14.11**	16.2	1e9	--	63.9 [100]	--	--	0
	+ meloxicam	99.1	**9.28**	20.6	9e7	--	79.7 [100]	--	--	0

Supplementary Table 11: 24h stability of a selected set of nanoparticles, measured after storage at room temperature.

Sample	**Radius [nm], 0h**	**PDI [%], 0h**	**Radius [nm], 24h**	**PDI [%], 24h**
Sorafenib – Congo Red	54.5	15.1	61.5	16.5
Sorafenib – Indomethacin	65.2	16.2	78.1	3.0
Sorafenib – Glycyrrhizin	42.1	17.0	57.2	23.9
Terbinafine – Taurocholic acid	134.2	23.1	317.2	23.7

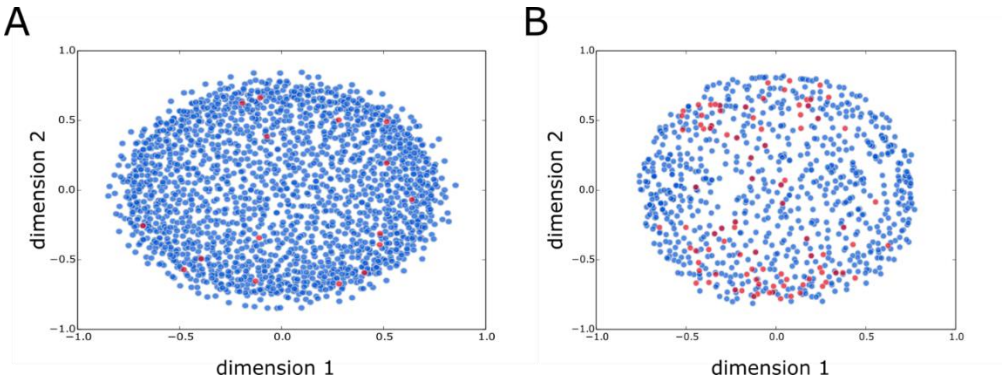
Supplementary Table 12: Drug loading and encapsulation efficiency data for sorafenib-glycyrrhizin particles. Concentrations are in ug/ml and were determined through high-performance liquid chromatography.

Sample	Conc 1	Conc 2	Average			
Total sorafenib	64.5	62.9	63.7	**Encapsulation 1** 60.0 / 64.5 = 93.0	**Encapsulation 2** 58.0 / 62.9 = 92.2	**Encapsulation** 92.6 ± 0.6
Particle sorafenib	60.0	58.0	59.0			
Particle glycyrrhizin	2.7	3.7	3.2	**Loading 1** 60 / 62.7 = 95.7	**Loading 2** 58 / 61.7 = 94.0	**Loading** 94.9 ± 1.2

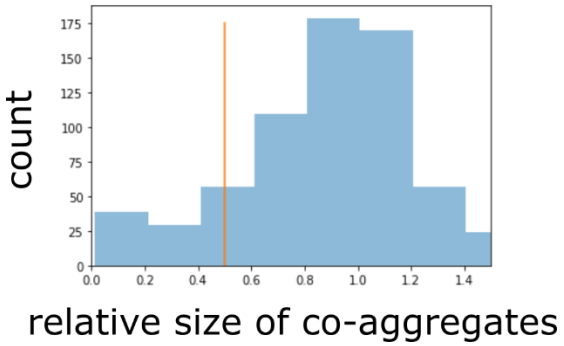
Supplementary Table 13: Serum chemistry assessment of liver toxicity associated markers in two healthy female 6 week old FVB/N mice. Mice received three injections of elevated dosages of 60 mg/kg sorafenib-glycyrrhizin particles within one week and showed no adverse symptoms.

Maker	Mouse 1	Mouse 2	Reference
ALT [IU/L]	42.0	45.0	17.0-77.0
AST [IU/L]	46.0	54.0	54.0-298.0
ALP [IU/L]	96.0	84.0	35.0-96.0
Albumin [g/dL]	3.0	2.7	2.5-4.8
GGT [IU/L]	0.0	0.0	--
Total protein [g/dL]	5.2	5.0	3.5-7.2
Direct bilirubin [mg/dL]	0.0	0.0	--

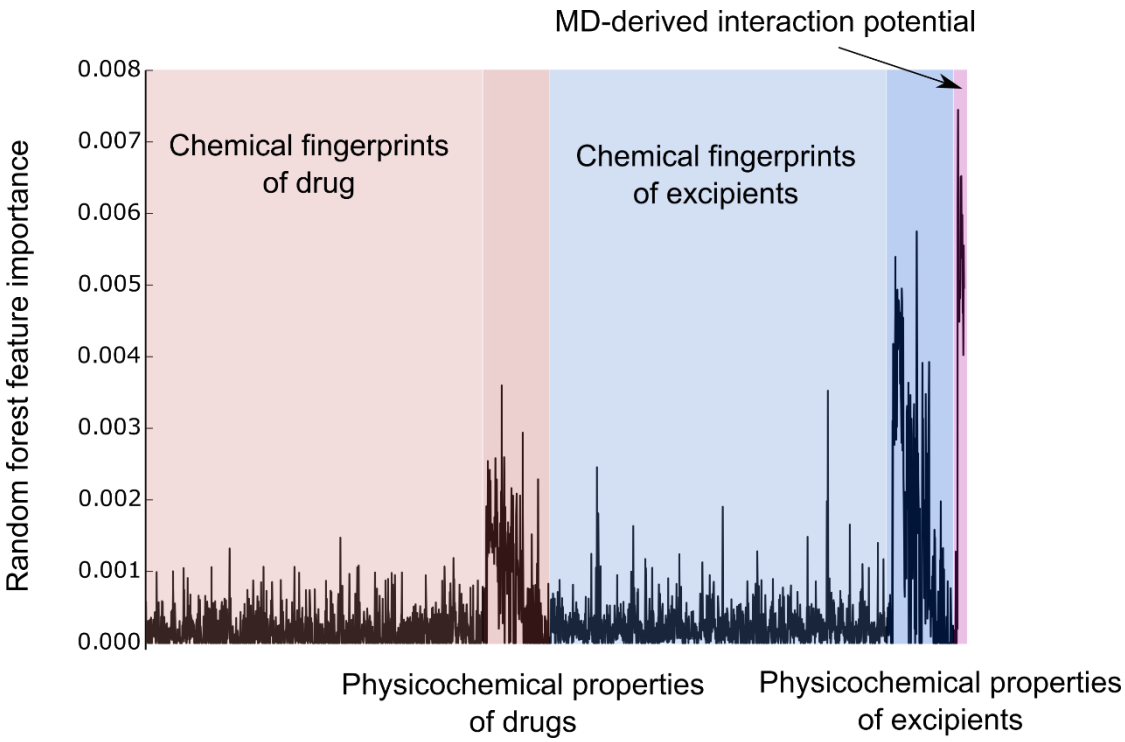
Supplementary Figures



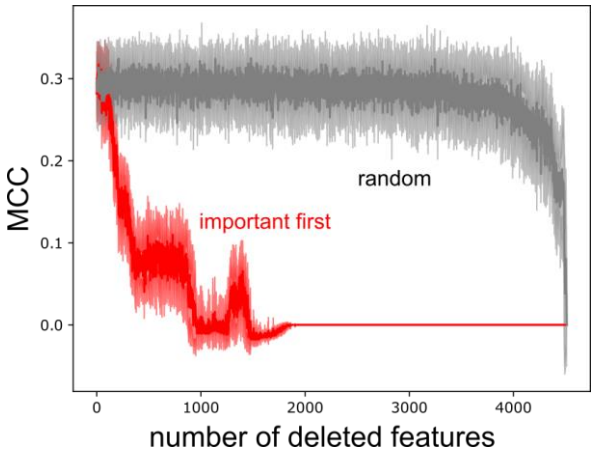
Supplementary Figure 1: Multi-dimensional scaling of all FDA-approved small molecular drugs (A) and all FDA approved inactive ingredients and “generally recognized as safe” compounds (B) in blue. The red dots indicate the position of the selected drugs (A) or excipients (B) that we subsequently used for high-throughput screening and machine learning training data generation.



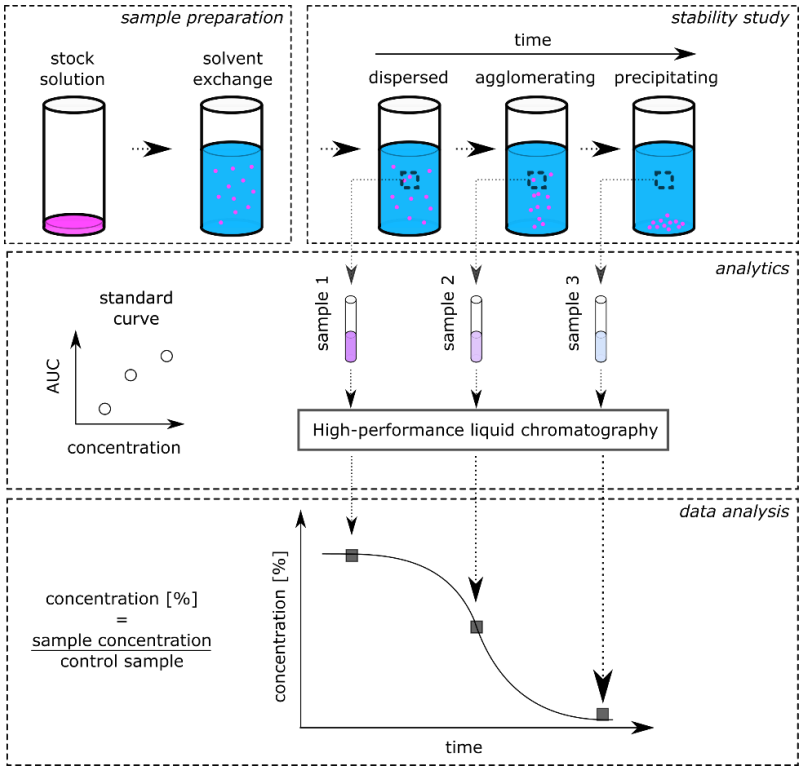
Supplementary Figure 2: Relative size distribution of observed co-aggregates in high-throughput experiment. Red line indicates the 50% threshold that was applied for further analysis of co-aggregates.



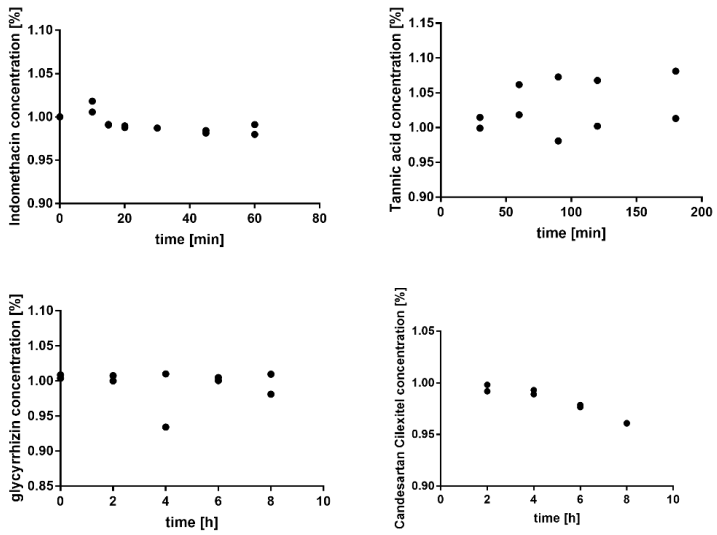
Supplementary Figure 3: Random forest feature importance of the final model for the 2460 numerical descriptors of the novel formulations. As can be seen here, the model considers both chemical and physicochemical properties of both the drug and the excipient as well as the MD-derived interaction potentials.



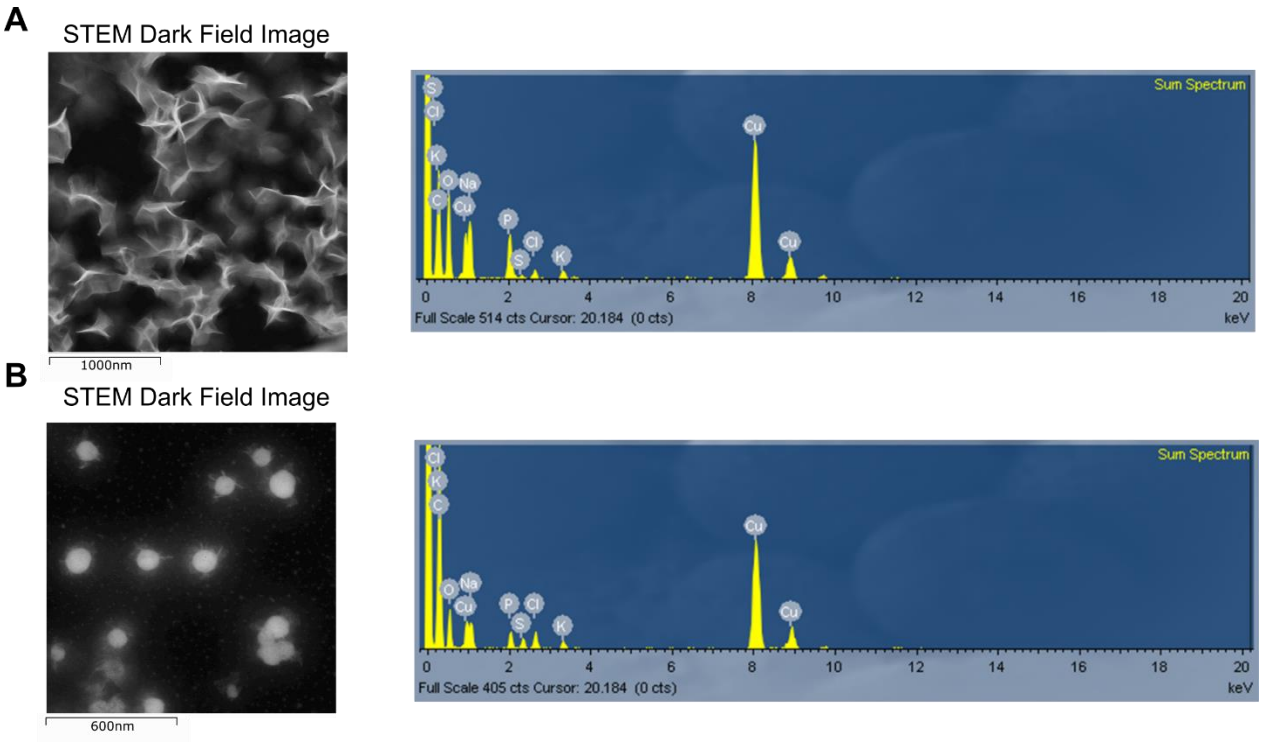
Supplementary Figure 4: Feature ablation experiment. Deletion of the most important features (red) results in a rapid increase in out-of-bag error as measured by loss of Matthews Correlation Coefficient (MCC) compared to randomly deleting features iteratively (grey). Shown are average performances (bold line) of ten independent runs and their standard deviation (shading).



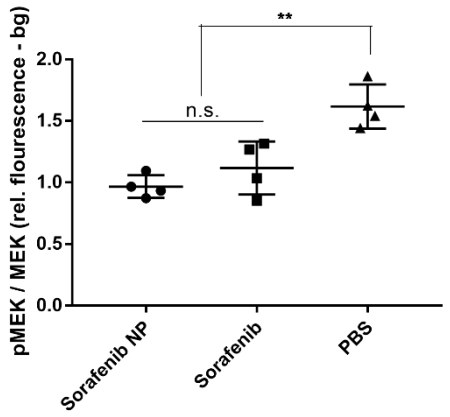
Supplementary Figure 5: Schematic explaining the experimental workflow to determine dispersion stability. Nanoparticles were generated through solvent exchange. The nanoparticle dispersions were stored at room temperature in a sealed vial. At pre-determined timepoints, we extracted samples and diluted these in an organic solvent to ensure full solubility of the nanoparticles prior to injection onto the HPLC to determine drug concentrations. Relative concentration was determined through standard curves and by comparing determined concentration by analysing the control sample consisting of equivalently diluted stock solution.



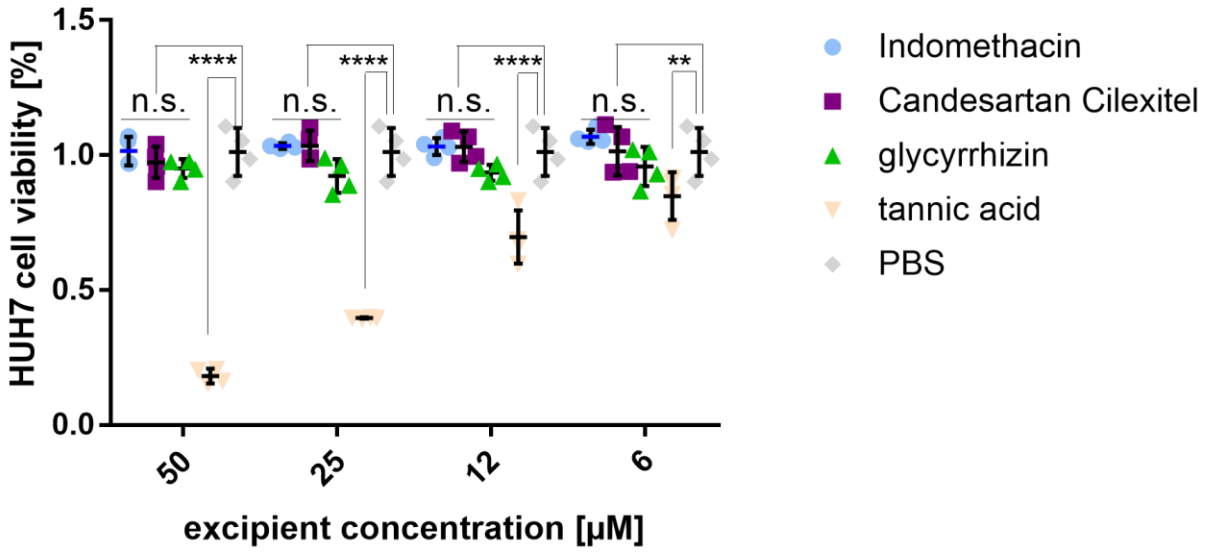
Supplementary Figure 6: High-performance liquid chromatography data for excipient concentrations during the precipitation kinetics experiments. n = 2 independent samples.



Supplementary Figure 7: Scanning Transmission Electron Microscopy-Energy Dispersive Spectroscopy (STEM-EDS) shows that terbinafine-taurocholic acid nanoparticles are rich in sulphur from taurocholic acid, both in cases of early aggregates (**A**) as well as for late-stage aggregates **(B).** Scale bar corresponds to 1000nm and 600nm in A and B respectively. Representative images are shown from ten acquisitions generated through two independent experiments, all images reproduced the here depicted behaviour.

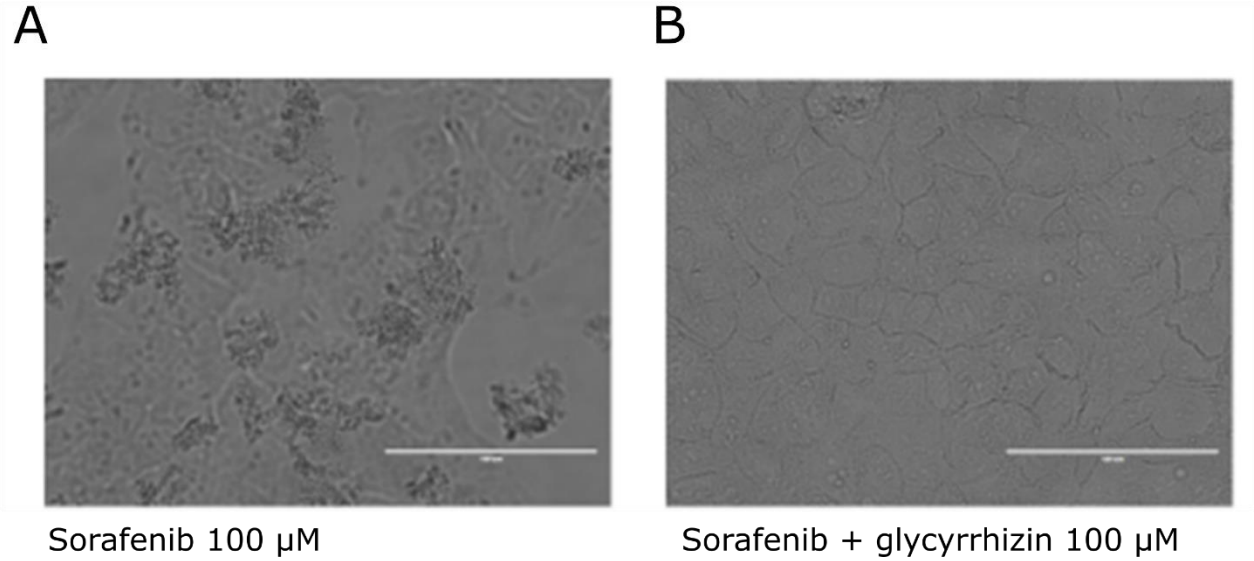


Supplementary Figure 8: Inhibition of MEK phosphorylation of the sorafenib nanoparticles. Inhibition of phosphorylation is equivalent to inhibition by free sorafenib and reduces the relative amount of pMEK by 50% after 5h of treatment in HUH7 HCC cells. n = 4 independent samples. Treatment concentration 500 µM. $p = 0.001$ one-way ANOVA, $p < 0.006$ Tukey's multiple comparisons test for sorafenib NP or sorafenib vs. PBS ($p < 0.006$ Tukey's multiple comparisons test), difference between sorafenib NP and sorafenib is not significant ($p = 0.45$ Tukey's multiple comparisons test). Lines corresponds to mean values; error bars represent one standard deviation.

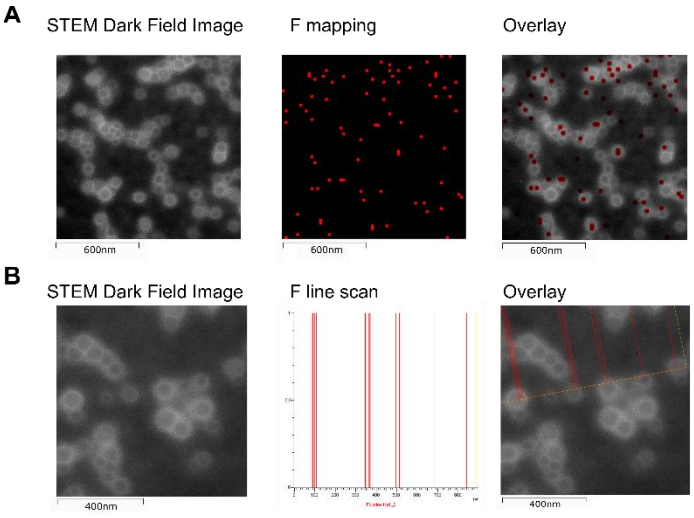


Supplementary Figure 9: HUH7 viability after 48 h of treatment with nanoparticle-forming excipients alone at different concentrations. $p < 0.0001$ two-way ANOVA. Dunnett's multiple comparisons test for comparison between excipient and PBS with 1% DMSO vehicle control. PBS treatment is not

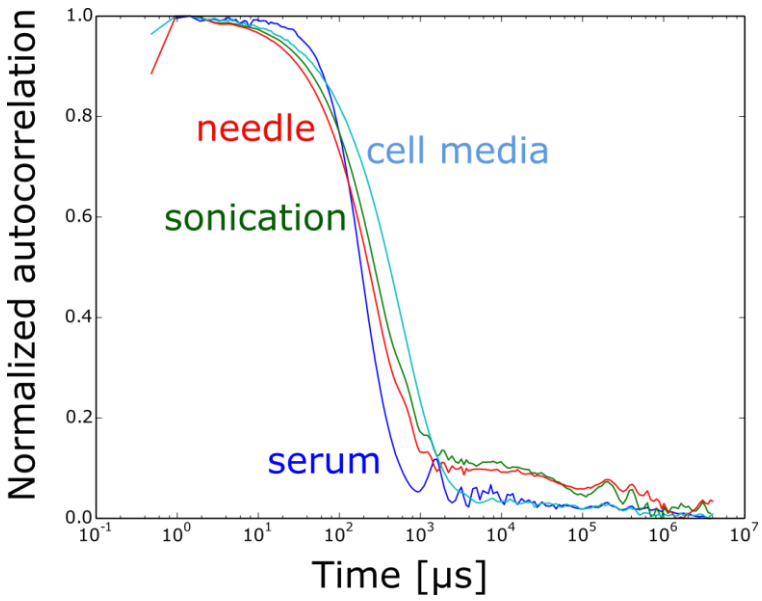
concentration dependent and was duplicated across the plot for readability. n = 4 independent samples. Lines corresponds to mean values; error bars represent one standard deviation.



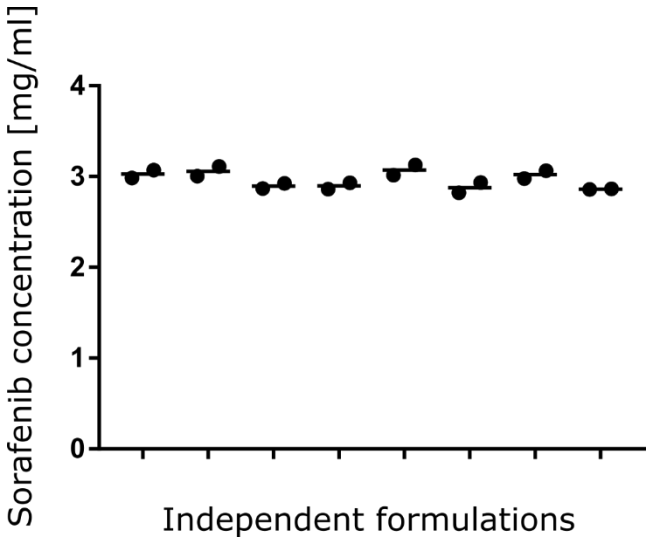
Supplementary Figure 10: Bright field microscopy images of HUH7 cells with 100 µM sorafenib (A) vs. sorafenib-glycyrrhizin (B). Scale bar corresponds to 100 µm. Representative images are shown from ten acquisitions generated through two independent experiments, all images reproduced the here depicted behaviour.



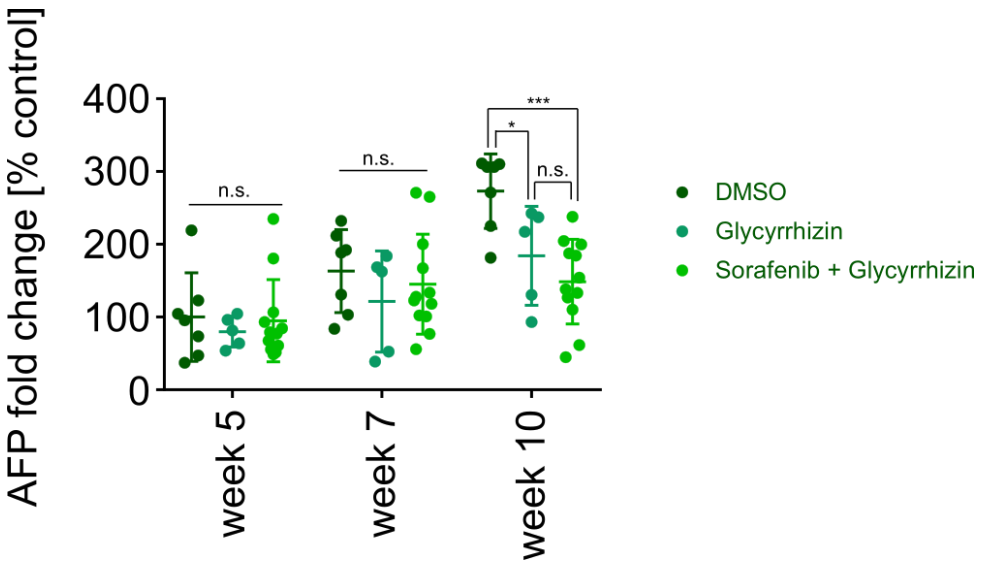
Supplementary Figure 11: STEM-EDS shows that sorafenib-glycyrrhizin nanoparticles are rich in fluorine from sorafenib. Representative images are shown from ten acquisitions generated through two independent experiments, all images reproduced the here depicted behaviour. **A** F mapping analysis to visualize 2D distribution of fluorine in particles. Scale bar corresponds to 600nm. **B** Line scan highlights specific enrichment of drug in the particles and lack of drug in the surrounding sample. Scale bar corresponds to 400nm.



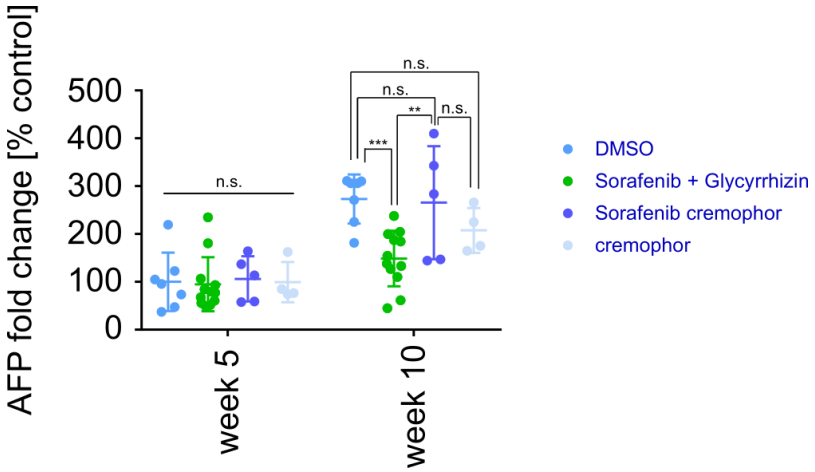
Supplementary Figure 12: Stability of sorafenib nanoparticles after different perturbations or in different environments. Shown are DLS autocorrelation functions of sorafenib-glycyrrhizin nanoparticles in serum (blue) or cell media (light blue) or after centrifugation and subsequent sonication for re-dispersion (green) or after pushing the nanoparticles through a 29 Gauge needle (red).



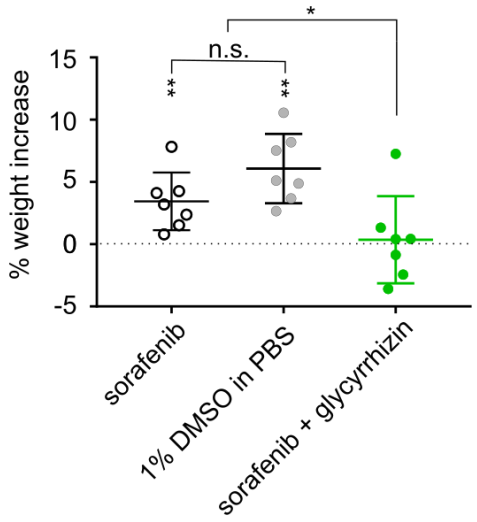
Supplementary Figure 13: Consistency of formulation strategy. HPLC analytics were performed in duplicates on eight independent nanoparticle formulations created to assess consistency of formulation protocol. n = 2 independent samples, Lines corresponds to mean values.



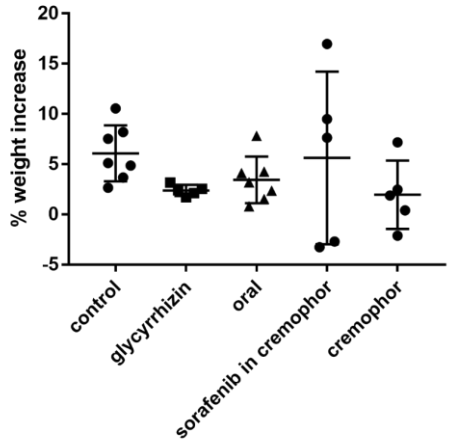
Supplementary Figure 14: AFP levels of mice treated with sorafenib-glycyrrhizin and glycyrrhizin-only compared to mice receiving injections of 1% DMSO in PBS. Sorafenib-glycyrrhizin and glycyrrhizin alone show similar reduction in AFP (p = 0.5) compared to 1% DMSO in PBS (p = 0.03 and p = 0.001, Tukey's multiple comparisons test). p = 0.016 two-way ANOVA. n = 12 independent animals for sorafenib-glycyrrhizin, n = 7 independent animals for DMSO, n = 5 independent animals for glycyrrhizin. Lines corresponds to mean values; error bars represent one standard deviation.



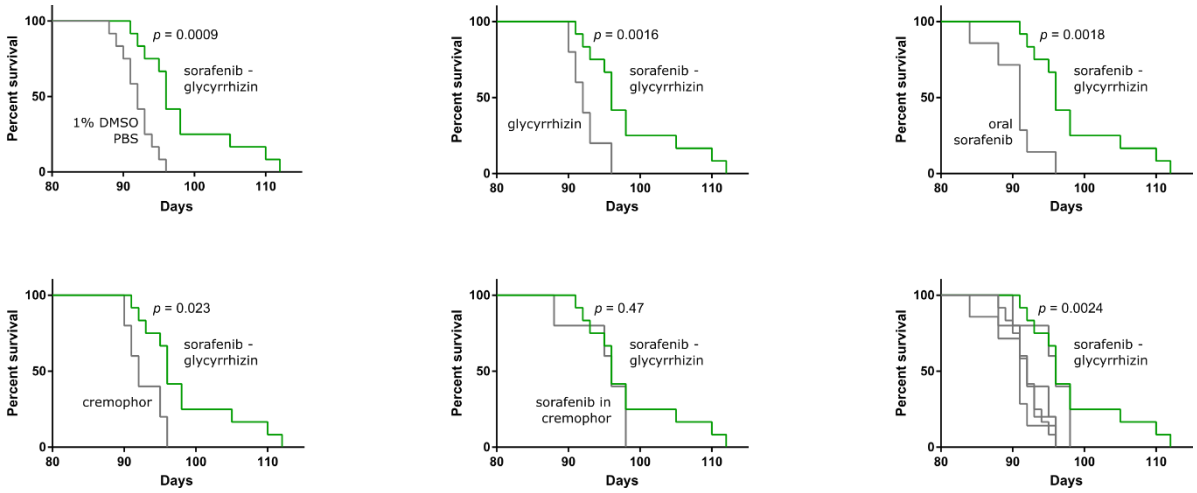
Supplementary Figure 15: AFP levels of sorafenib-cremophor control treatment and cremophor only control. No significant difference was observed between 1% DMSO in PBS and sorafenib-cremophor (p = 0.99) or cremophor only (p = 0.35, Tukey's multiple comparisons test). The sorafenib-glycyrrhizin nanoparticles performed superior compared to the sorafenib-cremophor control treatment (p = 0.0052). Two-way ANOVA p=0.038. n = 12 independent animals for sorafenib-glycyrrhizin, n = 7 independent animals for DMSO, n = 5 independent animals for sorafenib-cremophor and cremophor. Lines corresponds to mean values; error bars represent one standard deviation.



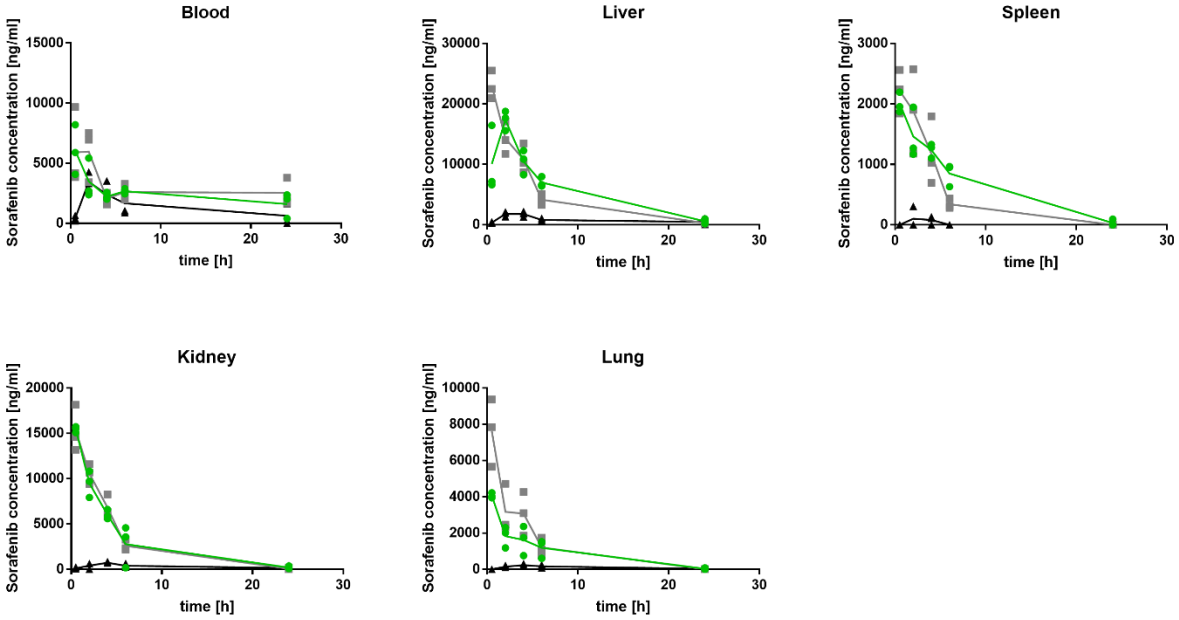
Supplementary Figure 16: Mice treated with sorafenib-glycyrrhizin particles had no significant tumour growth as indicated by no increase in bodyweight after four weeks of treatment (p = 0.79, one-sample, two tailed T test) while mice treated with oral sorafenib or 1% DMSO in PBS control showed significantly increased tumour burden (*p* = 0.001 and *p* = 0.007, one-sample, two-tailed T-test), p = 0.0063 one-way ANOVA, p = 0.0032 Dunnett's multiple comparisons test, n = 7 independent animals. Lines corresponds to mean values; error bars represent one standard deviation.



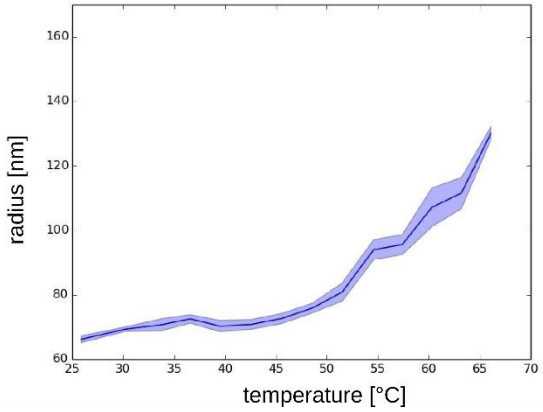
Supplementary Figure 17: Percentage weight increase due to tumour burden in control treatments. No significant difference was observed. One-way ANOVA *p*=0.27, *p* = 0.38, 0.59, 0.99, 0.29 for glycyrrhizin, oral, sorafenib in cremophor, cremophor according to Dunnett's multiple comparisons test against 1% DMSO in PBS. n = 7 independent animals for control and oral, n = 5 independent animals for glycyrrhizin, sorafenib in cremophor, and cremophor. Lines corresponds to mean values; error bars represent one standard deviation.



Supplementary Figure 18: Survival curves for control treatments compared to sorafenib-glycyrrhizin. Sorafenib-glycyrrhizin significantly extended morbidity-free survival (Logrank test $p < 0.05$) compared to all controls except for 30 mg / kg sorafenib dissolved in cremophor-ethanol IV injected. Last panel shows all controls in one figure for comparison. n = 12 independent animals for sorafenib-glycyrrhizin and 1%DMSO in PBS, n = 7 independent animals for oral sorafenib, n=5 independent animals for cremophor and sorafenib in cremophor.



Supplementary Figure 19: Sorafenib tissue accumulation in healthy mice after a single injection of 30 mg / kg sorafenib-glycyrrhizin (green), sorafenib dissolved in cremophor-ethanol (grey) or oral sorafenib (black). n=3 independent animals per treatment and timepoint. Lines corresponds to mean values.



Supplementary Figure 20: Temperature-dependent growth of sorafenib-indomethacin particles. Line indicates mean of n=3 independent samples. Shaded area corresponds to one standard deviation.