

Supporting Information:

**Non-specific yet selective interactions contribute
to small molecule condensate partitioning
behavior**

Cong Wang,[†] Henry R. Kilgore,^{*,†,¶} Andrew P. Latham,[‡] and Bin Zhang^{*,†}

[†]*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA*

[‡]*Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, Quantitative Biosciences Institute, University of California, San Francisco, San Francisco, CA, USA*

[¶]*Whitehead Institute for Biomedical Research, Cambridge, MA, USA*

E-mail: binz@mit.edu; hkilgore@wi.mit.edu

Contents

Details for Multiscale Simulations	S-4
MOFF simulations	S-4
MARTINI simulations	S-8
Atomistic simulations	S-9
Sequences of proteins	S-10
SMILES representations of small-molecule drugs	S-10
 Details for Interaction Fingerprints Analysis	 S-11
Identification of condensate-drug interactions	S-11
Construction of condensate-drug interaction fingerprints	S-11
Calculation of similarity matrices using index-based fingerprints	S-14
Low dimensional embedding of interaction fingerprints with UMAP	S-16
 Supplemental figures	 S-18
Figure S1	S-18
Figure S2	S-19
Figure S3	S-20
Figure S4	S-21
Figure S5	S-22
Figure S6	S-23
Figure S7	S-24
Figure S8	S-25
Figure S9	S-26
Figure S10	S-27
Figure S11	S-28
Figure S12	S-29
Figure S13	S-30

Figure S14	S-31
Figure S15	S-32
Figure S16	S-33
Figure S17	S-34
Figure S18	S-35
Figure S19	S-36
Figure S20	S-37

References	S-38
-------------------	-------------

Details for Multiscale Simulations

We followed a multiscale simulation protocol to prepare equilibrated configurations for condensate simulations with atomistic force fields. Three condensate-forming proteins, including HP1 α , NPM1, and MED1, were studied. The simulation details at the three stages of the multiscale protocol are detailed below.

MOFF simulations

MOFF is a transferable coarse-grained force field optimized for both ordered and disordered proteins, utilizing the maximum entropy principle and ordered protein folding landscape.^{S1} It represents each amino acid with one coarse-grained bead, with its position defined using the α -carbon atom. Setting up MOFF simulations requires an initial structure for the protein of interest, which we obtained using I-TASSER^{S2} or RaptorX.^{S3}

The force field containing the following six components:

$$U_{\text{MOFF}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{memory}} + U_{\text{contact}} + U_{\text{electrostatics}}. \quad (\text{S1})$$

1. The bonding potential, U_{bond} , consists of harmonic potentials for nearest neighbor beads:

$$U_{\text{bond}} = \sum_i \frac{1}{2} k_{\text{bond}} (r_i - r_0)^2, \quad (\text{S2})$$

where r_i is the distance between i -th and $(i + 1)$ -th bead. The equilibrium length $r_0 = 0.38 \text{ nm}$, and force constant $k_{\text{bond}} = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$.

2. The angular potential, U_{angle} , consists of harmonic potentials for angles between nearest neighbor bonds:

$$U_{\text{angle}} = \sum_i \frac{1}{2} k_{\text{angle}} (\theta_i - \theta_{i,0})^2, \quad (\text{S3})$$

where θ_i is the i -th angle. The equilibrium angles, $\theta_{i,0}$, are measured from the input

atomistic structure and force constant $k_{\text{angle}} = 120 \text{ kJ mol}^{-1} \text{ rad}^{-2}$.

3. The dihedral potential, U_{dihedral} is defined as:

$$U_{\text{dihedral}} = \sum_i k_{\text{dihedral}} [(1 - \cos(\phi_i - \phi_{i,0})) + 0.5(1 - \cos(3(\phi_i - \phi_{i,0})))] \quad (\text{S4})$$

where ϕ_i is the i -th dihedral. The equilibrium dihedrals, $\phi_{i,0}$, are measured from the input atomistic structure and force constant $k_{\text{dihedral}} = 3.0 \text{ kJ mol}^{-1} \text{ rad}^{-2}$.

4. The memory potential, U_{memory} , also known as native pair potential, stabilizes the secondary and tertiary structures found in the input atomistic configuration. It is defined as:

$$U_{\text{memory}} = \sum_{\text{native pairs}(i,j)} \epsilon \left[5 \left(\frac{r_{ij,0}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{ij,0}}{r_{ij}} \right)^{10} \right]. \quad (\text{S5})$$

U_{memory} is limited to native pairs identified using [SMOG 2](#).^{S4} $r_{ij,0}$ is the distance between C α atoms from residue i and j measured in the input structure. We used [STRIDE](#) to identify secondary structure segments.^{S5} We employed different ϵ values for the disordered and ordered regions of proteins: (1) For the disordered regions of proteins, we included all native pairs within a single secondary structure segment, and the parameter ϵ was set to 3 kJ mol^{-1} . (2) For the ordered regions of proteins, we included all native pairs within a single ordered region, and the parameter ϵ was increased to 6 kJ mol^{-1} . (3) Native pairs between a disordered region and an ordered region were all excluded.

U_{memory} was set up for HP1 α , MED1, and NPM1 according to the rules mentioned above. HP1 α monomer contains two ordered regions: the chromo domain (residues 17-72) and the chromoshadow domain (residues 115-176). The chromoshadow domain enables dimerization.^{S6S7S8} To stabilize the dimer interface for HP1 α , besides all the native pairs founded in individual monomers, we also included native pairs between the interface of two chromoshadow domains, with ϵ set to 6 kJ mol^{-1} . NPM1 contains

two ordered regions: residues 9-122 (PDB ID: 2P1B^{S9}) and residues 243-294 (PDB ID: 2VXD^{S10}). MED is considered as a pure disordered protein without any ordered regions.

5. The contact potential, U_{contact} , represents non-bonded interactions between pairs of amino acids. It involves a repulsion term for excluded volume effect and a specific term that measures the energetic cost of bringing a pair of amino acids into contact. It is defined as:

$$U_{\text{contact}} = \sum_{i < j} \left\{ \frac{|\epsilon_{IJ}| \sigma_{IJ}^{12}}{r_{ij}^{12}} - \frac{\epsilon_{IJ}}{2} [1 + \tanh(\eta(r_0 - r_{ij}))] \right\}, \quad (\text{S6})$$

where r_{ij} represents the distance between amino acid i and j . I and J are the amino acid types of bead i and j . $\eta = 7 \text{ nm}^{-1}$ and $r_0 = 0.8 \text{ nm}$. $\sigma_{IJ} = (\sigma_I + \sigma_J)/2$ where σ_I is the size of amino acid type I . Parameters σ_{IJ} are listed in Table S7 of Latham and Zhang.^{S1} Parameters ϵ_{IJ} are listed in Table S8 of Latham and Zhang.^{S1} 1-2, 1-3, 1-4 atom pairs, and native pairs are excluded from U_{contact} .

6. $U_{\text{electrostatics}}$ represents electrostatic interactions between charged residues. It is defined based on the Debye-Hückel theory as:

$$U_{\text{electrostatics}} = \sum_{i < j} \frac{q_I q_J}{4\pi\epsilon_0 r_{ij} \epsilon(r_{ij})} \exp(-r_{ij}/\lambda_D(r_{ij})), \quad (\text{S7})$$

with

$$\epsilon(r_{ij}) = A + \frac{B}{1 + \kappa \exp(-\zeta B r_{ij})}, \quad (\text{S8})$$

and

$$\lambda_D(r_{ij}) = \sqrt{\frac{k_B T \epsilon_0 \epsilon(r_{ij})}{2 N_A e^2}}. \quad (\text{S9})$$

r_{ij} represents the distance between amino acid i and j . I and J are the amino acid types of bead i and j . q_I and q_J denote the charges of residue type I and type J whose values

are listed in Table S1 of Latham and Zhang.^{S1} ϵ_0 is the vacuum permittivity. $\lambda_D(r_{ij})$ is the Debye screening length. In equation S8, $A = -8.5525$, $B = \epsilon_{\text{water}} - A$, $\kappa = 7.7839$, and $\zeta = 0.03627 \text{ nm}^{-1}$. $\epsilon_{\text{water}} = 78.4$ is the relative permittivity constant of water. In equation S9, the variables correspond to the Boltzmann constant k_B , temperature T , the Avogadro constant N_A , the monovalent salt concentration c , and the elementary charge e , respectively. 1-2, 1-3, and 1-4 pairs are excluded from $U_{\text{electrostatics}}$.

After setting up the force field from the input atomistic structure as described above, we carried out MOFF simulations as follows. We first randomly placed protein molecules in a cubic box with dimensions of $100 \text{ nm} \times 100 \text{ nm} \times 100 \text{ nm}$ using the GROMACS^{S11,S12} `insert-molecules` tool. Each box was separately populated with 10 HP1 α dimers, 14 NPM1 molecules, and 4 MED1 molecules, ensuring a protein density of approximately 130 mg mL^{-1} once the box was compressed to $15 \text{ nm} \times 15 \text{ nm} \times 15 \text{ nm}$. The simulations began with energy minimization followed by a $0.1 \mu\text{s}$ long constant pressure, constant temperature (NPT) simulation at 150 K and 1 bar, utilizing a timestep of 10 fs. A Parrinello-Rahman isotropic barostat with a time coupling constant of 1 ps was employed during this simulation, resulting in the compression of the simulation boxes. After the compression, the length of simulation box was reduced to about 11 nm. We then adjusted its length to 15 nm manually. Subsequently, energy minimization was performed again, followed by a $0.1 \mu\text{s}$ long constant volume, constant temperature (NVT) simulation with a timestep of 10 fs. During this NVT simulation, we implemented the GROMACS simulated annealing protocol to linearly raise the temperature from 150 K to 300 K, using a time coupling constant of 100 ps. Finally, we conducted the NVT production simulation for $2 \mu\text{s}$ with a timestep of 10 fs at 300 K, using a time coupling constant of 100 ps.

The MOFF simulations were performed using GROMACS 2018.^{S12} To prepare for the force field setup, we utilized scripts from <https://github.com/ZhangGroup-MITChemistry/MOFF>.^{S1} Recently, MOFF has been implemented in the Python library `OpenABC`,^{S13} providing a user-friendly and flexible approach for setting up MOFF and other coarse-grained

models with MD simulation engine [OpenMM](#).^{S14}

MARTINI simulations

We constructed initial configurations for MARTINI simulations using the following protocol. First, we utilized [REMO](#)^{S15} (REconstruct atomic MOdel from reduced representation) to reconstruct atomistic structures from the α -carbon configurations of condensates obtained from the final frames of MOFF simulations. We then converted the all-atom structure to a MARTINI3^{S16} configuration using [martinize2](#).^{S17} The resulting condensate was solvated with MARTINI water and 0.15 M NaCl to achieve a box of 15 nm $\times$ 15 nm $\times$ 15 nm. Additional ions were added to ensure charge neutrality.

To maintain protein secondary and tertiary structures during MARTINI simulations, we introduced elastic potentials as harmonic bonds between coarse-grained beads in ordered regions. These regions were identified based on secondary structure predictions from [STRIDE](#). For HP1 α , elastic potentials were applied to pairs of CG beads within the chromoshadow domains (residues 115-176) and chromo domains (residues 17-72). Additionally, elastic potentials were introduced between two chromoshadow domains within an HP1 α dimer to stabilize the interface. For NPM1, we applied elastic potentials to residues 9-122 (PDB ID: 2P1B) and residues 243-294 (PDB ID: 2VXD). We enforced neutral termini for each protein chain.

The system equilibration followed a three-step procedure. First, steepest descent energy minimization was performed, followed by a 5 ns NVT simulation with a 10 fs timestep. Temperature was maintained at 313.15 K using velocity-rescaling thermostats, with protein and solvent coupled separately with a 1 ps time constant. Second, an NPT simulation was conducted for 10 ns with a 20 fs timestep, maintaining the same thermostat scheme. Pressure was controlled using a Berendsen barostat set to 1 bar, with a compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$ and a 5 ps time constant. Finally, a 600 ns NPT simulation was performed with a 20 fs timestep. The temperature was maintained at 313.15 K with separate coupling for

protein and solvent using velocity-rescaling thermostats with a 1 ps time constant. Pressure control utilized a Parrinello-Rahman barostat with a 10 ps time constant, a compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$, and a reference pressure of 1 bar.

All simulations were conducted using GROMACS 2022.^{S12} The double-precision version of GROMACS was used for energy minimization to resolve atom clashes, while the mixed-precision version was employed for other simulation aspects.

Atomistic simulations

We constructed atomic structures of the condensates using the configurations obtained from the final frames of MARTINI simulations. The Backward tool^{S18} was used for structure reconstruction. We solvated the atomistic structures with CHARMM-modified TIP3P water^{S19} and 150 mM NaCl to reach a box size of $15 \times 15 \times 15 \text{ nm}^3$. Additional ions were added to achieve charge neutrality.

From the reconstructed structures, we equilibrated the condensate configurations through a three-step procedure. First, we performed steepest descent energy minimization, followed by NVT simulations for 100 ps with a timestep of 1 fs. To maintain the temperature at 313.15 K, we used a Nose-Hoover thermostat, separately coupling the protein and solvent with a time constant of 1.0 ps. During these steps, positional restraints were applied to the protein to avoid large-scale conformational changes from releasing steric clashes. Force constants of $400 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ and $40 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ were used for the backbone and side chains, respectively. Second, we performed an NPT equilibration for 5 ns with a timestep of 2 fs. The temperature was controlled as in the NVT simulations. For pressure control, we employed a Parrinello-Rahman barostat with a compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$, a reference pressure of 1 bar, and a time constant of 5 ps. Finally, we carried out 500 ns NPT simulations. The hydrogen mass repartitioning method^{S20} and the LINCS algorithm^{S21} were used to constrain all bonds, allowing for a larger timestep of 4 fs. Temperature and pressure were controlled as in the second step.

Upon equilibration, we randomly replaced water molecules with small molecules in the condensate systems using the GROMACS `insert-molecules` tool. A total of 80 small molecules were added to the simulation box to facilitate the exploration of the configurational space. Then we performed an NPT simulation for 2 μ s. A velocity-rescaling thermostat separately coupled the protein and solvent to 313.15 K with a time constant of 1.0 ps. Pressure was controlled using a Parrinello-Rahman barostat with a compressibility of 4.5×10^{-5} bar $^{-1}$, a reference pressure of 1 bar, and a coupling constant of 5 ps. Molecular configurations were stored every 2 ns.

All simulations were performed using GROMACS 2022^{S12} with the CHARMM36m force field.^{S22} The force fields for small molecules were parameterized using CGenFF.^{S23,S24} For energy minimization, we utilized the double-precision version of GROMACS to resolve atomic clashes. For all other simulations, we employed the mixed-precision version of GROMACS.

Sequences of proteins

The following human proteins or protein fragments were used:

- HP1 α (UniProt ID: F7IG39)

Full length, amino acids 1-191.

- MED1 (UniProt ID: Q15648)

Amino acids 600-1581.

- NPM1 (UniProt ID: P06748)

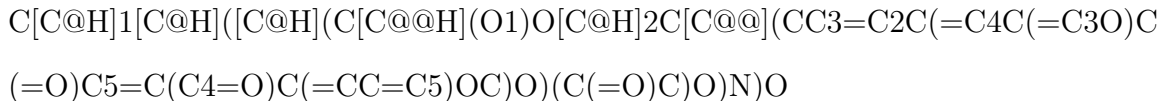
Full length, amino acids 1-294.

SMILES representations of small-molecule drugs

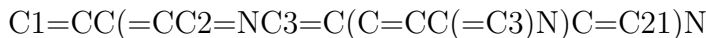
- Mitoxantrone

C1=CC(=C2C(=C1NCCNCCO)C(=O)C3=C(C=CC(=C3C2=O)O)O)NCCNCCO

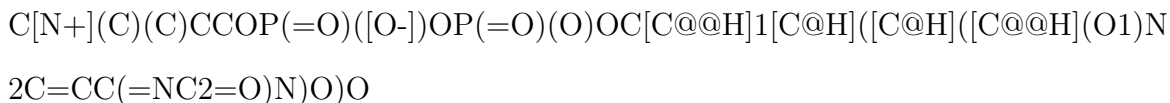
- Daunorubicin



- Proflavine



- Citcoline



Details for Interaction Fingerprints Analysis

Identification of condensate-drug interactions

For each condensate-drug system, we collected 500 snapshots uniformly spaced along the last 1 μ s segment of the atomistic simulation trajectory. As mentioned in the simulation details section, each system contains 80 small-molecule drugs, providing a total of 40,000 condensate-drug configurations for analysis. For each configuration, we determined whether any of the six interactions, including Hydrophobic, HBDonor, HBAcceptor, PiCation, PiStacking, and VdWContact, were formed between proteins and the small-molecule drugs using the Python library [ProLIF](#).^{S25} Interactions were identified if two groups of atoms in small-molecule drugs and proteins satisfy geometrical constraints based on distances and/or angles. Detailed definitions of each interaction type are listed in Table 2 of Bouysset and Fiorucci^{S25}.

Construction of condensate-drug interaction fingerprints

We defined a binary vector as the interaction fingerprint for each ligand-binding pose. This fingerprint records the types of interactions formed and identifies the set of amino acids in

contact with the small molecule. Two types of interaction fingerprints were established based on whether the residue index or residue type was recorded:

1. **Index-based fingerprints:** For the m -th condensate-drug configuration, we defined a $6 \times L$ interaction matrix, $\mathbf{M}_{\text{index},m}$, as follows:

$$\mathbf{M}_{\text{index},m} = \begin{pmatrix} n_{m,1,1} & \cdots & n_{m,1,L} \\ \vdots & \ddots & \vdots \\ n_{m,6,1} & \cdots & n_{m,6,L} \end{pmatrix} \quad 1 \leq m \leq M, \quad 1 \leq i \leq 6, \quad 1 \leq l \leq L. \quad (\text{S10})$$

Matrix element $n_{m,i,l}$ represents the number of the i -th type of interactions formed between the small-molecule and the l -th residue in the m -th configuration. Here 6 corresponds to the number of interaction types considered in the study, L is the length of the protein, and M is the total number of configurations. Notably, in condensate systems, a small molecule may form the same type of interactions with residues of the same index from different protein chains, resulting in matrix element $n_{m,i,l}$ being greater than 1.

We reshaped the interaction matrix $\mathbf{M}_{\text{index},m}$ into a one-dimensional vector $\mathbf{v}_m$ defined as:

$$\mathbf{v}_m = (n_{m,1}, \cdots, n_{m,j}, \cdots, n_{m,6L}) \quad 1 \leq j \leq 6L. \quad (\text{S11})$$

Then we binarized the vector $\mathbf{v}_m$ to obtain the final index-based fingerprint through the following steps:

- (1) For the j -th element $n_{m,j}$, we determined the maximum value, N_j , across all the vectors $\{\mathbf{v}_m\}_{m=1}^M$, as:

$$N_j = \max\{n_{m,j} : 1 \leq m \leq M\} \quad (\text{S12})$$

(2) We substituted $n_{m,j}$ in vector $\mathbf{v}_m$ with N_j binary bits, where the first $n_{m,j}$ bits are set to one, and the remaining $N_j - n_{m,j}$ bits are set to zero.

$$n_{m,j} \Rightarrow \overbrace{1, \dots, 1}^{n_{m,j}}, \overbrace{0, \dots, 0}^{N_j - n_{m,j}} \quad (\text{S13})$$

(3) We repeated step (1) and step (2) for all the elements in the vector $\mathbf{v}_m$ to determine the binary index-based fingerprints $\mathbf{fp}_m$ as:

$$\begin{aligned} \mathbf{fp}_m &= (\overbrace{1, \dots, 1}^{n_{m,1}}, \overbrace{0, \dots, 0}^{N_1 - n_{m,1}}, \dots, \overbrace{1, \dots, 1}^{n_{m,j}}, \overbrace{0, \dots, 0}^{N_j - n_{m,j}}, \dots, \overbrace{1, \dots, 1}^{n_{m,6L}}, \overbrace{0, \dots, 0}^{N_{6L} - n_{m,6L}}) \\ &= (d_{m,1}, \dots, d_{m,k}, \dots, d_{m, \sum_j^{6L} N_j}) \quad d_{m,k} = 0 \text{ or } 1, \quad 1 \leq k \leq \sum_j^{6L} N_j \end{aligned} \quad (\text{S14})$$

2. **Residue type-based fingerprints:** Similar to the index-based fingerprints, we first defined a 6×20 interaction matrix, $\mathbf{M}_{\mathbf{AA},m}$, for the m -th configuration as:

$$\mathbf{M}_{\mathbf{AA},m} = \begin{pmatrix} n_{m,1,1} & \cdots & n_{m,1,20} \\ \vdots & \ddots & \vdots \\ n_{m,6,1} & \cdots & n_{m,6,20} \end{pmatrix} \quad 1 \leq m \leq M, \quad 1 \leq i \leq 6, \quad 1 \leq t \leq 20 \quad (\text{S15})$$

The matrix element $n_{m,i,t}$ represents the number of i -th type of interactions formed between the small-molecule drug and amino acids of t -th type in the m -th configuration. Here 6 is the number of interaction types, 20 is the number of amino acid types, and M is the total number of configurations. A small molecule may form the same type of interactions with multiple residues of same type, resulting in matrix element $n_{m,i,t}$ being greater than 1. As before, we flattened matrices $\mathbf{M}_{\mathbf{AA},m}$ to a one-dimensional vector $\mathbf{v}_m$:

$$\mathbf{v}_m = (n_{m,1}, \dots, n_{m,j}, \dots, n_{m,120}) \quad 1 \leq j \leq 120. \quad (\text{S16})$$

Finally, we similarly binarized $\mathbf{v}_m$ and obtained the binary residue type-based fingerprints $\mathbf{fp}_m$ as:

$$\begin{aligned}\mathbf{fp}_m &= (\overbrace{1, \dots, 1}^{n_{m,1}}, \overbrace{0, \dots, 0}^{N_1 - n_{m,1}}, \dots, \overbrace{1, \dots, 1}^{n_{m,j}}, \overbrace{0, \dots, 0}^{N_j - n_{m,j}}, \dots, \overbrace{1, \dots, 1}^{n_{m,120}}, \overbrace{0, \dots, 0}^{N_{120} - n_{m,120}}) \\ &= (d_{m,1}, \dots, d_{m,k}, \dots, d_{m, \sum_j^{120} N_j}) \quad d_{m,k} = 0 \text{ or } 1, \quad 1 \leq k \leq \sum_j^{120} N_j\end{aligned}\tag{S17}$$

For residue type-based fingerprints, N_j was determined across all the condensate-drug systems to ensure a uniform fingerprint length, enabling comparison across all systems.

Calculation of similarity matrices using index-based fingerprints

To examine the dynamic nature of condensate-drug interactions, we calculated the similarity between index-based fingerprints at different timepoints. For a given small molecule, the similarity values along a simulation trajectory can be represented as a matrix. For the k -th small molecule, we computed the similarity matrix $\mathbf{S}_k$ across the trajectory as:

$$\mathbf{S}_k = \begin{pmatrix} s_{k,1,1} & \cdots & s_{k,1,T} \\ \vdots & \ddots & \vdots \\ s_{k,T,1} & \cdots & s_{k,T,T} \end{pmatrix} \quad 1 \leq k \leq 80, \quad 1 \leq i \leq T, \quad 1 \leq j \leq T, \tag{S18}$$

where $T = 500$ is the number of snapshots recorded along the simulation trajectory. The matrix element $s_{k,i,j}$ represents the Tanimoto similarity,^{S26} also known as the Jaccard similarity, between the fingerprints, $\mathbf{fp}_{k,i}$ and $\mathbf{fp}_{k,j}$ determined for the i -th and j -th snapshot. It is computed as:

$$s_{k,i,j} = \frac{\sum_{\text{all bits}} (\mathbf{fp}_{k,i} \wedge \mathbf{fp}_{k,j})}{\sum_{\text{all bits}} (\mathbf{fp}_{k,i} \vee \mathbf{fp}_{k,j})}, \tag{S19}$$

where $\mathbf{fp}_{k,i} \wedge \mathbf{fp}_{k,j}$ and $\mathbf{fp}_{k,i} \vee \mathbf{fp}_{k,j}$ represent the outcomes of the bitwise AND and OR operations between fingerprints $\mathbf{fp}_{k,i}$ and $\mathbf{fp}_{k,j}$, respectively. The value of $s_{k,i,j}$ ranges between

0 and 1, with larger values correspond to higher similarity between the fingerprints.

To quantitatively assess the number of binding sites a drug explores during a simulation, we applied a threshold to determine whether two binding sites in the i -th and j -th frames are identical:

$$s'_{k,i,j} = \begin{cases} 0 & \text{if } s_{k,i,j} < 0.6 \\ 1 & \text{if } s_{k,i,j} \geq 0.6. \end{cases} \quad (\text{S20})$$

We then computed the fraction of matrix elements with similarity values greater than 0.6 for the k -th drug molecule as follows:

$$f_k = \frac{1}{T^2} \sum_{i=1}^T \sum_{j=1}^T s'_{k,i,j}. \quad (\text{S21})$$

Numerical values for f_k range from $\frac{1}{T}$ to 1. A value of $f_k = 1$ indicates that the k -th small-molecule drug remained static in a single binding site, whereas a value of $f_k = \frac{1}{T}$ suggests that it traversed a distinct binding site in every snapshot.

To gain an intuitive understanding of the degree of localization or flexibility exhibited by a small molecule, we calculated the theoretical lower limit of f_k based on the number of binding sites n that the drug visited during the simulation. Assuming that the k -th small molecule explored n distinct binding sites over a T -frame simulation, with the drug's occupancy durations in these sites denoted as $\Delta t_1, \Delta t_2, \dots, \Delta t_n$. When the drug occupies different sites in frames i and j , $s'_{k,i,j}$ is 0; when it occupies the same site, $s'_{k,i,j}$ is 1. The

lower limit of f_k can be calculated as follows:

$$\begin{aligned}
f_k &= \frac{\Delta t_1^2 + \Delta t_2^2 + \cdots + \Delta t_n^2}{T^2} \\
&\geq \frac{\frac{1}{n}(\Delta t_1 + \Delta t_2 + \cdots + \Delta t_n)^2}{T^2} \\
&= \frac{\frac{1}{n}T^2}{T^2} \\
&= \frac{1}{n}.
\end{aligned}
\tag{S22}$$

By obtaining lower limits of f_k for different n , and given a value of f_k , we can deduce the minimum number of binding sites the drug likely explored. If $f_k < \frac{1}{n}$, the drug must have explored more than n unique binding sites. We depicted the lower limits of f_k for $n = 2, \cdots, 10$ in Figure S10. Across all condensate-drug systems, most drugs exhibited significant flexibility and did not remain confined to a few binding sites.

Low dimensional embedding of interaction fingerprints with UMAP

We utilized Uniform Manifold Approximation and Projection (UMAP)^{S27} to reduce the dimensionality of interaction fingerprints, thereby facilitating their visualization and statistical analysis. We concentrated on residue type-based interaction fingerprints due to their independence from the length of the protein chain, which enables comparisons across various condensate-drug systems.

The Python library [UMAP](#) was employed for the analysis, reducing the dimensionality to 2. We set the number of epochs for UMAP training to 1000. The Jaccard distance d_{Jaccard} , computed as 1 minus the Tanimoto similarity (Equation S19), was used to measure the distance between pairs of interaction fingerprints. Default values were maintained for other UMAP hyperparameters.

For each condensate-drug system, UMAP embedding was applied to 40,000 residue type-based fingerprints, as illustrated in Figure S12. To further investigate the impact of different drugs, we performed UMAP embedding on residue type-based fingerprints of the same pro-

tein with various small molecules, depicted in Figure S15, Figure S16, and Figure S17. Additionally, we examined the effect of proteins by applying UMAP embedding to residue type-based fingerprints of the same drug in different protein condensates, as shown in Figure 6 of the main text, Figure S18, and Figure S19.

Supplemental figures

Figure S1

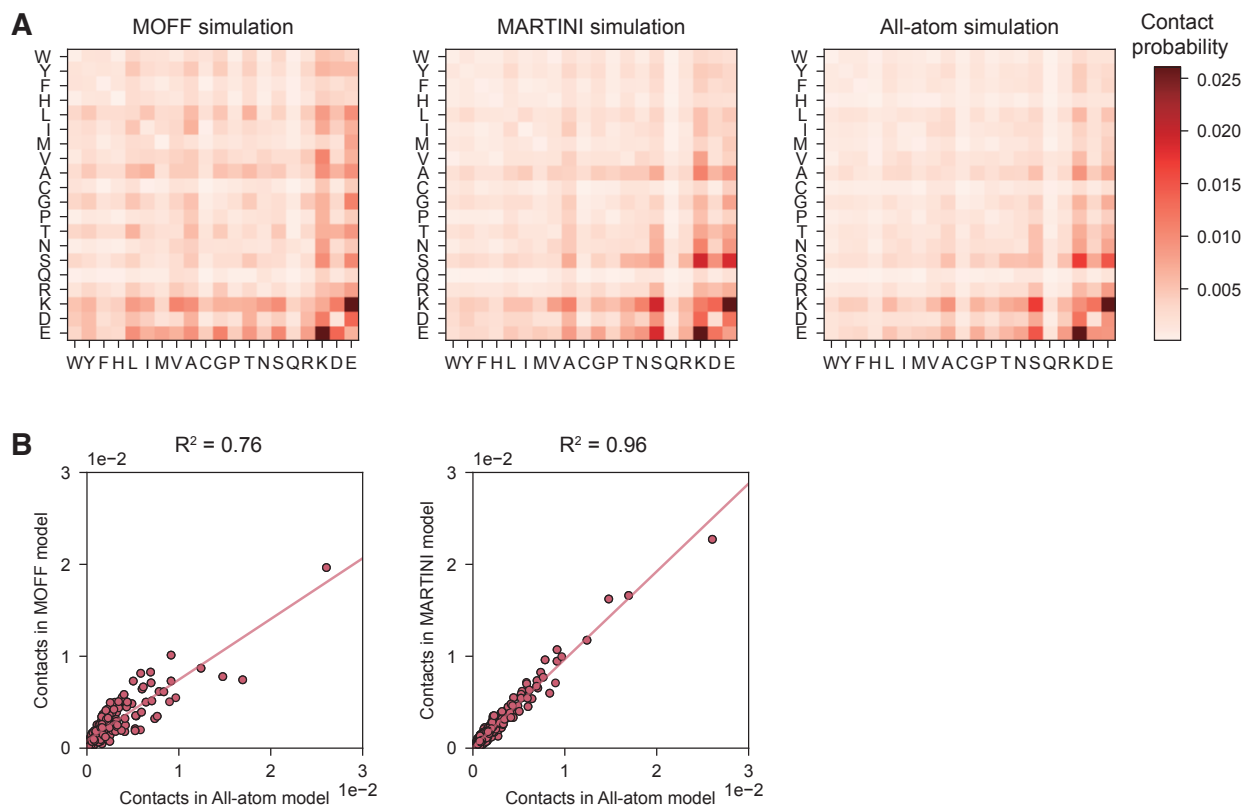


Figure S1: **Multiscale simulations for the HP1 α condensate at different resolutions produce consistent amino acid contacts.** (A) Interchain amino acid contact maps computed from simulations at different resolutions. (B) Scatter plots of interchain contacts computed from atomistic simulations versus those determined from MOFF simulations (left) and MARTINI simulations (right). The red lines are linear fits to the data, with the R^2 coefficients of the fit shown in the title.

Figure S2

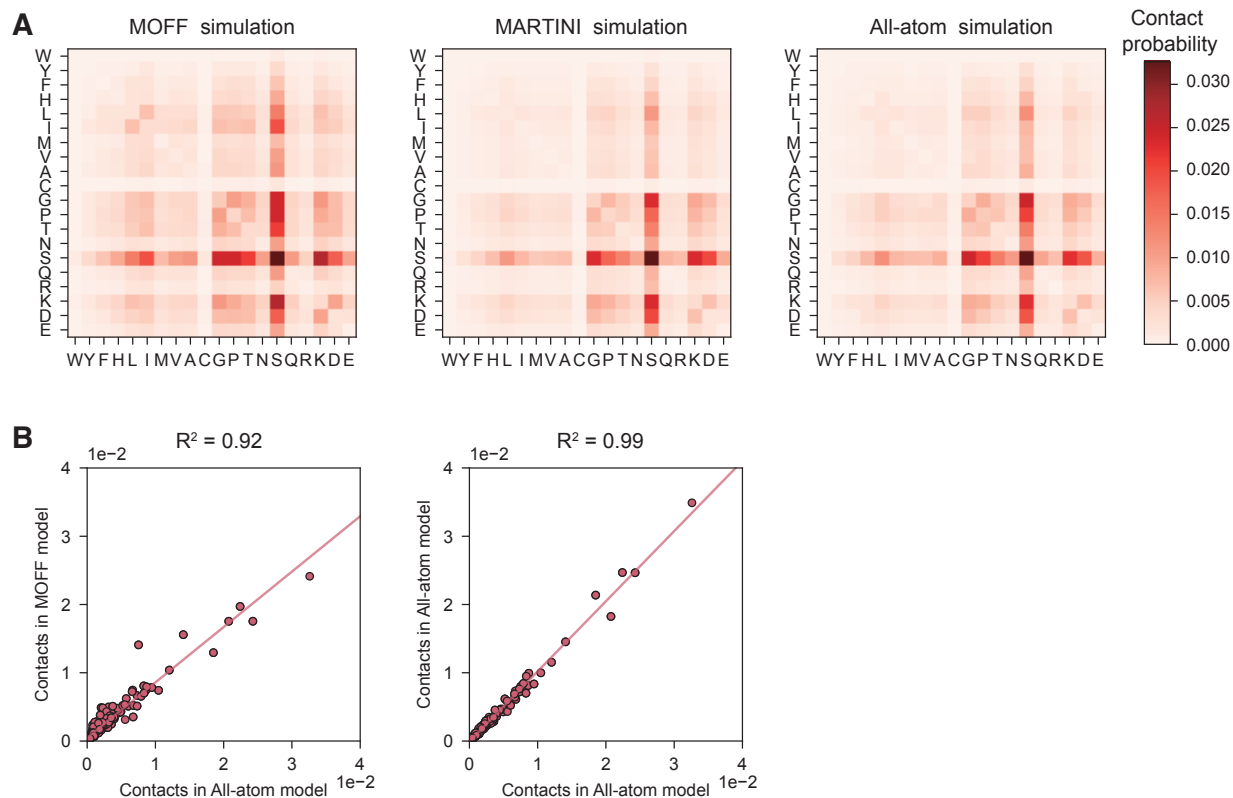


Figure S2: Multiscale simulations for the MED1 condensate at different resolutions produce consistent amino acid contacts. (A) Interchain amino acid contact maps computed from simulations at different resolutions. (B) Scatter plots of interchain contacts computed from atomistic simulations versus those determined from MOFF simulations (left) and MARTINI simulations (right). The red lines are linear fits to the data, with the R^2 coefficients of the fit shown in the title.

Figure S3

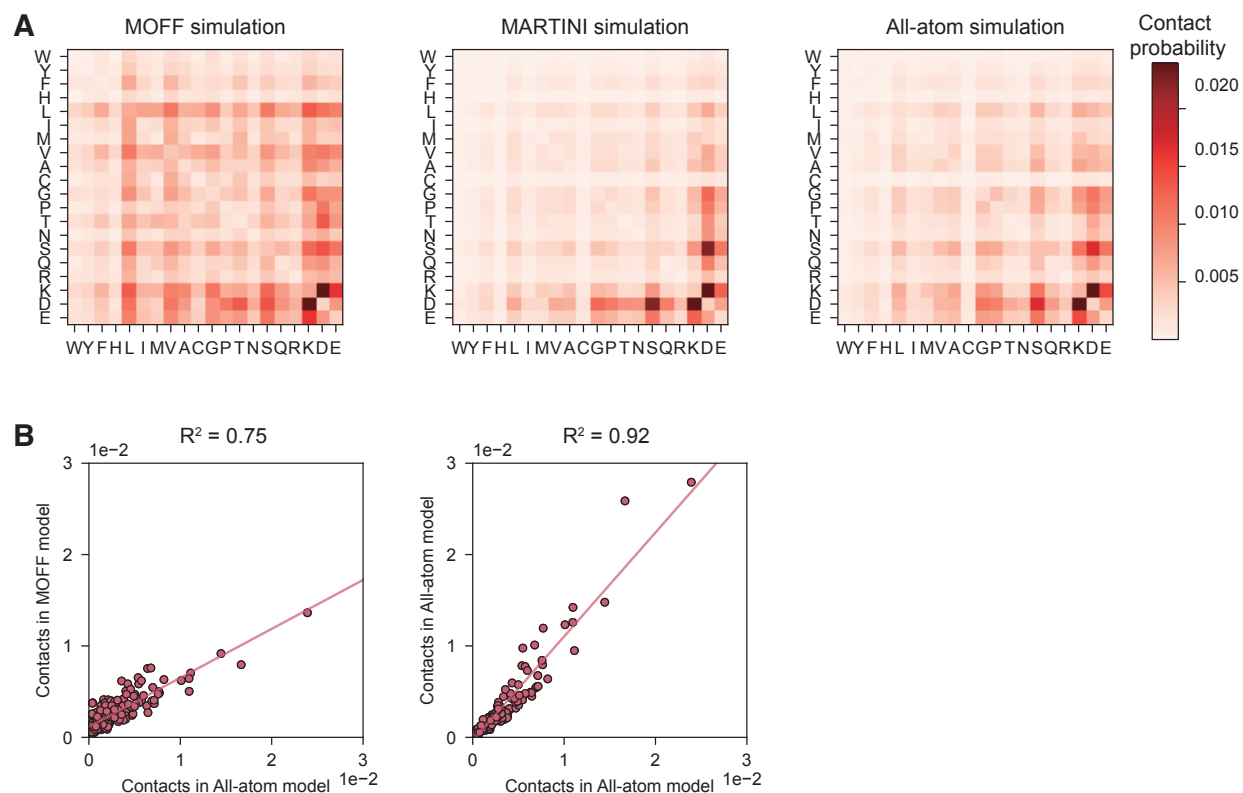


Figure S3: Multiscale simulations for the NPM1 condensate at different resolutions produce consistent amino acid contacts. (A) Interchain amino acid contact maps computed from simulations at different resolutions. (B) Scatter plots of interchain contacts computed from atomistic simulations versus those determined from MOFF simulations (left) and MARTINI simulations (right). The red lines are linear fits to the data, with the R^2 coefficients of the fit shown in the title.

Figure S4

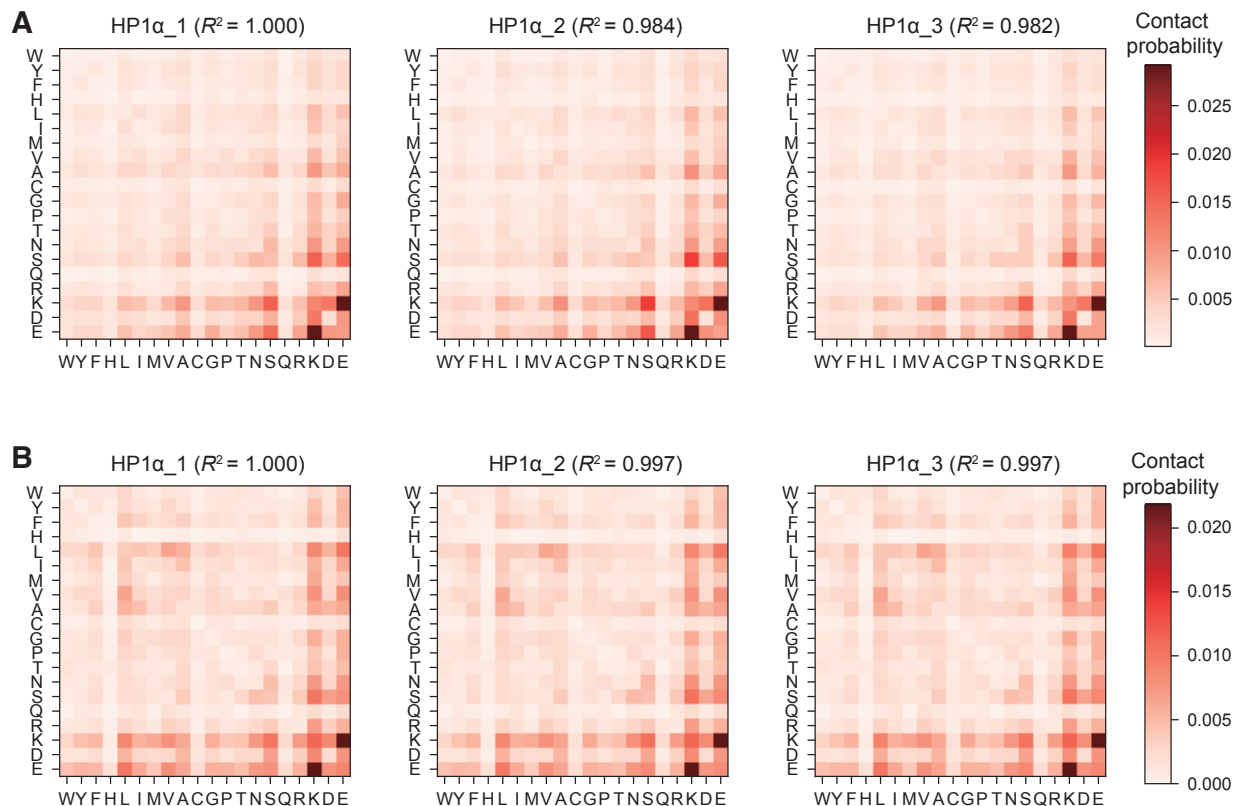


Figure S4: **Three independent atomistic simulations of the HP1 α condensate produce consistent intra and inter-chain contacts.** These simulations were initialized using different configurations produced along the MOFF simulation trajectory. (A) Inter-chain amino acid contact maps determined from three independent simulations. R^2 values of the linear fits between the respective contacts and those from the first simulation are provided in the titles. (B) Intrachain amino acid contact maps determined from three independent simulations, with the R^2 coefficients of the fit with contacts computed from the first simulation shown in the titles.

Figure S5

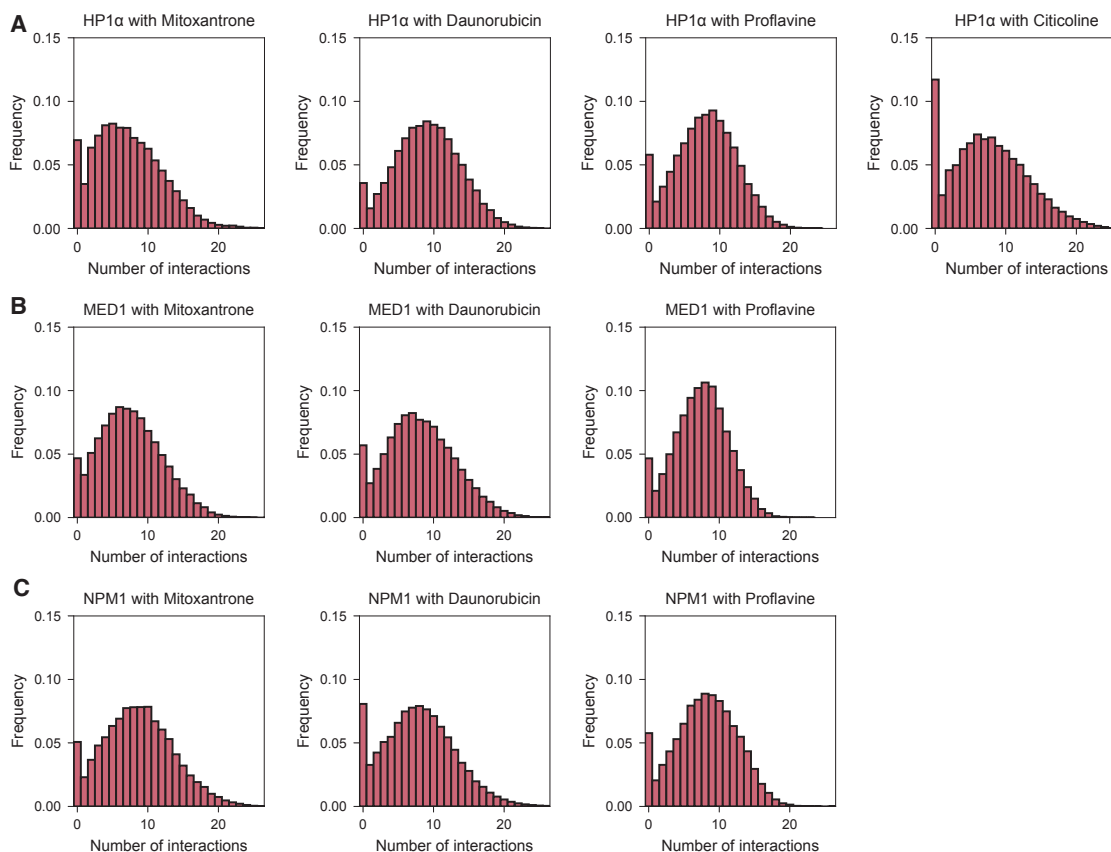


Figure S5: **Simulations uncovered diverse binding poses of small molecules inside condensates.** (A) From left to right, the plots display the probability distribution of the number of interactions detected in all binding poses uncovered from simulations of the HP1 α condensate with Mitoxantrone, Daunorubicin, Proflavine, and Citicoline. (B) From left to right, the plots display the probability distribution of the number of interactions detected in all binding poses uncovered from simulations of the MED1 condensate with Mitoxantrone, Daunorubicin, and Proflavine. (C) From left to right, the plots display the probability distribution of the number of interactions detected in all binding poses uncovered from simulations of the NPM1 condensate with Mitoxantrone, Daunorubicin, and Proflavine.

Figure S6

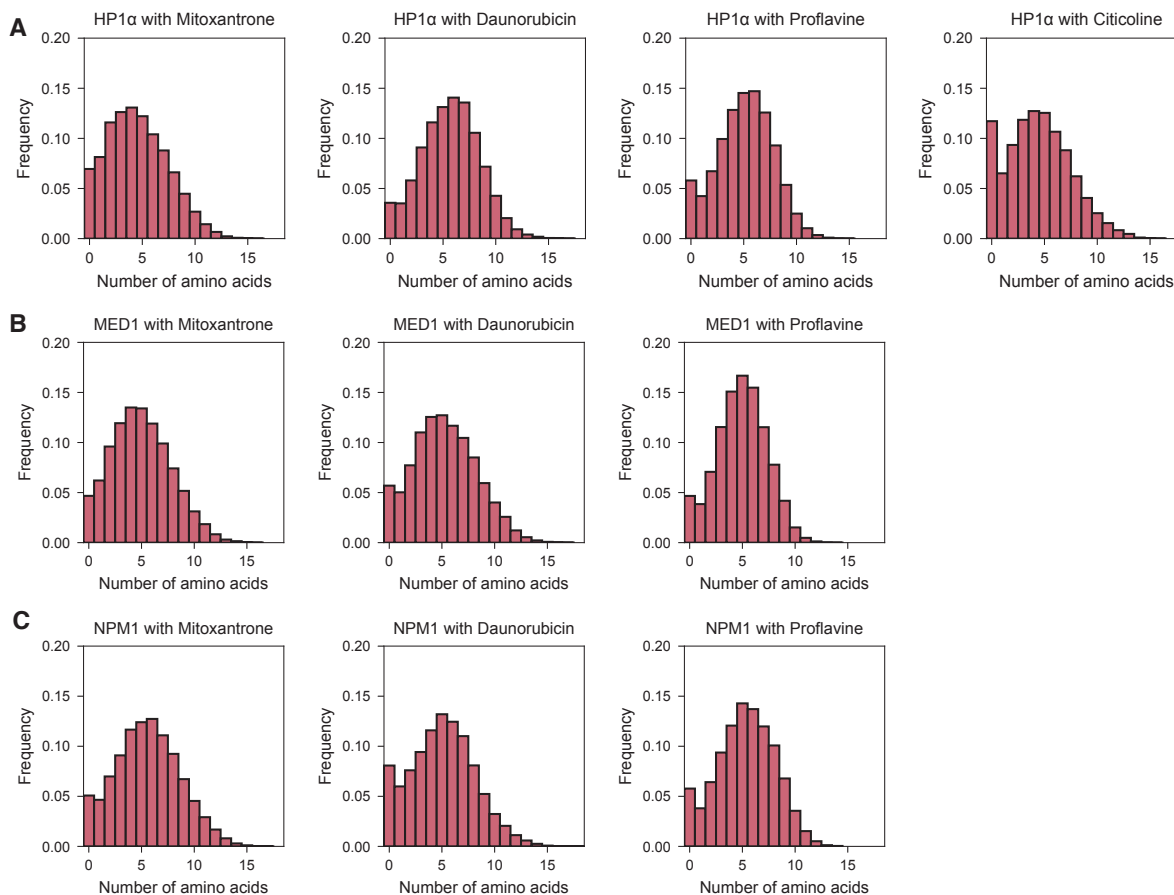


Figure S6: Simulations uncovered varying number of amino acids contribute to small molecule binding inside condensates. (A) From left to right, the plots display the probability distribution of the number of amino acids detected in all binding poses uncovered from simulations of the HP1 α condensate with Mitoxantrone, Daunorubicin, Proflavine, and Citicoline. (B) From left to right, the plots display the probability distribution of the number of amino acids detected in all binding poses uncovered from simulations of the MED1 condensate with Mitoxantrone, Daunorubicin, and Proflavine. (C) From left to right, the plots display the probability distribution of the number of amino acids detected in all binding poses uncovered from simulations of the NPM1 condensate with Mitoxantrone, Daunorubicin, and Proflavine.

Figure S7

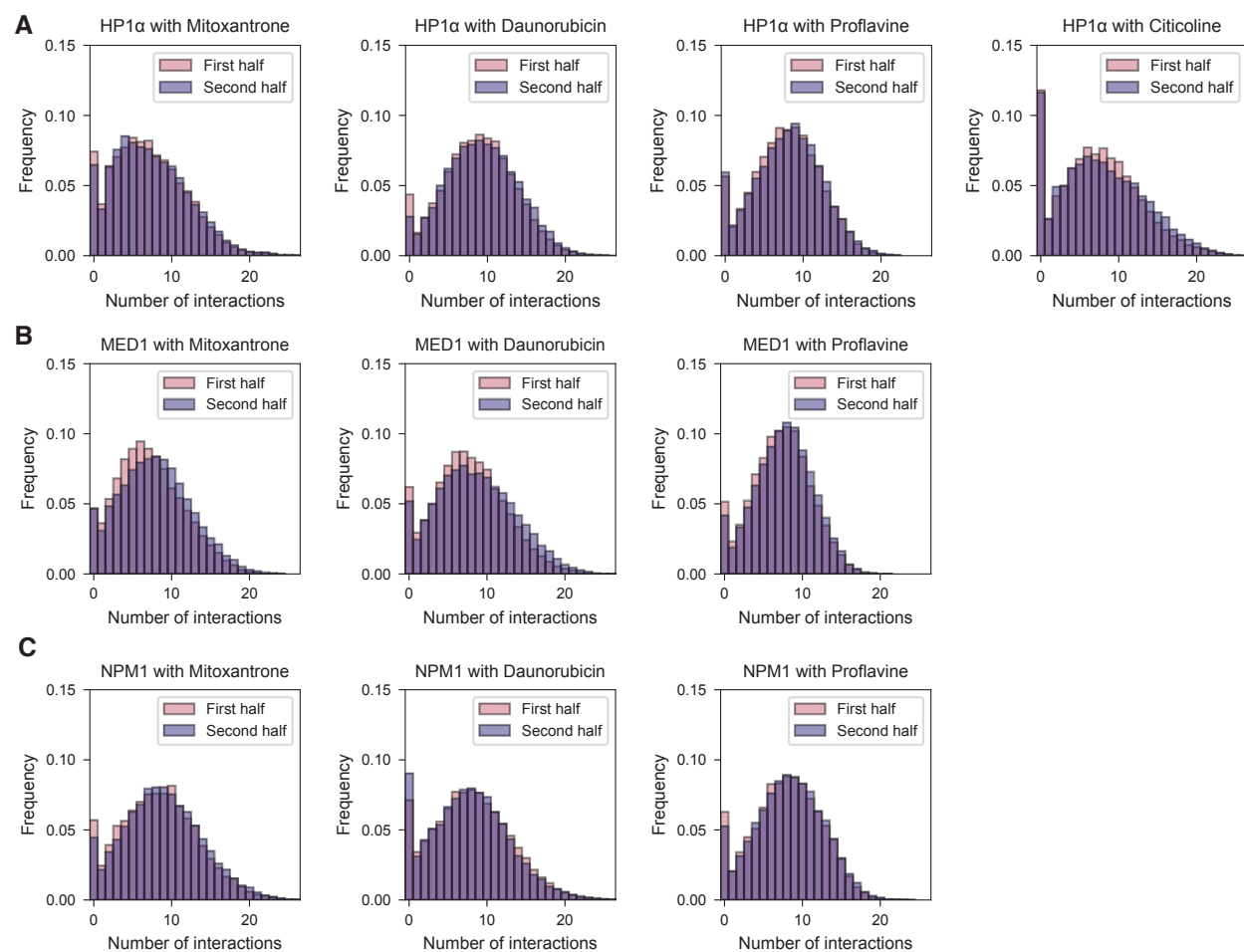


Figure S7: Probability distributions of the number of interactions in binding poses determined at different simulation times support statistical convergence. The plots are equivalent to those represented in Figure S5, except instead of the entire simulation trajectory, they were determined using the first (pink) and second (purple) half of the trajectory.

Figure S8

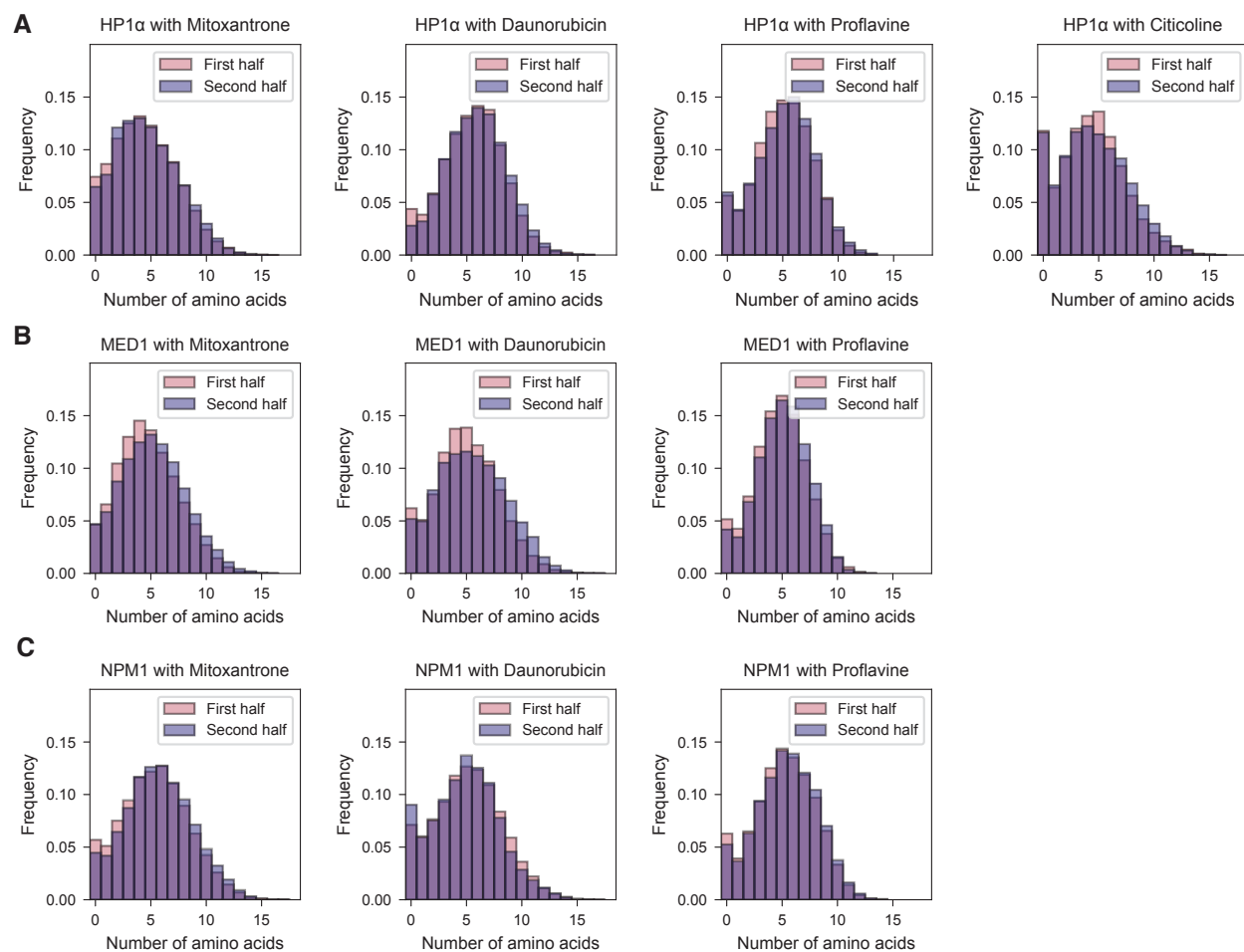


Figure S8: Probability distributions of the number of amino acids in binding poses determined at different simulation times support statistical convergence. The plots are equivalent to those represented in Figure S6, except instead of the entire simulation trajectory, they were determined using the first (pink) and second (purple) half of the trajectory.

Figure S9



Figure S9: Similarity matrices of interaction fingerprints for different Mitoxrantrone molecules in the HP1 α condensate.

Figure S10

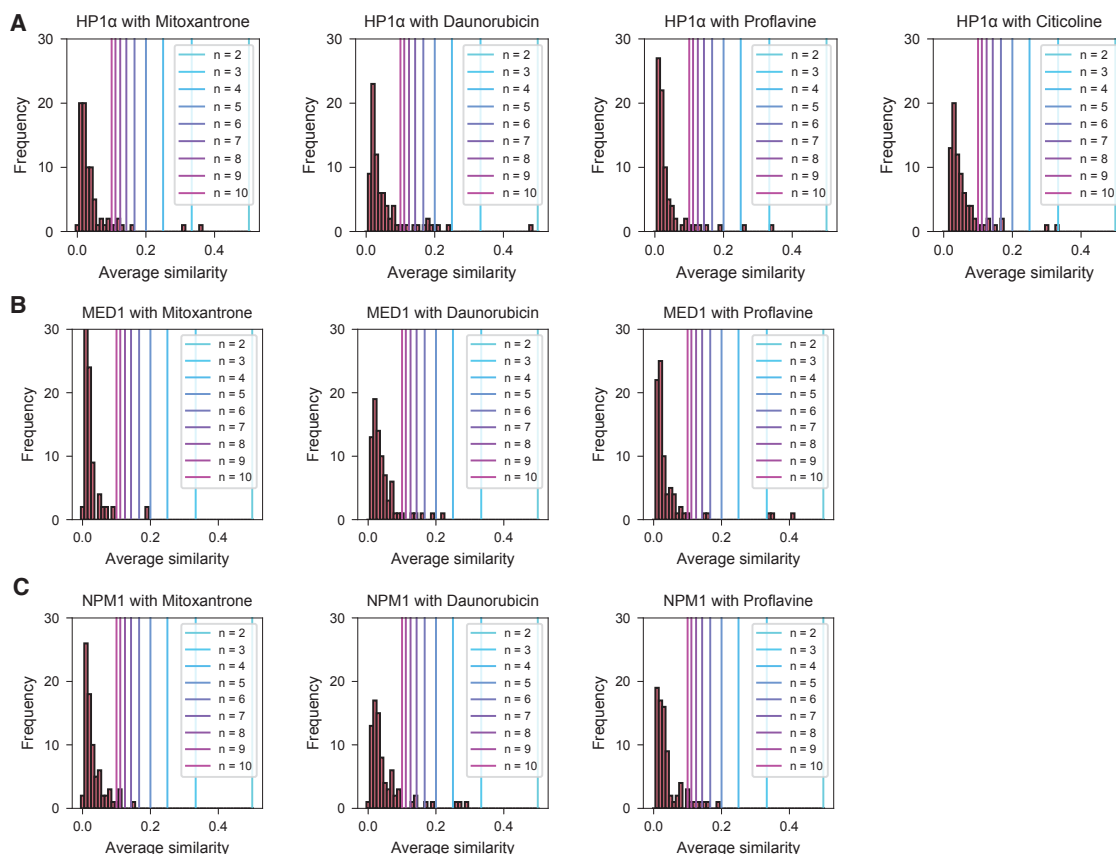


Figure S10: **The averages of similarity matrices support that small molecules explore multiple binding sites during the atomistic simulations.** The vertical lines denote theoretical lower limits of this average for a ligand exploring n unique binding poses. (A) From left to right, the plots display the probability distribution of the average value of the similarity matrix computed using interaction fingerprints for the HP1 α condensate with Mitoxantrone, Daunorubicin, Proflavine, and Citicoline. (B) From left to right, the plots display the probability distribution of the average value of the similarity matrix computed using interaction fingerprints for the MED1 condensate with Mitoxantrone, Daunorubicin, and Proflavine. (C) From left to right, the plots display the probability distribution of the average value of the similarity matrix computed using interaction fingerprints for the NPM1 condensate with Mitoxantrone, Daunorubicin, and Proflavine. See text *Section: Calculation of similarity matrices using index-based fingerprints* for further discussion.

Figure S11

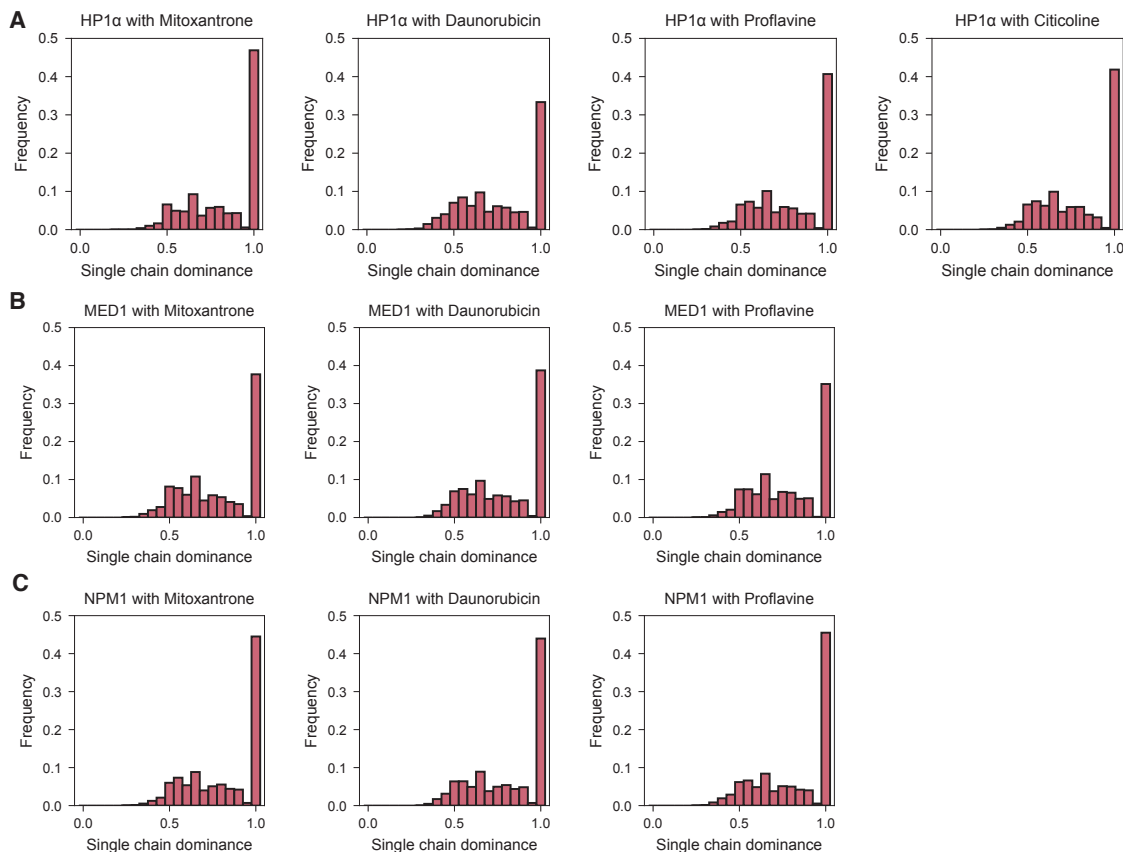


Figure S11: **Microenvironments formed by single and multi chains both contribute to condensate-ligand interactions.** (A) From left to right, the plots display the probability distribution of the single chain dominance computed using interaction fingerprints for the HP1 α condensate with Mitoxantrone, Daunorubicin, Proflavine, and Citicoline. (B) From left to right, the plots display the probability distribution of the single chain dominance computed using interaction fingerprints for the MED1 condensate with Mitoxantrone, Daunorubicin, and Proflavine. (C) From left to right, the plots display the probability distribution of the single chain dominance computed using interaction fingerprints for the NPM1 condensate with Mitoxantrone, Daunorubicin, and Proflavine.

Figure S12

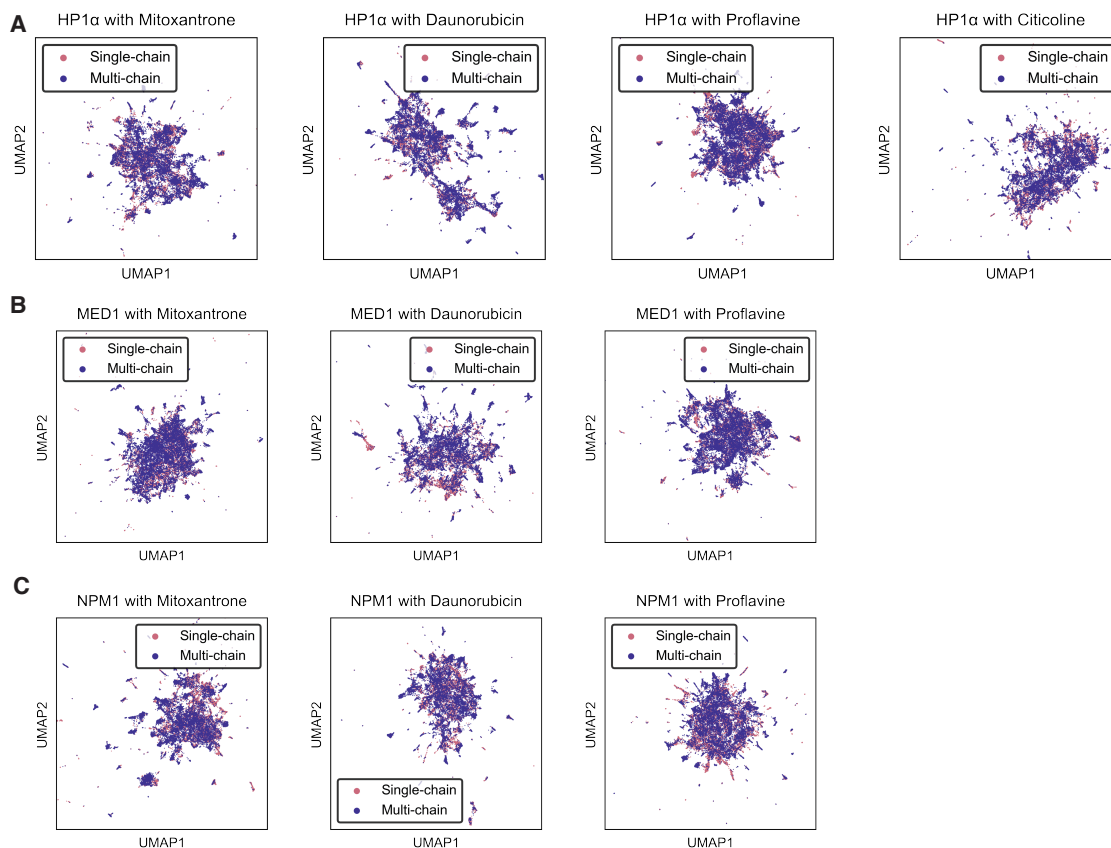


Figure S12: **Binding environments in the scaffolds formed by amino acids from the same and different chains show significant overlaps.** Binding poses with single and multiple chains are shown in pink and blue, respectively. (A) Scatter plot on the two dimensional UMAP embeddings of the interaction fingerprints for the HP1 α condensate with Mitoxantrone, Daunorubicin, Proflavine, and Citicoline. (B) Scatter plot on the two dimensional UMAP embeddings of the interaction fingerprints for the MED1 condensate with Mitoxantrone, Daunorubicin, and Proflavine. (C) Scatter plot on the two dimensional UMAP embeddings of the interaction fingerprints for the NPM1 condensate with Mitoxantrone, Daunorubicin, and Proflavine.

Figure S13

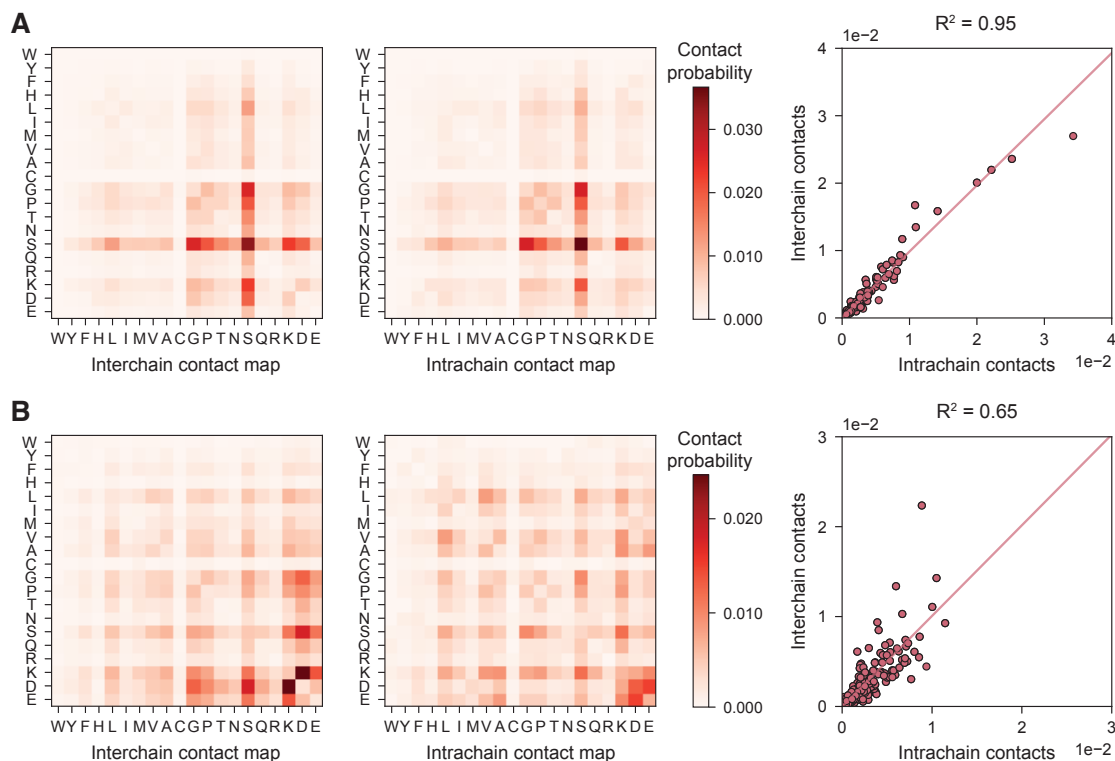


Figure S13: **Intra and inter-chain amino acid contact maps show significant correlation.** (A) Contact maps between amino acids from different (left) and the same (middle) chain computed using the MED1 condensate simulations. The right panel shows the correlation between the intra- and inter-chain amino acid contacts. The red line represents a linear fit to the data, with the R^2 coefficient for the fitting shown in the title. (B) Similar plots as in part A but for the NPM1 condensate.

Figure S14

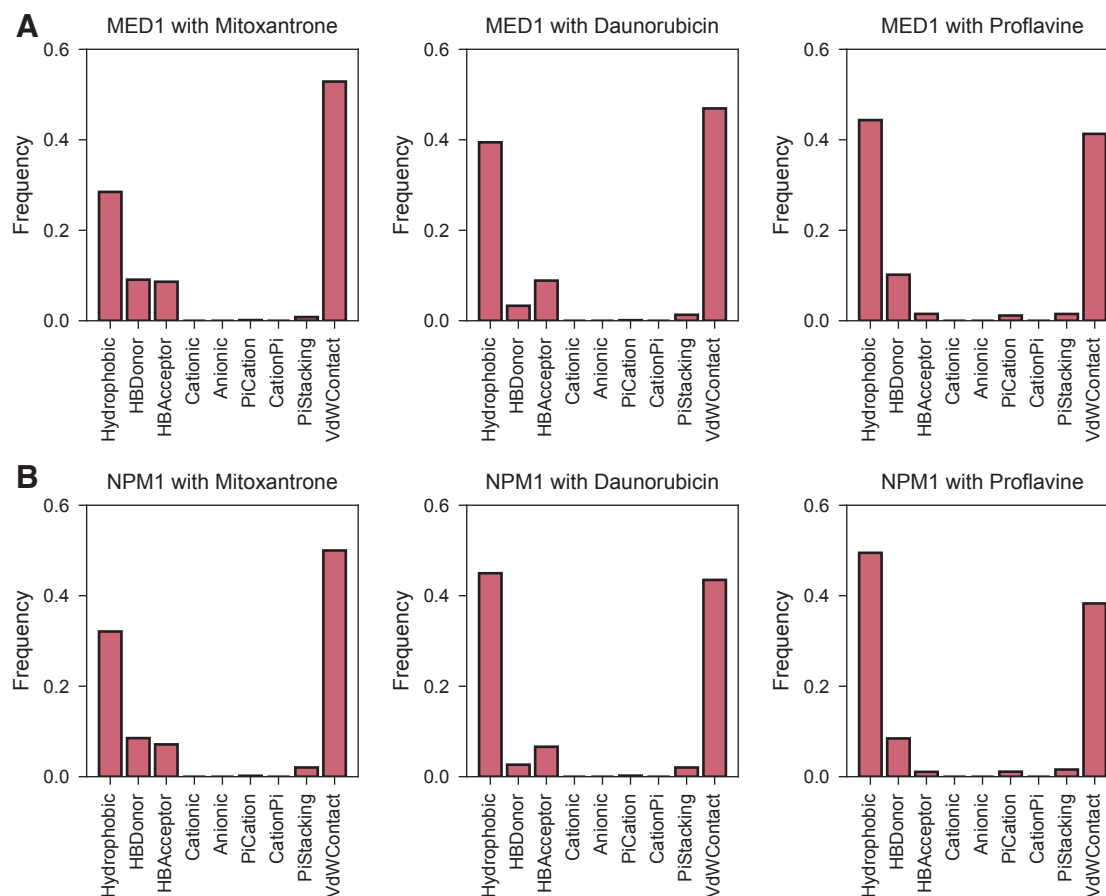


Figure S14: **Non-specific interactions contribute the most to ligand binding inside condensates.** (A) Frequency of various interactions detected in the binding pose between MED1 and different small molecules. (B) Frequency of various interactions detected in the binding pose between NPM1 and different small molecules.

Figure S15

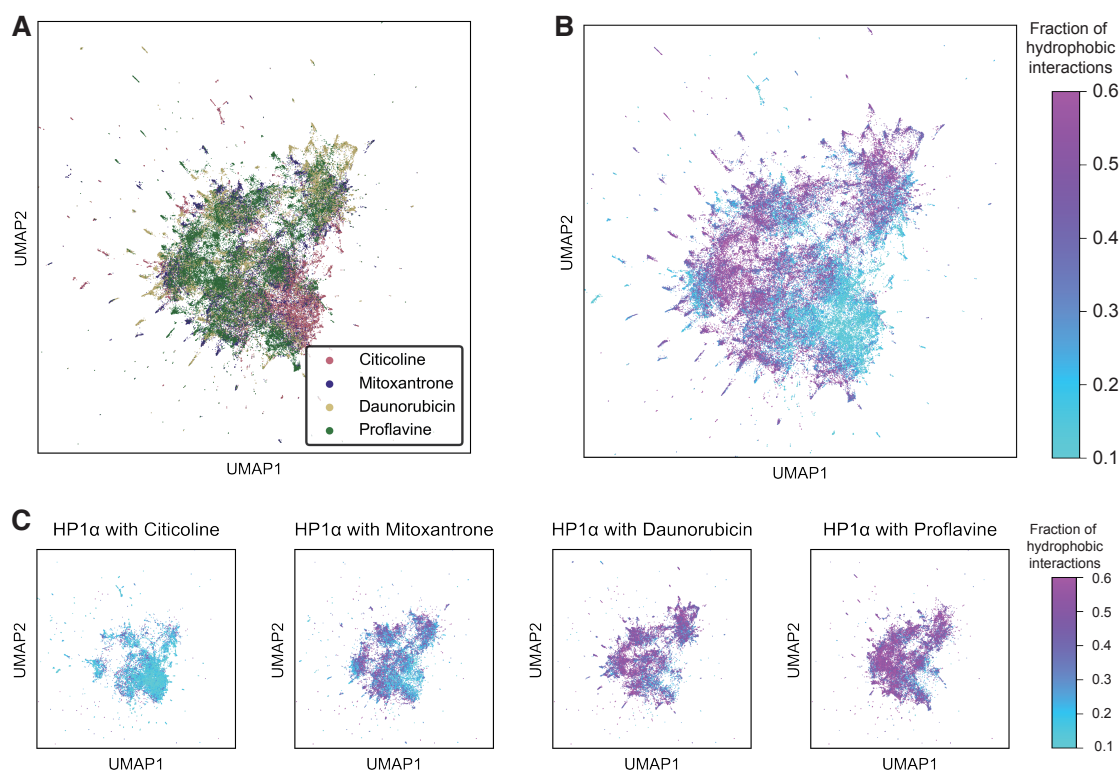


Figure S15: **Small molecules selectively interact with the heterogeneous set of microenvironments provided by the HP1 α condensate.** (A) Scatter plot on the two dimensional UMAP embeddings of the interaction fingerprints for the four small molecules in the HP1 α condensate. (B) Same plot as in part A but with the interaction fingerprints color-coded according to their fraction of hydrophobic interactions. (C) Scatter plot of the same UMAP embedding as in part B but shown separately for different small molecules.

Figure S16

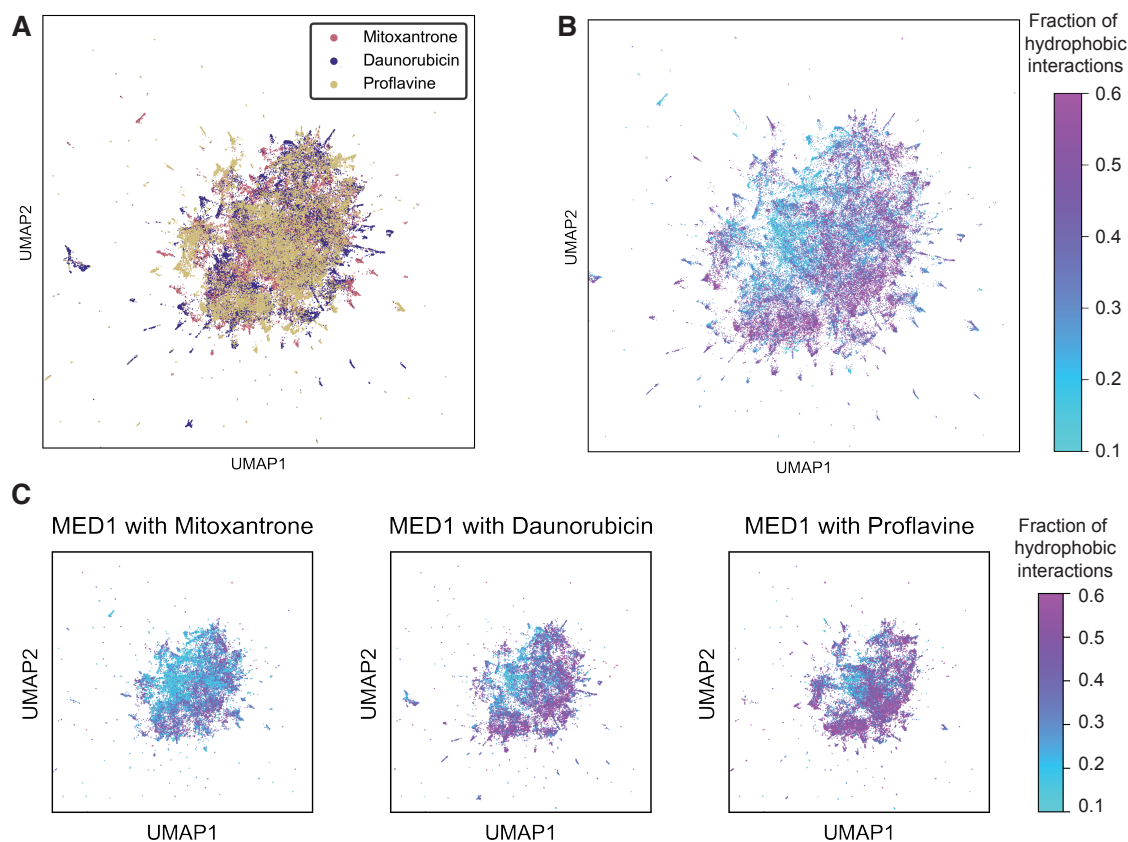


Figure S16: **Small molecules selectively interact with the heterogeneous set of microenvironments provided by the MED1 condensate.** (A) Scatter plot on the two dimensional UMAP embeddings of the interaction fingerprints for the three small molecules in the MED1 condensate. (B) Same plot as in part A but with the interaction fingerprints color-coded according to their fraction of hydrophobic interactions. (C) Scatter plot of the same UMAP embedding as in part B but shown separately for different small molecules.

Figure S17

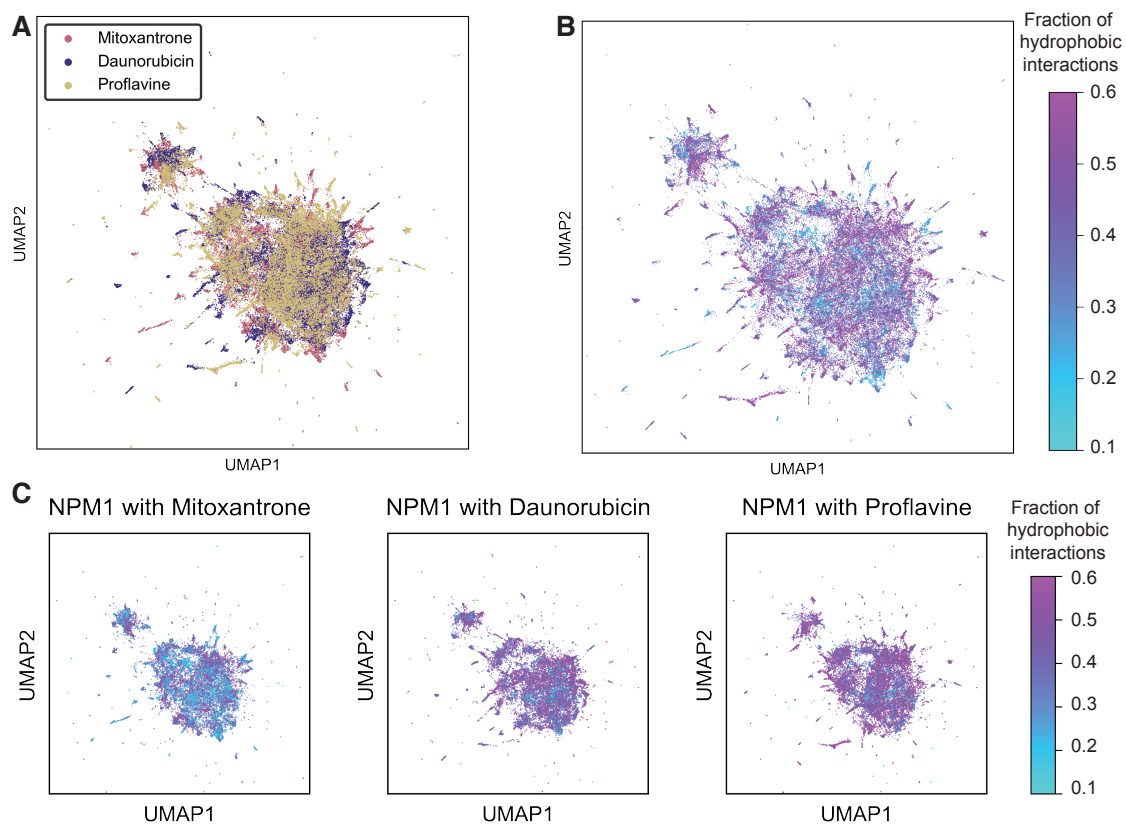
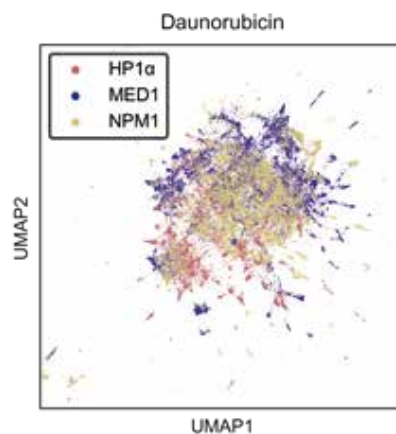


Figure S17: Similar plots as in Figure S16 but for the NPM1 condensate.

Figure S18

A



B

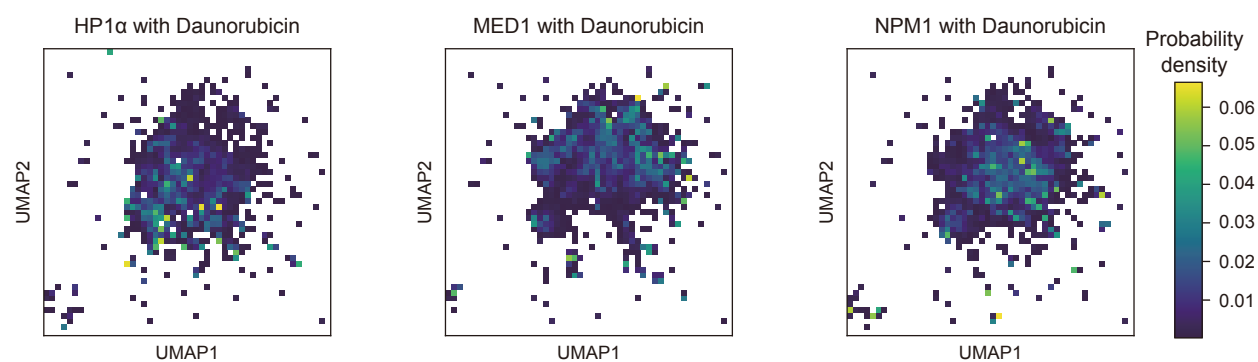
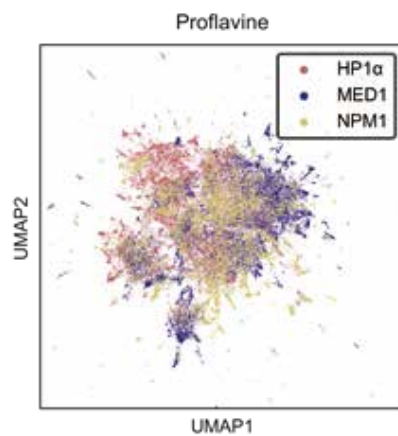


Figure S18: Similar plots as in Figure 6 of the main text but for the small molecule, Daunorubicin.

Figure S19

A



B

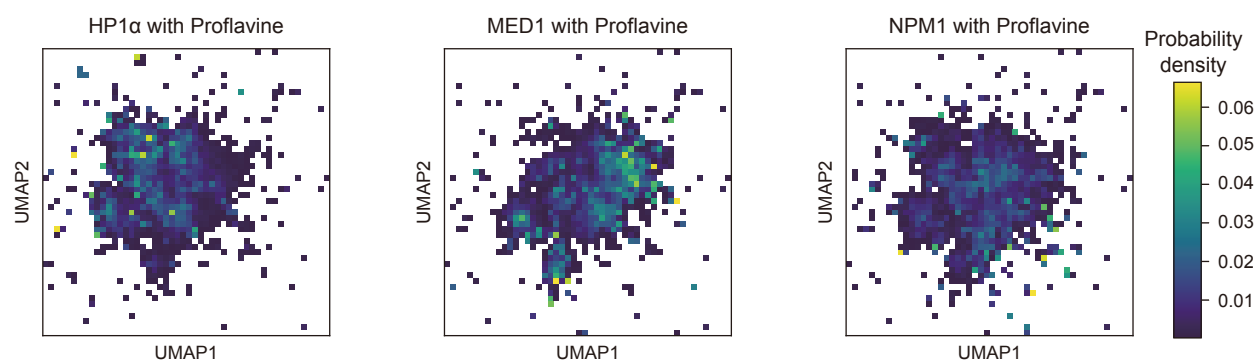


Figure S19: Similar plots as in Figure 6 of the main text but for the small molecule, **Proflavine**.

Figure S20

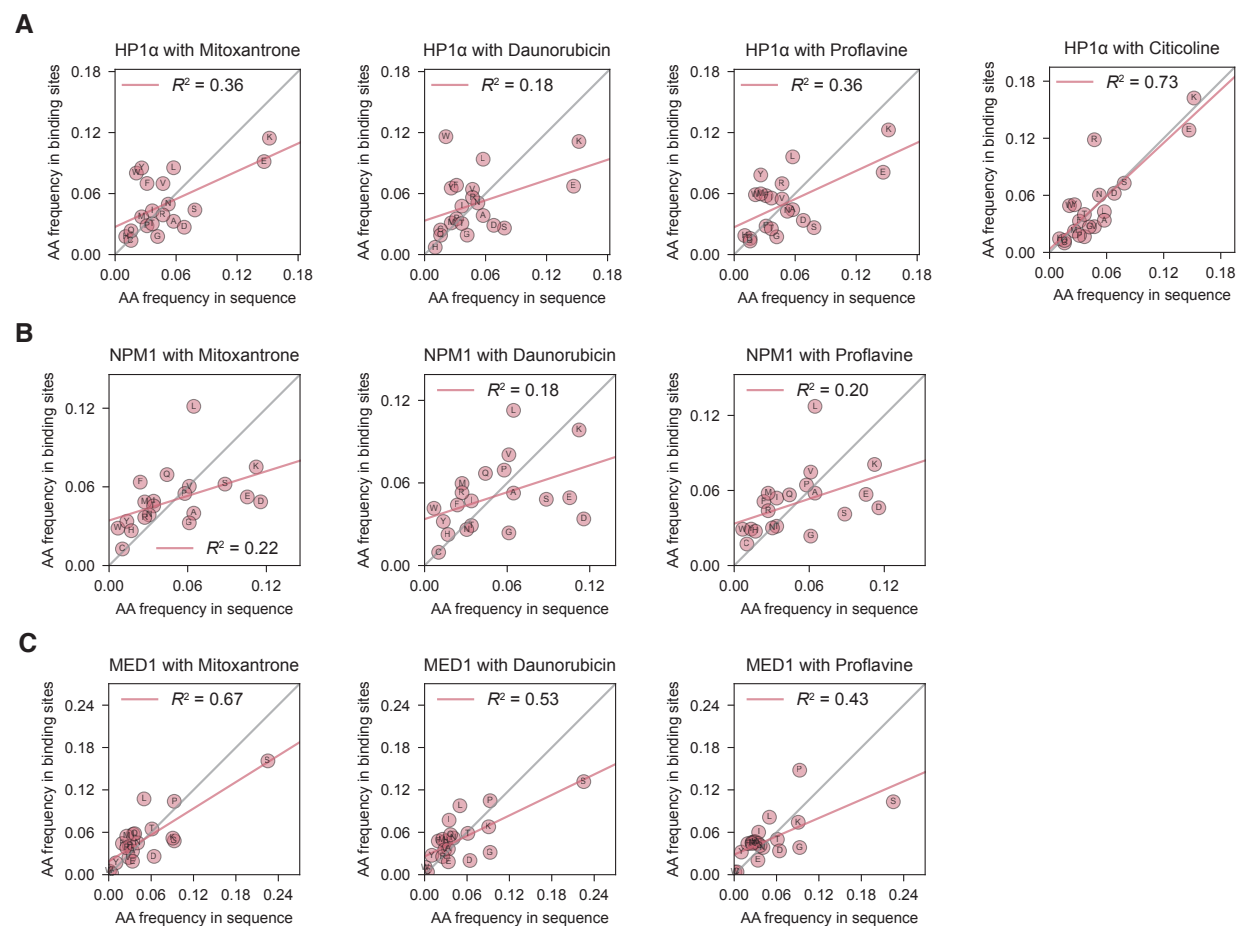


Figure S20: The frequency of amino acids in the binding poses correlates and the frequency computed directly from the protein sequence. The red lines are linear fits to the data, with their R^2 values indicated in the legends. The gray lines are diagonal references lines for visual guidance.

References

- (S1) Latham, A. P.; Zhang, B. Consistent force field captures homologue-resolved hp1 phase separation. *Journal of chemical theory and computation* **2021**, *17*, 3134–3144.
- (S2) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: protein structure and function prediction. *Nature methods* **2015**, *12*, 7–8.
- (S3) Wang, S.; Li, W.; Liu, S.; Xu, J. RaptorX-Property: a web server for protein structure property prediction. *Nucleic acids research* **2016**, *44*, W430–W435.
- (S4) Noel, J. K.; Levi, M.; Raghunathan, M.; Lammert, H.; Hayes, R. L.; Onuchic, J. N.; Whitford, P. C. SMOG 2: a versatile software package for generating structure-based models. *PLoS computational biology* **2016**, *12*, e1004794.
- (S5) Heinig, M.; Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic acids research* **2004**, *32*, W500–W502.
- (S6) Cowieson, N. P.; Partridge, J. F.; Allshire, R. C.; McLaughlin, P. J. Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. *Current Biology* **2000**, *10*, 517–525.
- (S7) Brasher, S. V.; Smith, B. O.; Fogh, R. H.; Nietlispach, D.; Thiru, A.; Nielsen, P. R.; Broadhurst, R. W.; Ball, L. J.; Murzina, N. V.; Laue, E. D. The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. *The EMBO journal* **2000**, *19*, 1587–1597.
- (S8) Thiru, A.; Nietlispach, D.; Mott, H. R.; Okuwaki, M.; Lyon, D.; Nielsen, P. R.; Hirshberg, M.; Verreault, A.; Murzina, N. V.; Laue, E. D. Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. *The EMBO journal* **2004**, *23*, 489–499.

- (S9) Lee, H. H.; Kim, H. S.; Kang, J. Y.; Lee, B. I.; Ha, J. Y.; Yoon, H. J.; Lim, S. O.; Jung, G.; Suh, S. W. Crystal structure of human nucleophosmin-core reveals plasticity of the pentamer-pentamer interface. *Proteins*. **2007-11-15**, 69.
- (S10) Grummitt, C. G.; Townsley, F. M.; Johnson, C. M.; Warren, A. J.; Bycroft, M. Structural consequences of nucleophosmin mutations in acute myeloid leukemia. *Journal of Biological Chemistry* **2008**, 283, 23326–23332.
- (S11) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer physics communications* **1995**, 91, 43–56.
- (S12) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, 1, 19–25.
- (S13) Liu, S.; Wang, C.; Latham, A. P.; Ding, X.; Zhang, B. OpenABC enables flexible, simplified, and efficient GPU accelerated simulations of biomolecular condensates. *PLoS Computational Biology* **2023**, 19, e1011442.
- (S14) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology* **2017**, 13, e1005659.
- (S15) Li, Y.; Zhang, Y. REMO: A new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks. *Proteins: Structure, Function, and Bioinformatics* **2009**, 76, 665–676.
- (S16) Souza, P. C.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M.; Wassenaar, T. A., et al. Martini 3:

- a general purpose force field for coarse-grained molecular dynamics. *Nature methods* **2021**, *18*, 382–388.
- (S17) Kroon, P. C.; Grünewald, F.; Barnoud, J.; van Tilburg, M.; Souza, P. C.; Wassenaar, T. A.; Marrink, S.-J. Martinize2 and vermouth: Unified framework for topology generation. *arXiv preprint arXiv:2212.01191* **2022**,
- (S18) Wassenaar, T. A.; Pluhackova, K.; Bockmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of chemical theory and computation* **2014**, *10*, 676–690.
- (S19) MacKerell Jr, A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B* **1998**, *102*, 3586–3616.
- (S20) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *Journal of chemical theory and computation* **2015**, *11*, 1864–1874.
- (S21) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: A linear constraint solver for molecular simulations. *Journal of computational chemistry* **1997**, *18*, 1463–1472.
- (S22) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods* **2017**, *14*, 71–73.
- (S23) Vanommeslaeghe, K.; MacKerell Jr, A. D. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *Journal of chemical information and modeling* **2012**, *52*, 3144–3154.

- (S24) Vanommeslaeghe, K.; Raman, E. P.; MacKerell Jr, A. D. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *Journal of chemical information and modeling* **2012**, *52*, 3155–3168.
- (S25) Bouysset, C.; Fiorucci, S. ProLIF: a library to encode molecular interactions as fingerprints. *Journal of cheminformatics* **2021**, *13*, 72.
- (S26) Jaccard, P. The distribution of the flora in the alpine zone. 1. *New phytologist* **1912**, *11*, 37–50.
- (S27) Trozzi, F.; Wang, X.; Tao, P. UMAP as a Dimensionality Reduction Tool for Molecular Dynamics Simulations of Biomacromolecules: A Comparison Study. *Journal of Physical Chemistry B* **2021**, *125*, 5022–5034.