nature machine intelligence

Article

https://doi.org/10.1038/s42256-022-00564-7

A multilevel generative framework with hierarchical self-contrasting for bias control and transparency in structure-based ligand design

In the format provided by the authors and unedited

SUPPLEMENTARY INFORMATION

A multilevel generative framework with hierarchical self-contrasting for bias control and transparency in structure-based ligand design

Lucian Chan,[1] Rajendra Kumar,[1] Marcel Verdonk,[1] and Carl Poelking[1,*]

[1]*Astex Pharmaceuticals, Cambridge, UK*

Below we provide additional details on the implementation of the framework, in particular regarding the neural network architectures used to represent the 2D and 3D context models (Sections 1-4). We furthermore extend the discussion of how the 1D, 2D and 3D context increase the information stored in the posterior (Sections 5-6). Finally we evaluate, using structural precedents, the realism of several of the suggestions made by the framework in relation to the experimental structure-activity relationships discussed in the main text (Section 7).

Notation. The following definitions are used: Vector embeddings are denoted by $\mathbf{x}$ and $\mathbf{h}$, which, except for the input feature vectors, are elements of $\mathbb{R}^{n\times d}$ with batch dimension n and hidden dimension $d = 256$; $\mathbf{e}$ are edge vectors describing covalent edges used in 2D convolutions; $\mathbf{t}$ are hypernode vectors representing triangular relationships between atoms in 3D. Indices a, a', a'' denote atoms of the query molecule incorporating the growth vector $\mathbf{u}$. Indices $b, b', \dots$ correspond to atoms of the 3D environment, which could belong either to the query ligand or surrounding protein. Indices i, j, k qualify motif atoms and the motif vector $\mathbf{v}$. Graph convolution operations are indicated by square brackets $G[\dots]$. Round brackets are used for both point-wise function evaluation as well as row-wise convolutions; $\circ$ denotes point-wise multiplication; $||$ means concatenation of vectors along the non-batch axis. $\mathbf{W}$ and $\mathbf{c}$ are learnable parameters of the model; γ's are attention weights.

Nonlinear activation functions are referred to using the terms: ELU (Elastic Linear Unit), ReLU (Rectifying Linear Unit), LeakyReLU (Leaky Rectifying Linear Unit), Sigmoid, SoftPlus. These are defined as follows:

$$\mathrm{ELU}(x) = \begin{cases} x & \text{if } x > 0, \\ \exp(x) - 1 & \text{else,} \end{cases} \tag{1}$$

$$\mathrm{ReLU}(x) = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{else,} \end{cases} \tag{2}$$

$$\mathrm{LeakyReLU}(x) = \begin{cases} x & \text{if } x > 0, \\ 0.01x & \text{else,} \end{cases} \tag{3}$$

$$\mathrm{Sigmoid}(x) = \frac{1}{1+\exp(-x)} \tag{4}$$

$$\mathrm{SoftPlus}(x) = \log(1+\exp x) \tag{5}$$

* carl.poelking@astx.com

1. 1D CONTEXT MODEL

The pseudo-generative 1D-conditional model G_p is defined by the set of motif vectors $\{\mathbf{v}_i\}$ with frequencies $\{f(\mathbf{v}_{i,1})\}$ that constitute the vocabulary $\mathcal{V}$:

$$\mathbf{v}_i \sim p^{\mathrm{1D}}(\mathbf{v}_i|\mathcal{V}) = \frac{f(\mathbf{v}_{i,1})}{\sum_j f(\mathbf{v}_{j,1})}. \tag{6}$$

Here the representation $\mathbf{v}_{i,1}$ of motif m_i is simply a hash of the motif's 2D structure, including a label indicating the hinge atom. Note that this model samples motifs independently of the 2D structure of the query molecule and growth vector.

Null metrics. G_p achieves considerable enrichment over a uniform baseline G_0 as illustrated in tables S1 and S2. Beyond serving as a contrastive baseline when training the 2D context model, G_p thus provides us with a set of *null* metrics against which 2D and 3D performance should be offset.

	AUC
ERD3	0.9884 ± 0.0005
PDBL	0.9893 ± 0.0007

Table S1. **Null enrichment.** AUC measured for the 1D pseudo-generative model G_p over 10000 reconstructions in the ERD3 (Enamine REAL) and PDBL (PDB ligand) sets. The negative baseline examples were drawn from the uniform G_0. Note that the AUC is independent of how many ($n \geq 1$) negative baseline samples are considered per reconstruction.

	Top-1^a	Top-1^b	Top-8^a	Top-8^b
ERD3	28.7	10.9 ± 0.2	63.2	42.4 ± 0.4
PDBL	22.9	11.6 ± 0.2	69.2	48.7 ± 0.2

Table S2. **Null accuracy.** Top-k% accuracy ($k \in 1, 8$) and associated standard error measured for (a): a "top-k motifs" model that statically ranks motifs based on their frequencies; and (b) the pseudo-generative 1D model (Eq. 6).

2. 2D CONTEXT MODEL

The 2D-conditional generative model consists of three components: a growth-vector prediction stage that samples a growth-vector atom a given the query ligand structure; a motif sampling stage that generates a motif $\mathbf{v}_i$ given a specific

growth vector $\mathbf{u}_a$; and a bonding model that samples a bond order/degree d (single, double, triple):

$$a \sim p(a|\mathbf{x}_l, \mathbf{e}_l), \tag{7}$$

$$\mathbf{v}_i \sim \frac{f(\mathbf{v}_{i,1})}{\sum_j f(\mathbf{v}_{j,1})} \times \frac{\alpha_2(\mathbf{v}_{i,2}; \mathbf{u}_{a,2})}{1 - \alpha^{2\mathrm{D}}(\mathbf{v}_{i,2}; \mathbf{u}_{a,2})}, \tag{8}$$

$$d \sim p(d|\mathbf{v}_i^b, \mathbf{u}_a^b). \tag{9}$$

Here, $\mathbf{x}, \mathbf{e}$ denote the input features and edge attributes of the ligand; $\mathbf{v}^b$ and $\mathbf{u}^b$ are growth- and motif-vector encodings specific to the bonding model. This section focuses on how the likelihood factor scores α_2 are represented, whereas the growth-vector and bond-order prediction models are described further below, Sec. .

Input features. Atoms (nodes) and bonds (edges) of the molecular graph are featurized using a set of chemical properties as summarized in Tab. S3 and S4. Categorical features are encoded in a one-hot format, resulting in an overall input dimension of 36 for the atom and 6 for the bond feature vectors.

Feature	Description	Dim.
Atom type	Element (one-hot)	12
Atom degree	Neighbour count (one-hot)	7
Radical electrons	Radical-electron count	1
Formal charge	Charge in e	1
Hybridization	s, sp^1-sp^3, sp^3d, sp^3d^2, other	7
Aromaticity	Aromatic indicator	1
Hydrogen count	Explicit + implicit	5
Chirality	CIP R and S indicators	2

Table S3. **Atomic input features.** Atom node attributes used by all 2D embedding networks.

Feature	Description	Dim.
Bond type	Single – triple, aromatic, conjugated	5
Ring bond	Ring indicator	1

Table S4. **Bond input features.** Covalent edge attributes used by all 2D embedding networks.

Graph architecture. Starting from the atom and bond features $\mathbf{x}$ and $\mathbf{e}$, an $\mathrm{AtomEncoder}(\mathbf{x}, \mathbf{e})$ builds 2D latent-vector representations using the following set of convolutions:

$$\mathbf{x}^{(0)} = \mathrm{LeakyReLU}\left(\mathrm{Linear}\left(\mathbf{x}\right)\right), \tag{10}$$

$$\mathbf{h}^{(1)} = \mathrm{ELU}\left(\mathrm{GA}_0\left[\mathbf{x}^{(0)}, \mathbf{e}^{(0)}\right]\right), \tag{11}$$

$$\mathbf{x}^{(1)} = \mathrm{ReLU}\left(\mathrm{GRU}\left(\mathbf{h}^{(1)}, \mathbf{x}^{(0)}\right)\right), \tag{12}$$

$$\mathbf{h}^{(n)} = \mathrm{ELU}\left(\mathrm{GA}_1\left[\mathbf{x}^{(n-1)}\right]\right), \tag{13}$$

$$\mathbf{x}^{(n)} = \mathrm{ReLU}\left(\mathrm{GRU}\left(\mathbf{h}^{(n)}, \mathbf{x}^{(n-1)}\right)\right), \tag{14}$$

$$\mathbf{x}_{2\mathrm{D}} = \mathrm{LayerNorm}\left(\mathbf{x}^{(4)}\right). \tag{15}$$

Here $\mathrm{Linear}(\mathbf{x}) = \mathbf{W}\mathbf{x} + \mathbf{b}$ is a dense linear layer with weights $\mathbf{W}$ and bias $\mathbf{b}$; $n = 2, \ldots, 4$ indexes successive convolutions over atomic neighbourhood shells; the graph operations GA_0, GA_1 are static attention mechanisms combined with a gated recurrent unit GRU (see below), as originally suggested by Xiong and co-workers for predictive modelling [1, 2]; and LayerNorm is defined by

$$\mathrm{LayerNorm}(\mathbf{x}) = \frac{\mathbf{x} - \langle \mathbf{x} \rangle}{\sqrt{\langle \mathbf{x}^2 \rangle - \langle \mathbf{x} \rangle^2}}, \tag{16}$$

where $\langle \ldots \rangle$ denotes averaging along the non-batch dimension.

The first graph operator $\mathrm{GA}_0[\mathbf{x}^{(0)}, \mathbf{e}^{(0)}]$ encapsulates bonding information within the first neighbourhood shell of an atom a:

$$\mathbf{x}^{(0)}_{\mathcal{N}(a)} = \sum_{a' \in \mathcal{N}(a)} \gamma_{aa'} \mathbf{W}' \mathbf{x}^{(0)}_{a'}. \tag{17}$$

The attention weights are calculated via

$$\mathbf{x}_{aa'} = \mathrm{LeakyReLU}\left(\mathbf{W}\left(\mathbf{x}^{(0)}_{a'} || \mathbf{e}^{(0)}_{aa'}\right)\right), \tag{18}$$

$$z_{aa'} = \mathrm{LeakyReLU}\left(\mathbf{c}_1^t \mathbf{x}^{(0)}_a + \mathbf{c}_2^t \mathbf{x}_{aa'}\right), \tag{19}$$

$$\gamma_{aa'} = \frac{\exp\left(z_{aa'}\right)}{\sum_{a'' \in \mathcal{N}(a)} \exp\left(z_{aa''}\right)}, \tag{20}$$

where $\mathbf{c}_1$ and $\mathbf{c}_2$ are network parameters.

With bonding information absorbed into the node vectors, the second graph operator $\mathrm{GA}_1[\mathbf{x}^{(0)}]$ implements convolutions along the (now unlabeled) covalent edges:

$$\mathbf{x}^{(0)\prime}_a = \gamma_{aa} \mathbf{W} \mathbf{x}^{(0)}_a + \sum_{a' \in \mathcal{N}(a)} \gamma_{aa'} \mathbf{W} \mathbf{x}^{(0)}_{a'} \tag{21}$$

The attention weights are again defined statically:

$$z_{aa'} = \mathrm{LeakyReLU}\left(\mathbf{c}^t \left(\mathbf{W}\mathbf{x}^{(0)}_a || \mathbf{W}\mathbf{x}^{(0)}_{a'}\right)\right), \tag{22}$$

$$\gamma_{aa'} = \frac{\exp\left(z_{aa'}\right)}{\sum_{a'' \in \mathcal{N}(a) \cup \{a\}} \exp\left(z_{aa''}\right)}. \tag{23}$$

A gated recurrent unit $\mathrm{GRU}(\mathbf{h}^{(n)}, \mathbf{x}^{(n-1)})$ is used to update the atom node vectors $\mathbf{x}$ based on the current node embeddings and the output of the graph convolution operators [3]:

$$\mathbf{r} = \mathrm{Sigmoid}\left(\mathrm{Linear}\left(\mathbf{h}^{(n)}\right) + \mathrm{Linear}\left(\mathbf{x}^{(n-1)}\right)\right),$$

$$\mathbf{s} = \mathrm{Sigmoid}\left(\mathrm{Linear}\left(\mathbf{h}^{(n)}\right) + \mathrm{Linear}\left(\mathbf{x}^{(n-1)}\right)\right),$$

$$\mathbf{t} = \tanh\left(\mathrm{Linear}\left(\mathbf{h}^{(n)}\right) + r \circ \mathrm{Linear}\left(\mathbf{x}^{(n-1)}\right)\right),$$

$$\mathbf{x}^{(n)} = (1 - \mathbf{s}) \circ \mathbf{t} + \mathbf{s} \circ \mathbf{x}^{(n-1)}.$$

The output of this topological atom encoder (Eq. 10-15) are d-dimensional atom embeddings that are further processed

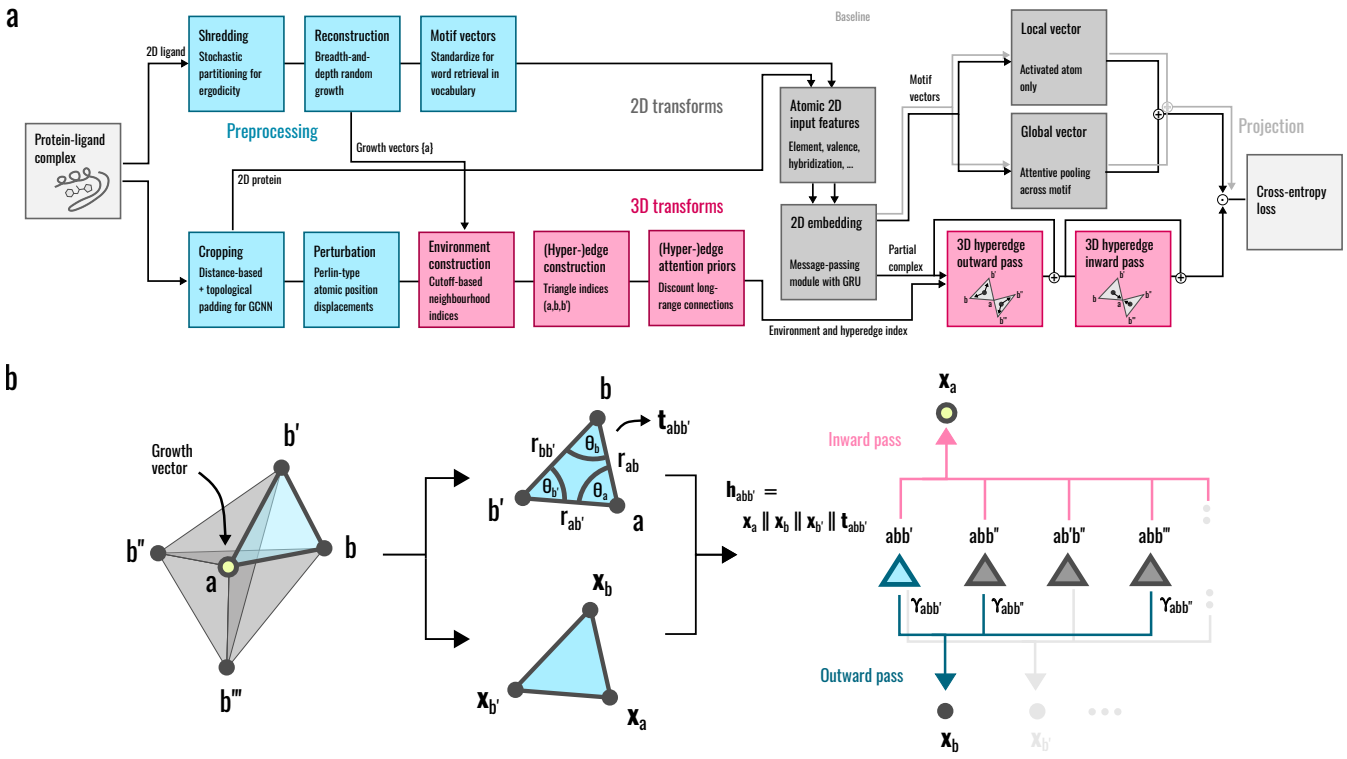

Figure S1. **Module R training flow and triangular attention mechanism.** (a) Preprocessing, 2D and 3D operations and transforms of the generative architecture during batch training of the network that encodes the 3D context $\mathbf{u}_{a,3}$. (b) Triplet construction (left), featurization of triangular (hyper-)nodes describing atom triplets (a, b, b') (center), and outward and inward attentive message passing from atom triplets to individual atoms (right). The central growth vector atom is denoted by a, environment atoms by b, b' etc.

into 2D growth and motif vectors:

$$\mathbf{x}_0^{(0)} = \mathbf{x}_{2D} \tag{24}$$

$$\mathbf{x}_0^{(k)} = \text{SoftPlus}\left(\text{Linear}\left(\mathbf{x}_0^{(k-1)}\right)\right), \tag{25}$$

$$\mathbf{x}_0 = \text{Linear}\left(\mathbf{x}_0^{(2)}\right), \tag{26}$$

$$\mathbf{x}_1^{(0)} = \mathbf{x}_{2D} \tag{27}$$

$$\mathbf{x}_1^{(k)} = \text{SoftPlus}\left(\text{Linear}\left(\mathbf{x}_1^{(k-1)}\right)\right), \tag{28}$$

$$\mathbf{x}_1 = \text{Linear}\left(\mathbf{x}_1^{(2)}\right), \tag{29}$$

$$\mathbf{u}_{a,2} = \mathbf{x}_{a,0}||\mathbf{x}_{a,1}, \tag{30}$$

$$\mathbf{v}_{i,2} = \mathbf{x}_{i,1}||\mathbf{x}_{i,0}, \tag{31}$$

where $k = 1, \ldots, 3$ is an iteration index; the subscripts 0, 1 indicate two coexisting representations, which are ultimately concatenated so as to obtain a symmetric scalar product between $\mathbf{u}_{a,2}$ and $\mathbf{v}_{i,2}$:

$$\alpha_2(\mathbf{v}_{i,2}; \mathbf{u}_{a,2}) = \text{Sigmoid}\left(\frac{\mathbf{v}_{i,2} \cdot \mathbf{u}_{a,2}}{2\sqrt{d}}\right). \tag{32}$$

This symmetrized projection treats the 2D elaboration step as a "reaction" A + B $\rightarrow$ AB with a commutative structure. The symmetric inner product means that the 2D context model is flexible enough to cross-link not only molecular fragments with motifs, but also fragments with other fragments – for example, for the purpose of fragment merging / linking.

3. 3D CONTEXT MODEL

The 3D-conditional generative model augments the 2D posterior with an additional likelihood factor that accounts for 3D structural and chemical propensities. Given a growth vector a, motif vectors are sampled from the full PQR posterior,

$$\mathbf{v}_i \sim \frac{f(\mathbf{v}_{i,1})}{\sum_j f(\mathbf{v}_{j,1})} \times \frac{\alpha^{2D}(\mathbf{v}_{i,2}; \mathbf{u}_{a,2})}{1 - \alpha^{2D}(\mathbf{v}_{i,2}; \mathbf{u}_{a,2})} \times \\ \times \frac{\alpha^{3D}(\mathbf{v}_{i,3}; \mathbf{u}_{a,3})}{1 - \alpha^{3D}(\mathbf{v}_{i,3}; \mathbf{u}_{a,3})}. \tag{33}$$

Even though various other implementations are conceivable, the architecture used here to represent the likelihood score function $\alpha_3(\mathbf{v}_{i,3}; \mathbf{u}_{a,3})$ is based on a simple hypergraph mechanism that performs attentive convolutions over 3D atom triplets. An overview is provided in Fig. S1a.

Input features. The 2D molecular graph is featurized in the same way as described above for the 2D graph-convolutional architecture (Tabs. S3 and S4). In addition to atom nodes and covalent bond edges, triangular hypernodes are generated

within a spherical volume around each growth vector (see *Methods* of the main text). These atom triplets are featurized using basic geometric features as summarized in Tab. S5. Convolutions over a local neighbourhood hypergraph, consisting of atomic nodes b within a spherical volume around a growth vector a, as well as atom triplets (a, b, b') within that same volume, are used to construct the 3D context vector $\mathbf{u}_{a,3}$ on which the distribution over motifs $\mathbf{v}_i$ is conditioned.

Feature	Description	Dim.
Node type	$b = b'$ vs $b \neq b'$	1
r_{ab}	Distance encoding (RBF)	$\lceil r_{\max}/\sigma \rceil + 1$
$r_{ab'}$	Distance encoding (RBF)	$\lceil r_{\max}/\sigma \rceil + 1$
$r_{bb'}$	Distance encoding (RBF)	$\lceil r_{\max}/\sigma \rceil + 1$
$\cos(\theta_a)$, $\sin(\theta_a)$	Angular phase	2
$\cos(\theta_b)$, $\sin(\theta_b)$	Angular phase	2
$\cos(\theta_{b'})$, $\sin(\theta_{b'})$	Angular phase	2

Table S5. **Hypernode features.** Attributes describing atom triplets (a, b, b') used by the 3D growth-vector embedding network. The growth-vector atom corresponds to atom index a. Triplets with $b \neq b'$ are accounted for twice, as (a, b, b') and (a, b', b) to guarantee permutational invariance. Here we used equispaced Gaussian radial basis functions (RBFs) for the distance encoding with width $\sigma = 1$Å, and cutoff $r_{\max} = 7.5$Å.

Hypergraph architecture. Topological embeddings are generated using the $\mathrm{AtomEncoder}$ sequence defined in Eqs. 10 to 15:

$$\mathbf{x}_{2\mathrm{D}} = \mathrm{AtomEncoder}\left[\mathbf{x}, \mathbf{e}\right] \tag{34}$$

Note, however, that even though the architecture of the 2D encoder is reused here, its weights are not shared across the model stages.

An environment index is constructed around every growth vector atom using a spherical cutoff. The 2D embeddings of the atoms of those spherical neighbourhoods are referred to by $\mathbf{x}_{\mathrm{env}}^{(0)}$, obtained by slicing $\mathbf{x}_{2\mathrm{D}}$ along the batch dimension. Furthermore denote by $\mathbf{t}$ the feature vectors (see Tab. S5) associated with the hypernodes within the atom-centered neighbhourhoods. 3D context vectors are then constructed using an $\mathrm{AtomEncoder3D}(\mathbf{x}_{\mathrm{env}}^{(0)}, \mathbf{t})$, with output $\mathbf{x}_{3\mathrm{D}}$, which consists of the following set of convolutions:

$$\mathbf{x}_{\mathrm{env}}^{(0)\,\prime} = \mathrm{ResTrans}\left(\mathbf{x}_{\mathrm{env}}^{(0)}\right), \tag{35}$$

$$\mathbf{x}_{\mathrm{env}}^{(1)} = \mathrm{TA}_{\mathrm{outward}}\left[\mathbf{x}_{\mathrm{env}}^{(0)\,\prime}, \mathbf{t}\right], \tag{36}$$

$$\mathbf{x}_{\mathrm{env}}^{(2)} = \mathrm{TA}_{\mathrm{inward}}\left[\mathbf{x}_{\mathrm{env}}^{(1)}, \mathbf{t}\right], \tag{37}$$

$$\mathbf{x}_{\mathrm{env}}^{(2)\,\prime} = \mathrm{ResTrans}\left(\mathbf{x}_{\mathrm{env}}^{(2)}\right), \tag{38}$$

$$\mathbf{x}_{3\mathrm{D}} = \mathbf{W}\mathbf{x}_{\mathrm{env}}^{(2)\,\prime}. \tag{39}$$

Here $\mathrm{ResTrans}$ is a residual transfer block, defined by

$$\mathbf{x}' = \mathrm{ELU}\left(\mathrm{Linear}\left(\mathbf{x}\right)\right), \tag{40}$$

$$\tilde{\mathbf{x}}' = \mathrm{LayerNorm}\left(\mathbf{x}'\right), \tag{41}$$

$$\mathbf{x}_{\mathrm{out}} = \mathrm{ELU}\left(\mathbf{x} + \mathrm{Linear}\left(\tilde{\mathbf{x}}'\right)\right). \tag{42}$$

$\mathrm{TA}_{\mathrm{outward}}$ and $\mathrm{TA}_{\mathrm{inward}}$ are hypergraph operators, corresponding to an *outward* and *inward* pass over atom triplets, as shown schematically in Fig. S1b. Specifically, $\mathrm{TA}_{\mathrm{outward}}$ transmits information from the ordered triplets (a, b, b') to the outermost vertex b' via:

$$\mathbf{x}_{\mu} = \mathrm{Linear}(\mathbf{x}) \text{ with } \mu \in \{0, 1, 2\}, \tag{43}$$

$$\mathbf{h}_{abb'} = \mathbf{x}_{a,0} \,||\, \mathbf{x}_{b,1} \,||\, \mathbf{x}_{b',2} \,||\, \mathbf{t}_{abb'}, \tag{44}$$

$$\mathbf{h}' = \mathrm{HGA}_{\mathrm{outward}}\left[\mathbf{x}, \mathbf{h}, \mathbf{w}\right], \tag{45}$$

$$\mathbf{x}_{\mathrm{out}} = \mathrm{ELU}(\mathbf{x} + \mathbf{h}'). \tag{46}$$

HGA is the attention mechanism that controls this information flow, relying on three inputs: the 2D atom embeddings $\mathbf{x}$, triangular hypernode embeddings $\mathbf{h}$, and hypernode-to-atom attention priors $\mathbf{w}$:

$$\mathbf{h}' = \sum_{ab \in \mathcal{H}(b')} \tilde{\gamma}_{abb'} \mathbf{W}\mathbf{h}_{abb'}, \tag{47}$$

$$\mathbf{h}_{b',abb'} = \mathrm{LeakyReLU}\left(\mathbf{W}'\mathbf{x}_{b'} + \mathbf{W}''\mathbf{h}_{abb'}\right), \tag{48}$$

$$z_{abb'} = \mathrm{LeakyReLU}\left(\mathbf{c}^t \mathbf{h}_{b',abb'}\right), \tag{49}$$

$$\gamma_{abb'} = \frac{\exp\left(z_{abb'}\right)}{\sum_{\bar{a}\bar{b} \in \mathcal{N}(b')} \exp\left(z_{\bar{a}\bar{b}b'}\right)}, \tag{50}$$

$$\tilde{\gamma}_{abb'} = \frac{\gamma_{abb'} w_{abb'}}{\sum_{\bar{a}\bar{b} \in \mathcal{H}(b')} \gamma_{\bar{a}\bar{b}b'} w_{\bar{a}\bar{b}b'}}. \tag{51}$$

$\mathcal{H}(b')$ denotes all pairs of atoms (a, b) that form a triangular hypernode (a, b, b') with atom b' as the terminal vertex. The attention priors $w_{abb'}$ impose smoothness on the attention weights as atoms leave the spherical neighbourhood shell, and furthermore discount *remote* over *close* triplets (consistent with the Jacobian $1/r^2$ of the 3D space in which we are operating):

$$w_{abb'} = f_{\mathrm{cut}}(r_{ab}) f_{\mathrm{cut}}(r_{ab'}) g(r_{ab}) g(r_{ab'}), \tag{52}$$

$$f_{\mathrm{cut}}(r) = \begin{cases} 1 & \text{if } r \leq r_{\mathrm{cut}} - \Delta, \\ 0 & \text{if } r \geq r_{\mathrm{cut}}, \\ \frac{1 + \cos\left(\pi \frac{r - r_{\mathrm{cut}} + \Delta}{\Delta}\right)}{2} & \text{else}, \end{cases} \tag{53}$$

$$g(r_{ab}) = \frac{1 - \omega(r_{ab})}{r_{ab'}^2} + \omega(r_{ab}) w_0, \tag{54}$$

$$\omega(r) = \mathrm{Sigmoid}\left(-\beta(r^2 - \rho^2)\right). \tag{55}$$

The hyperparameters of this prior are: the transition width of the cutoff function $\Delta = 0.5$Å; limiting weight factor $w_0 = 1$; radial pivot $\rho = r_{\mathrm{cut}} - 2\Delta$; and decay constant $\beta = 0.1$.

The inward pass $\mathrm{TA}_{\mathrm{inward}}$ inverts the information flow towards the central vertex a (i.e., the growth vector). $\mathrm{TA}_{\mathrm{inward}}$ resembles $\mathrm{TA}_{\mathrm{outward}}$ above, with, however, Eq. 45 replaced by

$$\mathbf{h}' = \mathrm{HGA}_{\mathrm{inward}}\left[\mathbf{x}, \mathbf{h}, \mathbf{w}\right]. \tag{56}$$

$\mathrm{HGA}_{\mathrm{inward}}\left[\mathbf{x}, \mathbf{h}, \mathbf{w}\right]$ closely resembles $\mathrm{HGA}_{\mathrm{outward}}$, with in-

dex relationships subtly rearranged according to:

$$\mathbf{x}_a = \sum_{bb' \in \mathcal{H}(a)} \tilde{\gamma}_{abb'} \mathbf{W} \mathbf{h}_{abb'}, \tag{57}$$

$$\mathbf{h}_{a,abb'} = \mathrm{LeakyReLU}\left(\mathbf{W}\mathbf{x}_a + \mathbf{W}'\mathbf{h}_{abb'}\right), \tag{58}$$

$$z_{abb'} = \mathrm{LeakyReLU}\left(\mathbf{c}^t \mathbf{h}_{a,abb'}\right), \tag{59}$$

$$\gamma_{abb'} = \frac{\exp\left(z_{abb'}\right)}{\sum_{\bar{b}\bar{b}' \in \mathcal{N}(a)} \exp\left(z_{a\bar{b}\bar{b}'}\right)}, \tag{60}$$

$$\tilde{\gamma}_{abb'} = \frac{\gamma_{abb'} w_{abb'}}{\sum_{\bar{b}\bar{b}' \in \mathcal{H}(a)} \gamma_{a\bar{b}\bar{b}'} w_{a\bar{b}\bar{b}'}}, \tag{61}$$

where the sum over $\mathcal{H}(a)$ runs over all tuples (b, b') that form triplets (a, b, b') with the growth vector a.

Motif encoding. The motif encodings $\mathbf{v}_{i,3}$ featuring in the 3D likelihood factor are again based on the 2D atom encoder sequence from Eqs. 10-15, with the weights contained therein shared across the growth and motif embedder networks of the 3D stage, but not across the context stages:

$$\mathbf{x}_{2\mathrm{D}} = \mathrm{AtomEncoder}\left[\mathbf{x}, \mathbf{e}\right]. \tag{62}$$

From $\mathbf{x}_{2\mathrm{D}}$, we obtain $\mathbf{x}_{2\mathrm{D}}^{\mathrm{vec}}$ as the subset of 2D vectors that correspond to the linker atoms on the motifs; and $\mathbf{x}_{2\mathrm{D}}^{\mathrm{env}}$, that comprises all 2D atom vectors associated with the motif as a whole. Convolutions placed thereon give rise to the final vector that combines the local information reflected by $\mathbf{x}_{2\mathrm{D}}^{\mathrm{vec}}$ with a global motif vector $\mathbf{x}^{\mathrm{env}}$ derived from $\mathbf{x}_{2\mathrm{D}}^{\mathrm{env}}$ using attentive pooling:

$$\mathbf{x}^{\mathrm{vec}} = \mathrm{ResTrans}(\mathbf{x}_{2\mathrm{D}}^{\mathrm{vec}}), \tag{63}$$

$$\mathbf{x}^{\mathrm{env}} = \mathrm{Reduce}\left[\mathrm{ResTrans}(\mathbf{x}_{2\mathrm{D}}^{\mathrm{env}})\right], \tag{64}$$

$$\mathbf{x}_{2\mathrm{D}}^{\mathrm{motif}} = \mathrm{Linear}\left(\mathbf{x}^{\mathrm{vec}} + \mathbf{x}^{\mathrm{env}}\right). \tag{65}$$

Here $\mathrm{ResTrans}$ is a residual transfer block as defined above; Reduce is an attentive pooling operator consisting of:

$$\mathbf{x}_m^{\mathrm{pool}} = \sum_{j \in \mathcal{M}(m)} \mathbf{x}_{2\mathrm{D},mj}^{\mathrm{env}}, \tag{66}$$

$$\Delta\mathbf{x}_m^{\mathrm{pool}} = \sum_{j \in \mathcal{M}(m)} \gamma_{mj} \mathbf{W} \mathbf{x}_m^{\mathrm{pool}}, \tag{67}$$

$$\mathbf{x}_m^{\mathrm{pool}} = \mathrm{ELU}\left(\mathbf{x}_m^{\mathrm{pool}} + \Delta\mathbf{x}_m^{\mathrm{pool}}\right), \tag{68}$$

$$\mathbf{x}_m^{\mathrm{env}} = \mathrm{LayerNorm}\left(\mathbf{x}_m^{\mathrm{pool}}\right), \tag{69}$$

with $\{j \in \mathcal{M}(m)\}$ comprising all atoms j participating in motif m (not just the linker atom). The attention weights are evaluated statically:

$$\mathbf{h}_{mj} = \mathrm{LeakyReLU}\left(\mathbf{W}_0 \mathbf{x}_{mj} + \mathbf{W}_1 \mathbf{x}_m^{\mathrm{pool}}\right) \tag{70}$$

$$z_{mj} = \mathrm{LeakyReLU}\left(\mathbf{c}^t \mathbf{h}_{mj}\right) \tag{71}$$

$$\gamma_{mj} = \frac{\exp(z_{mj})}{\sum_{j \in \mathcal{M}(m)} \exp(z_{mj})} \tag{72}$$

We note that the motif embeddings $\mathbf{v}_{i,3}$ used by the 3D-conditional model are thus constructed in a manner that is distinct from that of the 2D-conditional model. We decided against re-using the motif embeddings $\mathbf{v}_{i,2}$ (Eq. 31) for the 3D stage as we expected that the information that needs to be captured at each stage is rather different: $\mathbf{v}_{i,2}$ is likely concerned with chemical compatibility, motif sociability, and reactivity constraints, whereas $\mathbf{v}_{i,3}$ needs to capture the type of interactions a motif can form, its shape and surface electrostatics, its conformational preferences, etc. Shape and surface properties are, specifically, the motivation behind the attentive pooling operator in Eq. 67 designed to capture non-local structure beyond the central hinge atom of a motif.

3D likelihood factor. We identify the 3D growth vector $\mathbf{u}_{a,3}$ and motif vector $\mathbf{v}_{i,3}$ with the following tensor outputs of the hypergraph and graph architectures above:

$$\mathbf{u}_{a,3} \equiv \mathbf{x}_{a,3\mathrm{D}}, \tag{73}$$

$$\mathbf{v}_{i,3} \equiv \mathbf{x}_{i,2\mathrm{D}}^{\mathrm{motif}}. \tag{74}$$

Note therefore that $\mathbf{v}_{i,3}$ does not incorporate any 3D information, and that the subscript is used only to differentiate it from the motif vector $\mathbf{v}_{i,2}$ on which the 2D factor q is based. The likelihood score function α_3 that gives rise to the 3D likelihood factor $r = \alpha_3/1 - \alpha_3$ is, finally, obtained from the activated inner product

$$\alpha_3(\mathbf{v}_{i,3}; \mathbf{u}_{a,3}) = \mathrm{Sigmoid}\left(\frac{\mathbf{v}_{i,3} \cdot \mathbf{u}_{a,3}}{\sqrt{d}}\right). \tag{75}$$

4. GROWTH-VECTOR AND BOND SAMPLING

Growth-vector and bond-order sampling are necessary to embed the PQR framework into a complete generative cycle. Starting from an input complex, the growth-vector model first selects the atom that is to be modified, the PQR model samples a motif tailored to that atom, and a bonding model selects an appropriate bond type (single $b = 1$, double $b = 2$, triple $b = 3$). Both the growth-vector and bond-order model are trained on the same set of reconstructions as the PQR model. At every reconstruction step, all atoms of the substrate modified during this or any subsequent step are marked with a growth label $z = 1$, all remaining atoms with growth labels $z = 0$. The atom pair consisting of the growth-vector atom and the hinge atom of the ground-truth motif is labelled with a three-dimensional target vector $\mathbf{b} = (b_1, b_2, b_3)$, where $b_d = 1$ if the bond order b is equal to d, else 0.

Both the growth-vector and bond-order model are implemented using graph-convolutional neural networks trained using a binary cross-entropy loss function. Given atom features $\mathbf{x}$ and edge features $\mathbf{e}$, the growth-vector models predicts a growth label $\hat{z} \in [0, 1]^{n \times 1}$ (with n the number of atoms in the batch / input molecule),

$$\mathbf{x}_{2\mathrm{D}} = \mathrm{AtomEncoder}(\mathbf{x}, \mathbf{e}), \tag{76}$$

$$\hat{\mathbf{z}} = \mathrm{Sigmoid}\left(\mathrm{Linear}(\mathbf{x}_{2\mathrm{D}})\right), \tag{77}$$

where $\mathrm{AtomEncoder}$ again refers to the convolution operations described in Eqs. 10-15. The sampling distribution

$p(a|\mathbf{x}, \mathbf{e})$ from Eq. 7 is thus given by $p(a|\mathbf{x}, \mathbf{e}) = \hat{z}_a / \sum_{a'} \hat{z}_{a'}$, where the sum over a' runs over all atoms of the query compound.

The bond-order model predicts a bond-order score $\hat{\mathbf{b}} \in [0, 1]^{n\times 3}$ starting from atom encodings constructed according to

$$\mathbf{x}^{(0)} = \text{AtomEncoder}(\mathbf{x}, \mathbf{e}), \tag{78}$$

$$\mathbf{x}^{(1)} = \text{SoftPlus}\left(\text{Linear}(\mathbf{x}^{(0)}\right), \tag{79}$$

$$\mathbf{x}_{\text{2D}} = \text{Linear}(\mathbf{x}^{(1)}). \tag{80}$$

Given pairs of atom embeddings $\mathbf{v}^b = \mathbf{x}_{\text{2D}}$ (associated with a motif vector) and $\mathbf{u}^b = \mathbf{x}'_{\text{2D}}$ (associated with a growth vector), so that rows i of $\mathbf{v}^b$ and $\mathbf{u}^b$ correspond to pairs i, the bond-order scores are evaluated using

$$\mathbf{x} = \text{ELU}\left(\text{Linear}_{\text{in}}(\mathbf{v}^b||\mathbf{u}^b)\right), \tag{81}$$

$$\mathbf{x}' = \text{ELU}\left(\text{Linear}_{\text{in}}(\mathbf{u}^b||\mathbf{v}^b)\right), \tag{82}$$

$$\mathbf{x}_{\text{sym}} = \text{LayerNorm}\left(\mathbf{x} + \mathbf{x}'\right), \tag{83}$$

$$\hat{\mathbf{b}} = \text{SoftMax}\left(\text{Linear}(\mathbf{x}_{\text{sym}})\right), \tag{84}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are evaluated using shared weights for $\text{Linear}_{\text{in}}$ to guarantee that $\hat{\mathbf{b}}$ is invariant with respect to atom pair reordering. The sampling distribution from Eq. 9 is therefore given by $p(d|\mathbf{v}^b_i, \mathbf{u}^b_a) = b_d / \sum_{d'=1}^{3} b_{d'}$, where b_d are the components of a single row of $\hat{\mathbf{b}}$ above. Note that neither the growth-vector nor bond-order model share any of their weights with other components of the architecture. This is because training data is abundant, and little if any synergy is expected across the different prediction endpoints.

The models can be evaluated using standard AUROC metrics over the growth and bond labels and scores. The test-set performance is measured at AUC $= 0.960 \pm 0.002$ for the growth-vector prediction, which is thought to be close to optimal considering the intrinsic variability of data. The bond-order prediction is virtually unambiguous, with AUCs of 0.999 ± 0.0005, 0.999 ± 0.0004 and 0.999 ± 0.002, respectively, for the three output channels (single-, double- and triple-bond).

As a result of the binary cross-entropy loss function and training label assignment, both the growth-label and bond-order scores have the meaning of a rate or probability. Therefore, stochastic selection of a growth vector / bond order is achieved by sampling a growth vector and bond order proportionally to the predicted scores $\hat{\mathbf{z}}$ and $\hat{\mathbf{b}}$, respectively.

5. ENTROPIC SHIFT

Beyond measuring enrichment, we can estimate the relative information gain achieved by incorporating the 1D, 2D and 3D context by tracking how the entropy of the posterior changes in the QRP sequence (i.e., when conditioning first on the 2D, then 3D, then 1D context). To this end, we evaluate a normalized

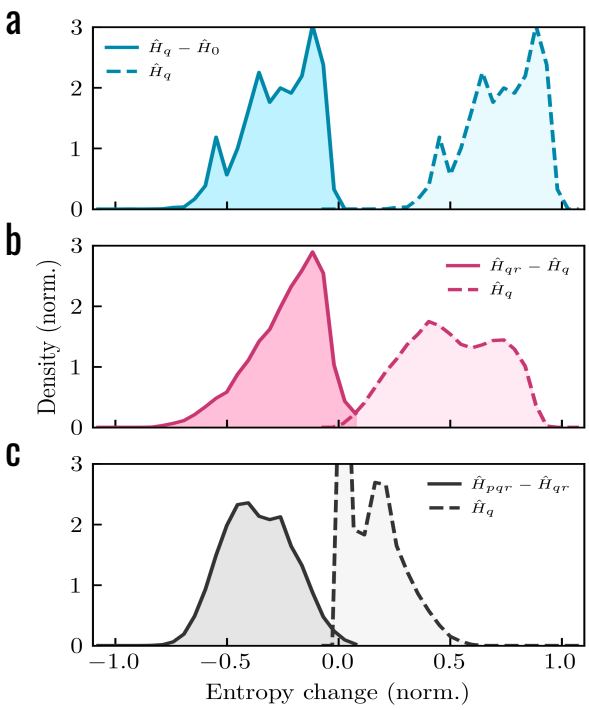

Figure S2. **Sequential information in the QRP model.** Effect of conditioning on the (a) 2D context, (b) 3D context and (c) 1D vocabulary, as quantified by the entropy (solid lines) and entropy change (dashed lines) of the posterior. The normalization applied to the entropy is such that $\hat{H}_0 = 1$ corresponds to the entropy of the uniform distribution over the vocabulary.

Shannon entropy $\hat{H}$ over the vocabulary $\mathcal{V}$,

$$\hat{H}[p_a] = -\frac{1}{\log|\mathcal{V}|} \sum_{\mathbf{v}_i \in \mathcal{V}} p_{i|a} \log p_{i|a}, \tag{85}$$

where $|\mathcal{V}|$ is the size of the vocabulary. With the chosen normalization, a uniform prior will result in the maximal possible $\hat{H}_0(a) = 1$. The likelihoods $p_{i|a}$ are given by, initially, the 2D likelihood factors $q_{i|a}$ (or, more precisely, the corresponding normalized probability density); then the combined 2D/3D likelihood factors $q_{i|a} \times r_{i|a}$; and finally the full posterior $p_{i|a} \times q_{i|a} \times r_{i|a}$.

We denote the corresponding entropies by $\hat{H}_q$, $\hat{H}_{qr}$, $\hat{H}_{pqr}$, respectively, distributions of which are shown in Fig. S2a-c (dashed lines). The same figure also reflects the entropy changes $\hat{H}_q - \hat{H}_0$, $\hat{H}_{qr} - \hat{H}_q$ and $\hat{H}_{pqr} - \hat{H}_{qr}$ (solid lines). Strongly lowering entropy, the 1D and 2D likelihood factors play the largest role in guiding the generative model towards a particular set of motifs, reflecting fragment sociability [4] and synthetic accessibility. The effect of the 3D factor r is sizeable but still less pronounced than of p and q. This is partially the result of the 2D-conditional model occasionally constraining the probability density to a small set of chemically and sterically more or less equivalent choices – but also a consequence of the fact that the 3D – as opposed to the 2D – model most likely still operates relatively far from its

$\langle\hat{\rho}\rangle$	$\mathbf{v}_{i,2}$	$\mathbf{v}_{i,3}$	$\mathbf{v}_{i,23}$	$\mathbf{v}_{i,\mathrm{ecfp}}$	$\mathbf{v}_{i,\mathrm{gylm}}$
$\mathbf{v}_{i,2}$	1.000	0.250	0.929	0.242	0.217
$\mathbf{v}_{i,3}$		1.000	0.555	0.035	0.228
$\mathbf{v}_{i,23}$			1.000	0.210	0.262
$\langle\rho\rangle$	$\mathbf{v}_{i,2}$	$\mathbf{v}_{i,3}$	$\mathbf{v}_{i,23}$	$\mathbf{v}_{i,\mathrm{ecfp}}$	$\mathbf{v}_{i,\mathrm{gylm}}$
$\mathbf{v}_{i,2}$	1.000	0.328	0.930	0.285	0.196
$\mathbf{v}_{i,3}$		1.000	0.626	0.136	0.202
$\mathbf{v}_{i,23}$			1.000	0.274	0.228

Table S6. **Correlation among the metric spaces induced by different motif-vector embeddings.** $\mathbf{v}_{i,2}$, $\mathbf{v}_{i,3}$ and $\mathbf{v}_{i,23}$ are the 2D and 3D, and joint 2D-3D embeddings learnt from reconstruction data. $\mathbf{v}_{i,\mathrm{ecfp}}$ and $\mathbf{v}_{i,\mathrm{gylm}}$ are static 2D and 3D fingerprints, respectively. $\hat{\rho}$ and ρ are the Spearman rank and Pearson correlation coefficient evaluated over the entries of the upper triangle of the motif-motif similarity matrix for each representation (Eqs. 87-91). The correlation between the dynamic and static encodings is very poor, ranging between 0.035 and 0.285.

theoretical peak performance.

6. MOTIF EMBEDDINGS

Visualization of the posterior over motifs. The 2D projections in Fig. 4 of the main text are based on a kernel k_{ij} defined between any two motif vectors $\mathbf{v}_i$ and $\mathbf{v}_j$, derived from the cross-correlation matrix of the 2D network scores $\alpha_n(\mathbf{v}_{i,n}; \mathbf{u}_{a,n})$, sampled over a large number N of randomly selected growth vectors a:

$$k_{ij,n} = \frac{1}{N}\sum_a \hat{\alpha}_n(\mathbf{v}_{i,n}; \mathbf{u}_{a,n})\,\hat{\alpha}_n(\mathbf{v}_{j,n}; \mathbf{u}_{a,n}),$$
$$k_{ij} = \frac{1}{2}(k_{ij,2} + k_{ij,3}), \tag{86}$$

where $\hat{\alpha}$ denotes the normalised (whitened) scores. We construct a distance metric d_{ij} from k_{ij} using $d_{ij} = 1 - k_{ij}$, and obtain a 2D projection of the vocabulary via stochastic neighbour embedding (t-SNE) of the resulting distance matrix. Projecting the values of the posterior onto this 2D map, we can visualize how the probability density changes as we incorporate 2D, 3D and frequency information.

Dynamics vs static motif embeddings. In our implementation of the 2D and 3D context modules Q and R, the growth-vector and motif-vector embeddings $\mathbf{u}_{a,n}$ and $\mathbf{v}_{i,n}$ are both represented by neural networks, trained to co-adapt dynamically in order to maximize the probability of the observed pairings. An alternative implementation could use static encodings in the form of hard-coded (pre-defined) fingerprints for the motif vectors in the projections of Eqs. 32 and 75. Here we decided against this approach, because we expected that hard-coded fingerprints are insufficient to model the complex physical / chemical and reactivity constraints that determine the likelihood of a motif given a certain growth-vector context. Additionally, accidental structural overlap could subject the posterior to undesired cross-talk among chemically disparate moieties.

Here we explore the properties of the embedding space represented by the model by assessing how the metric space (distance geometry) induced by the learnt network embeddings correlates with that based on static, pre-defined fingerprints. For the static fingerprints, we consider (a) a topological 1024-bit Extended Connectivity Fingerprint (ECFP) $\mathbf{v}_{i,\mathrm{ecfp}}$, and (b) an atom-centered 3D basis-function-based geometric fingerprint (GYLM), $\mathbf{v}_{i,\mathrm{gylm}}$ [5]. For the former, we use the ECFP6 implementation in RDKit, with a post-processing step that centers the fingerprint on the hinge atom of the motif by removing bits associated with features outside of a 3-bond neighbourhood around that atom. For the latter, we resort to the implementation in the gylmxx package, with default parameters as provided by the BenchML suite [5].

To quantify the similarity among the metric spaces underlying the learnt and static embeddings, we evaluate the Spearman rank correlation $\hat{\rho}$ and the Pearson correlation coefficient ρ between the upper triangles of the motif-motif similarity matrices K, whose components are defined as follows:

$$K_{ij,2} = \exp(-\lambda_2||\mathbf{v}_{i,2} - \mathbf{v}_{j,2}||), \tag{87}$$
$$K_{ij,3} = \exp(-\lambda_3||\mathbf{v}_{i,3} - \mathbf{v}_{j,3}||), \tag{88}$$
$$K_{ij,23} = \exp(-\lambda_{23}||\mathbf{v}_{i,23} - \mathbf{v}_{j,23}||), \tag{89}$$
$$K_{ij,\mathrm{ecfp}} = \exp(-\lambda_{\mathrm{ecfp}}||\mathbf{v}_{i,\mathrm{ecfp}} - \mathbf{v}_{j,\mathrm{ecfp}})||, \tag{90}$$
$$K_{ij,\mathrm{gylm}} = \mathbf{v}_{i,\mathrm{gylm}} \cdot \mathbf{v}_{j,\mathrm{gylm}}. \tag{91}$$

For the learnt embeddings and 2D fingerprint, these expressions translate the Euclidean distance $||\ldots||$ between the vectors into a similarity function using an exponential with length scale λ. For the 3D fingerprint, by contrast, we use a dot-product kernel, which is the native kernel for this particular type of neighbourhood expansion [6]. The length-scale parameters are here taken to be the mean of the Euclidean distance matrix for the respective representation; $\mathbf{v}_{i,23} = \mathbf{v}_{i,2}||\mathbf{v}_{i,3}$ is the concatenation of the 2D and 3D motif vector embedding.

The correlation coefficients $\hat{\rho}$ and ρ are shown in Tab. S6. Each correlation is evaluated over the $M(M-1)/2$ paired entries of the upper triangle of K, with $M = 7513$ the size of the motif library underlying PDBL. It can be seen that the metric spaces induced by the static fingerprints correlate very poorly with those derived from the learnt embeddings, with correlation coefficients of less than 0.3. We note that the magnitude of the correlation is insensitive to the choice of parameters of the static fingerprint, such as the topological radius (ECFP4 vs ECFP6), the number of bits (512, 1024, 2048), as well as to the choice of distance length scale. The lowest correlation of 0.035 is measured, perhaps not unsurprisingly, between the 2D fingerprint $\mathbf{v}_{i,\mathrm{ecfp}}$ and 3D motif-vector embedding $\mathbf{v}_{i,3}$. We also note that the metric spaces induced by the 2D and 3D embedding vectors themselves do not correlate strongly, with a Spearman rank and Pearson correlation coefficient of 0.250 and 0.328 respectively – indicating that conditioning separately on the 2D and 3D context is a meaningful approach.

Naturally the poor correlation among the different metric spaces is not unexpected. Consider, e.g., single-atom halogen motifs such as bromine, chlorine, iodine, that, as far as these

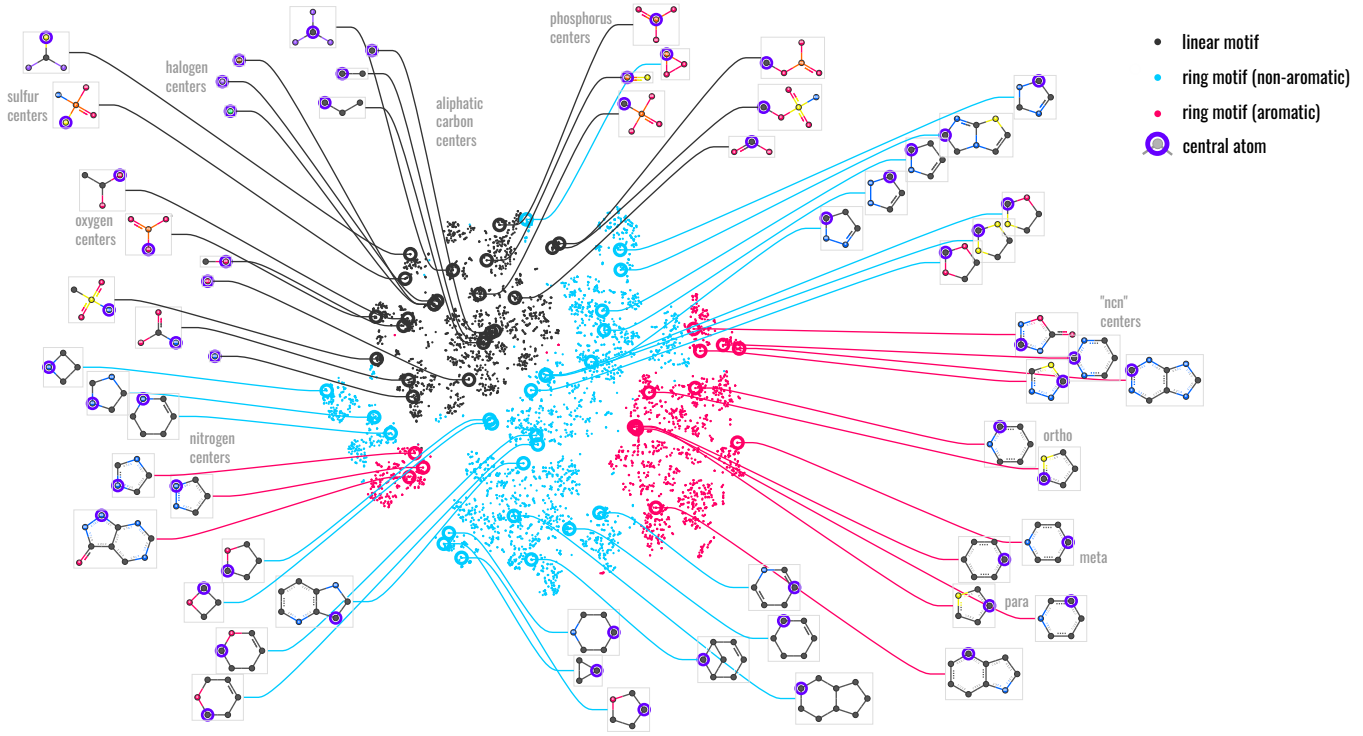

Figure S3. **Low-dimensional projection of chemical motifs.** Location of several probe motifs on a t-SNE cluster map based on the similarity metric of Eq. 86 (see *Methods*, main text). Each point represents a single motif, with linear motifs in black, ring motifs with non-aromatic centers in blue, and ring motifs with aromatic centers in red. For the highlighted motifs, the central (i.e., hinge) atom is indicated with a thick purple outline.

static fingerprints are concerned, are as remote from each other as carbon and oxygen, carbon and nitrogen, or nitrogen and fluorine – when clearly chemically and structurally these larger halogens are largely interchangeable. Still, there are a range chemical features and properties that we would expect to significantly affect the context-dependent likelihood amplification. We glean further insight into these features by visualizing in more detail how the chemical map from Fig. 4 of the main text is organized. To this end, Fig. S3 illustrates the relative location of a number of probe motifs across this map. It can be seen that the global organization is dominated by linear (black) vs non-aromatic ring (blue) vs aromatic ring (red) motifs. Linear motifs with nitrogen and oxygen centers (two key electronegative elements) are located from bottom to top on the left-hand side of the map, with a sulfur cluster directly adjacent to the oxygen cluster, in line with the grouping on the periodic table. Halogen centers are located close by the lipophilic carbocentric ring motifs. Aromatic rings appear locally ordered according to ortho, meta and para substitution – trends that evidently reflect the at times stark difference in reactivity across these sites. Furthermore, within the ring domains, the local structure appears to be linked to enumerations around ring scaffolds, such as ring-size variation (four- to five- to six-membered rings) or changes of specific positions in a ring from one element to another (see, e.g., the OCO- vs OCS- vs SCS-containing five membered rings close to the center of the map).

7. ELABORATION EXAMPLES (FIG. 5)

Validating the output of generative models using computational routes only is typically challenging. In the examples provided in Fig. 5 of the main text, the experimental SAR was recovered in three out of four cases. The lower-rank suggestions are, however, not trivially correct. In example (a: aminopeptidase) the halogen motifs are consistent with an SAR observed by Helgren *et al*. [7]. In example (b: yellow lupin), the top-ranked bromo motif is consistent with the groud-truth iodo, with otherwise only fluoro achieving significant amplification.

In example (d: hCA II), substitution of thiophene (rank 1) with other aromatic five-membered rings (ranks 2 and 4) appears broadly plausible. Pyrazole and triazole are, however, distinct from thiophene regarding their electrostatic and lipophilic properties. We have therefore scanned the PDB for precedents of these motifs, with three examples (Fig. S4a) overlayed with the reference structure 1ydb – thus indicating that there is some support for introducing aromatic nitrogen into the region of the growth vector. The thiol motifs (ranks 3 and 5) are, however, precedented only as thiones (Fig. S4b). Even though the 2D bonding model can assign a double bond to the link between the thiol and the growth vector (thus converting thiol to thione), the resulting thio-aldehyde is a known PAINS structure [8] and therefore to be avoided.

The 3D ranking failed to recover the butyl motif from example (c: SHBG). Carbyne motifs, notably, the three-atom

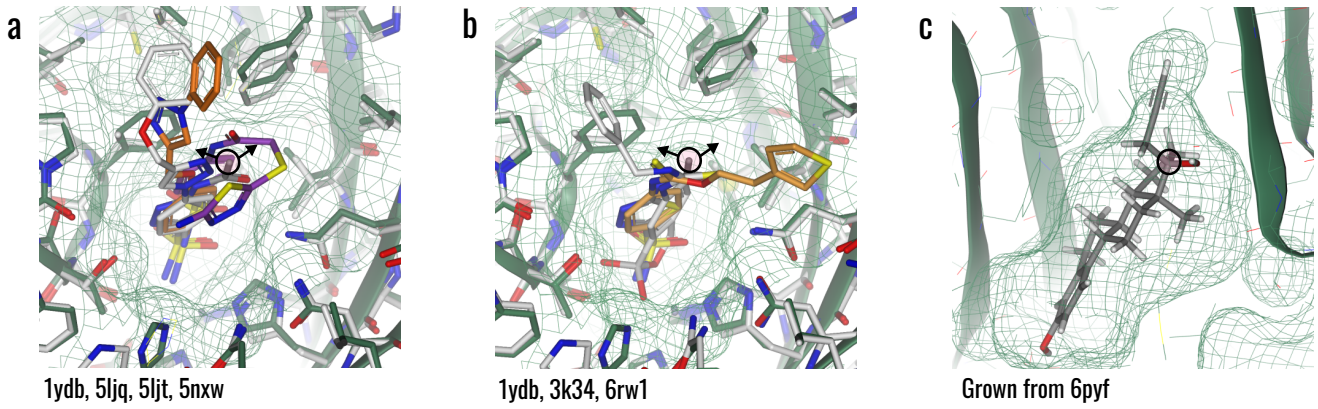

Figure S4. **Motif precedents and template-based docking.** (a) Potential precedents for observation of pyrazole and triazole in the neighbourhood of the growth vector of 1ydb (black circle with arrows). (b) Thione motifs in 3k34 and 6rw1 in proximity to the 1ydb growth vector. (c) Template-based docking of a carbyne motif CC≡C attached to the 6pyf growth vector.

vector CC≡C can be easily incorporated into the environment, resulting in an adequate fit as demonstrated by template-based docking (Fig. S4c).

[1] Xiong, Z. *et al.* Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of Medicinal Chemistry* **63**, 8749–8760 (2020). URL `https://pubs.acs.org/doi/10.1021/acs.jmedchem.9b00959`.

[2] Fey, M. & Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds* (2019).

[3] Paszke, A. *et al.* PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *NeurIPS*, 8024–8035 (2019).

[4] St. Denis, J. D., Hall, R. J., Murray, C. W., Heightman, T. D. & Rees, D. C. Fragment-based drug discovery: opportunities for organic synthesis. *RSC Med. Chem.* **12**, 321–329 (2021).

[5] Poelking, C., Faber, F. A. & Cheng, B. Benchml: an extensible pipelining framework for benchmarking representations of materials and molecules at scale (2021). URL `http://arxiv.org/abs/2112.02287`. ArXiv:2112.02287 [cond-mat, physics:physics].

[6] Poelking, C. *et al.* Meaningful machine learning models and machine-learned pharmacophores from fragment screening campaigns (2022). URL `http://arxiv.org/abs/2204.06348`. ArXiv:2204.06348 [cs, q-bio].

[7] Helgren, T. R. *et al.* Rickettsia prowazekii methionine aminopeptidase as a promising target for the development of antibacterial agents. *Bioorg. Med. Chem.* **25**, 813–824 (2017).

[8] Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (pains) from screening libraries and for their exclusion in bioassays. *Journal of Medicinal Chemistry* **53**, 2719–2740 (2010). URL `https://pubs.acs.org/doi/10.1021/jm901137j`.