Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Geometry-faithful fragment diffusion for physically valid, multi-target structure-based drug design

Michael Akinola Oluwasola
Universiti Putra Malaysia
Noor Dina Muhd Noor
Universiti Putra Malaysia
Adam Thean Chor Leow
adamleow@upm.edu.my

Universiti Putra Malaysia

Research Article

Keywords:

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11779928/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Geometry-faithful fragment diffusion for physically valid, multi-target structure-based drug design

Michael Akinola Oluwasola[1,2], Noor Dina Muhd Noor[1,3], Adam Thean Chor Leow[1,2,4*]

[1]Enzyme and Microbial Technology Research Center, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, UPM Serdang, 43400, Selangor, Malaysia.
[2]Department of Cell Biology and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, UPM Serdang, 43400, Selangor, Malaysia.
[3]Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, UPM Serdang, 43400, Selangor, Malaysia.
[4]Institute of Bioscience, Universiti Putra Malaysia, UPM Serdang, 43400, Selangor, Malaysia.

*Corresponding author(s). E-mail(s): adamleow@upm.edu.my;
Contributing authors: gs68662@student.upm.edu.my;
dina@upm.edu.my;

Abstract

Generative models can design ligands directly inside protein binding pockets, but errors in molecular geometry and connectivity limit how far their outputs can be trusted. We present GEMINI-Mol, an SE(3)-equivariant diffusion framework that projects fragment coordinates onto canonical templates during denoising, assembles bonds from template connectivity and valence constraints, relaxes each ligand in the presence of the protein, and combines conditioning information from several pockets within a single shared model. Benchmarked against four representative generators on the CrossDocked2020 100-pocket test set through one evaluation pipeline applied identically to every method, GEMINI-Mol reaches 94% intramolecular PoseBusters validity (Pocket2Mol 89%, TargetDiff 58%),

fabricates no three- or four-membered rings, and more closely reproduces the reference bond-angle and torsion-angle distributions, while its docking scores trail the strongest baselines. A controlled ablation that holds the trained network and every sampling setting fixed and substitutes distance-based bond perception for the valence-constrained assembly returns validity from 92% to 63%, reintroduces spurious three-membered rings in 26% of molecules and disconnects 13% of them, locating the improvement in the reconstruction stage rather than in the learned coordinates. In an exploratory dual-target application to pancreatic α-amylase and intestinal maltase-glucoamylase, generated molecules match the clinical inhibitor acarbose in docking score against maltase-glucoamylase (–7.00 against –6.97 kcal mol^{-1}) but not against α-amylase, and 500 ns of unbiased molecular dynamics shows that the docked pose is only partly persistent, in close contact with the protein for 80–82% of the trajectory against 96–100% for acarbose.

Scientific Contribution: We show that the physical validity of a pocket-conditioned generative model is governed less by its learned coordinates than by how connectivity is inferred from them after sampling, and we replace distance-based bond perception with a deterministic template- and valence-constrained assembly followed by protein-aware relaxation, which needs no retraining and raises PoseBusters validity above all four benchmarked generators. A single shared model conditions on several pockets simultaneously, so that multi-target engagement and multi-objective developability become design variables rather than filters applied to a finished pool. We further document two systematic pitfalls in the evaluation of generated structures: interaction fingerprints computed on unrelaxed poses are inflated rather than merely noisy, and a favourable docking or free-energy score is not evidence that the scored pose is the one the molecule occupies.

Keywords: Structure-based drug design, Denoising diffusion, SE(3)-equivariance, Molecular reconstruction, Fragment-based molecular generation, Multi-target design, Pocket conditioning, Physical validity

1 Introduction

The development of a new medicine remains one of the most protracted and uncertain undertakings in applied science. Bringing a single drug from concept to approval spans on average more than a decade and costs on the order of a billion US dollars, with the great majority of candidates failing before they reach patients [14, 77]. Attrition is not distributed evenly across that pipeline. Failures traced to poor absorption, distribution, metabolism, excretion and toxicity (ADMET) properties, to insufficient selectivity, or to intractable chemistry are largely determined by decisions taken at

the earliest stage, when a lead compound is chosen [75]. The identification of that lead therefore exerts a disproportionate influence over everything that follows. Historically it depended on serendipity and on the exhaustive screening of compound collections, an approach that is difficult to direct and harder still to reproduce. Advances in structural biology, cheminformatics and artificial intelligence have progressively shifted the field from screening toward rational design, in which candidate molecules are conceived against an explicit hypothesis about the biological target.

Rational design proceeds along two complementary routes. Ligand-based drug design derives new molecules from the structures of known actives and is valuable when the target structure is unavailable, but it cannot exploit the geometry of the binding site and is of limited use for targets with few characterised ligands. Structure-based drug design (SBDD) instead reasons directly from the three-dimensional structure of the target, an approach that has become far more broadly applicable now that accurate structures can be predicted for essentially any sequence [40]. SBDD is pursued in two ways. Virtual screening ranks existing compounds by a physics- or data-driven scoring function [52, 71], but even on-demand libraries of 10^{10}–10^{15} compounds sample a vanishing fraction of the 10^{60} or more drug-like molecules that could be made, and at that scale the imperfections of any scoring function are amplified into large absolute numbers of false positives [76]. De novo molecular generation avoids the enumeration entirely by constructing molecules inside the pocket, and it is this route that deep learning has transformed.

Early generative models for SBDD represented the pocket as a voxelised density map and coupled three-dimensional convolutional networks to conditional variational autoencoders [62]. Such models are not equivariant to rotation and translation of the complex, and their memory cost grows steeply with pocket size. Subsequent work represented molecules as three-dimensional graphs and built SE(3) equivariance into the architecture, generating atoms autoregressively: sampling a new atom conditioned

on the pocket and on the atoms placed so far [37, 48, 51, 59, 79]. Autoregressive construction, however, imposes an atom ordering that has no chemical meaning, allows errors made early in the sequence to propagate through everything placed afterwards, and is prone to terminating prematurely and returning fragments rather than complete ligands.

Diffusion models address these shortcomings by generating every atom simultaneously through iterative denoising, and have become the dominant paradigm for pocket-conditioned design [21, 22, 30, 33, 35, 36, 67]. Two problems nonetheless persist, and they bound what the current generation of models can contribute to a real design campaign. The first is that the molecules are frequently not physically realisable. Generated structures contain three- and four-membered rings, inappropriately fused and extra-large rings, distorted aromatic systems and ligand atoms overlapping protein atoms; systematic audits find that a substantial fraction fail elementary physical checks [4, 9, 25]. The cause lies in how a molecule is assembled once sampling ends. The dominant approach emits atomic coordinates and element types but not bonds, so connectivity is recovered afterwards by perceiving bonds from interatomic distances, conventionally with Open Babel [56]; bond diffusion, which predicts connectivity jointly with coordinates, is a more recent alternative we return to in the Discussion [33]. A coordinate error of a few tenths of an Ångström is negligible against the noise the model is trained to remove, yet decisive at the threshold separating a bonded from a non-bonded contact, and errors of that size fabricate rings and sever molecules. The consequence is not cosmetic. Rescoring, molecular dynamics, free-energy calculation and medicinal chemistry triage all presuppose a chemically real structure, and a docking score computed on a strained or disconnected one measures the artefact rather than the binding; redocking, the standard scoring protocol, conceals much of this by discarding the generated conformer in favour of one the docking program finds for itself [25].

The second problem is that the objective being optimised is far narrower than the one drug discovery poses. Almost without exception these models maximise predicted affinity to a single pocket. Yet a viable lead is defined by the simultaneous satisfaction of several competing requirements. It must engage the correct set of targets, sparing anti-targets to achieve selectivity, or deliberately engaging several at once to realise the polypharmacology underlying many modern therapeutics [2, 32, 61, 63]. It must be drug-like and synthetically accessible [5, 19]. And it must avoid the ADMET liabilities (hERG-mediated cardiotoxicity, mutagenicity, hepatotoxicity, cytochrome P450 interference) responsible for much of the attrition noted above. Where these criteria are addressed at all, they enter as light property guidance or as filters applied to a finished pool [33]; comprehensive ADMET safety is not modelled, and engagement across multiple targets has no representation in any pocket-conditioned generator, so neither designed selectivity nor designed polypharmacology can be expressed. Lead design is intrinsically a multi-target, multi-objective problem, and it is this coupled structure, rather than potency alone, that makes it difficult.

In this work we present GEMINI-Mol, an SE(3)-equivariant generative framework for structure-based design that addresses both problems together. It replaces unconstrained, distance-based bond perception with a template-guided, valence-constrained assembly, so that generated molecules are connected, valence-correct and free of spurious small rings for all but a rare failure mode rather than by chance, and it optimises the pose in the presence of the pocket so that internal geometry and protein fit are resolved jointly. It further treats multi-target engagement and molecular developability as design variables rather than as downstream filters: a single trained model designs against several pockets simultaneously without retraining for each target combination, and binding affinity, drug-likeness, synthetic accessibility and a predicted ADMET-safety score, drawn from a 22-endpoint Chemprop ADMET suite, are controlled together during and after generation. Benchmarked against four representative

generators on the CrossDocked2020 100-pocket test set through a single evaluation pipeline applied identically to every method, GEMINI-Mol attains the highest physical validity of the five methods, yields no fabricated small rings (by construction), and reproduces the crystallographic reference distributions of bond angles and torsional angles, while its mean docking score ranks third of the five methods, behind the two affinity-specialised baselines and ahead of the other two. We then apply multi-target design to a therapeutic problem whose rationale is established in the clinic: postprandial glycaemic control in type 2 diabetes, in which pancreatic α-amylase and intestinal maltase-glucoamylase hydrolyse dietary carbohydrate at consecutive steps and the standard-of-care inhibitor acarbose inhibits both [11, 43]. Dual-target designs match acarbose against maltase-glucoamylase under an identical docking protocol while being markedly more drug-like, and a subset reproduces the interaction pattern of the crystallographic acarbose complex rather than merely its score.

2 Methods

2.1 Framework overview

GEMINI-Mol is a pocket-conditioned, SE(3)-equivariant denoising diffusion model with three features that distinguish it from atomwise generators (Fig. 1). First, the generative state is organised around rigid molecular substructures. Each ligand is first decomposed into fragments: every atom is assigned either to one instance of 26 fragment types (rings and common functional groups) or to the linker, the remaining atoms that join the fragments. During the reverse process, at every step, the atoms of each fragment instance are projected onto that fragment's canonical template by optimal rigid alignment [41]. Ring systems therefore retain ideal internal geometry throughout sampling, even when the sampled coordinates are imperfect (Fig. 1a).

Second, the molecule is assembled by a deterministic, valence-constrained procedure rather than by unconstrained distance-based bond perception (Fig. 1b).

Intra-fragment bonds, including aromaticity, are taken directly from the templates; remaining connectivity is completed by adding only bonds that are within covalent range, leave both atoms with free valence, and whose two atoms are at least four bonds apart in the current graph. This is a deliberately conservative rule: it means that linker bonds close only rings of five or more members, which excludes the strained rings that distance-based perception fabricates from noisy geometry, but it also excludes small rings that do occur in therapeutics. Any residual disconnection is repaired by joining the nearest valence-admissible pair across components, provided one exists; when none does, the sample is returned as a validity failure rather than a malformed molecule. Subject to that rare failure mode, the procedure returns a connected molecule in which every ring is either a template ring or has at least five members (guarantees stated and proved in Additional file 1, Note S3). The assembled ligand is then relaxed in the presence of the pocket, so that internal geometry and steric fit with the protein are optimised against one another rather than in sequence.

Third, control is exercised in score space, which makes multi-pocket design and property optimisation compositional rather than architectural (Fig. 1c). Because the coordinate score is an affine function of the predicted clean coordinates, sampling from a $\boldsymbol{\beta}$-weighted product of single-pocket conditionals reduces exactly to averaging the clean-coordinate predictions of the same network evaluated once per pocket, so one trained model generates for any set of pockets and any engagement vector $\boldsymbol{\beta}$ without retraining. Binding affinity, being differentiable in the atomic coordinates, is optimised inside the sampling loop by gradient guidance from a learned affinity network, damped by that network's own predictive variance. Drug-likeness, synthetic accessibility and a 22-endpoint ADMET-safety profile are computed on the molecular graph and are not differentiable in the coordinates; these are optimised instead by uncertainty-gated, diversity-aware selection over the generated pool. The subsections below describe each component in detail.

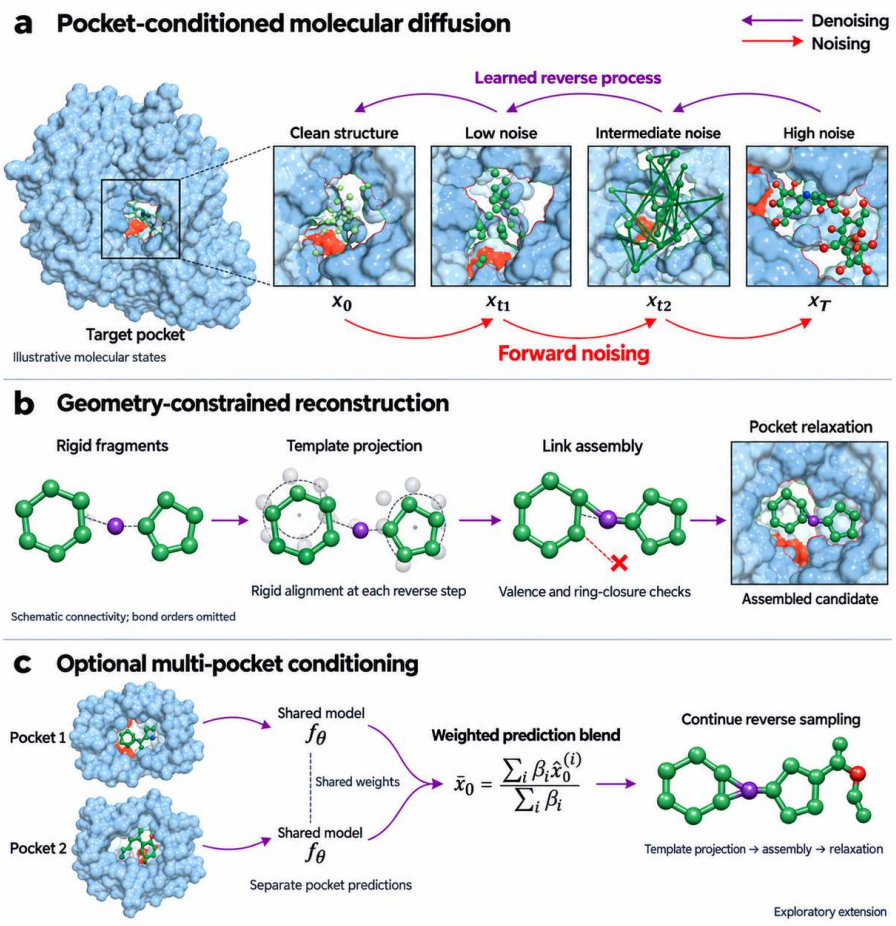

Fig. 1 Overview of the GEMINI-Mol framework. a, Schematic pocket-conditioned diffusion, with forward noising and learned reverse denoising. **b**, Geometry-constrained reconstruction: fragment coordinates are aligned to rigid templates during reverse sampling; subsequent graph assembly retains template bonds and screens candidate links using valence and ring-closure rules before pocket relaxation. **c**, Optional multi-pocket conditioning: the shared network is evaluated separately for each pocket, and its predictions are combined by a weighted average before reverse sampling continues. The equation illustrates coordinate averaging. Molecular structures, protein surfaces and intermediate states are illustrative artwork, not atomically validated structures or recorded sampling trajectories; the connectivity drawings omit bond orders.

2.2 Problem formulation

A protein pocket is a set of atoms $\mathcal{P} = \{(a_j, \mathbf{r}_j)\}_{j=1}^{M}$ with types a_j and coordinates $\mathbf{r}_j \in \mathbb{R}^3$. A ligand is a tuple $\mathbf{z} = (\mathbf{x}, \mathbf{v}, \mathbf{b})$ comprising atom coordinates $\mathbf{x} \in \mathbb{R}^{N\times 3}$, atom types $\mathbf{v} \in \{1, \ldots, C\}^N$ (element and aromaticity) and pairwise bond types

$\mathbf{b} \in \{0, \dots, B\}^{N \times N}$. Single-target design requires learning and sampling from the pocket-conditioned distribution $p(\mathbf{z} \mid \mathcal{P})$. We formulate two generalisations. *Multi-objective* design augments the target distribution with a set of molecular property functionals $\{g_k\}$ (binding affinity, drug-likeness, synthesisability and ADMET safety) and samples from a property-tilted distribution (Two-tier multi-objective control). *Multi-target* design conditions on a set of pockets $\{\mathcal{P}_i\}_{i=1}^{K}$ with a controllable engagement vector $\boldsymbol{\beta} \in \mathbb{R}^{K}_{\geq 0}$ and samples from $p_{\boldsymbol{\beta}}(\mathbf{z} \mid \mathcal{P}_1, \dots, \mathcal{P}_K)$ (Controllable multi-target design). All three problems are solved by a single diffusion model whose *score* is composed at sampling time; we therefore fix the backbone and its score first, then build the controllers on top of it.

2.3 SE(3)-equivariant diffusion backbone

We adopt a denoising diffusion probabilistic model [30, 69] over the joint ligand state $\mathbf{z}$, following the pocket-conditioned formulation of Guan et al. [21]. The forward process perturbs coordinates by a continuous Gaussian diffusion in a pocket-centred, scale-normalised frame,

$$q(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}\big(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0,\, (1-\bar{\alpha}_t)\mathbf{I}\big), \qquad \bar{\alpha}_t = \prod_{s \leq t} (1-\beta_s), \tag{1}$$

and perturbs the discrete atom types by a uniform categorical (D3PM) diffusion [3],

$$q(\mathbf{v}_t \mid \mathbf{v}_0) = \mathrm{Cat}\big(\mathbf{v}_t;\, \bar{\alpha}^v_t\,\mathbf{v}_0 + (1-\bar{\alpha}^v_t)\,\tfrac{1}{C}\mathbf{1}\big), \tag{2}$$

with an independent schedule $\bar{\alpha}^v_t$ that decays faster than the coordinate schedule, so that types resolve before fine geometry [60]. A denoising network ϕ_θ maps the noised ligand and pocket to a prediction of the clean state, $(\hat{\mathbf{x}}_0, \hat{\mathbf{v}}_0) = \phi_\theta(\mathbf{z}_t, t, \mathcal{P})$, using the $\mathbf{x}_0$-parameterisation: the network emits a per-atom *displacement* $\boldsymbol{\Delta}_\theta$ and we set

$\hat{\mathbf{x}}_0 = \mathbf{x}_t + \boldsymbol{\Delta}_\theta$. Predicting the displacement rather than the noise $\boldsymbol{\varepsilon}$ makes the low-t target $\approx \mathbf{0}$, the regime in which bond-scale geometry is fixed and $\boldsymbol{\varepsilon}$-prediction is ill-conditioned. Bond types $\hat{\mathbf{b}}$ are read from a pairwise head on the final atom features and clean geometry (Network architecture and featurisation).

The construction that makes control composable is the identification of the coordinate *score* through Tweedie's formula [17]: because Eq. (1) is Gaussian, the conditional score of the marginal $p_t(\mathbf{x}_t \mid \mathcal{P})$ is an affine function of the clean-coordinate prediction,

$$\mathbf{s}_\theta(\mathbf{x}_t, t \mid \mathcal{P}) \;\equiv\; \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathcal{P}) \;=\; \frac{\sqrt{\bar{\alpha}_t}\,\hat{\mathbf{x}}_0(\mathbf{z}_t, \mathcal{P}) - \mathbf{x}_t}{1 - \bar{\alpha}_t}. \tag{3}$$

This is the standard denoising-score-matching identity [70, 73]. Pocket conditioning is trained with classifier-free guidance [29]: the pocket is stochastically masked with probability 0.1 during training so that an unconditional branch $\mathbf{s}_\theta(\mathbf{x}_t, t \mid \varnothing)$ is available by hiding $\mathcal{P}$ as a message source, and the conditional score is amplified as $\mathbf{s}^w = \mathbf{s}_\theta(\cdot \mid \varnothing) + w\,[\mathbf{s}_\theta(\cdot \mid \mathcal{P}) - \mathbf{s}_\theta(\cdot \mid \varnothing)]$. Equation (3) and the guided and multi-target scores below are expressed in the same coordinate-score space, so control operations compose additively.

2.4 Network architecture and featurisation

Featurisation.

Ligand atoms carry a dual encoding: a functional-group stream (26 fragment types, embedded to 128 dimensions) and an atom stream (13 element-plus-aromaticity classes, embedded to 64), concatenated. Pocket atoms are encoded by a categorical residue and element embedding (21 residue types), with each residue's side chain capped at 15 heavy atoms for the encoder. This limit is separate from the ligand size

caps (32 heavy atoms, 16 linker atoms) described below. A sinusoidal timestep embedding is concatenated to every node feature. Protein and ligand atoms form a single joint graph; a ligand mask $\mathbf{m} \in \{0,1\}^{N+M}$ marks the movable atoms.

Equivariant message passing.

The backbone is an $E(n)$-equivariant graph neural network [31, 66], reimplemented with a built-in short-range repulsion for collapse stability. It operates on a local k-nearest-neighbour graph ($k = 32$) restricted to a cutoff radius r_{cut} in the normalised frame; edge features are a 20-dimensional Gaussian radial basis $\psi_k(d) = \exp\big(-(d-\mu_k)^2/2\delta^2\big)$ with centres μ_k spanning $[0, r_{\text{cut}}]$. For a node i with neighbours $\mathcal{N}(i)$, each layer computes an edge message, a scalar (invariant) update and a coordinate (equivariant) update:

$$\mathbf{m}_{ij} = \phi_e\big([\,\mathbf{h}_i \,\|\, \mathbf{h}_j \,\|\, \psi(\|\mathbf{x}_i - \mathbf{x}_j\|)\,]\big) \odot \sigma\big(\phi_g(\cdot)\big), \tag{4}$$

$$\mathbf{h}_i \leftarrow \mathrm{LN}\Big(\mathbf{h}_i + \phi_h\big[\,\mathbf{h}_i \,\|\, \tfrac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \mathbf{m}_{ij}\,\big]\Big), \tag{5}$$

$$\mathbf{x}_i \leftarrow \mathbf{x}_i + m_i\Big[\tfrac{1}{\sqrt{|\mathcal{N}(i)|}}\sum_{j\in\mathcal{N}(i)} \frac{\mathbf{x}_i - \mathbf{x}_j}{\|\mathbf{x}_i - \mathbf{x}_j\| + \epsilon}\,\tanh\big(\phi_x(\mathbf{m}_{ij})\big) + \mathbf{f}_i^{\text{rep}}\Big], \tag{6}$$

where $\phi_\bullet$ are multilayer perceptrons, LN is layer normalisation, m_i is the movable-atom mask (so pocket coordinates are held fixed), and $\mathbf{f}_i^{\text{rep}} = c\sum_j \frac{\mathbf{x}_i-\mathbf{x}_j}{\|\mathbf{x}_i-\mathbf{x}_j\|}\,(r_0 - \|\mathbf{x}_i-\mathbf{x}_j\|)_+$ is a fixed outward force active only for pairs closer than r_0, which prevents collapse. The coordinate multilayer perceptron ϕ_x is zero-initialised, so each layer begins at the identity and learns geometry as a residual. Because messages Eq. (4) depend only on interatomic *distances*, $\mathbf{h}$ is SE(3)-invariant; because Eq. (6) updates coordinates by a sum of relative vectors, $\mathbf{x}$ is SE(3)-equivariant. We stack 6 such layers at hidden width 256. The per-atom displacement is $\boldsymbol{\Delta}_\theta = \mathbf{x}^{(L)} - \mathbf{x}^{(0)}$; atom-type and bond-type logits are read from the final scalar features $\mathbf{h}^{(L)}$, the latter over ligand atom pairs.

2.5 Training objective and optimisation

Noise schedules.

We use $T = 1000$ diffusion steps. Coordinates are diffused in the pocket-centred, scale-normalised frame of Eq. (1) under a sigmoid β schedule interpolating between $\beta_1 = 10^{-7}$ and $\beta_T = 10^{-2}$,

$$\beta_t = \beta_1 + (\beta_T - \beta_1)\,\frac{\varsigma\big(\kappa(2t/T - 1)\big) - \varsigma(-\kappa)}{\varsigma(\kappa) - \varsigma(-\kappa)}, \qquad \varsigma(u) = \big(1 + e^{-u}\big)^{-1}, \tag{7}$$

with shape parameter $\kappa = 6$ controlling the steepness of the transition, and $\bar{\alpha}_t = \prod_{s \le t}(1 - \beta_s)$. The schedule is deliberately shallow at small t and steep at large t: it holds signal-to-noise high through the late reverse steps, where bond-scale geometry is fixed, and reaches $\bar{\alpha}_T \approx 6.6 \times 10^{-3}$. The terminal marginal is therefore close to, but not exactly, $\mathcal{N}(\mathbf{0}, \mathbf{I})$: the residual conditional-mean coefficient $\sqrt{\bar{\alpha}_T} \approx 0.081$ is small but non-zero, so initialising sampling from $\mathcal{N}(\mathbf{0}, \mathbf{I})$ is an approximation, with error scaling as $\sqrt{\bar{\alpha}_T}\,\|\mathbf{x}_0\|$ in the normalised coordinate frame, rather than an exact match.

Atom types are diffused by a uniform-transition D3PM chain [3] with transition matrices

$$\mathbf{Q}_t = (1 - \beta_t^v)\,\mathbf{I} + \beta_t^v\,\tfrac{1}{C}\mathbf{1}\mathbf{1}^\top, \qquad \bar{\mathbf{Q}}_t = \textstyle\prod_{s \le t}\mathbf{Q}_s = \bar{\alpha}_t^v\,\mathbf{I} + (1 - \bar{\alpha}_t^v)\,\tfrac{1}{C}\mathbf{1}\mathbf{1}^\top, \tag{8}$$

where the second equality follows because uniform-transition matrices are closed under multiplication, so the cumulative chain is characterised by the single scalar $\bar{\alpha}_t^v$. This scalar decays faster than the coordinate schedule $\bar{\alpha}_t$, so that element identities are resolved earlier in the reverse trajectory than fine geometry [60]; committing to types first is what allows the low-t steps to optimise coordinates against a settled chemical composition rather than a shifting one.

Total loss.

The network is trained to denoise all modalities jointly,

$$\mathcal{L} = w_x\,\mathcal{L}_{\text{coord}} + \mathcal{L}_{\text{type}} + \mathcal{L}_{\text{bond}} + w_\sigma\,\mathcal{L}_{\text{alea}}, \qquad w_x = 3.0, \tag{9}$$

with the timestep drawn as $t \sim \text{Uniform}\{1,\dots,T\}$ independently per training example.

Coordinate term.

The coordinate loss is a min-SNR-weighted $\mathbf{x}_0$ mean-squared error [24],

$$\mathcal{L}_{\text{coord}} = \mathbb{E}_{t,\mathbf{z}_0,\boldsymbol{\varepsilon}}\Big[\,\omega(t)\,\big\|\hat{\mathbf{x}}_0(\mathbf{z}_t,t,\mathcal{P})-\mathbf{x}_0\big\|_2^2\Big], \qquad \omega(t) = \frac{\min\big(\text{SNR}(t),\gamma\big)}{\mathbb{E}_{t'\in\mathcal{B}}\big[\min\big(\text{SNR}(t'),\gamma\big)\big]}, \tag{10}$$

where $\text{SNR}(t) = \bar{\alpha}_t/(1-\bar{\alpha}_t)$ and $\gamma = 5$. Two properties of this weighting matter here. Capping the weight at γ prevents the easy, high-SNR low-t samples from dominating the gradient, which under an unweighted $\mathbf{x}_0$ objective they otherwise do by orders of magnitude, since $\text{SNR}(t)\to\infty$ as $t\to 0$. Normalising ω to unit mean over the timesteps actually drawn in the minibatch $\mathcal{B}$ makes the effective step size invariant to that draw, so gradient magnitude does not fluctuate with which timesteps happen to be sampled. The loss is evaluated only on ligand atoms; pocket atoms are fixed and contribute no gradient.

Atom-type term.

The atom-type loss combines the categorical variational bound with an auxiliary clean-type cross-entropy,

$$\mathcal{L}_{\text{type}} = \mathcal{L}_{\text{VLB}} + \lambda_{\text{ce}}\,\text{CE}\big(\hat{\mathbf{v}}_0,\mathbf{v}_0\big), \qquad \lambda_{\text{ce}} = 0.3, \tag{11}$$

$$\mathcal{L}_{\mathrm{VLB}} = \mathbb{E}_q\Big[\underbrace{D_{\mathrm{KL}}\big(q(\mathbf{v}_T \mid \mathbf{v}_0)\,\|\,p(\mathbf{v}_T)\big)}_{\text{prior}} + \sum_{t=2}^{T} D_{\mathrm{KL}}\big(q(\mathbf{v}_{t-1} \mid \mathbf{v}_t, \mathbf{v}_0)\,\|\,p_\theta(\mathbf{v}_{t-1} \mid \mathbf{v}_t)\big) - \log p_\theta(\mathbf{v}_0 \mid \mathbf{v}_1)\Big], \quad (12)$$

where the forward posterior is available in closed form for uniform transitions,

$$q(\mathbf{v}_{t-1} \mid \mathbf{v}_t, \mathbf{v}_0) = \mathrm{Cat}\left(\mathbf{v}_{t-1};\ \frac{\big(\mathbf{v}_t \mathbf{Q}_t^\top\big) \odot \big(\mathbf{v}_0 \bar{\mathbf{Q}}_{t-1}\big)}{\mathbf{v}_0 \bar{\mathbf{Q}}_t \mathbf{v}_t^\top}\right), \quad (13)$$

and p_θ is obtained by substituting the predicted $\hat{\mathbf{v}}_0$ for $\mathbf{v}_0$ in Eq. (13). The auxiliary cross-entropy is not redundant with the bound: the bound weights each timestep by its posterior sharpness and therefore supplies little gradient at large t, where the categorical posterior is nearly uniform, whereas the direct term keeps a usable signal on $\hat{\mathbf{v}}_0$ across the whole trajectory.

Bond term.

$\mathcal{L}_{\mathrm{bond}}$ is a cross-entropy over unordered ligand atom pairs $\{i, j\}$, $i < j$, against $B + 1$ classes (no bond, single, double, triple, aromatic), read from the pairwise head on the final scalar features and the *clean* coordinate prediction $\hat{\mathbf{x}}_0$ rather than the noised coordinates $\mathbf{x}_t$. Evaluating on $\hat{\mathbf{x}}_0$ matches the condition at generation time, where bonds are determined from near-final geometry. Non-bonded pairs outnumber bonded ones by roughly N to 1 ($\sim$92% of pairs), so the term requires rebalancing if it is not to be minimised by predicting "no bond" everywhere; this is handled by class weighting rather than negative subsampling, downweighting the no-bond class to 0.1 against 1.0 for each of the four bonded classes in the cross-entropy.

Aleatoric term.

The final term optionally trains a per-atom variance head by heteroscedastic Gaussian negative log-likelihood [42],

$$\mathcal{L}_{\text{alea}} = \mathbb{E}\Big[\tfrac{1}{2}\big\|\text{sg}\big[\hat{\mathbf{x}}_0 - \mathbf{x}_0\big]\big\|_2^2/\sigma^2 + \tfrac{3}{2}\log\sigma^2\Big], \tag{14}$$

where sg[·] is the stop-gradient operator, so that σ^2 learns the per-atom coordinate uncertainty without back-propagating into, and thereby smoothing, the geometry itself. **In the model reported throughout this paper this head is disabled.** The predictive variance that damps affinity guidance described in Two-tier multi-objective control is that of the separate affinity network, not of this head.

Optimisation.

Ligand coordinates are augmented during training with isotropic Gaussian jitter of standard deviation 0.1 Å applied before diffusion, which discourages the network from exploiting the exact crystallographic pose of the training ligand. Models were implemented in PyTorch [58] and optimised with AdamW [50] (learning rate 2×10^{-4}, weight decay 10^{-2}, $\epsilon = 10^{-8}$) for 60 epochs, with one warmup epoch followed by cosine annealing to $\eta_{\min} = 10^{-6}$ over the remaining epochs. Gradients are clipped to global norm 1.0 and training runs in mixed precision. The effective batch size of 48 is realised as a micro-batch of 6 complexes with 8 gradient-accumulation steps, a split imposed by pocket-graph memory rather than by optimisation considerations. An exponential moving average of the weights (decay 0.9999) is maintained throughout, and *all* reported molecules are sampled from the EMA weights rather than the raw ones. The random seed is fixed at 42. Training used two remote NVIDIA A40 GPUs at approximately 45 minutes per epoch, for a nominal total of ∼45 hours over the 60 configured epochs.

2.6 Fragment-based generation and valence-constrained reconstruction

GEMINI-Mol generates over rigid molecular fragments (ring systems and common functional groups) rather than independent atoms, an inductive bias with long precedent both in medicinal chemistry [18, 54] and in molecular generative modelling [22, 38]. This subsection specifies the fragment vocabulary and its templates, the generation plan and its prior, the projection operator applied during sampling, the assembly procedure that converts a sampled point cloud into a molecule, and the final relaxation.

Fragment vocabulary and templates.

The vocabulary comprises $C_f = 26$ fragment types, enumerated from the training ligands: the ring systems obtained by taking each molecule's smallest set of smallest rings and merging rings that share a bond, so that fused systems such as indole are a single type rather than two, together with the recurrent acyclic functional groups (carboxyl, amide, sulfonamide, nitro and related). Ring-system types were retained by frequency in the training set, and every atom not belonging to a retained type is assigned to the linker class. The two halves of the vocabulary are therefore built differently, and we state this explicitly because it bears on reproduction: the ring systems are enumerated from the training ligands and thresholded by frequency, whereas the acyclic functional groups are a fixed list of groups common in medicinal chemistry rather than a frequency-selected one. Together they make up the 26 types, which are enumerated in full in the released code. Each type f carries a template

$$\mathcal{T}_f = \big(\mathbf{c}_f,\ \mathbf{v}_f,\ E_f\big), \qquad \mathbf{c}_f \in \mathbb{R}^{n_f \times 3}, \tag{15}$$

where $\mathbf{c}_f$ are canonical local coordinates, $\mathbf{v}_f$ the element and aromaticity labels, and E_f the explicit internal bonds with their orders, aromatic bonds included. Templates are *not* estimated from training-set instances: each is a precomputed, idealised reference geometry for its functional group (fixed bond lengths and angles, e.g. 1.424 Å C–C for the benzene template), stored centred at the origin with a fixed, graph-canonical atom ordering. The ordering is what makes the atom-to-template correspondence used below well defined: without it the alignment in Eq. (16) would not be a well-posed problem, since the objective is defined only once each sampled atom is paired with a specific template atom.

Generation plans and the plan prior.

A generation *plan* π fixes the composition of a molecule before sampling begins. It specifies a multiset of fragment instances drawn from the vocabulary, a number of linker atoms with their element types, and consequently the total atom count N; it induces an ownership map $o : \{1, \ldots, N\} \to \{\text{fragment instances}\} \cup \{\text{linker}\}$ assigning every ligand atom either to one fragment instance or to the linker set. Plans are drawn from an empirical prior extracted from the training ligands, so that fragment co-occurrence statistics and the ratio of fragment to linker atoms match those of real molecules rather than being chosen independently. Sampled plans are then filtered on two independent criteria before any diffusion is run (Additional file 1, Note S2). *Composition*: linkers containing non-drug elements, more than one phosphorus, or an oxygen-plus-phosphorus fraction above 0.4 are rejected, since these reconstruct into peroxide and polyphosphate cages. *Size*: plans exceeding 32 heavy atoms or 16 linker atoms are rejected. Linker atom count rather than total atom count is the flexibility control, because fragment atoms are held inside rigid templates and therefore contribute essentially no rotatable bonds; the uncapped prior reaches 69 heavy and 57 linker atoms, giving molecules of 25–33 rotatable bonds that are neither drug-like nor practically dockable.

Template projection during sampling.

At every reverse-diffusion step, the atoms of every fragment instance are replaced by a rigidly transformed copy of their template: writing $\mathbf{x}_f \in \mathbb{R}^{n_f \times 3}$ for the current coordinates of the instance's atoms in template order, the transform is the closed-form solution of the orthogonal Procrustes problem

$$\left(R^\star, \mathbf{t}^\star\right) = \underset{R \in \mathrm{SO}(3),\, \mathbf{t} \in \mathbb{R}^3}{\arg\min} \left\| R\,\mathbf{c}_f + \mathbf{t} - \mathbf{x}_f \right\|_F^2, \qquad \mathbf{x}_f \leftarrow R^\star \mathbf{c}_f + \mathbf{t}^\star, \tag{16}$$

given by the Kabsch construction [41] (SVD of $\mathbf{c}_f^\top \mathbf{x}_f$, with the usual determinant correction excluding reflections so fragment chirality cannot invert). The operator discards whatever internal distortion the sampler has introduced into the fragment while retaining its position and orientation, the degrees of freedom pocket conditioning should determine; because the alignment commutes with global rotation and translation, interleaving it with the equivariant update of Eqs. (4)–(6) leaves the whole reverse process SE(3)-equivariant. Linker atoms are never projected and remain fully free throughout.

Valence-constrained assembly.

Inferring bonds from interatomic distances (the near-universal final step in atomwise generators) fabricates strained small rings and disconnects molecules whenever the geometry is imperfect. We replace it with a deterministic procedure (Additional file 1, Note S3) that takes intra-fragment bonds directly from the templates and completes the remaining connectivity with a greedy, distance-ordered pass subject to valence and ring-size constraints, followed by a spanning pass that connects any remaining components.

Bond orders come from two sources only. Intra-fragment bonds inherit the order recorded in E_f, so aromatic rings are aromatic by construction and never require perception. Cross-fragment and linker bonds are added as single bonds, which is the

correct default for a linker connection and avoids assigning multiple-bond character to geometry that does not support it. The assembly recognises eleven elements (H, B, C, N, O, F, P, S, Cl, Br, I) with standard maximum valences and a covalent-radius-plus-tolerance distance rule ($\tau = 0.45$ Å); formal charges and implicit hydrogens are assigned by RDKit once the bond list is complete (full valence table and admissible-pair rules in Additional file 1, Note S3).

The procedure admits two guarantees that hold for every output, independently of sampling quality, stated and proved in Additional file 1, Note S3: if a valence-admissible cross-component pair always exists whenever the graph is disconnected, the returned molecule is connected; and no bond the procedure adds can create a three- or four-membered ring, so every ring in the output is either a template ring or has at least five members. This is a design restriction rather than a chemical one: cyclopropane, cyclobutane, oxetane, azetidine and epoxide motifs occur in approved drugs, and the restriction excludes them from the assembled linkers. We adopt it because a distance-based rule cannot tell a genuine small ring from one fabricated by imperfect geometry, and because spurious small rings were the dominant failure of distance-based perception in our ablation (Table 5). The rare case where no valence-admissible pair exists is returned as a failure rather than a malformed molecule; such samples are counted against the reported RDKit validity of 0.997 (Table 1) rather than removed silently. The whole procedure is deterministic (ties broken by ascending atom index) and computationally negligible next to a single denoising step (Additional file 1, Note S3).

Protein-aware relaxation.

The final step minimises the assembled ligand in the presence of the pocket, so that internal geometry and protein fit are optimised together (full procedure in Additional file 1, Note S1). The natural alternative (gas-phase MMFF94 minimisation [23] followed by a rigid-body adjustment to resolve clashes) optimises them against each

other, since the force field cannot see the pocket and the rigid step that follows has only six degrees of freedom with which to repair the fit it has just lost; under that ordering the generated poses are net repulsive (Additional file 1, Note S1).

MMFF94 cannot be typed on a CrossDocked pocket record, which stores bare atoms without residue topology. The protein therefore enters as restraints rather than as part of the typed system: the ligand retains its full MMFF94 force field, and each protein atom within $r_{\mathrm{env}} = 5$ Å of any ligand atom is added as a *fixed* extra point. The minimised objective is

$$E(\mathbf{x}) = E_{\mathrm{MMFF}}(\mathbf{x}) + k \sum_{(i,j)\in\mathcal{C}} \Big[\max\big(0,\ d_{ij}^{0} - \|\mathbf{x}_i - \mathbf{r}_j\|\big) \Big]^2, \qquad d_{ij}^{0} = r_i^{\mathrm{vdW}} + r_j^{\mathrm{vdW}} - \delta, \tag{17}$$

with force constant $k = 12$ kcal mol^{-1} Å^{-2} and slack $\delta = 0.45$ Å. The restraint is one-sided: it penalises separations *below* contact distance and is exactly zero above it, so genuine close contacts, including hydrogen bonds, remain unpenalised and only real overlap is removed. Every ligand degree of freedom stays free, which is precisely what a rigid-body adjustment to resolve clashes cannot offer. Two further design choices, detailed in Additional file 1, Note S1, are load-bearing: the contact set is rebuilt every round rather than fixed once, since it goes stale as soon as atoms move, and the stage only accepts a round that does not increase clash fraction, so it can never return a pose worse than the one it received.

2.7 Two-tier multi-objective control

GEMINI-Mol controls four objectives (binding affinity, drug-likeness, synthetic accessibility and ADMET safety), of which only one is differentiable with respect to the continuous atomic coordinates the diffusion process operates on. Binding affinity is scored by an SE(3)-invariant neural network trained to reproduce docking scores and differentiable in $\mathbf{x}$. Drug-likeness (QED) [5] and synthetic accessibility (SA) [19] are

computed exactly from the discrete molecular graph and are not meaningfully differentiable in $\mathbf{x}$. ADMET properties are predicted by a Chemprop directed message-passing ensemble [27, 78], one endpoint-specific model per property, trained and evaluated on the 22 endpoints of the Therapeutics Data Commons ADMET Benchmark Group [34], spanning absorption, distribution, metabolism, excretion, toxicity, general pharmacokinetics and physicochemical properties on their official test splits; held-out performance for all 22 is given in Additional file 1, Table S2. The scalar ADMET-safety term used in selection (below) is the mean predicted risk over a 6-endpoint safety subset (hERG, Ames mutagenicity, DILI, and CYP2C9/2D6/3A4 inhibition) rather than all 22; the remaining 16 endpoints are trained and validated surrogates not currently folded into the selection objective. Because the objectives differ in this way, GEMINI-Mol handles them with two mechanisms: binding affinity, which is differentiable, steers the diffusion sampling directly through its gradient (in-loop guidance); the other three objectives, which are not, are applied afterwards by selecting among generated candidates (post-hoc selection).

In-loop guidance (binding affinity).

To bias generation toward higher affinity we sample not from $p(\mathbf{x} \mid \mathcal{P})$ but from the affinity-*tilted* distribution $\tilde{p}(\mathbf{x} \mid \mathcal{P}) \propto p(\mathbf{x} \mid \mathcal{P}) \exp\big(\lambda\, g_{\text{bind}}(\mathbf{x})\big)$, whose score is the classifier-guidance form [13] instantiated with the affinity network rather than a label classifier,

$$\nabla_{\mathbf{x}_t} \log \tilde{p}_t(\mathbf{x}_t) = \mathbf{s}_\theta(\mathbf{x}_t, t \mid \mathcal{P}) + \lambda\, \nabla_{\mathbf{x}}\, g_{\text{bind}}\big(\hat{\mathbf{x}}_0(\mathbf{z}_t, \mathcal{P})\big). \tag{18}$$

The gradient is obtained by backpropagating through the affinity network at each retained sampling step and added to the coordinate score of Eq. (3). Guidance strength is damped by the network's own predictive variance, $\lambda = \lambda_0/(1 + \beta\, \widehat{\mathrm{Var}})$, so that uncertain affinity predictions exert a weaker pull on the sampling trajectory than confident ones.

Post-hoc selection (drug-likeness, synthesisability, safety).

For the non-differentiable objectives GEMINI-Mol generates a candidate pool per pocket and selects over it. Each candidate receives a utility score

$$U(\mathbf{x}) = w_{\mathrm{b}}\,\bar{g}_{\mathrm{bind}}(\mathbf{x}) + w_{\mathrm{q}}\,g_{\mathrm{QED}}(\mathbf{x}) + w_{\mathrm{s}}\,\bar{g}_{\mathrm{SA}}(\mathbf{x}) + w_{\mathrm{a}}\big(1 - g_{\mathrm{ADMET}}(\mathbf{x})\big), \tag{19}$$

a weighted combination of normalised binding ($\bar{g}_{\mathrm{bind}} \in [0,1]$, clipped-linear in predicted affinity), QED, normalised SA and ADMET safety (1 − risk), with default weights $w = (0.4, 0.2, 0.2, 0.2)$. Selection is diversity-aware: candidates are ranked by a maximal-marginal-relevance score [10], $U(\mathbf{x}) - \lambda \max_{j \in \mathrm{picked}} \mathrm{Tanimoto}(\mathbf{x}, \mathbf{x}_j)$ on Morgan fingerprints [65] with $\lambda = 0.5$ by default, and accepted greedily under this penalised score rather than under U alone, so the selected set trades utility against redundancy with previously selected molecules instead of collapsing to molecules nearly identical to the top scorer. U additionally carries a propagated uncertainty σ_U, the weighted quadrature sum of the affinity network's epistemic standard deviation (propagated through the same clipped-linear normalisation as $\bar{g}_{\mathrm{bind}}$) and the Chemprop ensemble's mean endpoint uncertainty [44] over the 6-endpoint safety subset; candidates in the top fraction of the pool by σ_U are flagged for deferral to an orthogonal oracle (full docking, or experimental triage) rather than accepted on the surrogate score alone.

2.8 Controllable multi-target design

We define the multi-target distribution as a geometric mixture, or product of experts [15, 28, 49], of the single-target conditionals, tempered by an engagement vector $\boldsymbol{\beta} \in \mathbb{R}^K_{\geq 0}$,

$$p_{\boldsymbol{\beta}}(\mathbf{x} \mid \mathcal{P}_1, \ldots, \mathcal{P}_K) \;\propto\; \prod_{i=1}^{K} p(\mathbf{x} \mid \mathcal{P}_i)^{\beta_i}. \tag{20}$$

The gradient of its log-density is the $\boldsymbol{\beta}$-weighted sum of the single-target scores,

$$\nabla_{\mathbf{x}_t} \log p_{\boldsymbol{\beta},t}(\mathbf{x}_t) = \sum_{i=1}^{K} \beta_i \, \mathbf{s}_\theta(\mathbf{x}_t, t \mid \mathcal{P}_i), \tag{21}$$

because the data-independent normaliser of Eq. (20) vanishes under $\nabla_{\mathbf{x}}$. Substituting the score identity Eq. (3) and writing $\hat{\mathbf{x}}_0^{(i)} = \hat{\mathbf{x}}_0(\mathbf{z}_t, \mathcal{P}_i)$, the mixed score of Eq. (21) equals the score Eq. (3) of a single effective clean-coordinate prediction

$$\hat{\mathbf{x}}_0^{\boldsymbol{\beta}} = \frac{\sum_i \beta_i \, \hat{\mathbf{x}}_0^{(i)}}{\sum_i \beta_i}, \tag{22}$$

so multi-target sampling reduces exactly to running the backbone once per pocket per step and combining the clean-coordinate predictions by the normalised weights $\beta_i / \sum_j \beta_j$, with no retraining. All dual-target results reported here use the balanced setting $\beta_1 = \beta_2$.

Selectivity control law.

Equation (20) interpolates geometrically between the single-target distributions: as $\boldsymbol{\beta}$ moves along the simplex, the tempered density places relatively more mass on regions favoured by pockets with larger β_i. Writing the cross-target binding gap $\Delta\Delta G(\mathbf{x}) = E_1(\mathbf{x}) - E_2(\mathbf{x})$ for a two-target pair, the tempering acts as an exponential tilt on $\beta_1 E_1 + \beta_2 E_2$. Holding the total weight $\beta_1 + \beta_2 = 2\beta$ fixed and writing the imbalance as $\tau = (\beta_1 - \beta_2)/(\beta_1 + \beta_2)$, so that $\beta_1 E_1 + \beta_2 E_2 = \beta(E_1 + E_2) + \beta\tau \, \Delta\Delta G$, the family $p_{\boldsymbol{\beta}}$ is an exponential family in the natural parameter $-\beta\tau$ with $\Delta\Delta G$ as its sufficient statistic. Differentiating the log-partition function twice therefore gives

$$\frac{\partial}{\partial \tau} \, \mathbb{E}_{p_{\boldsymbol{\beta}}}[\Delta\Delta G] = -\beta \, \mathrm{Var}_{p_{\boldsymbol{\beta}}}(\Delta\Delta G) \;\leq\; 0, \tag{23}$$

so the expected gap is non-increasing in τ, and hence monotone in $\log(\beta_1/\beta_2)$, strictly so wherever $\Delta\Delta G$ is not almost surely constant. It vanishes at $\beta_1 = \beta_2$ by the symmetry of the two targets under exchange. This predicts a single continuous control knob from balanced polypharmacology to designed selectivity. We report the balanced case here; the monotonicity prediction across the simplex is not tested in this work.

2.9 Data curation, datasets and baselines

The backbone is trained and evaluated on CrossDocked2020 [20], using the split established by Luo et al. [51] and adopted throughout this literature: poses are refined by retaining those with binding-pose RMSD < 1 Å and the data are split at 30% protein-sequence identity, yielding ∼100,000 training pocket–ligand complexes and 100 held-out test pockets. This split is shared with all baselines, ensuring parity. The affinity predictor and the 22-endpoint Chemprop ADMET suite are trained separately on their respective labelled data and held frozen during generation; guidance uses only their gradients through Eq. (18).

We compare against four baselines spanning the main generative paradigms: TargetDiff [21], an SE(3)-equivariant diffusion model; Pocket2Mol [59], an autoregressive model with a frontier network; AR [51], autoregressive atom placement; and a 3D conditional variational autoencoder [62]. For each baseline we use the molecule sets released by its authors, re-evaluated through our pipeline rather than taken from the original reports. For evaluation we use a matched budget per pocket for every method, since the released baseline pools contain roughly 90 molecules per pocket and selecting each method's best over its full pool would give it a deeper draw than ours. Table 1 reports this matched-budget comparison; Additional file 1, Table S1 separately reports full-pool metrics under the identical pipeline and is not a like-for-like comparison with it, since GEMINI-Mol's own row there is likewise drawn from a separate full-scale run rather than from the matched-budget pool.

2.10 Evaluation pipeline

All methods are scored through a single, uniform pipeline. We report: (i) *validity and diversity*: RDKit [45] validity, uniqueness, novelty (fraction absent from the training SMILES), and connectivity; (ii) *physical validity*: PoseBusters validity [9], run in its ligand-only (intramolecular) mode: loading, sanitisation, InChI convertibility, connectivity, radicals, bond lengths, bond angles, internal steric clash, and ring and double-bond flatness (Additional file 1, Table S6 lists all eleven checks individually). This aggregate does *not* include PoseBusters' optional protein–ligand distance/overlap checks; protein–ligand steric clash is instead reported as its own, separate metric alongside it; (iii) *drug-likeness*: QED, SA and molecular size; (iv) *binding*: AutoDock Vina [16, 71] in three modes: the Vina score (the generated pose as placed), the Vina minimised score (after local optimisation) and the Vina redocking score (after redocking), together with the fraction of generated molecules beating the reference, the proportion whose Vina redocking score is at least as good as that of the pocket's crystallographic ligand (a relative threshold, not an absolute measure of affinity), and gnina CNNaffinity [52] for the multi-objective utility. Docking uses AutoDock Vina 1.2 [16] with a cubic box of side 22.5 Å centred on the pocket, exhaustiveness 8, and 5 output poses per redocking run, identical across every method and control; the best-scoring pose is retained for the redocking score. These modes are not interchangeable: only the Vina score reflects the pose the generator produced, while the redocking score searches again over torsions and rigid-body placement and therefore reports a pose chosen by the docking program. We report all three; and (v) *geometric realism*: per-type Jensen–Shannon distances [47] of bond-length, bond-angle and torsion-angle distributions. Writing the Jensen–Shannon divergence between two distributions as

$$D_{\mathrm{JS}}(P \,\|\, Q) = \tfrac{1}{2} D_{\mathrm{KL}}(P \,\|\, M) + \tfrac{1}{2} D_{\mathrm{KL}}(Q \,\|\, M), \qquad M = \tfrac{1}{2}(P + Q), \tag{24}$$

we report throughout the Jensen–Shannon *distance*, its square root,

$$\mathrm{JS}(P \,\|\, Q) = \sqrt{D_{\mathrm{JS}}(P \,\|\, Q)}, \tag{25}$$

which is a metric on distributions and, with natural logarithms, is bounded above by $\sqrt{\ln 2} \approx 0.833$. Every value tabulated as "JS" in this work is this distance, computed between each generated distribution P and a reference Q. We build the reference directly from the ligands of the CrossDocked2020 *test* set, which gives a reference-versus-reference distance of 0.000 for every bond, angle and torsion type. Because these values depend on both the reference and the binning, the values reported here are internally consistent across the five methods but are not directly comparable with values tabulated in earlier work that used a different reference or reported the divergence rather than the distance. Conformational strain is measured as the heavy-atom RMSD between each generated conformer and its own gas-phase MMFF94 minimum [23].

2.11 Antidiabetic case study

Receptor preparation.

Human pancreatic α-amylase was taken from PDB 1HNY [7] and the C-terminal catalytic subunit of human intestinal maltase-glucoamylase from PDB 3TOP [64], the latter reduced to chain A. Waters, ions and heteroatoms were removed, hydrogens added at pH 7.4, and receptors converted to PDBQT. Docking boxes were centred on the catalytic residues of each enzyme and held fixed across all runs.

Docking and controls.

One hundred molecules were generated under each of three conditions (conditioned on 1HNY alone, on 3TOP alone, and on the balanced dual-target ensemble ($\beta_1 = \beta_2$)) and every molecule was docked into *both* enzymes, so that no set is scored only against

the pocket it was designed for. Because docking scores are not comparable across receptor preparations, box definitions or exhaustiveness settings, the three marketed α-glucosidase inhibitors acarbose, voglibose and miglitol were docked through the identical pipeline, receptor files and boxes, in five independent replicates each differing only in random seed (Additional file 1, Table S15). All values reported are docking scores from an empirical scoring function and are not measured binding affinities.

Reference ligands and interaction analysis.

Both receptors were taken from the deposited structures. The 3TOP entry contains acarbose bound to chain A, and that pose was used directly as the experimental reference. For α-amylase, which is apo in 1HNY, PDB 1B2Y [55] (the same protein in complex with the acarbose-derived pseudo-pentasaccharide) was superposed (0.64 Å Cα RMSD over 495 residues) and its ligand transplanted. A docked acarbose was evaluated first and rejected as a reference: with nine rotatable bonds it poses poorly enough that only three to five interactions are detectable. Protein–ligand interaction fingerprints were computed with ProLIF [6] and compared by Tanimoto similarity; three-dimensional pharmacophore recovery was scored as the fraction of reference features (donor, acceptor, aromatic, hydrophobe, ionisable) reproduced within 2 Å. Interaction diagrams were produced with LigPlot+ [46]. All interaction analysis was performed on protein-aware-relaxed generated poses; pose quality before and after relaxation is reported in Additional file 1, Table S10. Note that position 1510 is an arginine in the deposited structure rather than the aspartate given in some residue lists, so catalytic-residue matching is performed on residue number.

2.12 Molecular dynamics and MM-PBSA free-energy calculation

System setup.

One dual-target candidate ("Compound 1", selected in Results) was carried forward for physics-based evaluation in each enzyme, using the same pose reported in the case study: the docked pose in α-amylase (1HNY) and the model's own generated pose in maltase-glucoamylase (3TOP). Each protein (1HNY; 3TOP chain A) was parameterised with the AMBER99SB-ILDN forcefield and the ligand with GAFF and AM1-BCC charges via acpype [68, 74]. Each complex was solvated in a TIP3P [39] truncated-octahedral box with a minimum 1.0 nm solute–wall distance and neutralised with Na^+/Cl^-. After steepest-descent minimisation, the system was equilibrated for 100 ps under NVT and 100 ps under NPT with position restraints on heavy atoms, followed by a 500 ns unrestrained production run (GROMACS 2025.2 [1]; 2 fs timestep, particle-mesh Ewald electrostatics, V-rescale thermostat [8] at 310.15 K, Parrinello–Rahman barostat [57] at 1 bar). Acarbose was run through the identical pipeline in the same two receptors, under the same force field, ligand parameterisation and temperature.

MM-PBSA.

Binding free energy was estimated with gmx_MMPBSA v1.6.4 [53, 72] on 501 frames per target, using the Poisson–Boltzmann solver in `sander` (internal and external dielectric constants 1.0 and 80.0, no added salt, 310.15 K). ΔG is reported as complex − receptor − ligand with the standard deviation across frames. Entropy was not estimated, because normal-mode and quasi-harmonic terms scale poorly to a system this size and would not have added reliable information given the pose instability reported in Results. Per-residue decomposition was computed for every residue within 6 Å of the ligand. The same protocol was applied to the acarbose reference trajectories (Table 8,

Additional file 1, Table S14). For maltase-glucoamylase, the acarbose receptor topology was renumbered from residue 1, so its per-residue indices do not match the full-length catalytic-residue numbers used for Compound 1; no residue-level comparison is drawn for that target.

2.13 Statistics and reproducibility

Sample sizes are stated in each table and figure legend. Benchmark comparisons use a matched budget of 12 molecules per pocket over the 100 test pockets ($n \approx 1{,}200$ per method), against 89 crystallographic reference ligands for which all checks could be computed. Reference-drug controls are reported as the mean ± standard deviation over five independent docking replicates differing only in random seed. Proportions, including PoseBusters validity, carry 95% Wilson score intervals, and differences between two proportions are tested by the two-proportion z test, or by Fisher's exact test where the counts are small. These intervals and tests treat the individual molecules as independent draws. They therefore describe molecule-level sampling uncertainty only and understate the uncertainty contributed by clustering within the 100 test pockets, each of which supplies 12 molecules; they should be read as a lower bound on the width of the true interval rather than as a pocket-level inference. Distributional comparisons are reported as medians with interquartile ranges, since the underlying property distributions are skewed and not adequately summarised by means. No data were excluded other than molecules that failed to load or sanitise, which are reported as validity failures rather than removed silently: the RDKit validity column of Table 1 and of Additional file 1, Table S1 therefore gives the retained fraction for each method, and the number of molecules entering the property and geometry statistics is that fraction of the 1,200 sampled candidates. All generation used the exponential-moving-average weights with fixed random seeds recorded in the released configuration.

3 Results

This section is organised as follows. We first compare the physical validity, geometric realism, binding scores and molecular properties of the ligands GEMINI-Mol generates against four representative baselines, all evaluated through a single pipeline. We next report an ablation that isolates the contribution of the reconstruction stage. Finally we apply multi-target design to a therapeutic target pair, α-amylase and intestinal maltase-glucoamylase, and analyse the resulting molecules against crystallographic and clinical references.

3.1 Physical validity and geometric realism

We evaluated GEMINI-Mol on the CrossDocked2020 100-pocket test set against four baselines spanning the main generative paradigms: TargetDiff [21] (equivariant diffusion), Pocket2Mol [59] (autoregressive with a frontier network), AR [51] (autoregressive atom placement) and a 3D conditional variational autoencoder (CVAE) [62]. Every method was scored through the uniform pipeline described in the Methods, at a matched budget per pocket, so that no value is inherited from a differently configured protocol.

Table 1 summarises the benchmark. GEMINI-Mol attains a PoseBusters validity [9] of 0.94, the highest of the five methods, ahead of Pocket2Mol (0.89) and substantially ahead of the diffusion baseline TargetDiff (0.58) and the CVAE (0.04), while retaining an RDKit validity of 0.997 and complete connectivity (Additional file 1, Table S1). The margin over the closest method is larger than the sampling uncertainty of the comparison: the 95% Wilson interval is 0.925–0.952 for GEMINI-Mol against 0.871–0.906 for Pocket2Mol, a difference of 0.05 (95% confidence interval 0.03–0.07, two-proportion $P = 1 \times 10^{-5}$). These intervals treat the molecules as independent and so take no account of clustering within pockets, which makes them a lower bound on the true uncertainty (Methods, Statistics and reproducibility).

Table 1 Benchmark on the CrossDocked2020 100-pocket test set under a single protocol.

Method	HA	QED↑	SA↑	Valid↑	PB-valid↑	Clash↓	Redock(all)↓	top50↓	best1↓
TargetDiff	24.5	0.48	0.59	**1.000**	0.58	**0.003**	**-7.30**	**-8.25**	**-9.48**
Pocket2Mol	17.7	**0.58**	**0.76**	**1.000**	0.89	0.004	-6.98	-7.82	-8.55
GEMINI-Mol	24.0	0.48	0.66	0.997	**0.94**	0.005	-6.41	-7.24	-8.20
AR	17.7	0.51	0.65	0.952	0.66	0.001	-6.20	-6.94	-7.75
CVAE	20.0	0.39	0.59	0.997	0.04	0.000	-6.23	-6.68	-7.19

Notes. HA, mean heavy atoms; SA, normalised synthetic accessibility (higher is better); PB-valid, PoseBusters validity; Clash, protein–ligand steric-clash fraction. Redock(all), top50 and best1 are the Vina *redocking* score (kcal mol^{-1}, lower is better) as the mean over all molecules, the mean of the per-pocket top 50% and the mean of the per-pocket best. All three therefore score a pose found by the docking search rather than the pose each generator placed; the scores of the generated poses themselves are reported separately in Additional file 1, Table S1 and discussed in the Discussion. Matched budget of 12 molecules per pocket ($n \approx 1{,}200$ per method). Best value per column in bold.

PoseBusters validity aggregates eleven independent physical tests into a single pass or fail, so a method failing one test badly is indistinguishable from one failing three mildly. Resolving the aggregate into its components separates them (Fig. 2), and we included the crystallographic ligands of the same pockets as a control, since real structures do not pass every check either and only the gap to that control is interpretable.

Two features emerge that the aggregate conceals. GEMINI-Mol's internal geometry is essentially exact: 0.0% of its molecules violate the bond-angle check and 0.1% the bond-length check, matching the crystallographic reference and bettering every other generator, whereas TargetDiff violates bond angles in 26.2% of its molecules, a failure not visible in its aggregate score of 0.58. This agrees with the distributional analysis below: two independent measures, one a continuous distance and one a physical pass/fail criterion, concur that template-constrained ring geometry reproduces correct internal coordinates. GEMINI-Mol's own residual deficit, by contrast, is confined to a single check. Internal steric clash fails for 3.8% of its molecules against 1.0% for Pocket2Mol and 1.1% for the reference, while every other check sits at or near the reference value. The deficit is therefore attributable to one specific criterion rather than

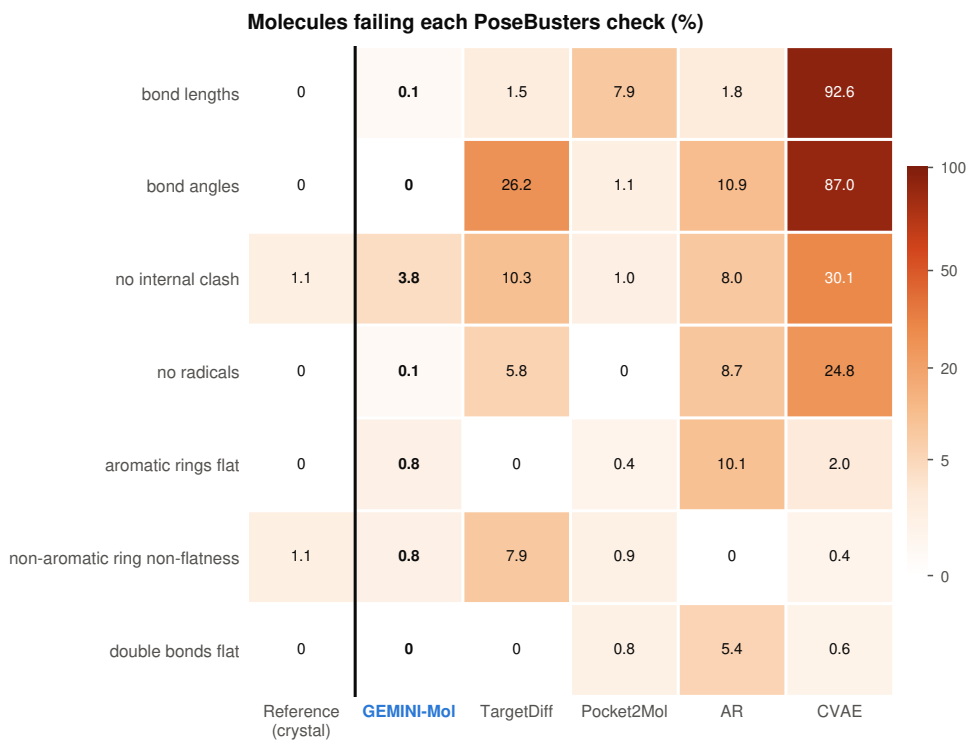

Fig. 2 PoseBusters checks resolved individually. Percentage of molecules failing each physical check, for every method and for the crystallographic reference ligands (black bars). Checks are ordered by how much they separate the methods. GEMINI-Mol sits at the reference on bond lengths and bond angles, the checks that dominate geometric realism, while TargetDiff fails bond angles in over a quarter of its molecules; the residual deficit of GEMINI-Mol is confined to internal steric clash. Four checks on which no method differs from the reference by more than 0.1 percentage points (loads, sanitises, InChI-convertible, all atoms connected) are omitted; complete values for all eleven checks are given in Additional file 1, Table S6. Matched budget of 12 molecules per pocket, $n \approx 1{,}200$ per method and 89 reference ligands.

diffuse: the protein-aware relaxation resolves *inter*molecular overlap, but its restraints do not act on contacts within the ligand itself.

Ring-size distribution.

Atom-wise generators characteristically distort ring statistics, either fabricating strained three-membered rings (AR and CVAE, 28–30% of all rings) or producing too many seven-membered rings (TargetDiff, 12%, against 1% in crystallographic ligands). The template-constrained construction produces fewer seven-or-larger rings

Table 2 Ring-size distribution of generated molecules (percentage of all rings).

	3-ring	5-ring	6-ring	7+ ring
Reference (crystallographic)	2	28	67	1
GEMINI-Mol	**0**	14	80	7
TargetDiff	0	30	49	12
Pocket2Mol	0	16	80	3
AR	30	16	51	2
CVAE	28	30	23	3

Notes. GEMINI-Mol avoids the large-ring excess of TargetDiff and the fabricated three-membered rings of AR and the CVAE. Its absence of three-membered rings is imposed by the assembly rule, not learned; the same rule also forbids four-membered rings in assembled linkers (Methods). The four-membered class is not tabulated, so rows need not sum to 100%, and percentages are rounded to the nearest integer.

than TargetDiff (7% against 12%) but still more than the reference (1%), and fabricates no three-membered rings (Table 2). It does not reproduce the reference distribution closely: five-membered rings are under-represented (14% against 28%) and six-membered rings over-represented (80% against 67%), a bias we have not yet traced to its source. The absence of three-membered rings follows by construction from the assembly rule (Methods) and is not independent evidence of realism: crystallographic ligands themselves contain a small fraction of such rings, which the rule excludes.

Bond-angle and torsional distributions.

Geometric realism is conventionally quantified as the Jensen–Shannon distance between the generated and reference distributions of each internal-coordinate type [47]: the square root of the Jensen–Shannon divergence, which is a metric on distributions and, in natural logarithms, is bounded above by $\sqrt{\ln 2} \approx 0.833$ (Methods, Eqs. (24)–(25)). Every value reported below and in Tables 3 and 4 is this distance. We computed the reference directly from the ligands of the CrossDocked2020 test set, which gives a reference-versus-reference distance of 0.000 for every bond, angle and torsion type, an

implementation sanity check confirming the computation collapses to zero for identical distributions, not evidence of statistical calibration; because these values depend on the reference and the binning, the values below are internally consistent across the five methods but are not directly comparable with those tabulated in earlier work.

GEMINI-Mol produces the most realistic bond angles and torsional angles of the five methods, achieving the lowest distance on four of five frequent angle types (Table 3) and on seven of eight frequent torsion types (Table 4). The margins are largest for aromatic geometry: the aromatic ring angle C:C:C reaches a distance of 0.07 against 0.20–0.56 for the baselines, and the aromatic torsion C:C:C:C reaches 0.10 against 0.42–0.72. Overlaying the distributions makes the origin of the difference visible (Fig. 3): the template constraint holds ring geometry at its ideal values, so the aromatic angle and torsion distributions remain sharp and superimpose on the reference, whereas the atomwise baselines broaden, and Pocket2Mol, AR and the CVAE lose aromatic planarity almost entirely. On bond *lengths* GEMINI-Mol is not uniformly ahead of the atomwise baselines: TargetDiff is best on seven of eight bond types (Additional file 1, Table S3, Additional file 1, Fig. S1), and GEMINI-Mol ranks second on four (C–C, C:C, C:N, O–P), fourth on three (C=C, C–O, C–N), and edges out TargetDiff on the eighth, C=O, by a negligible margin (0.455 against 0.456). Conformational strain is consistent with this picture: the heavy-atom RMSD between each generated conformer and its own gas-phase minimum has a median of 0.00 Å for GEMINI-Mol against 0.88 Å for the crystallographic reference and 1.13–1.38 Å for TargetDiff and the CVAE (Additional file 1, Table S8). A bound ligand is not at its own gas-phase minimum, so the reference value is non-zero by construction; GEMINI-Mol's near-zero median instead reflects that the reconstruction pipeline's own minimisation step already relaxes each molecule close to a local minimum, and should be read as a consequence of that procedure rather than as an energy-based strain estimate or as evidence of being more crystal-like than the crystal structures themselves.

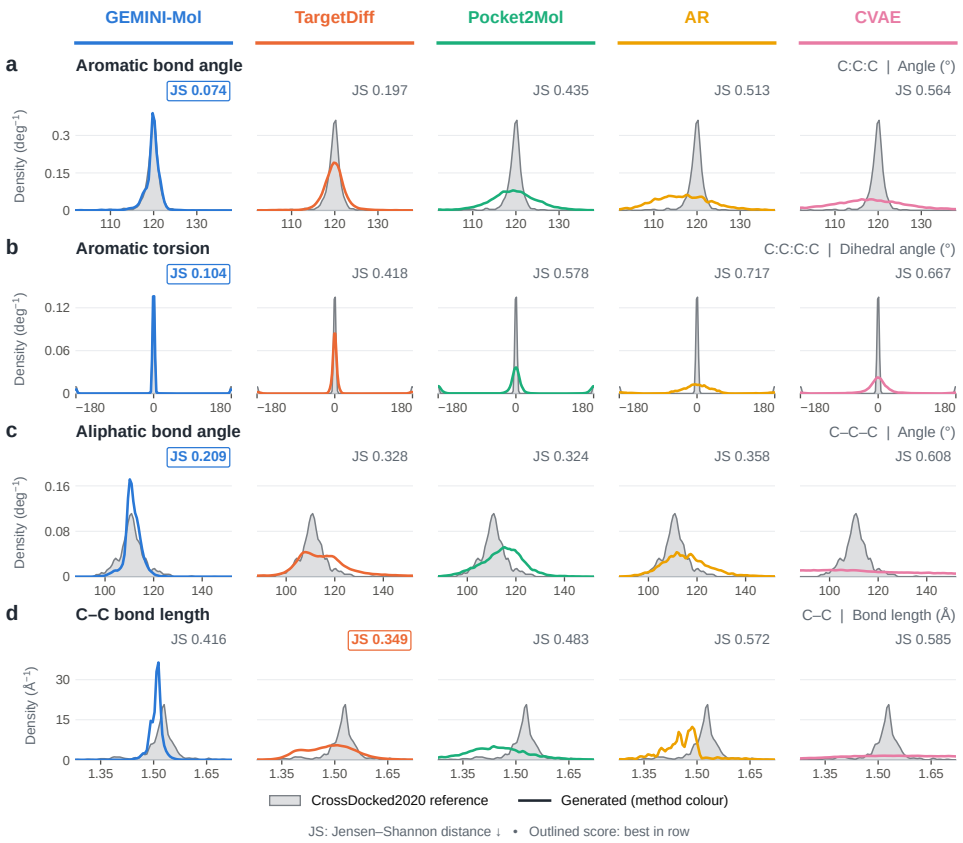

Fig. 3 GEMINI-Mol reproduces reference geometry distributions that atomwise generators broaden. a, Aromatic ring angle C:C:C. **b**, Aromatic torsion C:C:C:C. **c**, Aliphatic angle CCC. **d**, C–C bond length. Each of the five sub-panels in a row overlays one method's generated distribution (coloured) on the CrossDocked2020 test-set reference (grey fill) for that coordinate. JS denotes the Jensen–Shannon distance (the square root of the Jensen–Shannon divergence, using natural logarithms; Methods, Eq. (25)); the lowest score in each row is bold and outlined. Probability density has units of degrees^{-1} for angles and Å^{-1} for lengths, with a common y-axis scale within each row. Display curves use three-bin smoothing, with circular padding for torsions, and are normalised by the full number of observations; cropped tails are not renormalised. Scores use unsmoothed histograms on the canonical evaluation bins and retain the values reported in Tables 3 and 4. TargetDiff is closest to the reference for C–C bond length.

3.2 Binding scores and pose quality

On docking score, GEMINI-Mol is competitive but not leading (Table 1). Its mean Vina redocking score of -6.41 kcal mol^{-1} ranks third among the five methods, behind TargetDiff (-7.30 kcal mol^{-1}) and Pocket2Mol (-6.98 kcal mol^{-1}), and ahead of

Table 3 Bond-angle Jensen–Shannon distance per angle type.

Method	CCC	C:C:C	CCO	C:C:N	CCN
GEMINI-Mol	**0.209**	**0.074**	**0.178**	0.609	**0.166**
TargetDiff	0.328	0.197	0.379	**0.328**	0.329
Pocket2Mol	0.324	0.435	0.399	0.403	0.326
AR	0.358	0.513	0.443	0.442	0.386
CVAE	0.608	0.564	0.629	0.515	0.611

Notes. Jensen–Shannon distance from the CrossDocked2020 test-set reference (lower is better; Methods, Eq. (25)); best per column in bold. A colon denotes an aromatic bond and an equals sign a double bond.

Table 4 Torsion-angle Jensen–Shannon distance per torsion type.

Method	CCCC	C:C:C:C	CCOC	CCCO	CCNC	C:C:N:C	C:C:C:N	C:N:C:N
GEMINI-Mol	**0.233**	**0.104**	0.440	**0.246**	**0.370**	**0.195**	**0.186**	**0.328**
TargetDiff	0.307	0.418	0.404	0.271	0.407	0.463	0.484	0.536
Pocket2Mol	0.302	0.578	0.408	0.279	0.403	0.576	0.574	0.609
AR	0.346	0.717	0.404	0.308	0.398	0.710	0.691	0.779
CVAE	0.301	0.667	**0.393**	0.284	0.394	0.620	0.652	0.631

Notes. Jensen–Shannon distance from the CrossDocked2020 test-set reference (lower is better; Methods, Eq. (25)); best per column in bold. A colon denotes an aromatic bond.

CVAE (-6.23 kcal mol^{-1}) and AR (-6.20 kcal mol^{-1}). Retaining the best-scoring candidates per pocket, a routine step in any generate-and-filter campaign, raises this to -8.20 kcal mol^{-1}. The pattern across Table 1 is a trade-off rather than a deficiency: the two methods that dock better than GEMINI-Mol do so with molecules that fail physical checks more often, and a docking score computed on a structure that fails those checks is of limited practical value. Figure 4 makes the comparison concrete for one target. All methods place a ligand in the site, but they differ in what that ligand is: GEMINI-Mol's molecule is a compact, correctly aromatic scaffold, whereas the better-docking baselines achieve their scores with larger, more strained or highly fused structures. On this particular pocket both Pocket2Mol and TargetDiff out-dock GEMINI-Mol.

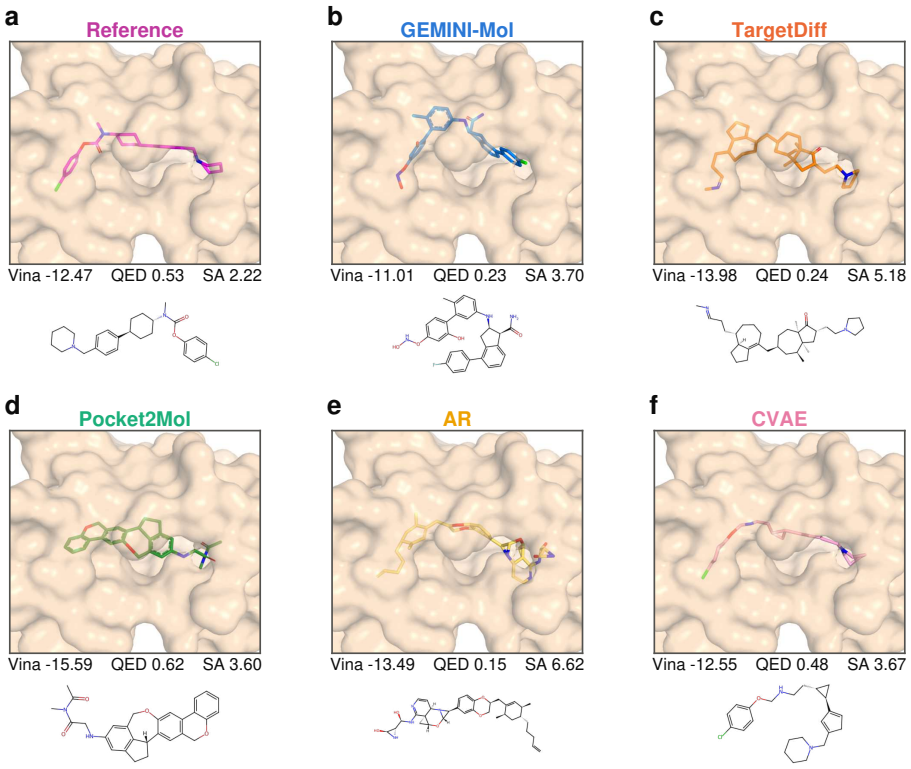

Fig. 4 Generated ligands in the binding site of a CrossDocked test pocket (squalene–hopene cyclase, pocket 52249). **a**, Crystallographic reference. **b–f**, GEMINI-Mol, TargetDiff, Pocket2Mol, AR and CVAE, respectively. For each method the best-docking molecule is shown inside the pocket surface with its Vina redocking score, QED and SA, and its two-dimensional structure beneath; the crystallographic ligand is in magenta. All panels share one camera, so poses are directly comparable. Every method is given the same 12-candidate budget, since the released baseline sets contain roughly 90 molecules per pocket and selecting each method's best over its full pool would give it a deeper draw; all molecules are re-docked here under the protocol described in the Methods rather than taken from any method's own reporting.

3.3 Molecular properties

GEMINI-Mol generates molecules with a mean QED of 0.48 and normalised SA of 0.66 (Table 1), matching TargetDiff on drug-likeness. Two considerations qualify how these summary statistics should be read. First, a diffusion model is trained to reproduce a data distribution and is not expected to exceed the training set on QED or SA; the relevant question is distributional fidelity rather than maximisation. Second,

comparing means presumes the underlying distributions have comparable shape, and here they do not (Fig. 5).

Set against the crystallographic reference, Pocket2Mol's higher QED accompanies a median of 15 heavy atoms and a single rotatable bond, where real ligands have 21 and 4; GEMINI-Mol's median of 23 heavy atoms is the closest of any method to the reference (Additional file 1, Table S7). QED rewards small, rigid molecules, so the drug-likeness gap in Table 1 reflects a size distribution shifted well below the crystallographic range at least as much as it reflects better molecules. This qualifies the comparison rather than reverses it: GEMINI-Mol's QED remains below Pocket2Mol's on the metric as defined.

The same view identifies two liabilities of GEMINI-Mol's own output. Its logP distribution is displaced upward, with a median of 2.65 against the reference's 1.22, so the molecules are more lipophilic than crystallographic ligands. And its rotatable-bond count runs high, with a median of 6 against 4. Because fragment atoms are held inside rigid templates, essentially every rotatable bond lies in the linker, so linker length is the direct control on flexibility; plans are consequently capped at 32 heavy atoms and 16 linker atoms (Additional file 1, Note S2). Molecular size is set directly by the fragment-plan distribution, which makes it a tunable design parameter rather than a fixed property of the model: restricting the plan prior to smaller molecules moves the mean heavy atom count from 28.0 to 24.0 and recovers most of the QED gap to the smaller-molecule baselines at a modest binding cost (per-pocket-best Vina redocking score from -8.43 kcal mol^{-1} to -8.20 kcal mol^{-1}), leaving the physical-validity and geometry results unchanged or improved.

3.4 The reconstruction stage is the source of the validity gains

To determine which component produces the physical-validity and geometry advantages, we ablated the reconstruction stage over the full 100-pocket benchmark,

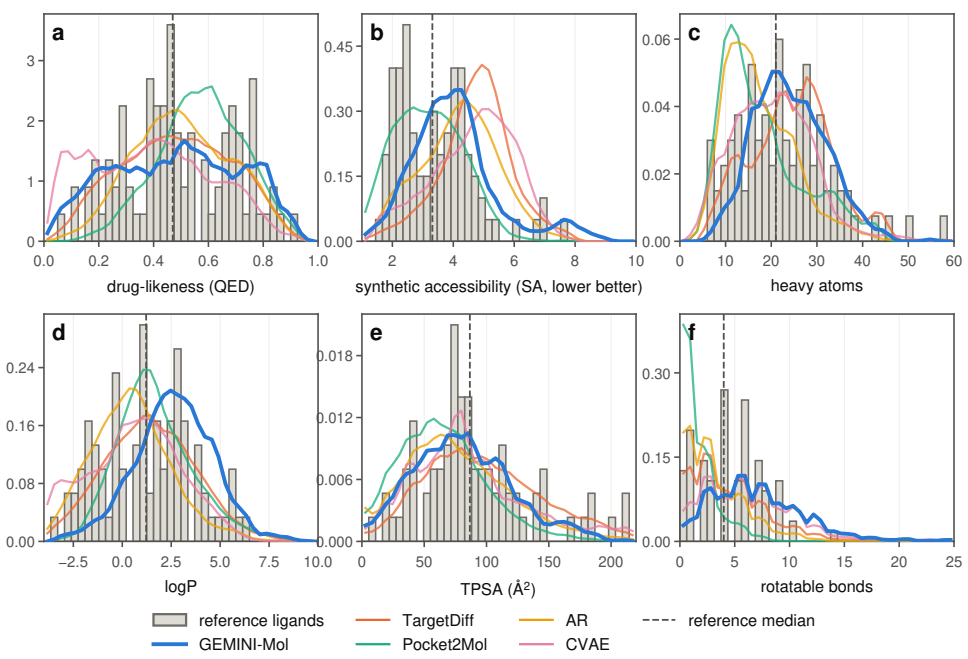

Fig. 5 Property distributions against the crystallographic reference. **a**, Drug-likeness (QED). **b**, Synthetic accessibility (SA, lower is better). **c**, Heavy atoms. **d**, logP. **e**, Topological polar surface area. **f**, Rotatable bonds. Each panel is a full distribution for one property, overlaid on the reference ligands (grey fill) with the reference median marked (dashed); y-axes are probability density and are not on a common scale between panels. Distribution shape, rather than the mean, carries the comparison: Pocket2Mol attains high QED with molecules far smaller and more rigid than any in the reference set, while GEMINI-Mol's size distribution tracks the reference most closely. The panels equally show GEMINI-Mol's liabilities: logP displaced upward, and a rotatable-bond tail extending to 25 in the uncapped configuration. Numerical values are given in Additional file 1, Table S7.

replacing the valence-constrained assembly with conventional distance-based bond perception and holding the trained backbone and all sampling settings fixed (Table 5). Distance-based perception fabricates small rings (26% of molecules acquire a spurious three-membered ring) and leaves 13% of molecules disconnected, capping PoseBusters validity at 0.63. The valence-constrained assembly removes both failure modes, giving no small-ring artefacts and complete connectivity, and raises PoseBusters validity to 0.92 with binding essentially unchanged (-6.79 against -6.68 kcal mol^{-1}). Both arms of the ablation are a separate sampling run from the one behind Table 1, which is why the valence-constrained arm reads 0.92 here against 0.94 at the matched budget of Table 1 and 0.943 for the full-scale run of Additional file 1, Table S1; the three runs

Table 5 Ablation of the reconstruction stage.

Reconstruction	3-ring artefact	Connectivity	RDKit-valid	PB-valid	Redock score ($kcal\,mol^{-1}$)
Distance-based perception	26%	0.87	0.86	0.63	-6.79
Valence-constrained (ours)	**0%**	**1.00**	**1.00**	**0.92**	-6.68

Notes. Full 100-pocket benchmark with identical trained backbone and sampling settings; the two arms differ only in the reconstruction procedure. Redock score is the mean Vina redocking score over all molecules. Both arms come from a sampling run separate from that of Table 1, so the valence-constrained arm reads 0.92 here against 0.94 there; the contrast between the two arms, not the absolute level, is what this table reports. The valence-constrained assembly delivers the physical-validity gains at no cost to binding.

differ only in sampling draw, and it is the within-ablation contrast, not the absolute level, that the comparison rests on. The reconstruction is therefore the decisive component behind these results, and because it is a deterministic post-sampling procedure it requires no retraining of the backbone and is transferable to other fragment-based generators.

3.5 Dual-target design against the antidiabetic enzymes

We applied multi-target design to a concrete therapeutic problem with an established dual-target rationale: postprandial glycaemic control in type 2 diabetes, in which inhibiting pancreatic α-amylase (PDB 1HNY [7]) and the C-terminal catalytic subunit of intestinal maltase-glucoamylase (PDB 3TOP [64]) together slows carbohydrate hydrolysis at two consecutive steps. Acarbose, the clinical standard, is itself a dual inhibitor [26, 43], so the design objective is one a marketed drug already satisfies, which makes this a demanding test rather than a favourable one.

We generated 100 molecules under each of three conditions (conditioned on α-amylase alone, on maltase-glucoamylase alone, and on the balanced dual-target ensemble) and docked every molecule into *both* enzymes, so that no set is scored only against the pocket it was designed for. Because docking scores are not comparable across box definitions, exhaustiveness settings or receptor preparations, quoting literature values for the clinical drugs would be uninformative; we instead docked all

three marketed α-glucosidase inhibitors (acarbose, miglitol and voglibose) through the identical pipeline, receptor file and box used for the generated molecules, in five independent replicates each (Table 6, Additional file 1, Table S15). The top-ranked per-molecule docking scores for all three generated sets, in both enzymes, are shown as 2D structures in Additional file 1, Fig. S5–S7.

The comparison is favourable but not uniformly so, and the asymmetry is informative. Figure 6 shows the paired per-molecule scores, the cross-target differences and the joint-threshold proportions for the three conditioning modes; Table 6 summarises them against the reference drugs. Against maltase-glucoamylase the dual-target set has a mean score of -7.00 kcal mol^{-1}, marginally better than acarbose (-6.97 ± 0.43 kcal mol^{-1}), with 56% of individual molecules scoring below the drug, and it exceeds miglitol (-4.95 kcal mol^{-1}) and voglibose (-5.59 kcal mol^{-1}) by a wide margin. Against α-amylase the same set reaches only -6.51 kcal mol^{-1} against acarbose's -7.94 ± 0.05 kcal mol^{-1}, and just 10% of molecules score below the drug. Acarbose is a 45-heavy-atom pseudotetrasaccharide that occupies the extended amylase cleft in a manner that the $\sim$23-heavy-atom designs do not reproduce: whatever their docking score, no generated molecule spans and engages that extended cleft the way acarbose does.

Two considerations qualify a purely score-based reading, in different directions. The generated molecules are far more drug-like by conventional medicinal-chemistry criteria than the carbohydrate-mimetic drugs they are compared against, with a mean QED of 0.54 for the dual set against 0.10 for acarbose, 0.25 for voglibose and 0.33 for miglitol. This is not straightforwardly an advantage for this indication, however: acarbose's oral bioavailability of $\sim$1–2% is a deliberate feature of α-glucosidase-inhibitor pharmacology rather than a liability, since the drug acts locally on brush-border enzymes in the intestinal lumen and low systemic absorption limits exposure elsewhere in the body [12]; a more membrane-permeant, higher-QED analogue is not obviously

Table 6 Docking scores against the antidiabetic enzymes with reference-drug controls.

Set	QED	1HNY mean	1HNY beat acarb.	3TOP mean	3TOP beat acarb.
Generated ($n = 94$–100)					
Conditioned on 1HNY	0.53	−6.21	5%	−6.78	40%
Conditioned on 3TOP	0.55	−6.67	12%	−6.94	46%
Dual-target ensemble	0.54	−6.51	10%	**−7.00**	**56%**
Reference drugs (5 replicates)					
Acarbose	0.10	**−7.94 ± 0.05**	—	−6.97 ± 0.43	—
Voglibose	0.25	−5.31 ± 0.06	—	−5.59 ± 0.03	—
Miglitol	0.33	−4.45 ± 0.07	—	−4.95 ± 0.04	—

Notes. All values are AutoDock Vina scores (kcal mol^{-1}, lower is better) computed through one pipeline with identical receptor files, box and exhaustiveness for generated molecules and controls alike. Controls are the three marketed α-glucosidase inhibitors, each docked in five independent replicates (mean ± s.d.). 1HNY is α-amylase and 3TOP maltase-glucoamylase. "Beat acarb." is the percentage of generated molecules scoring below the acarbose mean in that enzyme. These are docking scores from an empirical scoring function, not measured binding affinities.

preferable on that basis. The properties that would actually matter for a gut-localised inhibitor (intestinal solubility and stability, resistance to luminal degradation, confirmation that systemic absorption remains low) are not addressed by QED and are not measured here. Separately, the best individual designs reach −9.30 kcal mol^{-1} (α-amylase) and −8.86 kcal mol^{-1} (maltase-glucoamylase), so it is the population mean rather than the attainable ceiling that trails acarbose in the amylase site.

Interaction fingerprints and catalytic engagement.

Docking scores alone do not establish that a molecule binds like an inhibitor, so we compared each design's protein–ligand interaction fingerprint [6] with that of a crystallographic acarbose pose. We used the deposited structures. The 3TOP entry contains acarbose bound in chain A, so its pose was taken directly from the entry. 1HNY is apo, so for α-amylase we superposed 1B2Y, the same protein in complex with the acarbose-derived pseudo-pentasaccharide [55] (0.64 Å Cα RMSD over 495 residues), and transplanted its ligand. A docked acarbose was evaluated first and rejected as a reference: with nine rotatable bonds it poses poorly enough that only three to five interactions are detectable.

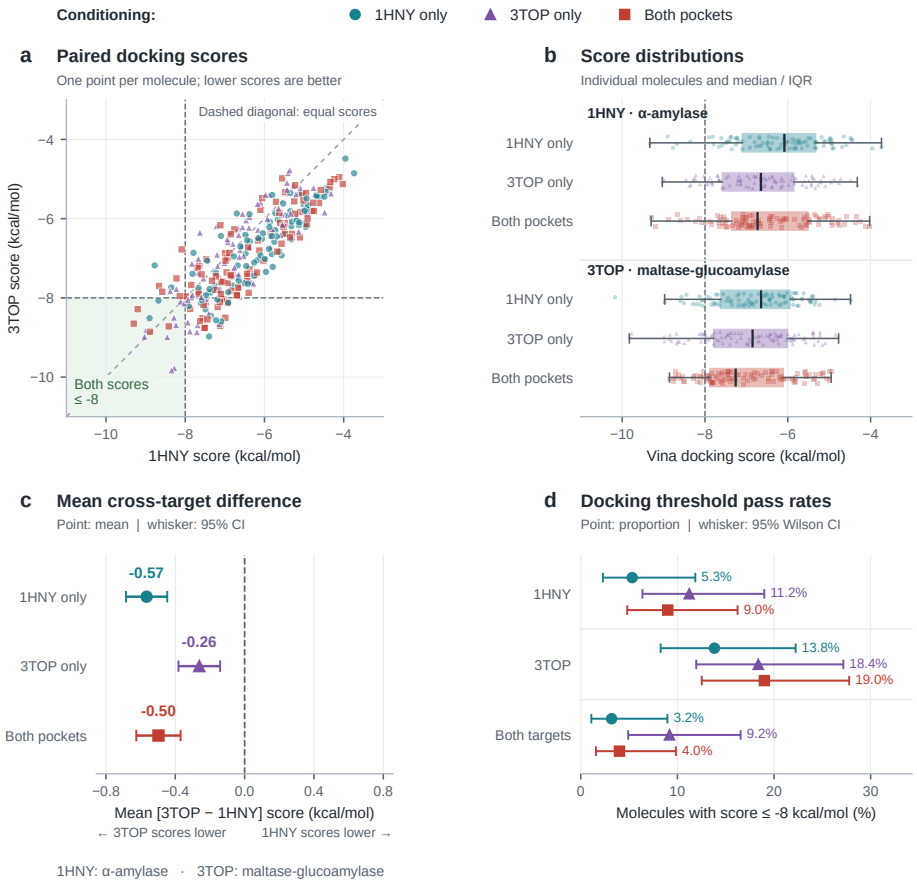

Fig. 6 Paired docking scores under single- and dual-pocket conditioning. Colours and symbols identify conditioning on 1HNY only, 3TOP only, or both pockets; each molecule is docked into both enzymes. 1HNY is α-amylase; 3TOP is maltase-glucoamylase. **a**, Paired Vina scores, with equal axis scales and lower scores toward the bottom left. Shading marks scores ≤ -8 kcal mol^{-1} in both targets; the diagonal marks equal numerical scores. **b**, Individual scores with boxplots showing the median, interquartile range and 1.5×IQR whiskers. **c**, Mean paired difference (3TOP minus 1HNY) with approximate 95% confidence intervals (mean ± 1.96 standard errors); negative values indicate lower scores in 3TOP. **d**, Proportions meeting the same threshold in each target or in both simultaneously, with 95% Wilson intervals. The analysis uses 94, 98 and 100 available paired records, respectively. These intervals describe molecule-level sampling uncertainty within the generated sets, not variation across independent generation runs. Docking thresholds and cross-target score differences do not establish biological activity or binding selectivity.

Against the experimental references, both the maltase-glucoamylase-conditioned and the dual-target sets reach a median interaction-fingerprint Tanimoto of 0.62, against 0.18 for the α-amylase-conditioned set. Catalytic-site engagement follows the same split: 43% and 39% of maltase-glucoamylase and dual molecules hydrogen-bond

at least one catalytic residue and 11% and 10% engage two or more, which is parity with crystallographic acarbose, which hydrogen-bonds exactly two in each enzyme; the corresponding figure for α-amylase is 13% (Additional file 1, Table S11, Additional file 1, Table S12). Trp1355 is contacted by 88–92% of molecules but has no detected hydrogen bonds in these poses, illustrating the distinction between proximity or nonpolar contacts and hydrogen bonds under the analysis criteria.

A candidate with predicted contacts in both sites.

Figure 7 follows one molecule, Compound 1, through both enzymes. Of the 100-molecule dual-target set, 32 pass a chemical-plausibility gate (RDKit-sanitisable, QED ≥ 0.45, no hypervalent or aromatic-sulfur centres, and no catenated triazo-type nitrogen runs) and engage at least one catalytic residue by predicted hydrogen bond in either enzyme. One of those 32 meets the catalytic-contact criterion in both sites. The selected contacts shown are between a ligand nitrogen and Asp300 OD2 in α-amylase (3.34 Å), and between a ligand nitrogen and Asp1526 OD2 in maltase-glucoamylase (2.72 Å); both are also marked by the archived LigPlot+ analysis. Additional N/O proximities detected by the distance-only 3D rendering utility are not treated as confirmed hydrogen bonds.

The two displayed poses have different provenance. Compound 1 was generated in the 3TOP coordinate frame and relaxed in the presence of that protein; its 1HNY pose was obtained by docking the same molecule into 1HNY. The stored docking runs give -7.55 and -7.32 kcal mol^{-1} for 1HNY and 3TOP, respectively, but the latter belongs to the separate docked 3TOP pose, not the generated pose shown here. The initial cross-docking screen used for ranking gave -7.65 and -7.18 kcal mol^{-1}, respectively. Candidate prioritisation considered chemical plausibility and predicted contacts alongside docking scores. Neither docking scores nor these contacts establish inhibition or persistent binding; Compound 1 was subsequently assessed by molecular dynamics.

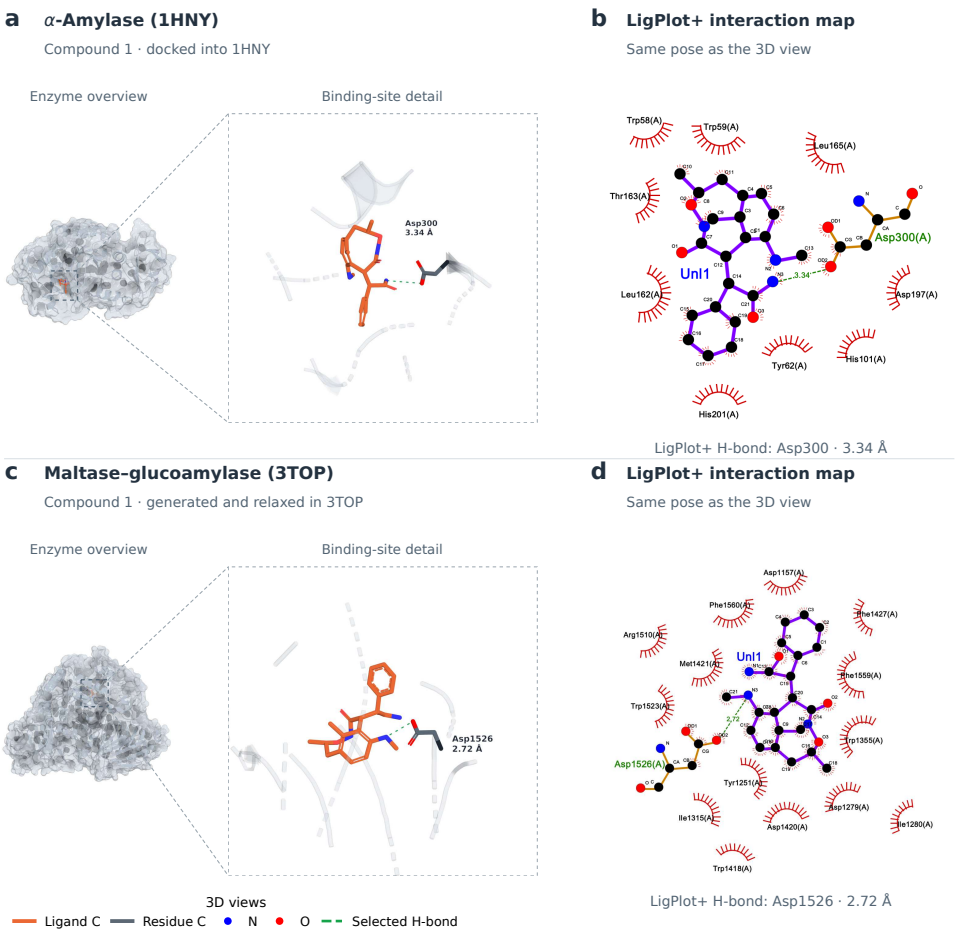

Fig. 7 Predicted binding poses of Compound 1 in two enzymes. a, α-Amylase (1HNY), with a docked ligand pose. **c**, Maltase-glucoamylase (3TOP), with the generated, protein-aware relaxed pose. Whole-enzyme views locate the binding sites; callouts show enlarged pocket views from the gemini-mol-suite PyMOL pipeline. **b**, **d**, LigPlot+ maps of the corresponding poses [46]. The selected LigPlot+ hydrogen bonds are shown in both views: Asp300 OD2–N, 3.34 Å, and Asp1526 OD2–N, 2.72 Å, respectively. Distances are between heavy atoms. Residue numbering and chain A follow the input receptors. These predicted contacts do not establish inhibition or persistent binding.

Taken together, these experiments show that pocket conditioning transfers from benchmark pockets to two real therapeutic targets, that the resulting molecules are competitive with the clinical standard against maltase-glucoamylase while being far more drug-like, and that a fraction of them reproduce the catalytic interaction pattern

of a crystallographic drug pose. They do not establish dual potency. The α-amylase arm trails acarbose on every measure reported (mean score, fingerprint similarity and catalytic engagement) and only 4% of dual-mode molecules clear -8 kcal mol^{-1} in both enzymes simultaneously, nominally below the 9% reached by the maltase-glucoamylase-only set scored against both enzymes after the fact, though the two rates are not statistically distinguishable at this sample size (Additional file 1, Table S9). Dual-conditioned generation is therefore not shown to raise simultaneous-hit yield over single-target generation screened post hoc; the case for it rests on the mean docking scores and interaction patterns, not on joint-hit rate. Docking scores are not affinities, and the case for these molecules rests on an empirical scoring function. We therefore subjected the compound in Fig. 7 to unbiased molecular dynamics and MM-PBSA in both enzymes, reported next; the result tempers rather than confirms the docking-based picture.

Unbiased dynamics do not confirm a persistently bound pose.

Each complex (the docked pose in α-amylase, the model's own generated pose in maltase-glucoamylase) was solvated, equilibrated and run for 500 ns without restraints (GROMACS 2025.2, AMBER99SB-ILDN/GAFF, TIP3P, 310.15 K; Methods, Molecular dynamics and MM-PBSA free-energy calculation). The receptor backbone is stable in both systems (Cα RMSD 1.9–2.2 Å, plateauing within 50 ns; per-residue RMSF is likewise similar with or without the generated ligand present; Table 7), so the protein itself is not the source of what follows. The ligand is not persistently retained in the catalytic cleft that docking placed it in. Minimum protein–ligand distance stays below the 0.35 nm close-contact threshold for only 82% of the α-amylase trajectory and 80% of the maltase-glucoamylase trajectory (Fig. 8, Fig. 9), against 96% and 100% for acarbose run through the same close-contact metric under the identical protocol (Methods, Molecular dynamics and MM-PBSA free-energy calculation). Direct hydrogen-bonding to the protein is correspondingly sparse and intermittent

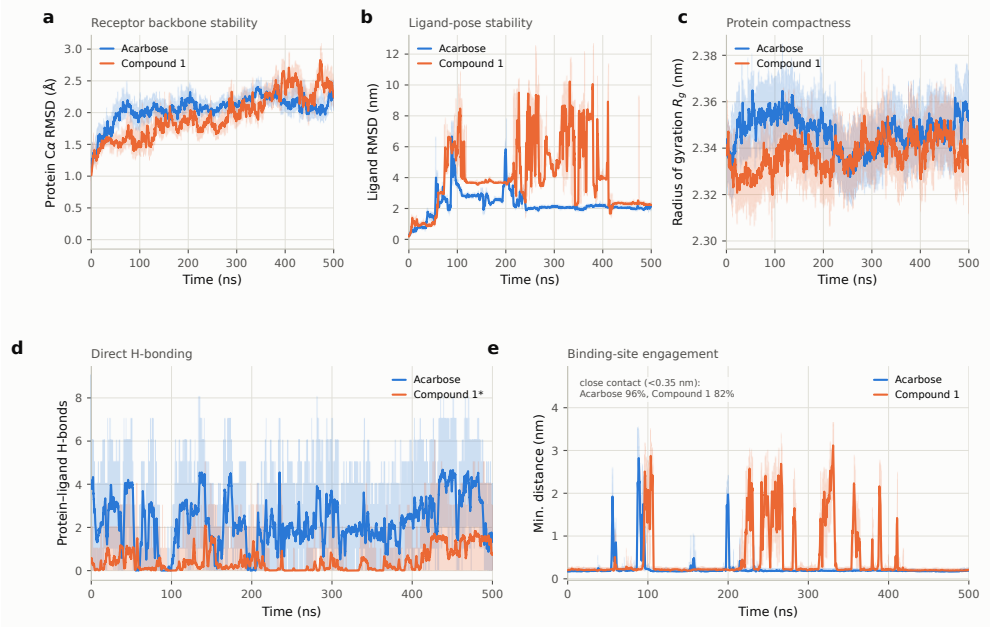

Fig. 8 Compound 1 versus acarbose in α-amylase (1HNY), 500 ns unbiased molecular dynamics. a, Protein Cα RMSD, comparable for both ligands. **b**, Ligand RMSD; Compound 1 (orange) shows sustained large-amplitude excursions from ∼100 ns that acarbose (blue) does not. **c**, Radius of gyration. **d**, Protein–ligand hydrogen bonds: Compound 1 sustains far fewer than acarbose (marked with an asterisk because the two use the same geometric criteria but sample very different bound populations, so the counts are not a strictly matched comparison). **e**, Minimum protein–ligand distance; close contact is maintained for 96% of the acarbose trajectory against 82% for Compound 1, both run under the identical protocol (Methods).

(typically 0–2 bonds against acarbose's sustained 2–5) and structural snapshots at 75, 250 and 425 ns show the ligand leaving the catalytic cleft in both enzymes: by 251 ns in α-amylase it is fully solvent-separated from the protein in the rendered frame, and by 425 ns in both enzymes it occupies a peripheral surface loop rather than the site docking placed it in.

MM-PBSA binding free energy over 501 frames per target (Poisson–Boltzmann solver, per-residue decomposition; [53, 72]; Methods, Molecular dynamics and MM-PBSA free-energy calculation) is nonetheless favourable on average: $\Delta G = -9.6 \pm 6.6$ kcal mol^{-1} in α-amylase and -8.4 ± 5.2 kcal mol^{-1} in maltase-glucoamylase (mean ± s.d. across frames; Table 8). The uncertainty is the same size as the mean, which

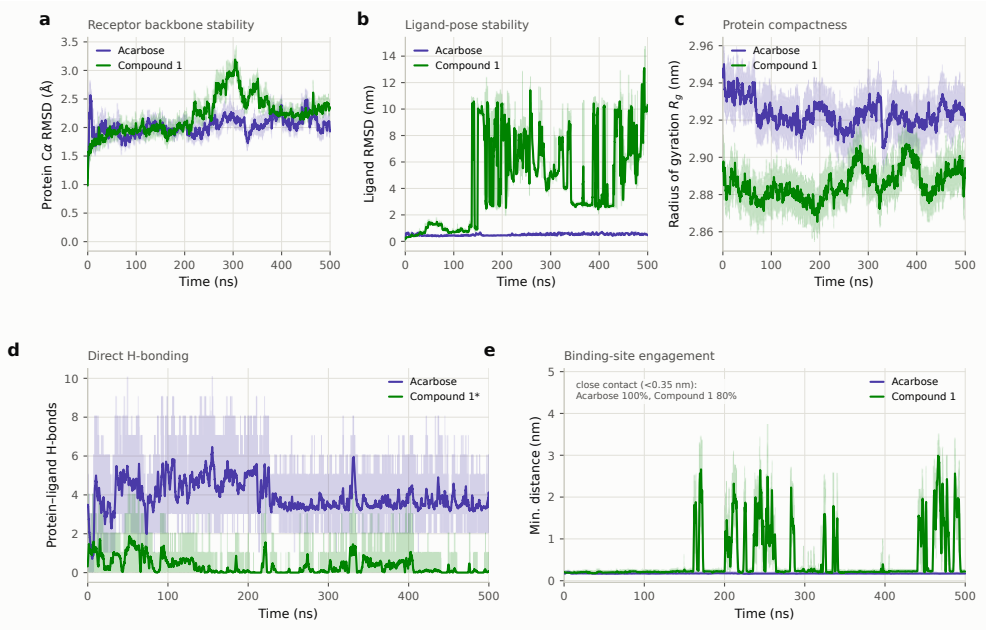

Fig. 9 Compound 1 versus acarbose in maltase-glucoamylase (3TOP), 500 **ns unbiased molecular dynamics.** Panels as in Fig. 8. Compound 1's ligand RMSD (panel **b**) departs from the acarbose baseline after ∼130 ns and does not return; close contact is maintained for 80% of the Compound 1 trajectory against 100% for acarbose (panel **e**).

is the thermodynamic signature of the instability just described rather than a separate observation: frames in which the ligand has left the pocket contribute weakly favourable or unfavourable energies alongside frames in which it has not, and the two populations are averaged together. Per-residue decomposition (Additional file 1, Table S13) makes the same point from the opposite direction: no single residue's mean contribution in either target exceeds 0.7 kcal mol^{-1} against a per-residue standard deviation of 8–28 kcal mol^{-1}, so nothing in the decomposition is distinguishable from noise, and the catalytic residues engaged in the static docked pose (Asp300 in α-amylase; Asp1420, Asp1157 and Arg1510 in maltase-glucoamylase) are not energetic hotspots. A favourable mean free energy computed this way should be read as directionally consistent with binding, not as confirmation of the specific catalytic engagement reported above.

Table 7 Molecular-dynamics trajectory statistics for Compound 1 and acarbose.

Metric	1HNY Compound 1	1HNY acarbose	3TOP Compound 1	3TOP acarbose
Protein RMSD (nm)	0.194 ± 0.036	0.206 ± 0.020	0.221 ± 0.033	0.202 ± 0.016
Ligand RMSD (nm)	4.020 ± 2.249	2.258 ± 0.784	4.595 ± 3.421	0.507 ± 0.081
Protein RMSF (nm)	0.105 ± 0.089	0.098 ± 0.067	0.107 ± 0.092	0.088 ± 0.067
Protein SASA (nm^2)	201.51 ± 3.73	200.33 ± 3.17	331.66 ± 4.74	333.91 ± 4.00
Ligand SASA (nm^2)	5.79 ± 0.18	8.62 ± 0.29	5.80 ± 0.18	8.52 ± 0.30
H-bonds	0.45 ± 0.74	2.23 ± 1.51	0.34 ± 0.63	3.95 ± 1.18
Radius of gyration (nm)	2.338 ± 0.009	2.347 ± 0.009	2.886 ± 0.009	2.923 ± 0.009

Protein RMSD (1.9–2.2 Å across all four systems), RMSF ($\approx$ 0.09–0.11 nm mean, similar spread throughout) and radius of gyration are essentially indistinguishable between Compound 1 and acarbose in either enzyme, so the large differences in ligand RMSD, ligand SASA and H-bond count are attributable to the ligand's behaviour in the pocket, not to receptor destabilisation. Compound 1's ligand RMSD (fit to the protein backbone, so it captures displacement from the binding site) is 4–20$\times$ acarbose's in both enzymes, and acarbose's higher SASA reflects its larger size: it is a 45-heavy-atom pseudotetrasaccharide, against the $\sim$23 heavy atoms of the dual-target designs (above). Acarbose's H-bond count in α-amylase (2.23 ± 1.51) is nonetheless the system in which its own MM-PBSA free energy is weakest and non-stationary (Table 8, below), underscoring that contact-based metrics and computed binding free energy are not interchangeable.

Notes. 1HNY is α-amylase and 3TOP maltase-glucoamylase. Mean $\pm$ s.d. over the full 500 ns trajectory per system (one trajectory per system; Methods). Ligand RMSD is fit to the protein backbone, so it reflects displacement from the binding site; RMSF is the mean $\pm$ s.d. of per-residue backbone values; H-bonds are the instantaneous protein–ligand count, averaged over the trajectory.

We subsequently ran the same MM-PBSA and decomposition protocol on the acarbose trajectories themselves, run under the identical simulation protocol as Compound 1 (Methods), and the two enzymes diverge sharply. In maltase-glucoamylase, acarbose's binding free energy is strongly favourable and stable for the full 500 ns ($\Delta G = -31.8 \pm 6.9$ kcal mol^{-1}; block-averaged over non-overlapping 50-frame windows it stays between -25 and -34 kcal mol^{-1} throughout), consistent with the sustained contact and hydrogen bonding already reported and with acarbose engaging this pocket more effectively than Compound 1 by every measure in this paper. In α-amylase, by contrast, acarbose's binding free energy is favourable only over roughly the first 150 ns (successive 50-frame block means of -17.5 and -5.9 kcal mol^{-1}) and then relaxes to a value indistinguishable from zero for the remainder of the trajectory, so the full-trajectory average ($\Delta G = -2.4 \pm 6.1$ kcal mol^{-1}; Table 8) is nominally weaker than Compound 1's own -9.6 ± 6.6 kcal mol^{-1} in the same pocket, despite acarbose

Table 8 MM-PBSA binding free energies and components for Compound 1 and acarbose.

Component	1HNY Compound 1	1HNY acarbose	3TOP Compound 1	3TOP acarbose
ΔVDWAALS	-16.36 ± 10.04	-3.53 ± 8.61	-14.74 ± 9.43	-31.97 ± 6.22
ΔEEL	-6.91 ± 7.19	-23.15 ± 37.89	-6.42 ± 8.59	-426.11 ± 26.83
ΔEPB	15.47 ± 11.26	24.76 ± 41.50	14.59 ± 13.86	430.56 ± 29.65
ΔENPOLAR	-1.81 ± 1.04	-0.46 ± 1.09	-1.84 ± 1.13	-4.32 ± 0.60
ΔGGAS	-23.27 ± 15.64	-26.68 ± 46.10	-21.16 ± 15.64	-458.08 ± 30.76
ΔGSOLV	13.65 ± 10.40	24.30 ± 40.45	12.75 ± 12.93	426.24 ± 29.17
ΔTOTAL	$\mathbf{-9.62 \pm 6.56}$	$\mathbf{-2.38 \pm 6.11}$	$\mathbf{-8.42 \pm 5.21}$	$\mathbf{-31.84 \pm 6.86}$

Poisson–Boltzmann solver in `sander`, 310.15 K, default dielectric and radii settings. Entropy not estimated (Methods). Acarbose's α-amylase total averages a non-stationary 500 ns series that starts favourable and relaxes to ≈ 0 after $\sim$150 ns (above); its maltase-glucoamylase total is stable throughout. In that maltase-glucoamylase system the two polar terms, ΔEEL and ΔEPB, are each roughly an order of magnitude larger than in the other three systems and cancel to within about 1% of their own magnitude, leaving a total set almost entirely by ΔVDWAALS. Neither polar term is interpretable on its own at that scale, and only the totals are compared in the text.

Notes. 1HNY is α-amylase and 3TOP maltase-glucoamylase. Values in kcal mol^{-1}, mean ± s.d. across 501 frames per system. ΔX is complex − receptor − ligand. GGAS is the sum of the gas-phase terms (van der Waals, electrostatic and internal); GSOLV is EPB + ENPOLAR. Per-residue decomposition is given in Additional file 1, Table S13 (Compound 1) and Additional file 1, Table S14 (acarbose), run under the identical simulation protocol (Methods).

maintaining close contact for far more of the trajectory (96% against 82%). This is the same qualitative lesson as the pose-instability result above, now visible in the reference compound: a favourable starting pose, even a crystallographic one, does not guarantee a favourable equilibrated binding free energy, and coarse contact metrics can stay satisfied while the underlying computed thermodynamics drifts. Per-residue decomposition of the acarbose runs shows the same noise-dominated pattern as Compound 1's: for the residues flagged in α-amylase, Asp300 and Glu233 remain indistinguishable from zero (-0.03 ± 21.3 and -0.09 ± 13.2 kcal mol^{-1}; Additional file 1, Table S14), and across all residues in both acarbose trajectories no single mean exceeds 2.0 kcal mol^{-1} in magnitude against a per-residue standard deviation of 3.6–35.0 kcal mol^{-1}.

This result does not overturn the docking-based case study: the static interactions in Fig. 7 are real geometric features of a real generated pose, and a molecule with no affinity at all would not average to a favourable MM-PBSA score across five

hundred nanoseconds of unbiased motion. But it reframes what that pose represents. A hydrogen bond present in one relaxed snapshot is a claim about a local energy minimum, not about residence time, and residence time is what unbiased dynamics measures directly. Read together, the docking and MD results say that Compound 1 finds the catalytic site and interacts with it in a manner geometrically consistent with a glycosidase inhibitor, but does not remain there under thermal motion on the hundreds-of-nanosecond timescale accessible here, a gap of the same character as the in-place docking gap already reported (the poses a scoring function favours are not always the poses that persist), now observed one level further down the validation chain. It sharpens rather than removes the case for the wet-lab work argued for in the Discussion.

4 Discussion

We set out to address two limitations that bound what current pocket-conditioned generators can contribute to a design campaign: that the molecules they produce are frequently not physically realisable, and that the objective they optimise is narrower than the one drug discovery poses. GEMINI-Mol attains the highest physical validity of the five methods benchmarked, the most accurate bond-angle and torsional geometry, and a ring-size distribution that avoids both the fabricated small rings of autoregressive models and the excess large rings of atomwise diffusion, while placing third of the five methods on mean docking score and while treating multi-target engagement and molecular developability as design variables rather than as filters.

The ablation localises the physical-validity advantage precisely. Holding the trained backbone and every sampling setting fixed and substituting distance-based bond perception for the valence-constrained assembly returns PoseBusters validity from 0.92 to 0.63, reintroduces spurious three-membered rings in 26% of molecules and disconnects 13% of them. For this backbone, it is therefore not the coordinate predictions

that determine physical validity but the way connectivity is inferred from them, an ablation that isolates one architecture and one pair of reconstruction procedures, not a claim about backbones in general. The reconstruction is a deterministic procedure applied after sampling. It requires no retraining and operates on atoms grouped into recognisable substructures. We expect its guarantees to transfer architecturally to other fragment-based generators, though we have not tested that transfer empirically. The broader implication is that physical realism in this setting is better obtained by constraint than by supervision. Bond diffusion, as introduced by DiffGui [33], attacks the same failure by learning bond types jointly with coordinates; the approach taken here reaches the same objective by removing that inference from the reconstruction step instead. At generation time, bond orders come only from the deterministic reconstruction procedure (Additional file 1, Note S3), which combines template bonds with single bonds between fragments and linker atoms. They do not come from a learned classifier. A pairwise bond-type head is trained as an auxiliary loss (Training objective and optimisation) to shape the network's geometric representations, but its predictions play no role in the final reconstruction, so no additional generative head is needed to *produce* the molecule. The two approaches to bond assignment are complementary rather than competing, and nothing prevents a model from using both.

GEMINI-Mol does not achieve the strongest docking scores, and we regard the trade-off as real rather than incidental. TargetDiff and Pocket2Mol reach lower mean Vina redocking scores, but the strongest-docking method, TargetDiff, has the lowest PoseBusters validity of the four diffusion and autoregressive generators (0.58 against 0.66 for AR, 0.89 for Pocket2Mol and 0.94 here), so a substantial share of those scores is computed on poses that fail elementary physical checks. Scored in the pose the generator places rather than after redocking, GEMINI-Mol's molecules reach only -0.07 kcal mol^{-1} against -5.1 to -6.5 for the baselines: placement is much improved by relaxing in the presence of the pocket, but it is not yet competitive, and we do

not present it as such. What GEMINI-Mol offers instead is candidates that survive the calculations that follow generation. For rescoring, molecular dynamics, free-energy estimation or medicinal chemistry triage, a physically valid and geometrically realistic molecule is substantially more actionable than a higher-scoring but strained or clashing one.

Applying multi-target design to α-amylase and intestinal maltase-glucoamylase tested the method where the target pair, the mechanism and the comparator are all fixed in advance by the clinic. Two findings survive that framing. Generated molecules reach similar docking scores to the clinical standard against maltase-glucoamylase under the stated protocol (the dual-target set has a mean of -7.00 kcal mol^{-1} against acarbose's -6.97, with 56% of individual molecules scoring below the drug), though acarbose's own docking search does not fully converge at the exhaustiveness matched to the generated set (replicate s.d. 0.43 against 0.03–0.07 elsewhere, Additional file 1, Table S15), so this is a similar-score comparison rather than a demonstrated equivalence. The generated set is also far more drug-like than the carbohydrate mimetics it would displace (QED 0.54 against 0.10), though that is not straightforwardly an advantage for this indication: acarbose's $\sim$1–2% oral bioavailability is a deliberate feature of α-glucosidase-inhibitor pharmacology, since the drug acts locally in the intestinal lumen, rather than a liability a more permeant, higher-QED analogue would improve on [12]. And a fraction of designs reproduce the *interaction pattern* of a crystallographic drug pose rather than merely scoring well: a median fingerprint Tanimoto of 0.62 to crystallographic acarbose, with 10–11% forming hydrogen bonds with two or more catalytic residues, which is parity with acarbose itself. The distinction is not decorative. Docking scores can be achieved by packing against a catalytic carboxylate without engaging it, and the best-scoring molecule in our dual set does exactly that, so interaction-level agreement is the more demanding claim and the one worth making.

That case study also exposed a pitfall that we suspect is not unique to this work. Interaction fingerprints computed on unrelaxed generated poses are not merely noisy but systematically inflated: catalytic engagement read 65% and 56% before relaxation and 43% and 39% after, because overlapping atoms satisfy distance criteria that a valid pose would not. A hydrogen bond of 0.7 Å is detectable by every standard tool and means nothing. Any interaction-level analysis of generated structures needs a pose-validity gate reported alongside it.

Scope and limitations.

Three constraints bound what these results establish.

The assembly rule forbids any new bond that would close a three- or four-membered ring. This is what removes the strained rings that distance-based perception fabricates, but it is a conservative restriction rather than a chemical principle: cyclopropane, cyclobutane, oxetane and azetidine motifs occur in approved drugs, and the present procedure cannot generate them outside the template library. Admitting them selectively, by adding small-ring fragment templates, is a straightforward extension.

Multi-target design is demonstrated asymmetrically. The dual-target ensemble matches acarbose against maltase-glucoamylase and engages its catalytic residues, but the α-amylase arm trails the drug on every measure reported. On simultaneous engagement the dual-mode and maltase-only sets are statistically indistinguishable at this sample size (4% against 9% clearing -8 kcal mol^{-1} in both enzymes, Fisher's exact $P = 0.16$, $n = 98$–100), so the evidence for dual conditioning rests on the mean docking scores and interaction patterns above rather than on this comparison. We therefore claim controllable multi-target *generation* with demonstrated engagement in one enzyme of the pair, not a validated selectivity control law: the tempering formulation predicts that the expected cross-target gap varies monotonically with $\log(\beta_1/\beta_2)$, and testing that prediction across the simplex remains outstanding.

As for all in silico structure-based design, the decisive test is experimental. No wet-lab validation is reported here, and docking scores, ADMET predictions and affinity-network surrogates remain approximations of the underlying biology.

Outlook.

Three directions follow directly from the results. On bond *lengths* GEMINI-Mol is second to TargetDiff, indicating that the templates and the relaxation reproduce angular and torsional geometry more faithfully than the precise bond-length distribution; a learned bond-length refinement is the natural remedy. Mean docking affinity trails the affinity-specialised baselines, and reinforcement-learning fine-tuning against a docking reward is the obvious next step. Because the reconstruction retains all generated atoms, molecular size is set by the fragment-plan distribution, which makes it tunable but means the residual QED gap to Pocket2Mol is not closed by that lever alone.

The results reported here rest largely on empirical scoring functions, and we subjected the leading dual-target candidate to physics-based evaluation as a first check: 500 ns of unbiased molecular dynamics in each enzyme, and MM-PBSA binding free-energy calculation with per-residue decomposition, to test whether the catalytic engagement seen in the static poses persists. It does not, cleanly. The receptor is stable, the mean MM-PBSA free energy is favourable in both enzymes, and the static interactions are real features of a real generated pose, but the ligand is in close contact with the protein for only 80–82% of the trajectory, direct hydrogen bonding to the catalytic site is intermittent, and structural snapshots show it relocating to peripheral surface sites rather than remaining in the pocket docking placed it in. The per-residue decomposition is consistent with this: no catalytic residue emerges as an energetic hotspot, and the free energy's uncertainty is the same size as its mean. Running the identical protocol on the crystallographic acarbose trajectories shows this is not an artefact particular to the generated molecule: acarbose's own free energy is stable and strongly favourable in maltase-glucoamylase, tracking every other measure of its performance

there, but in α-amylase it is non-stationary and, averaged over the full trajectory, weaker than Compound 1's own value despite far better raw contact and hydrogen bonding, and its per-residue decomposition is equally noise-dominated. Good contact metrics and a favourable computed free energy are therefore not interchangeable even for the reference drug, which argues that the instability documented here is a property of the pocket and the calculation's resolving power rather than a defect specific to this molecule. This is a second instance of the same problem raised for the pose-validity pitfall and the in-place docking gap: an empirically favourable score computed on a pose is not evidence that the pose is the one the molecule occupies at equilibrium. This discrepancy is now observed one level further down the validation chain, on physics-based rather than empirical grounds. It sharpens, rather than removes, the gap to experiment. More broadly, these results argue that progress in generative structure-based design will come not only from higher docking scores but from treating physical realism, multi-target engagement and multi-objective developability as explicit design criteria alongside affinity. A generator that reports strong affinity on structures that cannot exist has not advanced the problem; one that produces chemically real, developable candidates against the right set of targets has, even at a modest cost in predicted potency.

5 Conclusions

GEMINI-Mol demonstrates that explicit geometric and connectivity constraints can substantially improve the physical validity of pocket-conditioned molecular generation. By combining rigid-fragment diffusion, valence-constrained assembly and protein-aware relaxation, the framework achieved 94% intramolecular PoseBusters validity on CrossDocked2020 and more closely reproduced reference bond-angle and torsion distributions than the benchmarked alternatives. The controlled reconstruction ablation

identifies a concrete source of this improvement: with the trained backbone and sampling settings held fixed, valence-constrained assembly increased validity from 63% to 92% relative to distance-based bond perception. These findings establish molecular reconstruction as an important part of generative model design, with explicit chemical constraints complementing learned coordinate predictions.

The framework also combines conditioning information from multiple pockets within a single shared model, extending generation to design objectives that involve more than one target without retraining. The α-amylase and maltase-glucoamylase case study illustrates this capability and its current limits. Generated molecules matched acarbose's docking score against maltase-glucoamylase but not against α-amylase, and the experiments did not establish an improvement in simultaneous-hit yield over single-target generation followed by cross-screening. Moreover, 500-ns simulations of the selected candidate showed intermittent catalytic contacts and relocation from the initial binding sites. The case study therefore supports multi-target candidate generation, while sustained binding, selectivity and inhibitory activity remain to be established.

The principal contribution is a practical route to generating more physically plausible molecular structures while retaining flexible pocket conditioning. Further development should address the remaining docking-performance gap, broaden the assembly rules to accommodate small rings in linkers, and test target-weight control across a wider range of proteins. Prospective synthesis and biochemical assays will be needed to determine whether these computational gains translate into useful ligands. Together, the results support evaluating molecular generators through distinct, complementary criteria: structural validity, binding-site compatibility, pose persistence and experimentally measured activity.

List of abbreviations

ADMET: absorption, distribution, metabolism, excretion and toxicity; D3PM: discrete denoising diffusion probabilistic model; GAFF: generalised Amber force field; JS: Jensen–Shannon; KL: Kullback–Leibler; MD: molecular dynamics; MMFF94: Merck molecular force field 94; MM-PBSA: molecular mechanics Poisson–Boltzmann surface area; NPT: isothermal–isobaric ensemble; NVT: canonical ensemble; PDB: Protein Data Bank; QED: quantitative estimate of drug-likeness; RMSD: root-mean-square deviation; RMSF: root-mean-square fluctuation; SA: synthetic accessibility; SASA: solvent-accessible surface area; SBDD: structure-based drug design; SE(3): special Euclidean group in three dimensions; TIP3P: three-site transferable intermolecular potential.

Supplementary information

Additional file 1 (PDF). *Supplementary figures, tables and notes.* Per-type bond-length distributions; held-out performance of the 22-endpoint ADMET ensemble; per-residue flexibility and surface area for Compound 1 against acarbose in both enzymes; 2D structures and docking scores of the top 50 molecules in each of the three docking sets; full benchmark metrics; complete per-type bond-length, bond-angle and torsion-angle Jensen–Shannon distances; per-check physical validity; molecular-property and conformational-strain statistics; multi-target docking scores; generated-pose quality before and after protein-aware relaxation; catalytic-site engagement in each enzyme; per-residue MM-PBSA decompositions; raw per-replicate reference-drug docking scores; and three supplementary notes describing protein-aware relaxation, fragment-plan filtering and the valence-constrained assembly procedure.

Declarations

Availability of data and materials

All data, code and model weights supporting the conclusions of this article are openly available without registration, login or licence terms other than those stated below.

- **Project name:** GEMINI-Mol.
- **Project home page:** https://github.com/mikelangelo1/GEMINI-mol.
- **Archived version:** ¡Zenodo DOI of the archived release¿.
- **Operating system(s):** Linux (platform independent for analysis code).
- **Programming language:** Python ($\geq$ 3.8).
- **Other requirements:** PyTorch ($\geq$ 2.0) with CUDA, PyTorch Geometric, RDKit, Open Babel; for the reported evaluation and case study, AutoDock Vina 1.2, gnina, GROMACS 2025.2 and gmx_MMPBSA 1.6.4.
- **Licence:** MIT (source code); CC BY 4.0 (generated molecule sets, source data).
- **Any restrictions to use by non-academics:** none.

The repository contains the source code, the trained model weights, the fragment templates and generation-plan prior, the complete evaluation pipeline, the generated molecule sets for every reported experiment, the recomputed CrossDocked2020 test-set reference distributions, the per-molecule docking scores underlying every figure, and the source data for all main-text and additional-file figures and tables. The CrossDocked2020 dataset used for training and evaluation is publicly available from its original authors [20]. Crystal structures were obtained from the Protein Data Bank under accession codes 1HNY, 1B2Y and 3TOP (https://www.rcsb.org). The datasets supporting the conclusions of this article are also included within the article and Additional file 1.

Competing interests

The authors declare that they have no competing interests.

Funding

M.A.O. was supported by the Queen Elizabeth Commonwealth Scholarship (QECS), managed by the Association of Commonwealth Universities (ACU). The funder had no role in the conceptualisation or design of the study, in the collection, analysis or interpretation of data, in the decision to publish, or in the preparation of the manuscript.

Authors' contributions

M.A.O. conceived the method, designed and implemented the framework, performed all experiments and analyses, and drafted the manuscript. N.D.M.N. contributed to the design of the enzyme case study and to the interpretation of the structural analyses. A.T.C.L. supervised the study, secured resources and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors gratefully acknowledge the support of the Queen Elizabeth Commonwealth Scholarship (QECS) managed by the Association of Commonwealth Universities (ACU).

References

[1] Abraham MJ, Murtola T, Schulz R, et al (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001

[2] Anighoro A, Bajorath J, Rastelli G (2014) Polypharmacology: challenges and opportunities in drug discovery. Journal of Medicinal Chemistry 57(19):7874–7887. https://doi.org/10.1021/jm5006463

[3] Austin J, Johnson DD, Ho J, et al (2021) Structured denoising diffusion models in discrete state-spaces. In: Advances in Neural Information Processing Systems (NeurIPS), pp 17981–17993

[4] Baillif B, Cole J, McCabe P, et al (2024) Benchmarking structure-based three-dimensional molecular generative models using GenBench3D: ligand conformation quality matters. arXiv preprint arXiv:240704424

[5] Bickerton GR, Paolini GV, Besnard J, et al (2012) Quantifying the chemical beauty of drugs. Nature Chemistry 4(2):90–98. https://doi.org/10.1038/nchem.1243

[6] Bouysset C, Fiorucci S (2021) ProLIF: a library to encode molecular interactions as fingerprints. Journal of Cheminformatics 13(1):72. https://doi.org/10.1186/s13321-021-00548-6

[7] Brayer GD, Luo Y, Withers SG (1995) The structure of human pancreatic α-amylase at 1.8 å resolution and comparisons with related enzymes. Protein Science 4(9):1730–1742. https://doi.org/10.1002/pro.5560040908

[8] Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. The Journal of Chemical Physics 126(1):014101. https://doi.org/10.1063/1.2408420

[9] Buttenschön M, Morris GM, Deane CM (2024) PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. Chemical Science 15(9):3130–3139. https://doi.org/10.1039/D3SC04185A

[10] Carbonell J, Goldstein J (1998) The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp 335–336

[11] Chiasson JL, Josse RG, Gomis R, et al (2002) Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial. The Lancet 359(9323):2072–2077. https://doi.org/10.1016/S0140-6736(02)08905-5

[12] Clissold SP, Edwards C (1988) Acarbose: a preliminary review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential. Drugs 35(3):214–243. https://doi.org/10.2165/00003495-198835030-00003

[13] Dhariwal P, Nichol A (2021) Diffusion models beat GANs on image synthesis. In: Advances in Neural Information Processing Systems (NeurIPS), pp 8780–8794

[14] DiMasi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of Health Economics 47:20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012

[15] Du Y, Li S, Mordatch I (2020) Compositional visual generation with energy based models. In: Advances in Neural Information Processing Systems (NeurIPS), pp 6637–6647

[16] Eberhardt J, Santos-Martins D, Tillack AF, et al (2021) AutoDock Vina 1.2.0: New docking methods, expanded force field, and Python bindings. Journal of Chemical Information and Modeling 61(8):3891–3898. https://doi.org/10.1021/acs.jcim.1c00203

[17] Efron B (2011) Tweedie's formula and selection bias. Journal of the American Statistical Association 106(496):1602–1614. https://doi.org/10.1198/jasa.

2011.tm11181

[18] Erlanson DA, Fesik SW, Hubbard RE, et al (2016) Twenty years on: the impact of fragments on drug discovery. Nature Reviews Drug Discovery 15(9):605–619. https://doi.org/10.1038/nrd.2016.109

[19] Ertl P, Schuffenhauer A (2009) Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. Journal of Cheminformatics 1(1):8. https://doi.org/10.1186/1758-2946-1-8

[20] Francoeur PG, Masuda T, Sunseri J, et al (2020) Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. Journal of Chemical Information and Modeling 60(9):4200–4215. https://doi.org/10.1021/acs.jcim.0c00411

[21] Guan J, Qian WW, Peng X, et al (2023) 3D equivariant diffusion for target-aware molecule generation and affinity prediction. In: International Conference on Learning Representations (ICLR)

[22] Guan J, Zhou X, Yang Y, et al (2023) DecompDiff: Diffusion models with decomposed priors for structure-based drug design. In: International Conference on Machine Learning (ICML), pp 11827–11846

[23] Halgren TA (1996) Merck molecular force field. I. basis, form, scope, parameterization, and performance of MMFF94. Journal of Computational Chemistry 17(5-6):490–519. https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6⟨490::AID-JCC1⟩3.0.CO;2-P

[24] Hang T, Gu S, Li C, et al (2023) Efficient diffusion training via min-SNR weighting strategy. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp 7441–7451

[25] Harris C, Didi K, Jamasb AR, et al (2023) PoseCheck: Generative models for 3D structure-based drug design produce unrealistic poses. In: NeurIPS 2023 Workshop on Machine Learning in Structural Biology

[26] Hedrington MS, Davis SN (2019) Considerations when using alpha-glucosidase inhibitors in the treatment of type 2 diabetes. Expert Opinion on Pharmacotherapy 20(18):2229–2235. https://doi.org/10.1080/14656566.2019.1672660

[27] Heid E, Greenman KP, Chung Y, et al (2024) Chemprop: A machine learning package for chemical property prediction. Journal of Chemical Information and Modeling 64(1):9–17. https://doi.org/10.1021/acs.jcim.3c01250

[28] Hinton GE (2002) Training products of experts by minimizing contrastive divergence. Neural Computation 14(8):1771–1800. https://doi.org/10.1162/089976602760128018

[29] Ho J, Salimans T (2022) Classifier-free diffusion guidance. arXiv preprint arXiv:220712598

[30] Ho J, Jain A, Abbeel P (2020) Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems (NeurIPS), pp 6840–6851

[31] Hoogeboom E, Satorras VG, Vignac C, et al (2022) Equivariant diffusion for molecule generation in 3D. In: International Conference on Machine Learning (ICML), pp 8867–8887

[32] Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. Nature Chemical Biology 4(11):682–690. https://doi.org/10.1038/nchembio.118

[33] Hu Q, Sun C, He H, et al (2025) Target-aware 3D molecular generation based on guided equivariant diffusion. Nature Communications 16(1):7928. https://doi.

org/10.1038/s41467-025-63245-0

[34] Huang K, Fu T, Gao W, et al (2021) Therapeutics Data Commons: Machine learning datasets and tasks for drug discovery and development. Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks

[35] Huang L, Xu T, Yu Y, et al (2024) A dual diffusion model enables 3D molecule generation and lead optimization based on target pockets. Nature Communications 15(1):2657. https://doi.org/10.1038/s41467-024-46569-1

[36] Huang Z, Yang L, Zhou X, et al (2024) Protein-ligand interaction prior for binding-aware 3D molecule diffusion models. In: International Conference on Learning Representations (ICLR)

[37] Jiang Y, Zhang G, You J, et al (2024) PocketFlow is a data-and-knowledge-driven structure-based molecular generative model. Nature Machine Intelligence 6(3):326–337. https://doi.org/10.1038/s42256-024-00808-8

[38] Jin W, Barzilay R, Jaakkola T (2018) Junction tree variational autoencoder for molecular graph generation. In: International Conference on Machine Learning (ICML), pp 2323–2332

[39] Jorgensen WL, Chandrasekhar J, Madura JD, et al (1983) Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics 79(2):926–935. https://doi.org/10.1063/1.445869

[40] Jumper J, Evans R, Pritzel A, et al (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2

[41] Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. Acta Crystallographica Section A 32(5):922–923. https://doi.org/10.1107/S0567739476001873

[42] Kendall A, Gal Y (2017) What uncertainties do we need in Bayesian deep learning for computer vision? In: Advances in Neural Information Processing Systems (NeurIPS)

[43] van de Laar FA, Lucassen PL, Akkermans RP, et al (2005) α-glucosidase inhibitors for patients with type 2 diabetes: results from a Cochrane systematic review and meta-analysis. Diabetes Care 28(1):154–163. https://doi.org/10.2337/diacare.28.1.154

[44] Lakshminarayanan B, Pritzel A, Blundell C (2017) Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in Neural Information Processing Systems (NeurIPS) 30

[45] Landrum G, et al (2026) RDKit: Open-source cheminformatics. https://www.rdkit.org, accessed 2026

[46] Laskowski RA, Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. Journal of Chemical Information and Modeling 51(10):2778–2786. https://doi.org/10.1021/ci200227u

[47] Lin J (1991) Divergence measures based on the Shannon entropy. IEEE Transactions on Information Theory 37(1):145–151. https://doi.org/10.1109/18.61115

[48] Liu M, Luo Y, Uchino K, et al (2022) Generating 3D molecules for target protein binding. In: International Conference on Machine Learning (ICML), pp 13912–13924

[49] Liu N, Li S, Du Y, et al (2022) Compositional visual generation with composable diffusion models. In: European Conference on Computer Vision (ECCV), pp 423–439

[50] Loshchilov I, Hutter F (2019) Decoupled weight decay regularization. In: International Conference on Learning Representations (ICLR)

[51] Luo S, Guan J, Ma J, et al (2021) A 3D generative model for structure-based drug design. In: Advances in Neural Information Processing Systems (NeurIPS), pp 6229–6239

[52] McNutt AT, Francoeur P, Aggarwal R, et al (2021) GNINA 1.0: molecular docking with deep learning. Journal of Cheminformatics 13(1):43. https://doi.org/10.1186/s13321-021-00522-2

[53] Miller BR, McGee TD, Swails JM, et al (2012) MMPBSA.py: An efficient program for end-state free energy calculations. Journal of Chemical Theory and Computation 8(9):3314–3321. https://doi.org/10.1021/ct300418h

[54] Murray CW, Rees DC (2009) The rise of fragment-based drug discovery. Nature Chemistry 1(3):187–192. https://doi.org/10.1038/nchem.217

[55] Nahoum V, Roux G, Anton V, et al (2000) Crystal structures of human pancreatic α-amylase in complex with carbohydrate and proteinaceous inhibitors. Biochemical Journal 346(1):201–208. https://doi.org/10.1042/bj3460201

[56] O'Boyle NM, Banck M, James CA, et al (2011) Open Babel: An open chemical toolbox. Journal of Cheminformatics 3(1):33. https://doi.org/10.1186/1758-2946-3-33

[57] Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics 52(12):7182–7190. https://doi.org/10.1063/1.328693

[58] Paszke A, Gross S, Massa F, et al (2019) PyTorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems (NeurIPS)

[59] Peng X, Luo S, Guan J, et al (2022) Pocket2Mol: Efficient molecular sampling based on 3D protein pockets. In: International Conference on Machine Learning (ICML), pp 17644–17655

[60] Peng X, Guan J, Liu Q, et al (2023) MolDiff: Addressing the atom-bond inconsistency problem in 3D molecule diffusion generation. In: International Conference on Machine Learning (ICML), pp 27611–27629

[61] Proschak E, Stark H, Merk D (2019) Polypharmacology by design: A medicinal chemist's perspective on multitargeting compounds. Journal of Medicinal Chemistry 62(2):420–444. https://doi.org/10.1021/acs.jmedchem.8b00760

[62] Ragoza M, Masuda T, Koes DR (2022) Generating 3D molecules conditional on receptor binding sites with deep generative models. Chemical Science 13(9):2701–2713. https://doi.org/10.1039/D1SC05976A

[63] Ramsay RR, Popovic-Nikolic MR, Nikolic K, et al (2018) A perspective on multi-target drug discovery and design for complex diseases. Clinical and Translational Medicine 7(1):3. https://doi.org/10.1186/s40169-017-0181-2

[64] Ren L, Qin X, Cao X, et al (2011) Structural insight into substrate specificity of human intestinal maltase-glucoamylase. Protein & Cell 2(10):827–836. https://doi.org/10.1007/s13238-011-1105-3

[65] Rogers D, Hahn M (2010) Extended-connectivity fingerprints. Journal of Chemical Information and Modeling 50(5):742–754. https://doi.org/10.1021/ci100050t

[66] Satorras VG, Hoogeboom E, Welling M (2021) E(n) equivariant graph neural networks. In: International Conference on Machine Learning (ICML), pp 9323–9332

[67] Schneuing A, Harris C, Du Y, et al (2024) Structure-based drug design with equivariant diffusion models. Nature Computational Science 4(12):899–909. https://doi.org/10.1038/s43588-024-00737-x

[68] Sousa da Silva AW, Vranken WF (2012) ACPYPE – AnteChamber PYthon Parser interfacE. BMC Research Notes 5:367. https://doi.org/10.1186/1756-0500-5-367

[69] Sohl-Dickstein J, Weiss E, Maheswaranathan N, et al (2015) Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning (ICML), pp 2256–2265

[70] Song Y, Sohl-Dickstein J, Kingma DP, et al (2021) Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (ICLR)

[71] Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry 31(2):455–461. https://doi.org/10.1002/jcc.21334

[72] Valdés-Tresanco MS, Valdés-Tresanco ME, Valdés-Tresanco E, et al (2021) gmx_MMPBSA: A new tool to perform end-state free energy calculations with GROMACS. Journal of Chemical Theory and Computation 17(10):6281–6291.

https://doi.org/10.1021/acs.jctc.1c00645

[73] Vincent P (2011) A connection between score matching and denoising autoencoders. Neural Computation 23(7):1661–1674. https://doi.org/10.1162/NECO_a_00142

[74] Wang J, Wolf RM, Caldwell JW, et al (2004) Development and testing of a general amber force field. Journal of Computational Chemistry 25(9):1157–1174. https://doi.org/10.1002/jcc.20035

[75] Waring MJ, Arrowsmith J, Leach AR, et al (2015) An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nature Reviews Drug Discovery 14(7):475–486. https://doi.org/10.1038/nrd4609

[76] Warren GL, Andrews CW, Capelli AM, et al (2006) A critical assessment of docking programs and scoring functions. Journal of Medicinal Chemistry 49(20):5912–5931. https://doi.org/10.1021/jm050362n

[77] Wouters OJ, McKee M, Luyten J (2020) Estimated research and development investment needed to bring a new medicine to market, 2009–2018. JAMA 323(9):844–853. https://doi.org/10.1001/jama.2020.1166

[78] Yang K, Swanson K, Jin W, et al (2019) Analyzing learned molecular representations for property prediction. Journal of Chemical Information and Modeling 59(8):3370–3388. https://doi.org/10.1021/acs.jcim.9b00237

[79] Zhang O, Zhang J, Jin J, et al (2023) ResGen is a pocket-aware 3D molecular generation model based on parallel multiscale modelling. Nature Machine Intelligence 5(9):1020–1030. https://doi.org/10.1038/s42256-023-00712-7