

# Residue-Conditioned Molecular Design for Parologue Selectivity: A Computational Framework with Scaffold Preservation and Traceable Chemical Modifications

Ahmed Amine Sadouk

`ahmedamine.sadouk@octara.org`

---

Research Article

Keywords:

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11163933/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

---

# Residue-Conditioned Molecular Design for Parologue Selectivity: A Computational Framework with Scaffold Preservation and Traceable Chemical Modifications

---

Ahmed Amine SADOUK

ahmedamine.sadouk@octara.org

Independent Researcher

## **Abstract :**

Structure based generative models explore chemical spaces inside the protein active site , typically prioritizing candidates through empirical scoring functions. Few evaluate whether the result is selective for the intended protein over a close parologue , and none of them combines that evaluation with residue level conditioning and covalent bounds placement . In this paper we describe a design loop in which each round installs one chemical group aimed at one named pocket residue that a pairwise alignment marks as divergent in a chosen anti target subject while enforcing structural continuity by requiring each successive molecule to retain its predecessor as a substructure. The anti target pocket is built by mutating only the divergent position within the target's own structure , so the two differ at those residues and nowhere else .

Across 17 design campaigns the selected scaffold was retained in every final lead .

Comparison against 44 reference compounds from the literature that target 12 receptor pairs demonstrated preservation of established recognition chemistry in 19 comparisons , against a 12.6% unrelated active decoy rate . generated candidates the structure relationships with established inhibitors , including 78% shared core with imatinib and 59% with erlotinib while adding alternative chemical strategies . These included residue directed electrophile warheads such as a chloroacetamide targeting ABL1 Cys388, and the generation of a macrocyclic peptidomimetic architecture for KRAS G12C.

Together, these results demonstrate an integrated computational framework capable of translating therapeutic objectives and structural evidence into traceable medicinal chemistry design decisions across multiple protein targets.

## Introduction :

The rapid development of structure based generative model has expanded the possibilities of computational drug discovery by enabling the construction of molecular candidates directly with protein active binding pocket Pocket2Mol, TargetDiff, DecompDiff and DiffSBDD have demonstrated the potential of 3D molecular generation to explore chemical space while accounting for the geometrical constraints imposed by a protein target [1-4] . These methods have advanced molecular generation through improved structural representation , diffusion based sampling and scaffold aware design . However , molecular generation and medicinal chemistry reasoning remain a clouded area , candidates may satisfy the geometrical and physical requirement of a binding pocket without actually addressing the biological and chemical mechanisms that govern target recognition and selectivity [1] .

In medicinal chemistry molecular optimization is far complex than just improving predicted binding affinity . the selection and modification of chemical groups must account for established recognition features , residue specific interaction , structural difference between related protein and alternative binding mechanisms . These considerations become significantly more important when designing a selective inhibitor against closely related targets where a conserved binding pocket can limit the selectivity margin between intended protein and its parallels . Covalent inhibitor design introduces an additional dimension as electrophile functionalities must be selected and positioned with respect to suitable reactive residues while keeping the molecular interactions required for target recognition [5] .

Recent studies have begun incorporating medicinal chemistry knowledge more explicitly into molecular generation . Transpharmer showed that pharmacophore informed generation can preserve relevant molecular features while exploring structural distinct chemotypes with experimental validation of generated PLK1 inhibitors [6] . In parallel, agentic approaches such as LIDDIA have studied the use of large language models to coordinate molecular generation , optimization and computational evaluation across therapeutic targets [7] . These developments highlight a transition from isolated generative models toward integrated molecular design workflow . An important challenge remaining is the translation of target specific scientific evidence into explicit chemical modifications whose objectives and structural consequences can be seen throughout successive design rounds [6,7] .

Here we introduce our hybrid reasoning guided framework for sequential residue conditioned in pocket molecular design . This architecture integrates literature retrieval , pharmacophore - guided scaffold selection , protein structure analysis and target anti target alignment to formulate molecular design objective rather than treating molecular generation as an unconditioned search for favorable docking scores , it translate structural evidence into chemically defined modifications directed toward selected binding pocket residues . Each round proposes a specific interaction objective while specialized molecular

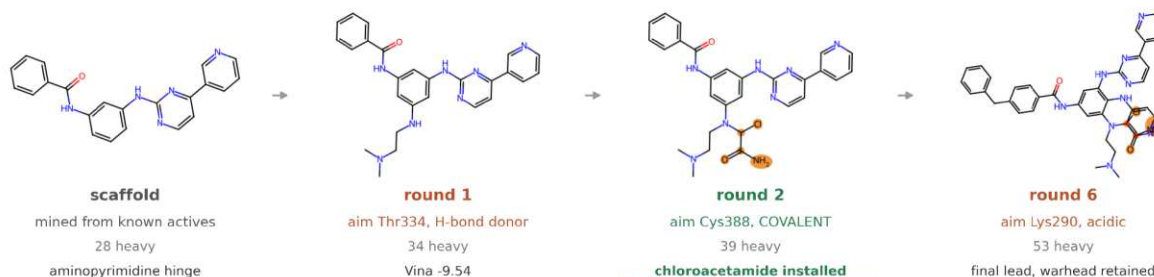
generation and evaluation components construct and assess the corresponding candidates . A structural containment operator persevere previously incorporated molecular features allowing successive modifications to accumulate without discarding the established scaffold

We evaluated this architecture across 18 receptor pairs covering kinases, a zinc metalloenzyme, a cyclooxygenase, a GTPase and several other enzyme families. 17 of them produced a final carried lead . We report what the system preserved, what it added, where the design decisions were directed and also where the computational evaluation methods could not establish the claim. Two methodological findings are reported as well, one concerning a receptor preparation defect that inflated every reference docking score in the same direction and one concerning the inability of non covalent docking to represent a covalent warhead. The whole design loop is summarised in Figure 1 .

#### a Resolution: a free-text objective becomes a measured pocket



#### b Sequential residue-conditioned growth, ABL1 over SRC, this study



the aminopyrimidine hinge motif of the scaffold is present in every later structure: containment refuses any candidate that does not contain the round before it

#### c What is learned, and what is not

|                                       |                                                                                                     |
|---------------------------------------|-----------------------------------------------------------------------------------------------------|
| ■ <b>Language model planner</b>       | states the residue, the interaction type and the reason, reading retrieved literature               |
| ■ <b>Message passing network</b>      | ADMET endpoints, Chemprop trained on Therapeutics Data Commons                                      |
| ■ <b>Random forest</b>                | activity prior on circular fingerprints, used only where cross-validated ROC AUC clears 0.65        |
| ■ <b>Containment operator</b>         | admits a candidate only if it contains the current lead as a subgraph, so no round can undo another |
| ■ <b>Pharmacophore scaffold miner</b> | the substituent most known actives share, kept where a Murcko frame would delete it                 |
| ■ <b>Same-frame counter-pocket</b>    | the anti-target built by mutating only the divergent positions of the target's own structure        |
| ■ <b>Geometry and force field</b>     | ETKDG embedding, parent coordinates transplanted, MMFF94 with every inherited atom pinned           |
| ■ <b>Empirical docking and typing</b> | AutoDock Vina at a fixed seed, ProLIF typed interactions, PoseBusters validity                      |
| ■ <b>Quantum chemistry (xTB)</b>      | GFN2-xTB reaction energies for the covalent rounds, reported apart from the docking margin          |

**Figure 1.** The design loop . Literature retrieval and pharmacophore mining select the starting scaffold , a pairwise alignment between the target and the chosen anti target identifies the divergent pocket positions, and each round selects one named residue and one interaction type. The molecular generation and evaluation components construct the candidates and the containment operator keeps the previous round inside the next one.

## Results :

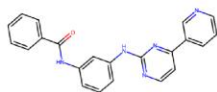
### The scaffold is retained across every campaign

The design loop begin from a scaffold mined from the target own known actives rather than from a Bemis-Murcko framework [8] . Murcko frames remove the exocyclic substituents, and for many series that remove the pharmacophore itself, a benzenesulfonamide is reduced to a benzene and the zinc binder is gone. On carbonic anhydrase II 36 of the 36 collected actives carry a primary sulfonamide , so the mined scaffold is benzenesulfonamide and not benzene.

The containment operator requires that the molecule of round N contain the molecule of round N-1 as a subgraph . This is checked before any scoring , so an exchange that replaces a group with a different group of the same size , which a heavy atom count can not detect, is refused . Across the 17 campaigns that produced a final lead the starting scaffold was present whole in the final molecule in **17 of 17** cases, verified as a subgraph match rather than inferred from the molecular size ( Wilson 95% interval 82% to 100% ). CA2 grew from 10 heavy atoms to 35 and the final lead still carried the primary sulfonamide that was installed at round 0 , EGFR grew from 17 to 59 heavy atoms and KRAS from 65 to 94. Figure 2 gives two of those trajectories round by round , and the two others are in the supplementary information .

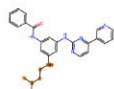
## ABL1 over SRC

6 rounds from a 28 heavy-atom scaffold to 53, scaffold intact in the final lead



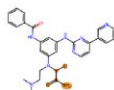
### round 0: scaffold, mined from the target's known actives

28 heavy atoms. This is what every later round must keep.



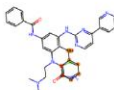
### round 1: add hbond\_donor aimed at Thr334

34 heavy atoms, 6 installed this round. contains the previous round  
engine logged this as: position is CONSERVED in the anti-target | counter-screen, reported  
not blocking: REJECTED - binds anti-target SRC MORE tightly than ABL1 (-0.48 kcal/mol)



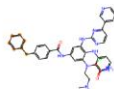
### round 2: add covalent aimed at Cys388

39 heavy atoms, 5 installed this round. contains the previous round  
engine logged this as: position DIFFERS in the anti-target | counter-screen, reported not  
blocking: REJECTED - binds anti-target SRC MORE tightly than ABL1 (-0.92 kcal/mol)



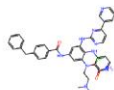
### round 3: add aromatic aimed at Tyr272

46 heavy atoms, 7 installed this round. contains the previous round  
engine logged this as: position DIFFERS in the anti-target | counter-screen, reported not  
blocking: REJECTED - binds anti-target SRC MORE tightly than ABL1 (-1.87 kcal/mol)



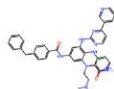
### round 4: add hydrophobic aimed at Ile333

53 heavy atoms, 7 installed this round. contains the previous round  
engine logged this as: position DIFFERS in the anti-target | counter-screen, reported not  
blocking: REJECTED - binds anti-target SRC MORE tightly than ABL1 (-0.98 kcal/mol)



### round 5: add hydrophobic aimed at Ile312

53 heavy atoms, 0 installed this round. contains the previous round  
engine logged this as: position DIFFERS in the anti-target | counter-screen, reported not  
blocking: no anti-target resolved a dockable pocket - selectivity undecidable



### round 6: add acidic aimed at Lys290

53 heavy atoms, 0 installed this round. contains the previous round  
engine logged this as: position is CONSERVED in the anti-target | counter-screen, reported  
not blocking: no anti-target resolved a dockable pocket - selectivity undecidable

## Carbonic anhydrase II over carbonic anhydrase I

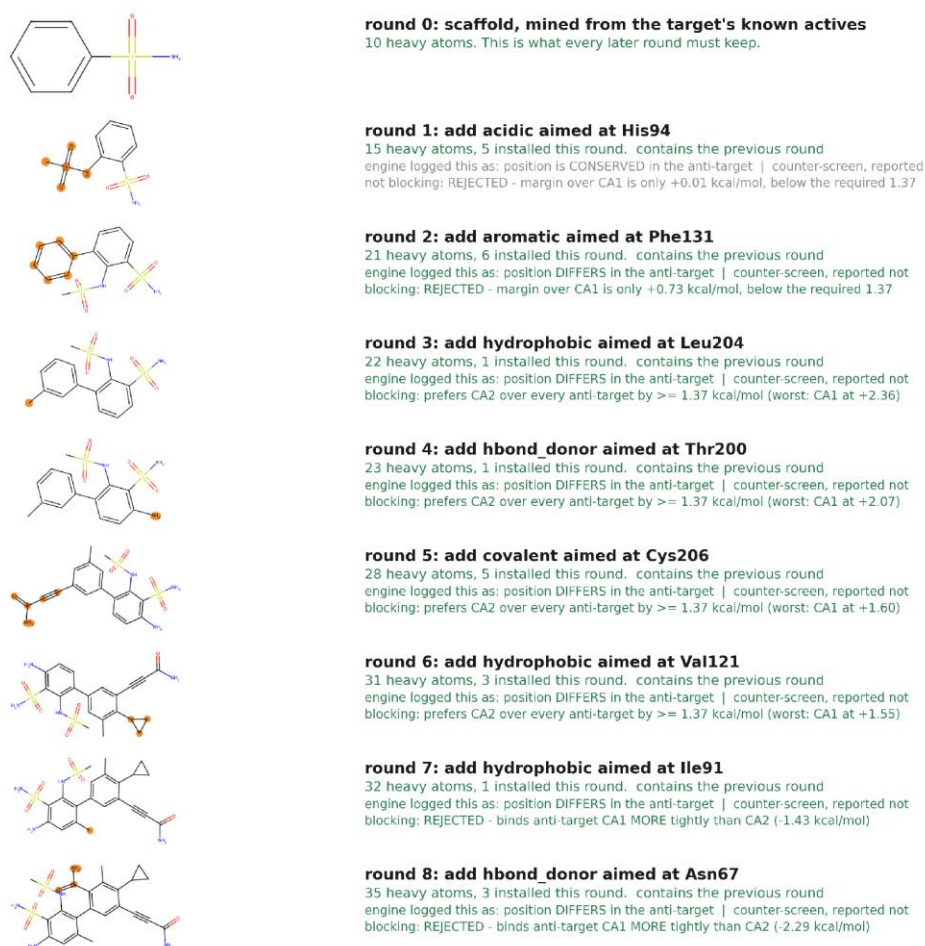


**Figure 2.** Sequential design, one row per round, for ABL1 over SRC ( a ) and carbonic anhydrase II over carbonic anhydrase I ( b ). The atoms installed in the current round are coloured in orange, everything in black is inherited from the previous round . Each row report the residue that was aimed at , the number of atoms installed , the engine's own classification of the position as divergent or conserved in the anti target, and the result of the per round counter screen. The EGFR and KRAS trajectories are given in the supplementary information ( Figures S1 and S2 ).

The trajectories also record the rounds where nothing was added. ABL1 rounds 5 and 6 and KRAS rounds 7 and 8 installed zero atoms , the front had only exchanges to offer and the containment operator refused them , so the previous lead was held . This is the scheme working rather than a failure , a round that can not add is a round that adds nothing.

## Design decisions are directed at named pocket residues

Each round receives the pocket residues annotated with three measured properties , whether the position differs in the anti target under a pairwise alignment, the distance from the nearest free attachment point on the current molecule , and whether the side chain is resolved in the structure . One residue and one feature type are selected and the engine then enumerates the chemistry that satisfies that objective.

## Sequential residue-conditioned design in the binding pocket

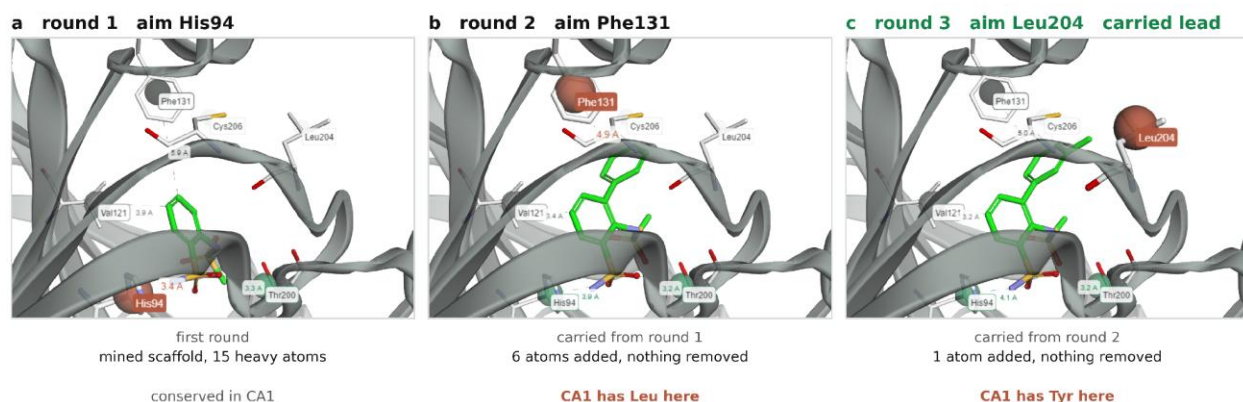


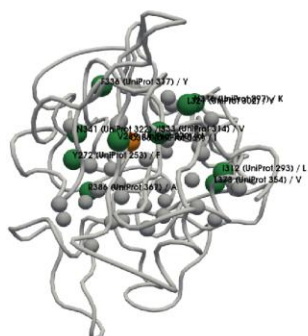
**Figure 3.** The application in pocket view for the three carried rounds of carbonic anhydrase II. The residue that each round is aimed at is filled and labelled in orange with its measured distance to the nearest ligand atom, all the panels share one camera so that only the chemistry changes between them. Round 1 aims at His94 which carbonic anhydrase I also carry, round 2 and 3 aim at Phe131 and Leu204 where the paralogue carries a leucine and a tyrosine. No docking score is shown, for the reasons given in the covalent section .

The distinction that Figure 3 makes is important and it is not favorable in every panel. His94 is the zinc coordinating histidine and it is conserved between the two isoforms, so engaging it buy affinity and can not buy selectivity. Phe131 and Leu204 are divergent and engaging them can. A figure that did not report the conserved case would leave the reader to assume that every round is a selectivity round .

## The counter pocket is constructed in the same structural frame

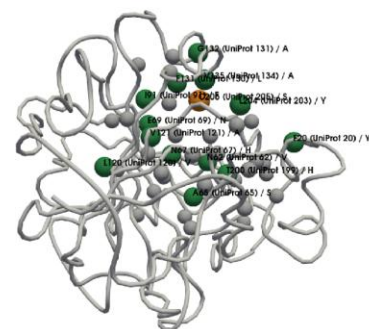
The anti target pocket is built by mutating only the divergent position inside the target's own structure, so the target and the counter differ at those residues and nowhere else. This removes the backbone and the preparation differences that dominate a comparison between two independently solved structures .

#### ABL1 against SRC



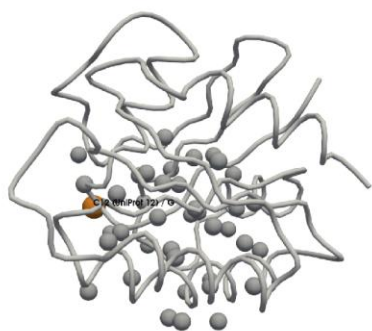
11 of 44 pocket positions differ. warhead at C369, which differs

#### CA2 against CA1



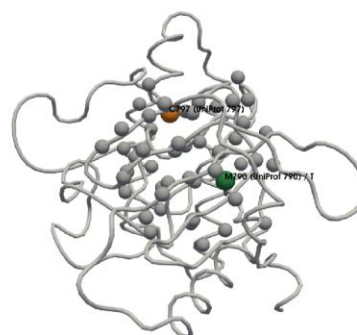
14 of 45 pocket positions differ. warhead at C205, which differs

#### KRAS against KRAS wild type



1 of 40 pocket positions differ. warhead at C12, which differs

#### EGFR against EGFR wild type



1 of 50 pocket positions differ. warhead at C797, which is CONSERVED

**Figure 4.** The pocket positions that a paralogue actually differ at, for the four campaigns that placed a covalent warhead. Green sphere, the anti target carries a different residue at this position, grey sphere, the position is conserved, orange sphere, the residue that the covalent round aimed at. Numbering is the deposited one with the canonical UniProt position in the bracket, recovered by alignment.

The number of usable positions is small and it is very different across the targets. ABL1 against SRC have 11 divergent position out of 44 pocket residues and CA2 against CA1 have 14 out of 45, but EGFR T790M against the wild type enzyme have **only one**, the gatekeeper position 790 itself, out of 50 pocket residues, and KRAS G12C against wild type KRAS also have only one, which is the G12C mutation. For those two mutant cases the entire selectivity argument rest on a single position and Figure 4 shows that plainly.

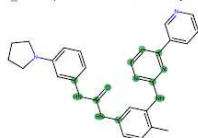
## **Recognition chemistry is preserved against reference compounds**

We compared each final lead against a curated set of marketed and reference compounds, 44 compounds spanning 12 of the receptor pairs, 41 of them approved, each structure resolved from PubChem by name so that no compound entered the comparison with a structure typed from memory. Five of the 17 campaigns have no marketed compound for their target and are not represented in this comparison.

Before anything is counted the comparison can simply be looked at. Figure 5 place the final lead of each campaign beside the best published compound of that target and draw the two of them as chemical structures, with the recognition chemistry coloured in green and the covalent warhead in orange on both molecules, so that what is shared and what is new can be seen directly instead of being read from a table. The percentage of every row is the maximum common substructure with that published compound, rings matched to rings. It is computed against the reference set of the campaign itself and not against the marketed compounds, so it is not the same number as the one of the next paragraph and the two should not be mixed.

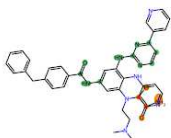
Read that way the campaigns do not all do the same thing. ABL1 share 63% of its core with the imatinib class reference and it keep the aminopyrimidine hinge motif and the anilide while adding a chloroacetamide that the published compound does not have. CA2 share 50% and keep the primary sulfonamide that was installed at round 0. BRAF reach 68% and 82% on the two structures of that target. EGFR share 31% and carry a chloroacetamide directed at the same cysteine that the published acrylamide is directed at. KRAS share only 16%, which is the lowest of the panel, and that is a property of its starting point rather than of the search since its scaffold was mined from peptidomimetic activities and it is not size comparable with a small molecule reference.

ABL1\_SRC published: 1-{4-Methyl-3-(4-pyridin-3

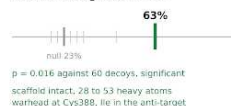


docking -12.41 vs -9.20 kcal/mol, context only

generated lead, 6 rounds shared core 63% of the published compound

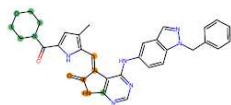


**shared core against chance**



kept aminopyrimidine, anilide / secondary amide. covalent chloroacetamide aimed at Cys388, which the published compound does not have

EGFR\_T790M published: Pyrrolo[2,3-d]pyrimidin-6-

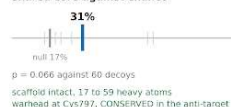


docking -12.46 vs -9.20 kcal/mol, context only

generated lead, 8 rounds shared core 31% of the published compound



**shared core against chance**



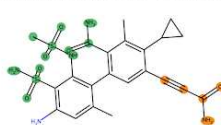
kept anilide / secondary amide. covalent chloroacetamide aimed at Cys797; the published compound uses acrylamide at the same residue

CA2\_CA1 published: 1,3,4-Thiadiazole-2-sulfon



docking -8.85 vs -6.95 kcal/mol, context only

generated lead, 8 rounds shared core 50% of the published compound

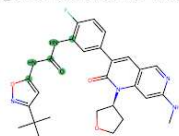


**shared core against chance**



kept primary sulfonamide, secondary sulfonamide. covalent propiolamide aimed at Cys206, which the published compound does not have

KRAS\_G12C published: 1-(3-tert-butyl-1,2-oxazol



docking -10.98 vs -8.40 kcal/mol, context only

generated lead, 8 rounds shared core 16% of the published compound

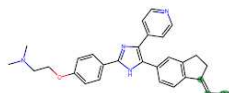


**shared core against chance**



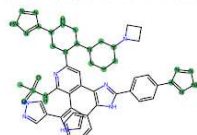
kept anilide / secondary amide. covalent acrylonitrile aimed at Cys12, which the published compound does not have

BRAF\_V600E published: SB590885



docking -10.56 kcal/mol

generated lead, 8 rounds shared core 68% of the published compound

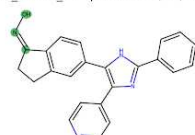


**shared core against chance**



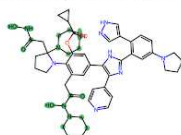
added methylsulfone, piperazine

BRAF\_V600E\_alt published: (NZ)-N-[5-(2-phenyl-5-pyri



docking -10.91 vs -9.11 kcal/mol, context only

generated lead, 8 rounds shared core 82% of the published compound

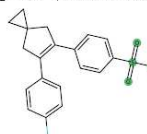


**shared core against chance**



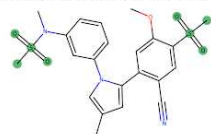
added hydroxamic acid, piperazine

PTGS2\_PTGS1 published: 5-(4-Fluorophenyl)-6-(4-me



docking -11.38 vs -3.39 kcal/mol, context only

generated lead, 8 rounds shared core 46% of the published compound

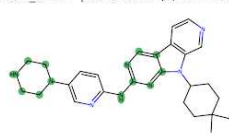


**shared core against chance**



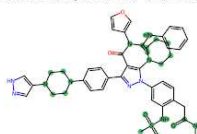
kept methylsulfone

CDK6\_CDK2 published: 8-(4,4-dimethylcyclohexyl)



docking -10.36 vs -7.80 kcal/mol, context only

generated lead, 8 rounds shared core 24% of the published compound

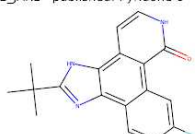


**shared core against chance**



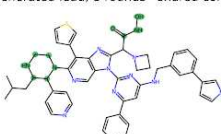
kept piperazine. added hydroxamic acid, methylsulfone

JAK2\_JAK1 published: Pyridone 6

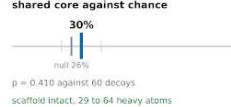


docking -11.37 vs -9.43 kcal/mol, context only

generated lead, 8 rounds shared core 30% of the published compound



**shared core against chance**



added hydroxamic acid, piperazine

**Figure 5.** The final lead of each campaign against the best published compound of that target, drawn as chemical structures. The recognition chemistry is coloured in green and the covalent warhead in orange on the two molecules so that they can be compared by eye. The strip at the right of every row give the decoy null of that exact comparison, the exact p value , whether the scaffold is still contained in the final lead and , where a warhead exist, the residue it was directed at and whether the anti target differs at that position.

Counting the recognition chemistry requires a decision about what counts . Preservation is measured by role and not by simple enumeration . A solubilising piperazine or a metabolic soft spot is classified as peripheral , while a hinge binding aminopyrimidine or a zinc binding sulfonamide is classified as recognition chemistry, so the absence of a solubilising group is not scored as the same event as the loss of a zinc binder . The generated leads retained every primary recognition motif of the reference compound in **19 of the 44 comparisons** , where an unrelated active from Another target does the same in **12.6%** of the cases ( binomial  $p < 0.0001$  ) . Figure 6 give every comparison of that set , with what was kept and what is missing listed for each of them .

## Every receptor against its named commercial compounds

final carried round of each campaign; structure decides, the docking score is context

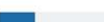
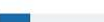
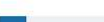
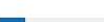
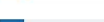
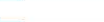
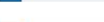
| receptor pair  | commercial compound | shared core                                                                                      | recognition kept         | recognition missing  | covalent warhead                                                | Vina, context |
|----------------|---------------------|--------------------------------------------------------------------------------------------------|--------------------------|----------------------|-----------------------------------------------------------------|---------------|
| ABL1_SRC       | imatinib            | approved  78%   | aminopyrimidine, anilide | none                 | chloroacetamide at Cys388                                       | -9.2 / -9.06  |
| EGFR_T790M     | erlotinib           | approved  59%   | aminoquinazoline, quinaz | none                 | chloroacetamide at Cys797                                       | -9.2 / -7.9   |
| EGFR_T790M     | gefitinib           | approved  55%   | aminoquinazoline, quinaz | none                 | chloroacetamide at Cys797                                       | -9.2 / -8.32  |
| EGFR_T790M     | dacomitinib         | approved  52%   | aminoquinazoline, anilid | none                 | chloroacetamide at Cys797<br>same residue as the drug's warhead | -9.2 / -9.48  |
| EGFR_T790M     | afatinib            | approved  50%   | aminoquinazoline, anilid | none                 | chloroacetamide at Cys797<br>same residue as the drug's warhead | -9.2 / -7.94  |
| PTGS2_PTGS1    | rofecoxib           |  50%            | methylsulfone            | none                 | -                                                               | -3.39 / n/a   |
| ABL1_SRC       | nilotinib           | approved  49%   | aminopyrimidine, anilide | none                 | chloroacetamide at Cys388                                       | -9.2 / n/a    |
| PTGS2_PTGS1    | etoricoxib          | approved  46%   | methylsulfone            | pyridine             | -                                                               | -3.39 / n/a   |
| PTGS2_PTGS1    | valdecoxib          |  45%            | none                     | primary sulfonamide  | -                                                               | -3.39 / n/a   |
| ABL1_SRC       | ponatinib           | approved  41%   | anilide / secondary amid | none                 | chloroacetamide at Cys388                                       | -9.2 / n/a    |
| CA2_CA1        | acetazolamide       | approved  38%   | primary sulfonamide      | anilide / secondary  | propiolamide at Cys206                                          | -6.95 / n/a   |
| ABL1_SRC       | bosutinib           | approved  36%   | pyridine                 | nitrile              | chloroacetamide at Cys388                                       | -9.2 / n/a    |
| CA2_CA1        | methazolamide       | approved  36%   | primary sulfonamide      | none                 | propiolamide at Cys206                                          | -6.95 / n/a   |
| JAK2_JAK1      | tofacitinib         | approved  35%   | none                     | nitrile              | -                                                               | -9.43 / -7.87 |
| PTGS2_PTGS1    | celecoxib           | approved  35%   | none                     | primary sulfonamide  | -                                                               | -3.39 / n/a   |
| ABL1_SRC       | dasatinib           | approved  30%   | anilide / secondary amid | none                 | chloroacetamide at Cys388                                       | -9.2 / n/a    |
| JAK2_JAK1      | ruxolitinib         | approved  30%  | none                     | nitrile              | -                                                               | -9.43 / n/a   |
| JAK2_JAK1      | baricitinib         | approved  27% | none                     | nitrile              | -                                                               | -9.43 / n/a   |
| CA2_CA1        | dorzolamide         | approved  26% | primary sulfonamide      | none                 | propiolamide at Cys206                                          | -6.95 / n/a   |
| PARP1_PARP2    | olaparib            | approved  25% | none                     | none                 | -                                                               | -10.29 / n/a  |
| PARP1_PARP2    | niraparib           | approved  25% | none                     | primary amide        | -                                                               | -10.29 / n/a  |
| AKT1_AKT2      | caviasertib         | approved  23% | none                     | anilide / secondary  | -                                                               | -9.77 / n/a   |
| PIK3CA_PIK3CD  | idelalisib          | approved  23% | none                     | imidazole, quinazoli | acrylamide, no residue                                          | -10.66 / n/a  |
| PIK3CA_PIK3CD  | duvelisib           | approved  23% | none                     | imidazole, pyridine  | acrylamide, no residue                                          | -10.66 / n/a  |
| CA2_CA1        | brinzolamide        | approved  22% | primary sulfonamide      | none                 | propiolamide at Cys206                                          | -6.95 / n/a   |
| CDK6_CDK2      | ribociclib          | approved  22% | none                     | aminopyrimidine, pyr | -                                                               | -7.8 / n/a    |
| EGFR_T790M     | osimertinib         | approved  22% | anilide / secondary amid | aminopyrimidine      | chloroacetamide at Cys797<br>same residue as the drug's warhead | -9.2 / -6.85  |
| BRAF_V600E     | vemurafenib         | approved  21% | pyridine, secondary sulf | none                 | -                                                               | n/a / n/a     |
| BRAF_V600E_alt | vemurafenib         | approved  21% | pyridine                 | secondary sulfonamid | -                                                               | -9.11 / n/a   |
| CDK6_CDK2      | palbociclib         | approved  21% | none                     | aminopyrimidine, pyr | -                                                               | -7.8 / n/a    |
| PARP1_PARP2    | talazoparib         | approved  21% | none                     | none                 | -                                                               | -10.29 / n/a  |
| BRAF_V600E     | dabrafenib          | approved  20% | secondary sulfonamide    | none                 | -                                                               | n/a / n/a     |
| BRAF_V600E_alt | dabrafenib          | approved  20% | none                     | secondary sulfonamid | -                                                               | -9.11 / n/a   |
| EGFR_T790M     | neratinib           | approved  20% | anilide / secondary amid | nitrile, pyridine    | chloroacetamide at Cys797<br>same residue as the drug's warhead | -9.2 / -5.81  |
| PIK3CA_PIK3CD  | alpelisib           | approved  20% | none                     | anilide / secondary  | acrylamide, no residue                                          | -10.66 / n/a  |
| AKT1_AKT2      | ipatasertib         |  19%          | none                     | none                 | -                                                               | -9.77 / n/a   |
| BRAF_V600E     | encorafenib         | approved  19% | methylsulfone, secondary | aminopyrimidine, ani | -                                                               | n/a / n/a     |
| BRAF_V600E_alt | encorafenib         | approved  19% | none                     | aminopyrimidine, ani | -                                                               | -9.11 / n/a   |
| CDK6_CDK2      | abemaciclib         | approved  19% | none                     | aminopyrimidine, pyr | -                                                               | -7.8 / n/a    |
| JAK2_JAK1      | fedratinib          | approved  19% | none                     | aminopyrimidine, sec | -                                                               | -9.43 / n/a   |
| KRAS_G12C      | sotorasib           | approved  17% | none                     | phenol, pyridine     | acrylonitrile at Cys12<br>same residue as the drug's warhead    | -8.4 / n/a    |
| PARP1_PARP2    | rucaparib           | approved  17% | anilide / secondary amid | none                 | -                                                               | -10.29 / n/a  |

**Figure 6.** Every receptor pairs against its named reference compounds. The shared core with the reference compound is drawn as a bar, the recognition chemistry that was kept and the recognition chemistry that is missing are listed by role, and the covalent warhead with the residue it was directed at is given where one exists. The docking values are placed last and in grey, as context only.

### Shared core has to be read against a scale

A shared core expressed as a percentage does not mean anything by itself, two arbitrary drug-like molecules share some substructure and a number like 78% only become a claim once it is known what an unrelated molecule score is. Two controls were used for that.

The first is a decoy null, built from 60 known activities of the other targets in the panel and scored in exactly the same way against the same reference compound. These are real, ring rich, synthesisable molecules that were simply never designed for this pocket, which make the null harder to beat than a random chemical library would be. 11 of the 44 comparisons beat this null at  $p < 0.05$ .

The second control is the one that we consider more informative. For every target we computed the shared core between each **pair of its own reference compounds**, 66 pairs in total. Two marketed compounds for the same target share a median of 33% of one another, afatinib and dacomitinib reach 79% because they are the same chemotype, acetazolamide and methazolamide reach 64%, while palbociclib and abemaciclib share only 9%. That spread is the natural unit, it is what two molecules that both work on the same target actually look like as a number. Figure 7 place the generated leads on that same axis.

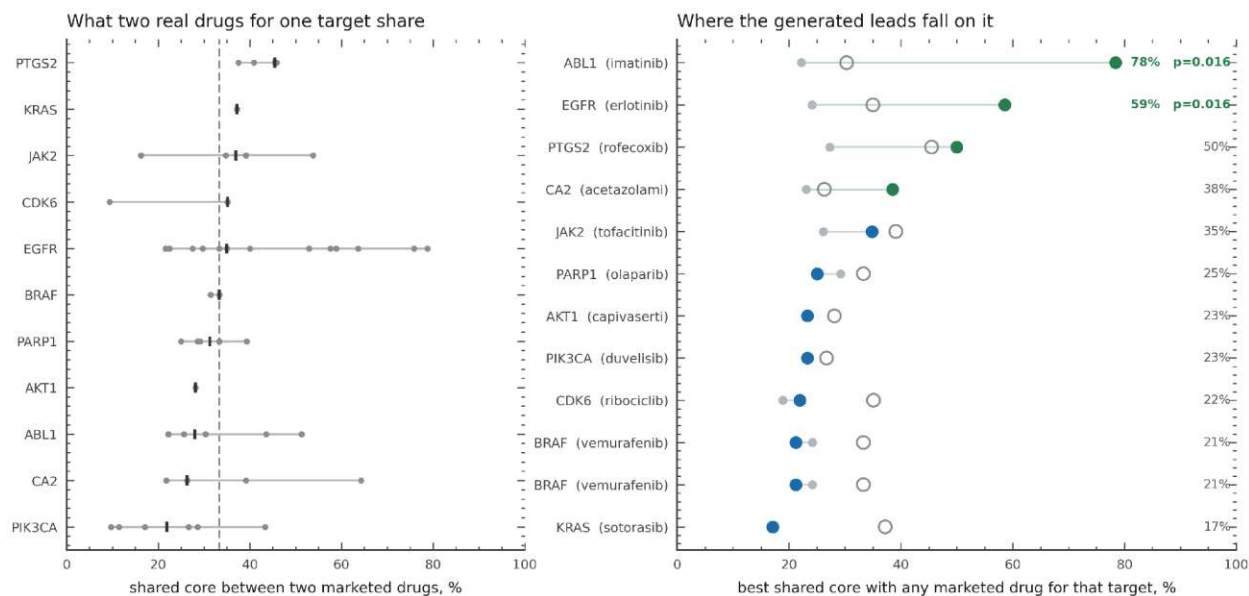


**Figure 7.** Left , how much the reference compounds of one target resemble each other , one dot per pair , the tick is the median for that target and the dashed line the median across all the 66 pairs. Right , the generated leads are placed on the same axis, against their matched decoy null and against the drug to drug median of their own target . The decoy null is computed with the identical best of set rule so that neither side is given an advantage the other does not have .

Read on this scale , 4 of the 12 targets produced a lead that recovered its nearest reference compound **more closely than two reference compounds of that same target recovered each other** . ABL1 reach 78% of imatinib against a drug to drug scale of 30%, EGFR reach 59% of erlotinib against a scale of 35% , PTGS2 reach 50% of rofecoxib and CA2 reach 38% of acetazolamide . ABL1 and EGFR also clear their matched decoy null at  $p = 0.016$ . The remaining 8 targets do not exceed their scale and they are drawn in the same plot at the same size rather than moved to a supplementary figure.

This scale also resolves a comparison that would otherwise be read as a failure . The BRAF lead shares 68% of its core with SB590885 , a published BRAF inhibitor, and only 21% with vemurafenib. Vemurafenib is an azaindole sulfonamide and SB590885 is an imidazole , they are two unrelated chemotypes for the same target, and requiring a single molecule to resemble both is not a chemically meaningful criterion because no marketed compound satisfy it either .

### **Covalent warheads are directed at named cysteines**

Five of the final leads carry a covalent warhead , which is the number that Figure 6 report at its foot . Four of them were installed by a round that explicitly set a covalent objective , ABL1 , CA2 , EGFR and KRAS , while the fifth, PIK3CA, acquired an acrylamide through ordinary growth and not through a directed covalent round , which we report as incidental rather than claiming it as a design result .

Of the four that were directed, three sit at a position that the alignment verify as divergent in the anti target. ABL1 Cys388 ( canonical Cys369 ) align to an isoleucine in SRC , CA2 Cys206 ( canonical Cys205 ) align to a serine in CA1 , and KRAS Cys12 exist only because of the G12C mutation itself so the wild type protein can not carry it by definition. The fourth, EGFR Cys797, is **conserved** between T790M and the wild type enzyme , and we state that explicitly rather than counting it under the divergence claim . Its selection is a reach and reactivity decision and not a sequence divergence one, and it remains interesting for a different reason , the reference compound for that target is an acrylamide and the generated lead is a chloroacetamide , which is two different electrophilic chemistries arriving independently at the same cysteine that the third generation EGFR inhibitors bond.

### **Non covalent docking can not evaluate a covalent round**

A docking score computed on a molecule whose intended binding mode is covalent measures the non covalent part of that molecule. AutoDock Vina contains no reactive term, it optimizes non covalent contacts and it has no mechanism that would bring an electrophilic carbon against a cysteine sulfur , so a

bond distance measured inside such a pose reports on the scoring function and not on the ligand . This is documented , the WIDOCK study showed that adding a reactivity scaled pseudo Lennard-Jones potential to AutoDock4 " induced a flip in the warhead structure , by placing the reactive beta-carbon in close proximity to the targeted cysteine sulfur ( 2.06 Å ) " , in direct contrast with the standard docking which predict an incorrect warhead geometry [9] . For this reason the covalent rounds are reported separately and no docking margin is claimed to measure them .

An endpoint reaction energy does not supply the missing term either . Computed at GFN2-xTB [10] with implicit water, with both residues modelled as the neutral species and with the same leaving group on both sides , the reaction energies favour **serine** for three of the four warheads, because a C-O bond is stronger than a C-S bond . Figure 8 give the four of them .

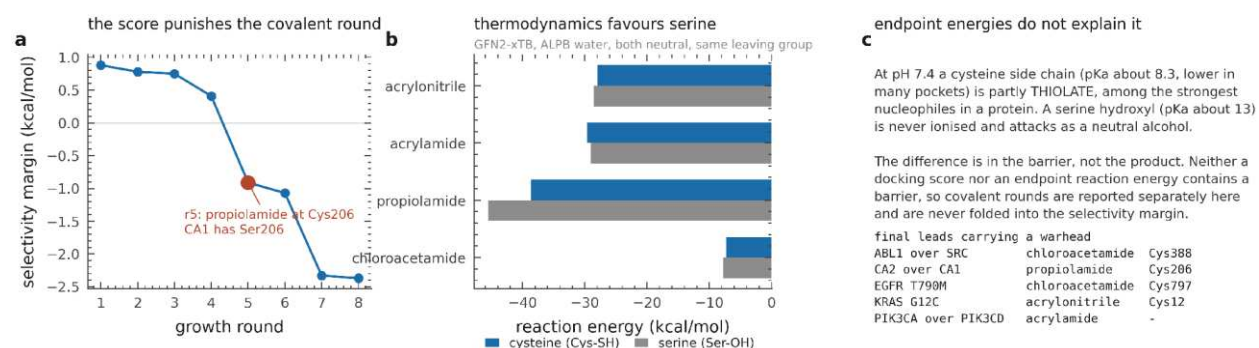


**Figure 8.** The covalent analysis. The reaction energies for the four warheads against a cysteine thiol and a serine hydroxyl , balanced like for like, and what a non covalent scoring function can and can not represent . The endpoint energies do not explain the cysteine selectivity of these warheads.

The usual explanation for cysteine selectivity is kinetic, at pH 7.4 a cysteine side chain with a pKa around 8.3 and often lower inside a pocket is partly present as a thiolate which is among the strongest nucleophile in a protein, while a serine hydroxyl with a pKa around 13 is never ionised . The difference would then be in the barrier and not in the product . That explanation is standard and it is consistent with what we measured, but **no barrier was computed in this work**, so it remains a mechanistic hypothesis and not a finding of this study. What this study establishes is the negative result , that the endpoint thermodynamics do not account for it .

## Comparison with published structure based generative models

We surveyed 32 published structure based generative methods [1-4,11-21] and classified them on the capabilities that this architecture claims . The comparison is on capability , which is checkable from each paper's own text, and not on a pooled docking column, because the target sets , the scoring protocols and the molecular sizes differ between those papers. Those methods are almost all reported on the

CrossDocked2020 split [28], whose reference ligand is the compound that happened to be co-crystallised with the pocket and not a compound that was selected for its potency

Three of the surveyed methods are 3D pocket conditioned and evaluate a paralogue, DiffSBDD [4] which perform an explicit negative design on one kinase pair, CByG [11] which add a selectivity guidance term and an on target versus off target benchmark, and CMD-GEN [12] which use selective pharmacophore points and reported a wet lab validated PARP1 over PARP2 selective compound. Two are 3D pocket conditioned and conditioned on a named residue, DeepICL [13] which condition the generation on an explicit interaction type per pocket atom, and DiffPharma [14] which use residue level interaction particles. A covalent generative model exists as well, CovaGEN [15], but it is conditioned on the sequence through a protein language model and not on the pocket geometry. **None of the 32 is 3D pocket conditioned and placed a covalent warhead**, and none of them combine the three.

| method           | year        | 3D pocket  | residue conditioned | paralogue evaluated | selectivity in search | covalent design | stated reason |
|------------------|-------------|------------|---------------------|---------------------|-----------------------|-----------------|---------------|
| liGAN            | 2022        | yes        | no                  | no                  | no                    | no              | no            |
| AR (3D-SBDD)     | 2021        | yes        | no                  | no                  | no                    | no              | no            |
| GraphBP          | 2022        | yes        | no                  | no                  | no                    | no              | no            |
| Pocket2Mol       | 2022        | yes        | no                  | no                  | no                    | no              | no            |
| FLAG             | 2023        | yes        | no                  | no                  | no                    | no              | no            |
| DrugGPS          | 2023        | yes        | no                  | no                  | no                    | no              | no            |
| SurfGen          | 2023        | yes        | no                  | no                  | no                    | no              | no            |
| TargetDiff       | 2023        | yes        | no                  | no                  | no                    | no              | no            |
| DecompDiff       | 2023        | yes        | no                  | no                  | no                    | no              | no            |
| ResGen           | 2023        | yes        | no                  | unknown             | no                    | no              | no            |
| D3FG             | 2023        | yes        | partial             | no                  | no                    | no              | no            |
| Lingo3DMol       | 2023        | yes        | partial             | no                  | no                    | no              | no            |
| DiffSBDD         | 2024        | yes        | no                  | yes                 | no                    | no              | no            |
| PMDM             | 2024        | yes        | no                  | no                  | no                    | no              | no            |
| PocketFlow       | 2024        | yes        | no                  | no                  | no                    | no              | no            |
| IPDiff           | 2024        | yes        | no                  | no                  | no                    | no              | no            |
| IRDiff           | 2024        | yes        | no                  | no                  | no                    | no              | no            |
| AliDiff          | 2024        | yes        | no                  | no                  | no                    | no              | no            |
| TacoGFN          | 2024        | yes        | no                  | no                  | no                    | no              | no            |
| FlexSBDD         | 2024        | yes        | no                  | no                  | no                    | no              | no            |
| DeepICL          | 2024        | yes        | yes                 | no                  | no                    | no              | partial       |
| PIDiff           | 2024        | yes        | no                  | partial             | no                    | no              | no            |
| BindGPT          | 2025        | yes        | no                  | no                  | no                    | no              | no            |
| MolPilot         | 2025        | yes        | no                  | no                  | no                    | no              | no            |
| DynamicFlow      | 2025        | yes        | no                  | no                  | no                    | no              | no            |
| CByG             | 2025        | yes        | no                  | yes                 | yes                   | no              | no            |
| CMD-GEN          | 2025        | yes        | partial             | yes                 | yes                   | no              | no            |
| DiffPharma       | 2026        | yes        | yes                 | no                  | no                    | no              | no            |
| MCTS kinase sel. | 2022        | no         | no                  | yes                 | yes                   | no              | no            |
| CogMol           | 2020        | no         | no                  | yes                 | yes                   | no              | no            |
| CovaGEN          | 2026        | no         | no                  | no                  | no                    | yes             | no            |
| Covalent MORL    | 2026        | no         | partial             | no                  | no                    | yes             | no            |
| <b>This work</b> | <b>2026</b> | <b>yes</b> | <b>yes</b>          | <b>yes</b>          | <b>yes</b>            | <b>yes</b>      | <b>yes</b>    |

**Table 1.** The 32 surveyed structure based generative methods, classified by whether they are 3D pocket conditioned, whether they condition on a named residue, whether they evaluate a paralogue, whether the selectivity term participate in the selection or is applied afterward as a filter, whether they place a covalent warhead, and whether the method emit a stated reason for each decision.

Table 1 gives the full survey . Where a docking comparison is made against these methods it is made on the instrument that they all report, and it is reported with the size confound visible rather than hidden. That comparison is Figure 9 .

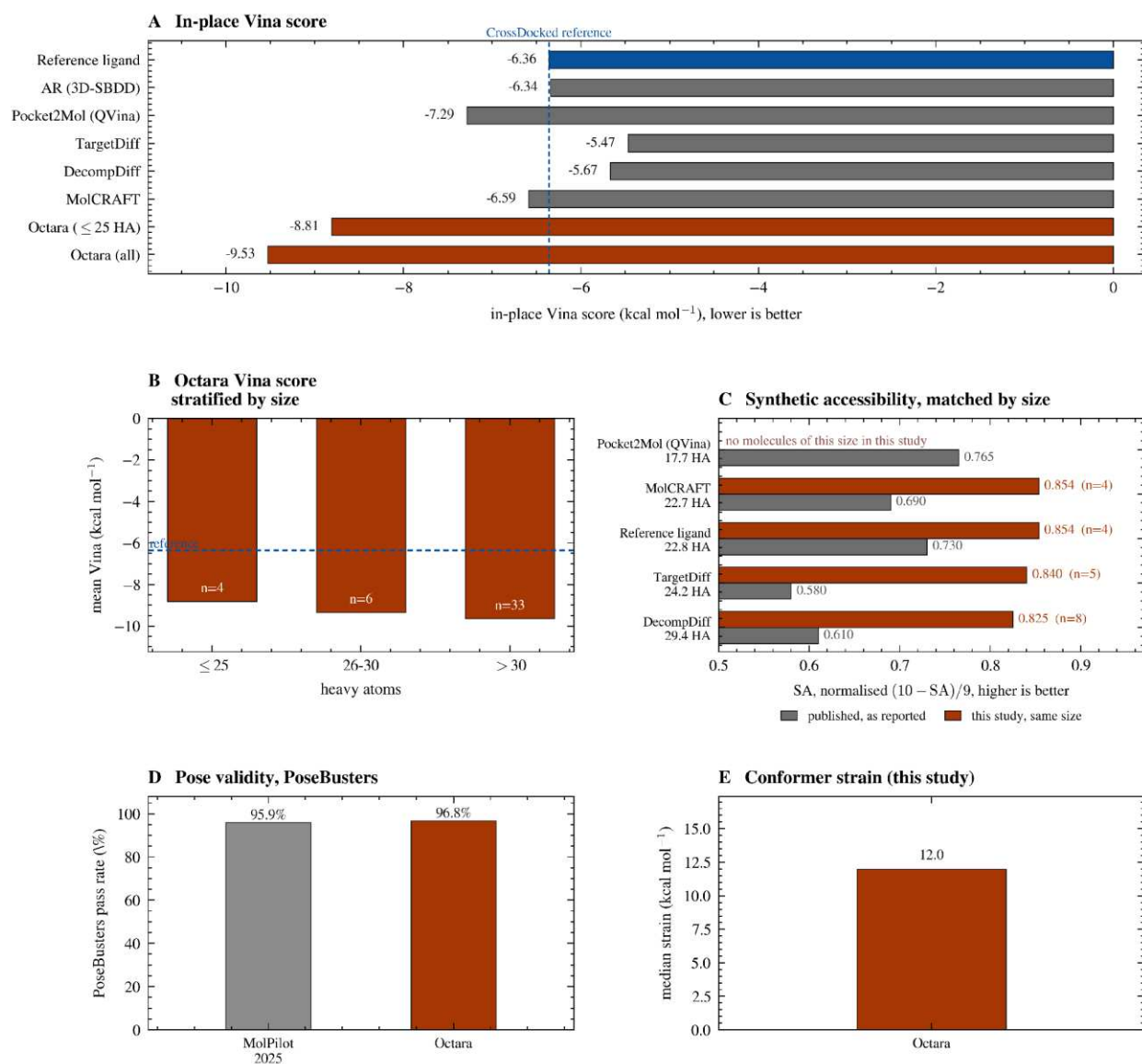


**Figure 9.** Comparison against the published generative models . ( a ) in place Vina score . ( b ) the same score for this work stratified by molecular size , since the score rise with the heavy atom count . ( c ) synthetic accessibility matched by size, which is the comparison that the size confound can not affect . ( d ) pose validity measured by PoseBusters against MolPilot which report the same metric. ( e ) conformer strain for this work.

The molecules generated here average 44.7 heavy atoms against 17.7 to 29.4 for the diffusion models, and a docking score rise with the heavy atom count [23,24] , so panel ( a ) is reported with panel ( b ) beside it and not alone . Panel ( c ) is the comparison that survive the size correction , matched at the same molecular size the generated molecules reach a normalised synthetic accessibility of 0.854 against 0.690

for MolCRAFT at 22.7 heavy atoms, 0.840 against 0.580 for TargetDiff at 24.2 and 0.825 against 0.610 for DecompDiff at 29.4. Panel ( d ) report a PoseBusters [22] pass rate of 96.8% against 95.9% for MolPilot [21], which is the current geometry state of the art, and we report that as a tie and not as a win.

### A receptor preparation defect that inflate every reference score

During this work we found a defect in the receptor preparation path that is not specific to this architecture and that we report because it affects any benchmark prepared the same way. Alternate conformations were stripped by blanking the slot characters, which live at column 17 of a PDB ATOM record, using a string replacement of the form `line.replace(line[16], " ", 1)`. A string replacement does not take a position, it replaces the first occurrence of that character anywhere in the record, and a PDB atom name very often contains its own altloc letter while sitting in columns 13 to 16, before the altloc. An atom named 'CB' with altloc 'B' therefore comes back named 'C', and the residue ends with two atoms called C roughly 2.5 Å apart, one of them a backbone carbonyl carbon and the other a side chain beta carbon. Figure 10 gives the mechanism on a real record and the number of records that each receptor of the panel loses to it.

#### The mechanism

`str.replace` takes a character, not a column. The atom name holds the altloc letter first.

#### deposited record

```
ATOM      38  N   ATYR A   7
```

^ ^ ^ ^ ^

atom name altloc, column 17

#### after the broken transform

```
TOM      38  N   ATYR A   7
```

the letter was removed from the ATOM NAME; the altloc is still there

#### after the corrected transform

```
ATOM      38  N   TYR A   7
```

the atom keeps its name; column 17 is blanked

Consequence: the residue carries two atoms with the same name about 2.5 Å apart, so the pocket gains contacts it does

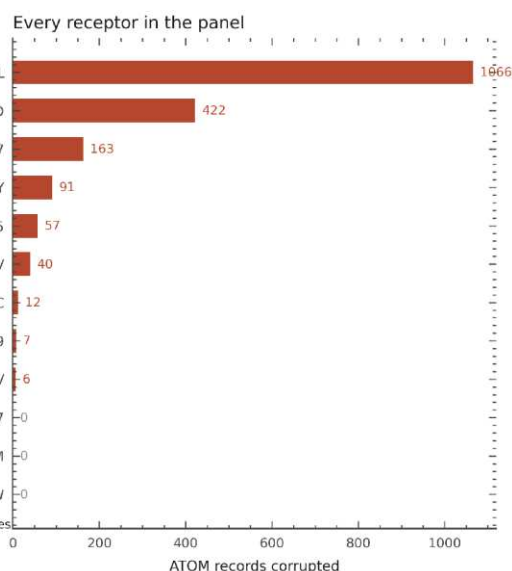


**Figure 10.** The receptor preparation defect. Left, the mechanism shown on a real record taken from 6SBL, with the character that the replacement hit marked. Right, the number of ATOM records that each receptor of the panel would lose to the defect.

In 6SBL this corrupted **1066 ATOM records** spread over 88 residues, and 9 of the 12 receptors of the panel are affected. Spurious atoms inside a pocket add contacts and improve the docking score of whatever is docked into it, and the compounds that are docked into these receptors to set the **reference bars** are the published actives, so the error push in one direction only, it make the bar that a new method

has to beat artificially high . On CA2 the stored reference value is -8.85 and the same compound redocked against a corrected receptor is -7.73 , which changes the sign of that comparison .

## **Methods :**

### **Scaffold selection**

Known activities for each target are retrieved and the Bemis-Murcko frameworks are counted. The scaffold is defined as the framework plus the substituent that the majority of the actives carry [8], which prevent the removal of the pharmacophore that a bare framework would produce . Candidate scaffolds are ranked by the absolute difference between their heavy atom count and the median heavy atom count of the reference compounds, so that the search starts from a framework of comparable scale to the compounds that it will be compared against. This is a comparability criterion and not a size objective, nothing in the growth reward a smaller molecule .

### **Residue conditioning**

For each round the pocket residues within 9 Å of the site centre are annotated with the divergence status against the anti target, the distance from the nearest free attachment point on the current molecule, and whether the side chain is resolved . One residue and one interaction type are selected per round . The engine then enumerates the chemistry that can satisfy that objective at that position , the selection component does not write molecules directly .

### **Containment**

A candidate is admitted only if the current lead is a subgraph of it , tested with an RDKit substructure match without chirality . The check runs before the scoring so that a candidate which exchanges a group for a different group of equal size is refused rather than detected afterward . When no candidate in the round contains the parent , the round stalls and the parent is held .

### **Anti target construction**

The divergent positions are taken from a pairwise alignment between the target and the anti target UniProt sequences, restricted to the positions whose numbering was verified against UniProt, since a mis-numbered mutation silently mutates the wrong residue . The mutations are applied to the target's own structure one at a time with a relaxation between each, which avoids the collision that an all at once build produces when several enlarged side chains are placed into the same tight pocket .

### **Residue numbering**

The deposited numbering and the canonical UniProt numbering do not agree for every entry and both are reported . 6SBL omit the initiator methionine so the deposited residue 206 is the canonical residue 205 ,

and 6NPV is numbered in the ABL1 isoform 1b convention which run 19 residues ahead of the canonical isoform 1a, so the deposited residue 388 is the canonical residue 369 . Every divergence statement in this work is recomputed from the structure and from the two alignments rather than read from a stored label, because a check that trusts the deposited numbering produces false negatives and a check that trust a stored label produces false positives.

### **Docking and scoring**

Docking is performed with AutoDock Vina 1.2.7 [25] into a box centred on the cognate ligand of the deposited structure. The receptors are prepared through the corrected altloc path described in the results . The docking values are used as one axis among several during the search, and they are not used to support the structural claims of this work , for the reasons given in the covalent and the preparation sections.

### **Statistics**

The decoy null for each comparison is built from 60 known actives of the other targets of the panel, drawn with a fixed seed , and scored against the same reference compound with the same maximum common substructure procedure, rings matched to rings and complete rings only. The synthetic accessibility is reported in the normalised convention of Ertl and Schuffenhauer [26], since the generative literature and the geometry benchmarks do not use the same one [27]. The reported p value is the exact fraction of that null which matches or beats the generated lead. For the best of set comparison the null is given the identical best of set rule so that the maximum is not an advantage handed to one side . The containment proportion is reported with a Wilson interval. No test in this work is computed on a docking score.

### **Populations**

Several distinct populations are used and they are not interchangeable. 18 receptor pairs were attempted and 17 produced a final carried lead . The reference comparison covers 12 of those 17, because five of the campaigns have no marketed compound for their target . The generative model comparison uses 43 generated molecules that passed the chemistry screen and are duplicated, which is not one final lead per campaign. The in pocket views of Figure 3 are three rounds of a single campaign, redocked for the figure .

### **Limitations :**

No compound reported in this work was synthesised and no compound was assayed. Every result here is a design result and the architecture is evaluated on what it preserved, what it directed and what it can be shown to have done, not on measured potency or measured selectivity. The selectivity margin used inside

the search is a difference between two docking scores computed on a chimeric structure and it is not a measurement of selectivity .

The number of campaigns is small . 17 campaigns and 44 reference comparisons support a method claim and they do not support a claim of general superiority. The reference comparisons within a single target are also not independent of each other, the six EGFR rows are six views of one lead and their p values should be read that way .

Three of the campaigns ran only two or three rounds , and containment holding across a two round campaign is a weaker test of the operator than containment holding across eight rounds .

The scaffold is mined from the target's own known activities, so landing near a published chemotype is partly a property of the starting point and not only of the search. This is the reason that the comparison in this work is made against the **marketed** compounds, which is a harder test than the campaign own reference set, and BRAF is the case where the two differ , 68% of a tool compound that is in the seed set against 21% of vemurafenib .

The KRAS scaffold was 937 Da before any growth because it was mined from peptidomimetic activities. That is appropriate for the target class but it means that the KRAS lead is not size comparable to the small molecule reference compounds, and its 17% shared core with sotorasib sit below both its decoy null and its drug to drug scale . We report that as a limitation of that campaign and not as an alternative chemotype strategy .

The counter screen that runs once per round reports a verdict on the lead that was already carried, it does not block it. The selectivity term enters the search earlier, through the divergent contacts axis of the Pareto front and through a margin based eligibility preference, and the per round counter screen is an independent audit of the carried lead rather than a gate .

MM-GBSA rescoring was computed for a subset only , and CA2 could not be rescored because the available force field lacked validated zinc parameters .

## **Conclusion :**

We described a design loop that translates a structural comparison between a target and a chosen paralogue into explicit chemical modifications directed at named pocket residues, and that keeps every modification through the following rounds by construction. Across 17 campaigns the scaffold was retained in every final lead, the recognition chemistry of the reference compounds was preserved in 19 of 44 comparisons against a 12.6% unrelated active rate , four targets produced a lead that recovered its

nearest reference compound more closely than two reference compounds of that target recover each other, and five leads carry a covalent warhead of which three sit at a verified divergent position.

We also report what the computational evaluation could not establish. A non covalent scoring function can not place a covalent warhead and can not evaluate a covalent round, the endpoint reaction energies do not explain the cysteine selectivity of the warheads, and a receptor preparation defect that we found and corrected had been inflating every reference docking score in the same direction across 9 of the 12 receptors of the panel .

The next step for this architecture is the synthesis of the EGFR lead and a kinact/Ki determination, which would test the one claim that the whole design loop is aimed at and that no computation in this work can settle .

### **Data availability :**

The campaign records of the 18 receptor pairs, the 44 reference compounds with their resolved structure and their InChIKey, every statistical output reported here and the code that compute They are available at <https://github.com/amsadouk/residue-conditioned-inpocket-benchmark> . Each number of this paper is produced by one command of that repository and the expected output of every command is listed in its README. The design engine is proprietary and it is not distributed, every analysis reported here is computed from the supplied records and it can be reproduced independently from them.

### **References**

1. Peng X, Luo S, Guan J, Xie Q, Peng J, Ma J. Pocket2Mol : Efficient Molecular Sampling Based on 3D Protein Pockets. ICML 2022. <https://proceedings.mlr.press/v162/peng22b>
2. Guan J, Qian WW, Peng X, Su Y, Peng J, Ma J. 3D Equivariant Diffusion for Target-Aware Molecule Generation and Affinity Prediction ( TargetDiff ). ICLR 2023. <https://arxiv.org/abs/2303.03543>
3. Guan J, Zhou X, Yang Y, et al. DecompDiff : Diffusion Models with Decomposed Priors for Structure-Based Drug Design. ICML 2023. <https://proceedings.mlr.press/v202/guan23a>
4. Schneuing A, Du Y, Harris C, et al. Structure-based drug design with equivariant diffusion models ( DiffSBDD ). <https://arxiv.org/abs/2210.13695>
5. Singh J, Petter RC, Baillie TA, Whitty A. The resurgence of covalent drugs. Nat Rev Drug Discov 2011. <https://pmc.ncbi.nlm.nih.gov/articles/PMC2136416>
6. Pharmacophore informed generative design with experimental validation of PLK1 inhibitors ( TransPharmer ). Nat Commun 2025. <https://www.nature.com/articles/s41467-025-56349-0>
7. LIDDIA , an agentic approach coordinating molecular generation, optimization and computational evaluation across therapeutic targets. <https://arxiv.org/abs/2502.13959>

8. Bemis GW, Murcko MA. The properties of known drugs. 1. Molecular frameworks. *J Med Chem* 1996 ; 39 : 2887-2893. <https://pubs.acs.org/doi/10.1021/jm9602928>
9. Scarpino A, Ferenczy GG, Keseru GM. WIDOCK : a reactive docking protocol for covalent binders. *J Comput Aided Mol Des* 2021 ; PMID 33458809. <https://pubmed.ncbi.nlm.nih.gov/33458809>
10. Bannwarth C, Ehlert S, Grimme S. GFN2-xTB , An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method. *J Chem Theory Comput* 2019 ; 15 : 1652-1671. <https://pubs.acs.org/doi/10.1021/acs.jctc.8b01176>
11. Choi S, et al. Bayesian Flow Network with gradient guidance and a selectivity benchmark ( CByG ). 2025. <https://arxiv.org/abs/2508.21468>
12. Zou Y, et al. Coarse-grained pharmacophore diffusion for selective inhibitor design ( CMD-GEN ), with a wet lab validated PARP1 over PARP2 selective compound. *Commun Biol* 2025. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11903766>
13. Zhung W, Kim H, Kim WY. 3D molecular generative framework for interaction-guided drug design ( DeepICL ). *Nat Commun* 2024 ; 15 : 2688. <https://www.nature.com/articles/s41467-024-47011-2>
14. Sako M, et al. Interaction constrained 3D molecular generation ( DiffPharma ). *npj Drug Discovery* 2026. <https://www.nature.com/articles/s44386-026-00040-x>
15. Zhang W, et al. Conditional latent diffusion for covalent inhibitor design ( CovaGEN ), validated on EGFR T790M and SARS-CoV-2 Mpro. *Commun Biol* 2026. <https://pmc.ncbi.nlm.nih.gov/articles/PMC13022440>
16. Ragoza M, Masuda T, Koes DR. Generating 3D molecules conditional on receptor binding sites with deep generative models ( liGAN ). *Chem Sci* 2022. <https://pubs.rsc.org/en/content/articlelanding/2022/sc/d1sc05976a>
17. Luo S, Guan J, Ma J, Peng J. A 3D Generative Model for Structure-Based Drug Design ( AR , 3D-SBDD ). *NeurIPS* 2021. <https://arxiv.org/abs/2203.10446>
18. Qu Y, Qiu K, Song Y, et al. MolCRAFT : Structure-Based Drug Design in Continuous Parameter Space. *ICML* 2024. <https://arxiv.org/abs/2404.12141>
19. Jiang Y, et al. PocketFlow , an autoregressive flow with chemical knowledge constraints and wet lab validated HAT1 and YTHDC1 hits. *Nat Mach Intell* 2024 ; 6 : 326-337. <https://www.nature.com/articles/s42256-024-00808-8>
20. Zhang Z, et al. FlexSBDD : Structure-Based Drug Design with Flexible Protein Modeling. *NeurIPS* 2024. <https://arxiv.org/abs/2409.19645>
21. Qiu K, et al. MolPilot , VLB-optimal scheduling for structure based drug design. *ICML* 2025. <https://icml.cc/virtual/2025/poster/44761>
22. Buttenschoen M, Morris GM, Deane CM. PoseBusters : AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem Sci* 2024 ; 15 : 3130-3139. <https://doi.org/10.1039/D3SC04185A>
23. Gao B, Ren M, Ni Y, et al. Rethinking Specificity in SBDD : Leveraging Delta Score and Energy-based Models. *ICML* 2024. <https://arxiv.org/abs/2403.12987>
24. Zheng K, Zhang Z, Fu T, Zitnik M, et al. Beyond Affinity : A Benchmark of 1D, 2D and 3D Methods Reveals Critical Trade-offs in Structure-Based Drug Design. *TMLR* 2026. <https://arxiv.org/abs/2601.14283>
25. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0 : New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model* 2021 ; 61 : 3891-3898. <https://pubs.acs.org/doi/10.1021/acs.jcim.1c00203>
26. Ertl P, Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J Cheminform* 2009 ; 1 : 8. <https://doi.org/10.1186/1758-2946-1-8>
27. Baillif B, Cole J, McCabe P, Bender A. Benchmarking structure-based three-dimensional molecular generative models using GenBench3D , ligand conformation quality matters. 2024. <https://arxiv.org/abs/2407.04424>
28. Francoeur PG, Masuda T, Sunseri J, et al. Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design ( CrossDocked2020 ). *J Chem Inf Model* 2020 ; 60 : 4200-4215. <https://pubs.acs.org/doi/10.1021/acs.jcim.0c00411>

## Supplementary Information

**Figure S1.** The EGFR T790M over wild type EGFR trajectory, 8 rounds, 17 to 59 heavy atoms .

`FIG\_trajectory\_EGFR\_T790M.pdf`

**Figure S2.** The KRAS G12C over wild type KRAS trajectory , 8 rounds, 65 to 94 heavy atoms .

`FIG\_trajectory\_KRAS\_G12C.pdf`

**Figure S3.** The final leads against the published compound of each target , with the decoy null , the exact p value, the containment status and the covalent residue for every row . `FIG\_structure\_panel.pdf`

**Table S1.** The 44 reference compounds with their case, approval status , InChIKey and PubChem CID .

`marketed\_drugs.csv`

**Table S2.** Per comparison shared core , decoy null median and exact p value . `STATS\_recovery.csv`

**Table S3.** The drug to drug scale per target and the best of set recovery of each lead .

`STATS\_chemotype\_scale.csv`

**Data S1.** The populations used by each figure , recomputed from the campaign records. `DATASETS.txt`

## Funding

This research received no external funding.

## Ethics declaration

Not applicable.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SUPPORTINGINFORMATION.docx](#)