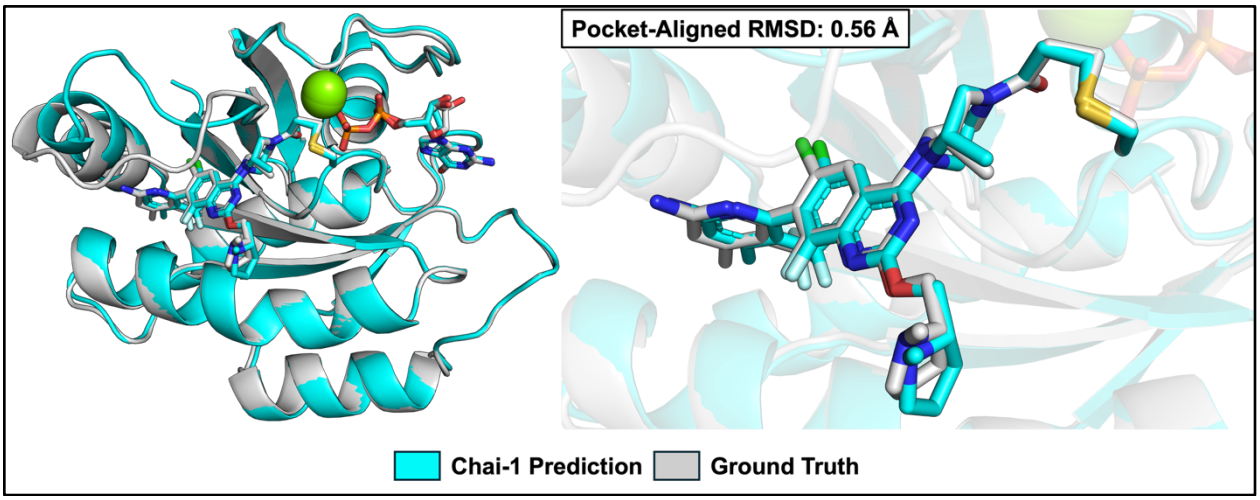
Single-sequence deep learning delivers crystal-quality models of covalent K-Ras G12 hotspot complexes

Sungwon Jung[a], Qinheng Zheng[b], and Kevan M. Shokat[a,c]

[a]Department of Cellular and Molecular Pharmacology and Howard Hughes Medical Institute, University of California, San Francisco, CA, USA; [b]Department of Biological Chemistry and Molecular Pharmacology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA; [c]Department of Chemistry, University of California, Berkeley, CA, USA



ABSTRACT

Structure-based design of covalent drugs has achieved tremendous success by understanding and leveraging the three-dimensional interactions between small-molecule drug candidates and their protein targets. However, this approach traditionally relies on high-resolution co-complex structures obtained by X-ray crystallography, NMR, or cryo-EM, methods that are time-consuming and resource-intensive. Here we show that Chai-1, a publicly available structure prediction tool that accepts user-defined ligands, is able to accurately predict covalent K-Ras(G12C) complexes without using a multiple sequence alignment (MSA). Chai-1 yields pocket-aligned RMSDs <2 Å for chemically diverse K-Ras(G12C) inhibitors, ranging from ARS-853 to BBO-8520. In addition to the conventional acrylamide-based covalent K-Ras(G12C) inhibitors, Chai-1 with a covalent-bond restraint successfully reproduced the binding poses of covalent K-Ras(G12D) and K-Ras(G12S) inhibitors, while showing limitations in capturing chemical details such as accounting for leaving-groups, bond properties, and stereochemistry. Chai-1 also provides ~40-fold higher throughput than state-of-the-art AlphaFold3 while maintaining comparable pose accuracy. Together, these findings establish Chai-1 as an accessible and computationally efficient tool for covalent protein-ligand co-complex structure prediction, with its covalent-restraint mode offering a unique solution to accelerate covalent drug discovery, especially for challenging targets beyond cysteine.

KEYWORDS

Deep Learning · Covalent Inhibitor · K-Ras(G12C) · K-Ras(G12D) · K-Ras(G12S) · Structure Prediction · Chai-1

INTRODUCTION

KRAS is the most frequently mutated oncogene that drives ~30% of human cancers[1]. The *KRAS* oncogene encodes a small GTPase K-Ras, which was historically considered to be an "undruggable" target for cancer therapy because of the lack of small molecule binding sites[2]. Acrylamide-based Compound 12[3] overturned that view by forming a covalent bond with Cys12 of one of the most common K-Ras missense mutations, K-Ras(G12C), revealing a previously unknown small molecule binding pocket. Now known as the switch-II pocket[3–5], the space between Switch-II and α3 helix has been exploited by medicinal chemists to deliver an increasing number of drug candidates, including two FDA-approved K-Ras(G12C) covalent inhibitors to treat non-small cell lung cancer, sotorasib and adagrasib.

Despite these breakthroughs, the drug development pipeline still relies on co-crystallography, a step that could lag chemical optimization by weeks or months[6]. This bottleneck is particularly pronounced for dynamic protein targets, presenting a significant hurdle for the rapid design and iteration of novel covalent inhibitors[7]. This rising demand for faster, more efficient structural insights has fueled the development of covalent docking methods. Deep-learning approaches are beginning to demonstrate success alongside traditional physics-based docking methods, often yielding more accurate structural predictions for complex covalent ligand-protein interactions[8–10].

Deep-learning-based biomolecular complex-structure prediction models like AlphaFold3 (AF3)[9] could in principle supplement crystallization or spectroscopy-based experimental structural determination. However, the public AlphaFold Server has clear restrictions: it can be used only for non-commercial purposes and prohibits the use of server outputs in docking or screening tools, while also accepting only a curated list of biologically common ligands rather than arbitrary, user-defined small molecules. These limitations significantly reduce its applicability in drug-discovery campaigns. In contrast, Chai-1[11] specifically fills this gap: its free web server accepts user-specified ligand SMILES along with any protein sequence and returns structure models in mmCIF (.cif) format, with the option to compute them in "single-sequence" mode without requiring multiple-sequence alignment (MSA) or templates. This makes it straightforward to move from sequence and SMILES to docking. In addition to being fully available for commercial use under an Apache 2.0 license, the open-source code supports covalent-bond restraints between the ligand and protein, enabling the use of the covalent bond as a constraint to aid in solving less common electrophile warheads and more challenging covalent amino acid targets such as aspartate and serine.

Our lab has been interested in developing covalent ligands targeting the somatically mutated oncogenic K-Ras variants, including G12C[3], G12S[12], G12R[13], and the most frequent G12D[14,15] mutation. Our work benefited greatly from an early co-crystal structure of a tethering compound we discovered covalently bound to K-Ras(G12C)[3], revealing a pocket not visible in previous PDB structures due to the dynamic nature of the Switch II loop. As our work evolved beyond G12C, we were often challenged by a lack of ability to obtain structures for ligands that reacted

too slowly or incompletely, making optimization difficult. Once more optimized compounds were obtained, we were able to solve helpful co-crystal structures. The recent explosion in machine learning based structure prediction attracted our attention, and we became curious if these methods could accurately predict structures we had recently solved (including one that had not yet been deposited when we initiated this work; PDB: 9DMM[16]). We thought this exercise would allow us to benchmark current ML methods, and since we are not a computational lab, we hoped our experience might be helpful to similar groups. Because the target pocket remained constant while only the nucleophilic residue varied (Cys, Asp, or Ser), we were able to attribute differences in prediction accuracy to the covalent features.

Here, we show that Chai-1 accurately predicts K-Ras(G12C) co-crystal structures, extends to difficult targets such as K-Ras(G12D) and K-Ras(G12S) using covalent restraints, and achieves superior computational efficiency compared to AlphaFold3 without substantial loss of accuracy. Notably, three of the hotspot mutations at codon 12 (G12C, G12D, and G12S) now have published covalent ligands, providing a controlled test in which the switch-II pocket geometry is well characterized and the primary variable is the electrophile chemistry. By benchmarking across these variants, we asked whether Chai-1 could faithfully reproduce the experimentally determined binding poses for diverse warheads, and in what contexts the predictions deviated. Such a framework will be particularly valuable as the field advances beyond cysteine toward more challenging covalent amino acid targets, where experimental structures remain scarce and computational predictions will likely provide the primary guidance for covalent ligand design.

RESULTS

Crystal-Quality Web Predictions for K-Ras(G12C) Switch-II Pocket Inhibitors

We first asked how Chai-1 would perform on our recently determined 1.90-Å K-Ras(G12C)–Divarasib (Figure 1) structure[16], which had not been available in the PDB at the time of prediction, serving as a true blind test. The Chai-1 web mode delivered an overlay with a pocket-aligned RMSD of 0.56 Å (Figure 2a), surpassing the 2.0 Å-threshold commonly used in deep learning–based docking benchmarks[9]. Consistent with its well-known conformational plasticity[17], the switch-II loop is the only region where the Chai-1 prediction substantially deviates from the crystal structure. Notably, this segment receives comparatively lower predicted local distance difference test (pLDDT) scores, although it remains within the high-confidence range (Figure 2b). Furthermore, while examining the outputs we noticed that in several restraint-free predictions, PyMOL displayed a covalent link between Cys12 and the acrylamide warhead even though no covalent bond restraint had been applied. Follow-up checks confirmed that this bond was not generated by Chai-1 itself, but was instead drawn by PyMOL's default heuristic, which infers connectivity from inter-atomic distances and element types when explicit CONECT records are missing in the input file. We nonetheless note that the appearance of such "spontaneous" bonds can be taken as a fortuitous hint of a strong intrinsic preference for covalent engagement and could serve as a qualitative cue when prioritizing covalent drug candidates.

In addition to divarasib, we wondered whether this method could be applied to other covalent K-Ras(G12C) complexes, especially those that were not used as the training set for Chai-1 (training cutoff: 12 Jan 2021). We assembled six additional G12C ligands with diverse chemotypes: ARS-853[18], ARS-1620[19], sotorasib, adagrasib, JDQ-443[20], and BBO-8520[21]

(Figure 1). Chai-1 reproduced all poses within 2.0 Å Pocket-Aligned RMSD, even including JDQ-443, which does not employ the traditional quinazoline scaffold. Remarkably, the model also captured the GTP-bound conformation of BBO-8520, a first-in-class dual-nucleotide-state inhibitor, with 0.99 Å RMSD, comparable to its GDP counterpart (0.95 Å) (Figure 2c). These results demonstrate that single-sequence language-model embeddings are sufficient to accurately position K-Ras(G12C) covalent inhibitors across diverse ligands and nucleotide states, without the need for multiple sequence alignments.

Covalent Restraint Enables Accurate Modeling of Unconventional K-Ras(G12D) and K-Ras(G12S) Covalent Chemistry

Recent success against K-Ras(G12C) has relied on Michael-acceptor acrylamides that react with the thiolate of Cys12 and dominate existing training sets for structure-prediction and docking algorithms. Pivoting to the more prevalent K-Ras(G12D) mutant[22] demands electrophiles that can engage an aspartate carboxylate, a markedly less nucleophilic target, which has motivated the discovery of uncommon warheads such as strain-released malolactones[14], α-diazoacetamide[15], tunable oxirane[23], and aziridine[24]. Because only a handful of aspartate-guided covalent complexes have been solved to date, deep-learning models trained primarily on cysteine-reactive ligands lack coverage of these rarer electrophiles. Benchmark studies show that covalent-docking accuracy falls steeply when the warhead class is under-represented in training data[25,26], underscoring the possible need to incorporate explicit bond restraints or retraining when tackling Asp-targeting inhibitors. Against this backdrop, we assessed whether Chai-1 could generalize beyond the prevalent cysteine-centric data.

Extending the protocol to the K-Ras(G12D) mutant[22] failed: in all web-based predictions the β-lactone-based covalent ligand (G12Di-1[14]) was completely displaced from the switch-II pocket (Figure 3a). We suspected that the high precision for G12C predictions might stem from the fact that acrylamide-based, cysteine-directed covalent complexes are substantially overrepresented across the PDB relative to aspartate-directed chemistries, likely biasing the model toward Cys-covalent recognition and geometries thus explaining the discrepancy. However, we did not confine our attention solely to the training data. We considered another factor: the interaction between the nucleophile (amino acid) and the electrophile (warhead). We hypothesized that this displacement could occur because the model typically predicts an energy-minimized local minimum, coupled with the fact that no covalent bond restraint was imposed. As a result, significant steric hindrance between the electrophile and nucleophile prior to covalent bond formation may have led to this displacement. In our G12C case, the relatively small steric clash between acrylamide and cysteine could allow the scaffold's directing effect to override this, resulting in an accurate prediction, even without a covalent bond restraint. Consistent with this view, a previously reported β-lactone K-Ras complex was crystallized with wild-type K-Ras rather than K-Ras(G12D)[27], suggesting that the Asp-directed early noncovalent pose is less stable without an explicit covalent restraint.

To overcome this, we ran the model locally with a covalent bond restraint between G12D and the ligand, which restored the ligand to its experimental pose with a pocket-aligned RMSD of 0.68 Å (Figure 3a, b). However, we obtained a poor consistency across multiple predictions of the switch-II loop region and further, the output is not completely correct in terms of the covalent linker chemistry; the newly formed ester bond has a strained C-O-C angle (measured angle:

86.5°) and an odd planar-like sp^3-hybridized carbon atom (Figure 3c). For G12C ligands, applying a covalent restraint maintained <2 Å accuracy (Figure 2c).

Encouraged by the success with K-Ras(G12D), we next tested whether the covalent restraint could similarly enable the accurate prediction of a ligand pose targeting the challenging nucleophile serine in the K-Ras(G12S) mutant. Serine is also a far less commonly targeted electrophile in covalent drug discovery. For this test, we used G12Si-5, a β-lactone–based covalent inhibitor[12]. Without the restraint, the model failed to form the desired covalent bond between the serine hydroxyl and the warhead in all five web-based predictions. However, unlike the complete displacement observed with the G12D inhibitor, the ligand scaffold itself was consistently well-situated within the Switch-II pocket. This suggests the model could predict the binding site accurately but failed to resolve the specific covalent chemistry. Upon applying the covalent restraint, the model successfully recapitulated the experimental binding pose, correctly forming the serine-carbon covalent bond (Figure 4a). Once again, despite this overall accuracy, the predictions revealed a substantial conformational difference within the ligand's piperidine moiety. Specifically, a near 180° rotation was observed about the C-N bond connecting the piperidine nitrogen to the pyridopyrimidine substituent. Also, the stereochemistry of the piperidine ring was inverted relative to the crystal structure (cis in the experimental pose versus trans in the prediction, Figure 4b). This reflects a known limitation of diffusion-based co-folding models in preserving stereochemical fidelity, highlighting an area for future improvement[28].

Chai-1 Matches AlphaFold3 Pose Accuracy while Delivering ~40-Fold Higher Throughput in Single-Sequence Mode

To contextualize Chai-1's computational advantages, we additionally benchmarked AlphaFold3 on the same K-Ras G12 mutant and covalent ligand complexes. We tested AlphaFold3 using MSA and template inputs, widely regarded as state-of-the-art for protein structure prediction due to the rich coevolutionary signal they provide[9], whereas Chai-1 was evaluated in single-sequence mode (i.e., without MSAs). According to the Chai-1 preprint, the model can run in single-sequence mode while preserving most of its performance and markedly reducing runtime[11]. Given this efficiency, Chai-1's single-sequence mode may be particularly advantageous for large-scale binding pose prediction campaigns involving extensive ligand-protein libraries.

Regarding the ligand–protein pairs predicted earlier by Chai-1, AlphaFold3 achieved a 0.15 ± 0.11 Å lower pocket-aligned RMSD than Chai-1, corresponding to an average improvement of 20 ± 8%. The median paired difference was –0.057 Å. Nevertheless, the difference was not statistically significant (Wilcoxon matched-pairs signed-rank, W = –31, P = 0.074). Importantly, when a 2 Å accuracy cutoff is applied, both models correctly predict all nine ligand poses, yielding 100 % success for this dataset (Figure 5a).

For computational efficiency, AlphaFold3 required dramatically more time per prediction than Chai-1 in single-sequence mode. Across nine matched K-Ras ligand complexes run on identical A40 (32 GB) GPUs, AF3 consumed ≈ 52 min of CPU wall-time per complex (3,098 ± 48 s), whereas Chai-1 finished in ≈ 1.4 min (84 ± 10 s). Peak RAM usage was modest for both methods (≤ 15 GB), yet AF3's longer runtime generated a larger memory-seconds integral (1.10×10^3 GB·s vs. 0.87×10^3 GB·s). Extrapolated to a 24-h GPU day, Chai-1 can therefore generate ~1,000 ligand–protein models, while AF3 is limited to fewer than 30 models (Figure

5b), which is a ~37-fold reduction in throughput that could pose a substantial challenge for large-scale virtual screening. We acknowledge that most of AF3's wall-clock time is spent generating MSAs and template features on CPUs rather than on GPU inference; once those files are cached, they can be reused for additional ligands bound to the same protein. However, medicinal-chemistry campaigns rarely focus only on a single invariant sequence. Screening panels of mutants, splice isoforms, or off-targets is routine. Because each new sequence requires a fresh MSA search, AF3's preprocessing cost re-emerges, whereas Chai-1's single-sequence mode workflow keeps its short runtime essentially unchanged for typical sequence lengths. In any setting that involves multiple variants or entirely different proteins, Chai-1 retains a decisive, order-of-magnitude efficiency advantage.

DISCUSSION

Our data validate that Chai-1 can predict covalent inhibitor poses for K-Ras(G12C), K-Ras(G12D), and K-Ras(G12S) at near-crystal resolution, yet several practical constraints remain. First, the model is biased toward the most abundant structures in the PDB and therefore fails to recognize alternative pockets such as the Switch I/II cleft, a transient allosteric groove between Switch-I and Switch-II. In our tests, Chai-1 did not reproduce the published poses for the G12C fragments targeting the Switch I/II cleft (Cpd3) or Switch I (Cpd4)[29] (Extended Data Fig. 1a, b). These pyridine compounds, which carry an acrylamide warhead, can cross-link Cys12 and occupy the Switch I/II cleft (Cpd3) or Switch I (Cpd4) rather than Switch II; however, Chai-1 predictions placed both Cpd3 and Cpd4 in the Switch II pocket. Another possible explanation could be fragment potency. Cpd3 and Cpd4 are low-affinity fragments (EC50 for MS modification: 23 μM and 11 μM, respectively)[29], so the model may struggle to stabilize the alternative pocket without a tight binding, fully optimized ligand. Potent, chemically optimized ligands impose stronger geometric constraints, guiding the model toward the correct binding mode. Thus, it would be worthwhile to test Chai-1 with higher-affinity Switch I/II ligands[30,31] to determine the exact cause.

Second, the model struggles with fine-grained chemical details. For the K-Ras(G12D) complex, the predicted covalent adduct displays improper bond angles (Figure 3c). A possible corrective workflow would be to generate an initial prediction and then run a short MD simulation or a Rosetta FastRelax[32] to obtain a more physically relevant structure. For the G12S mutant and its inhibitor G12Si-5, Chai-1 with a covalent bond restraint largely reproduced the binding pocket and ligand orientation (Figure 4a). However, it did not preserve the correct stereochemistry (Figure 4b) and missed a key hydrogen bond network in the ground truth structure[12], even though the input ligand was supplied with the proper stereochemical specification. This shortcoming likely stems from the network's inability to lock stereochemistry for reliable conformational analysis or to model reaction mechanisms and explicit crystallographic water molecules, which often stabilize the true, lower-energy state through hydrogen bonding. We regard this as a fundamental limitation of deep-learning structure predictors such as Chai-1 and AF3. Encouragingly, newer models have been developed to address these gaps. For example, Boltz-2x[33,34] adds steering potentials during inference and joint learning for structure and affinity to push poses toward physically consistent geometry. Separately, DiffDock-Glide[35] refines generative samples with Glide minimization to enforce local stereochemistry and relieve clashes. Finally, MIC[36] assigns positions for water and ions using deep learning to help restore stabilizing hydration networks.

Third, no current deep learning-based structure prediction tool automatically removes the leaving group after a covalent bond is formed. Our workaround for G12C inhibitors is to model acrylamide as propanamide; for G12Di-1 and G12Si-5, we input the ligand SMILES in its "post-reaction" form. Ideally, the software would apply covalent constraints that mimic authentic adducts, but that capability is not yet available. The challenge is even greater for arginine-specific warheads, which form two covalent bonds with a single residue[13]; to our knowledge, no existing pipeline can handle that chemistry. As an alternative, treating the ligand-modified residue as a non-canonical amino acid and predicting the covalently modified protein without an explicit ligand entry is possible, but the approach has not been optimized for high-throughput covalent docking.

Fourth, our validation set is limited to K-Ras G12 point mutants. Broader generality across other targets, including additional K-Ras cysteine variants (G13C, G61C) or even H- and N-Ras, remains untested. Finally, while Chai-1 is faster than MSA-based methods, each prediction still requires ~1–2 min of GPU time, making large-library screening slower and costlier than conventional docking. Even so, Chai-1's fully open, scriptable architecture should enable the community to iteratively close these gaps, positioning it as a routine and scalable engine for covalent lead optimization.

METHODS

Chai-1 web predictions. All jobs were submitted to the Chai-1 web portal at http://lab.chaidiscovery.com/. Predictions ran in single-sequence mode without using multiple sequence alignments, template structures, or any restraints.

Local GPU predictions with covalent restraint. GPU runs were carried out on the UCSF Wynton HPC (qb3-atgpu queue) using NVIDIA A40 GPUs. The source codes were drawn from the following GitHub repositories (Chai-1: https://github.com/chaidiscovery/chai-lab, AlphaFold3: https://github.com/google-deepmind/alphafold3). For ligand preparation, we used the SMILES string of the ligand as present in the crystal structure (after the covalent bond-forming reaction), with the leaving group already removed.

RMSD calculation. Pocket-aligned RMSDs were computed following the procedure described in the AlphaFold3 Supporting Information[9]. Briefly, the binding pocket is defined as all heavy-atom coordinates located within 10 Å of any heavy atom of the ligand in the reference crystal structure. The predicted model is then superposed onto the ground-truth structure by least-squares rigid-body alignment using only the Cα atoms of residues within this pocket. Finally, the RMSD is calculated over all heavy atoms of the ligand. A reproducible PyMOL/Python workflow and scripts are provided (alignment to a pocket selection, ligand extraction from ground truth and predictions, and RDKit-based RMSD with optional manual atom-pair overrides), together with example commands. Code and exact commands are included in the SI.

SUPPORTING INFORMATION

A single compressed archive, **Supplementary_Data.zip**, accompanies this Article. It contains three top-level directories:

1_Without Covalent Restraint_Web/

Chai-1 single-sequence web runs (no covalent restraint).

- **Chai-1 Inputs/** – protein sequence and ligand SMILES (FASTA files) used for inference.
- **Predicted Structures/** – raw model outputs (e.g., pred.model_idx_0 … pred.model_idx_4, in CIF as produced by the service).
- **PyMOL Sessions/** – .pse files showing the alignment of five predictions (single seed) to the reference structure for visualization.

2_With Covalent Restraint/

Local runs with explicit covalent restraints.

- **AlphaFold3/** – input files (JSON + ligand CIF), the corresponding predicted structures, and .pse files showing the alignment.
- **Chai-1/** – input files (FASTA + restraint file), the corresponding predicted structures, and .pse files showing the alignment.

3_Pocket-Aligned_RMSD/

Reproducible workflow, scripts, and raw results for the pocket-aligned RMSD procedure.

- **README.md** – step-by-step instructions (pocket definition in PyMOL, alignment, ligand extraction, and RMSD computation) consistent with the Methods text.
- **predictions_align_to_pocket.py** *(used in PyMOL)* – aligns pred.model_idx_0 … pred.model_idx_4 to the ground-truth **pocket** using only pocket **Cα** atoms.
- **save_ligands.py** *(used in PyMOL)* – extracts the ground-truth ligand (gt_ligand.pdb) and predicted ligands (pred.model_idx_#_ligand.pdb) based on user-set selections (edit gt_selection/selection inside the script).
- **pocket_aligned_ligand_RMSD.py** *(Python; RDKit + NumPy)* – computes heavy-atom ligand RMSDs between the reference and each pocket-aligned prediction; supports optional **manual atom-pair overrides**; writes a CSV summary with sanity checks.
- **Raw Pocket-Aligned RMSDs and Prediction Throughputs.xlsx** – raw RMSD values for all predictions and the measured prediction throughput figures used in the paper.

AUTHOR INFORMATION

Corresponding Author

Kevan M. Shokat - *Department of Cellular and Molecular Pharmacology and Howard Hughes Medical Institute, University of California, San Francisco, California 94158, United States*; *Department of Chemistry, University of California, Berkeley, California 94720, United States*; Email: kevan.shokat@ucsf.edu

ORCID

Sungwon Jung https://orcid.org/0009-0007-6284-2106

Qinheng Zheng https://orcid.org/0000-0002-8440-8673

Kevan M. Shokat https://orcid.org/0000-0001-8590-7741

Author Contributions

CRediT: **Sungwon Jung**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing; **Qinheng Zheng**: Investigation (suggested experiments), Writing – review & editing; **Kevan M. Shokat**: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

Declaration of interests

K.M.S. is an inventor on patents owned by University of California San Francisco covering KRAS targeting small molecules. K.M.S. has consulting agreements for the following companies, which involve monetary and/or stock compensation: AperTOR, BridGene Biosciences, Erasca, Exai, G Protein Therapeutics, Genentech, Initial Therapeutics, Kumquat Biosciences, Kura Oncology, Lyterian, Merck, Montara Therapeutics, Nested, Nextech, Revolution Medicines, Pfizer, Rezo, Totus, Type6 Therapeutics, Vevo, Vicinitas, Wellspring Biosciences (Araxes Pharma).

ACKNOWLEDGMENTS

Portions of this work were performed on the Wynton HPC Co-Op cluster, which is supported by UCSF research faculty and UCSF institutional funds. The authors wish to thank the UCSF Wynton team for their ongoing technical support of the Wynton environment. K.M.S. acknowledges support from the NIH (5R01CA244550), a gift from the Sjöberg Foundation, and the LRRK2 Investigative Therapeutics Exchange (LITE) program of the Michael J. Fox Foundation. S.J. acknowledges support from the Fulbright U.S.–Korea Presidential STEM Initiative Award, administered by the Korean American Educational Commission (Fulbright Korea). Q.Z. is the Connie and Bob Lurie Fellow of the Damon Runyon Cancer Research Foundation (DRG-2434-21) and acknowledges support from the UCSF Pancreas Center via a Mentored Scientist Award.

FIGURES

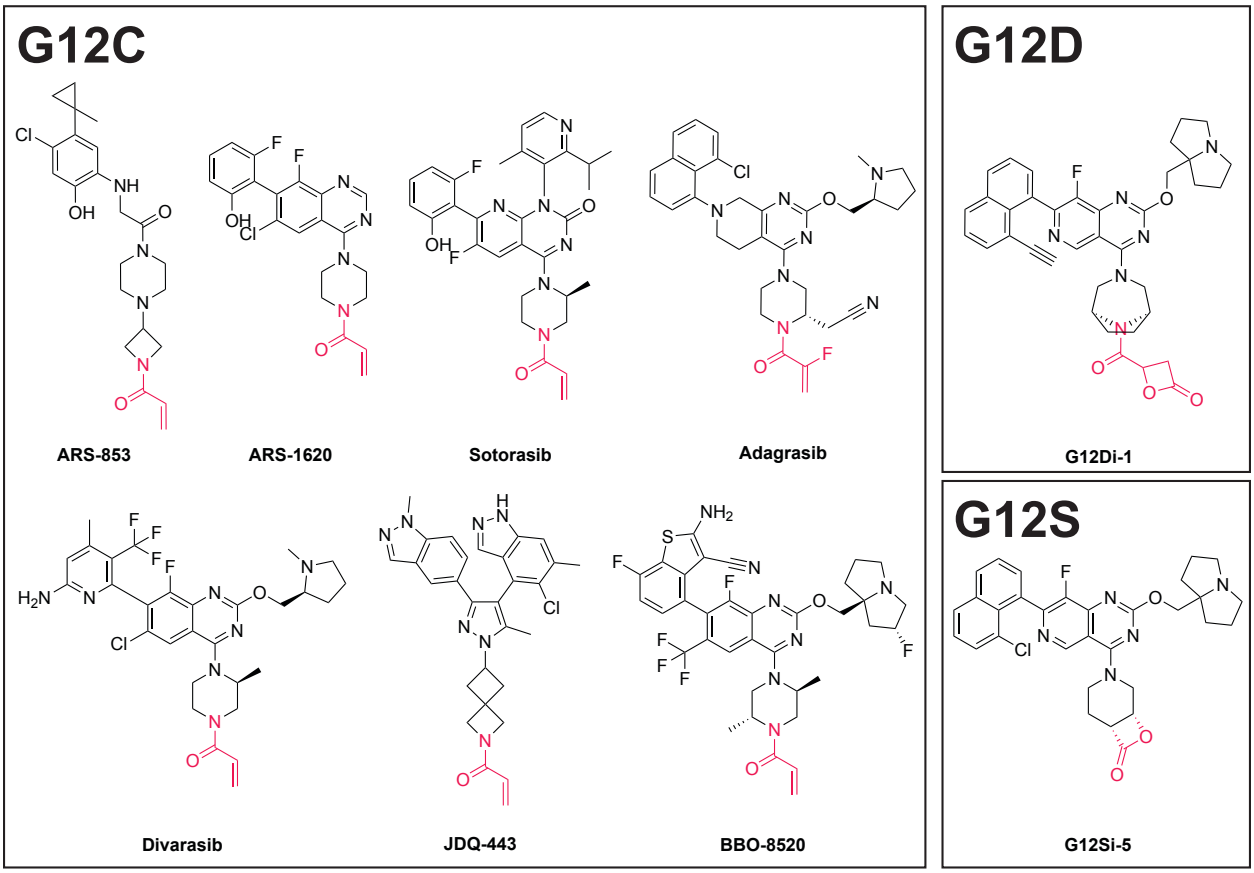


Figure 1 | Chemical structures of representative K-Ras(G12C) covalent inhibitors (ARS-853, ARS-1620, Sotorasib, Adagrasib, Divarasib, JDQ-443, and BBO-8520), K-Ras(G12D) covalent inhibitor (G12Di-1), and K-Ras(G12S) covalent inhibitor (G12Si-5), which were used for structure predictions. Only ARS-853 and ARS-1620 have K-Ras(G12C) co-crystal PDB entries released before the Chai-1 training cutoff (12 Jan 2021). The warhead is colored red.

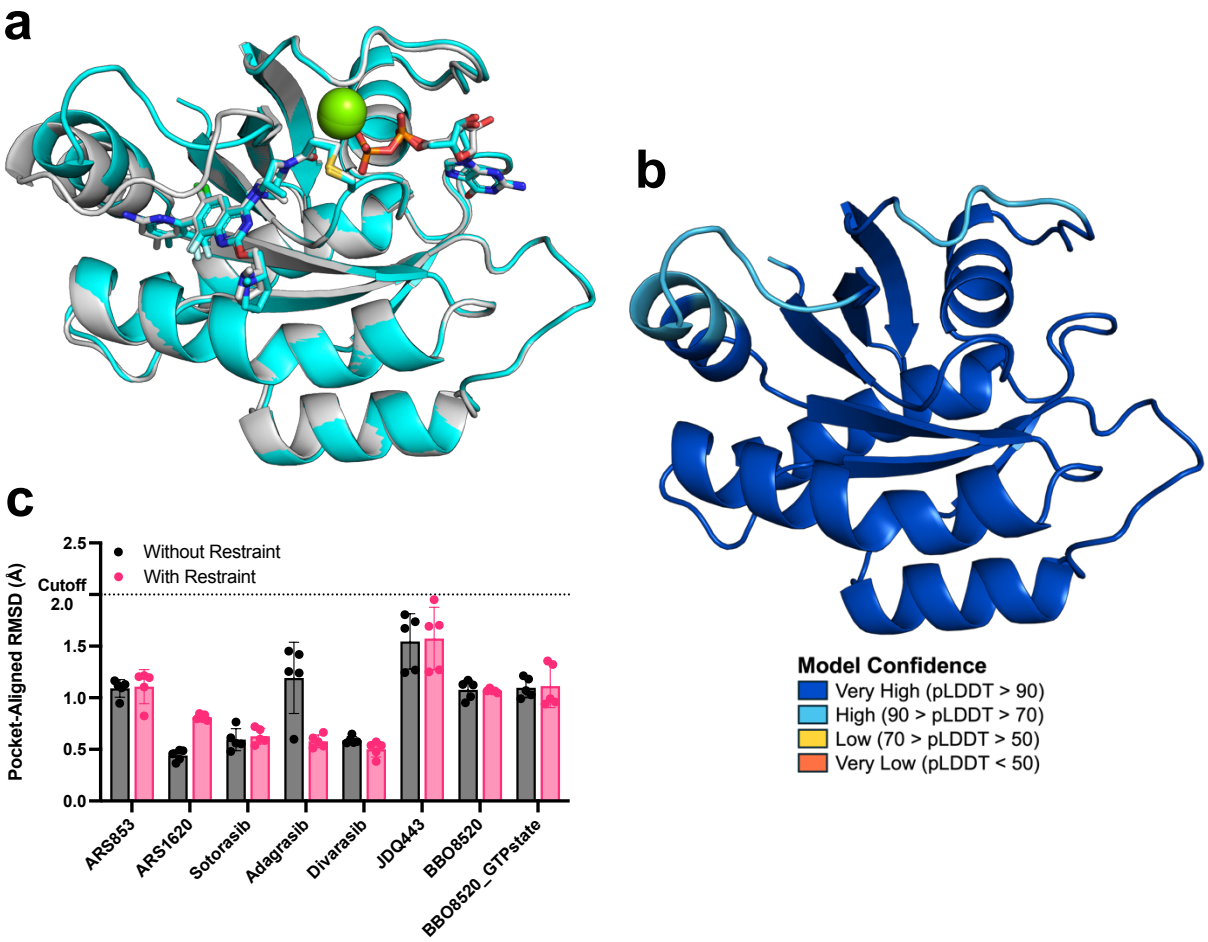


Figure 2 | ***Accurate prediction for covalent K-Ras(G12C) complexes.*** (a) Overlay of predicted (cyan) vs. crystal (grey) G12C–Divarasib complex; pocket-aligned RMSD 0.56 Å. (b) Predicted K-Ras(G12C)–Divarasib complex colored by per-residue pLDDT (predicted Local Distance Difference Test); switch-II loop exhibits relatively low confidence score. (c) Pocket-aligned RMSDs for K-Ras(G12C) inhibitors, comparing predictions without and with covalent restraint.

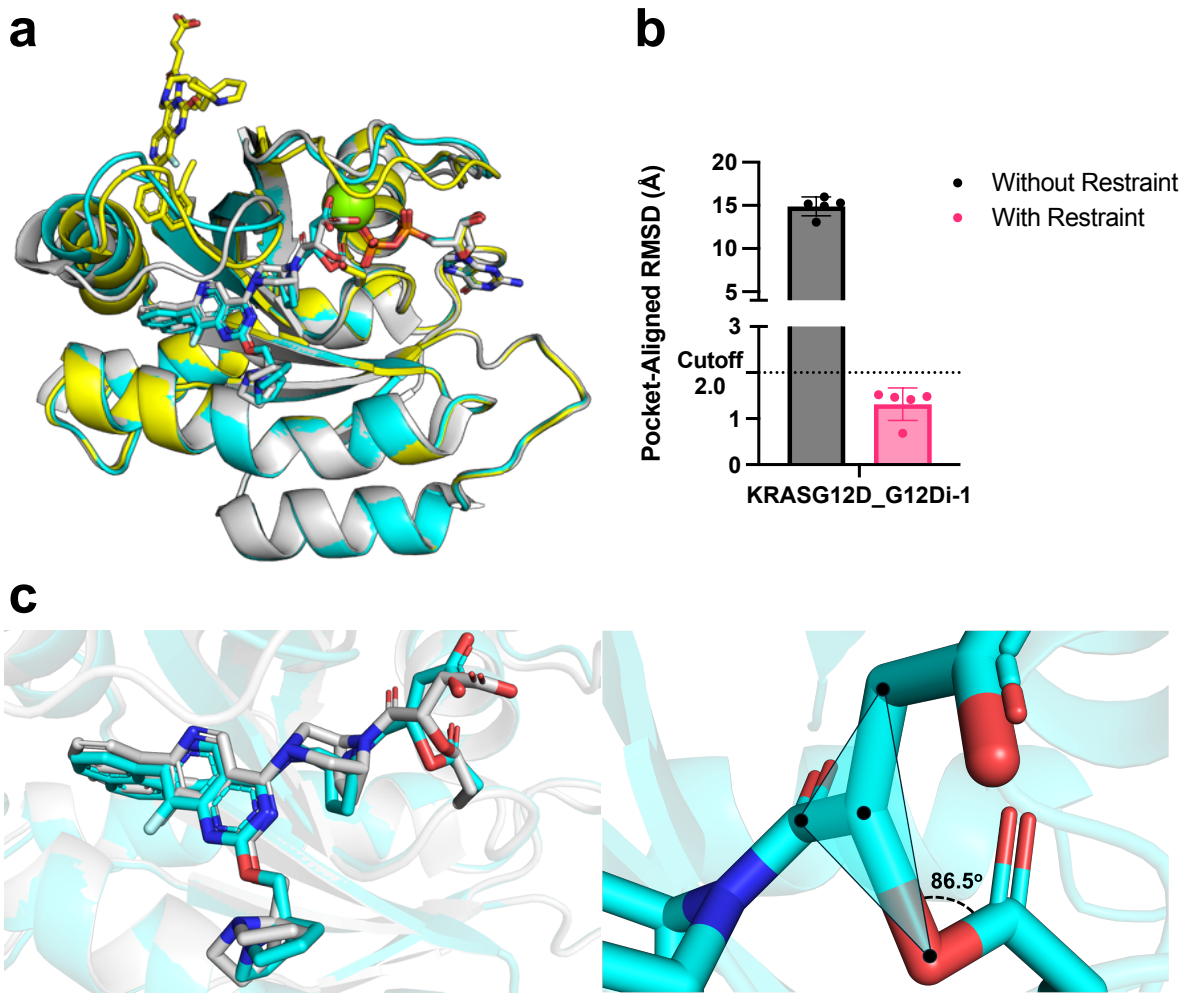


Figure 3 | *Effect of covalent restraint on K-Ras(G12D) complex.* (a) Overlay of naïve single-sequence prediction (yellow) shows ligand ejection from the switch-II pocket. Covalent restraint restores the correct pose (cyan, 0.68 Å Pocket-Aligned RMSD) relative to the crystal structure (grey). (b) Pocket-Aligned RMSD for K-Ras(G12D) inhibitor (G12Di-1), comparing predictions without and with covalent restraint. (c) Close-up overlay of the predicted G12Di-1 structure (cyan) and the crystal structure (grey). The newly formed covalent ester bond exhibits a strained C–O–C angle, and the sp^3-hybridized carbon within this C–O–C linkage adopts an unusually planar geometry.

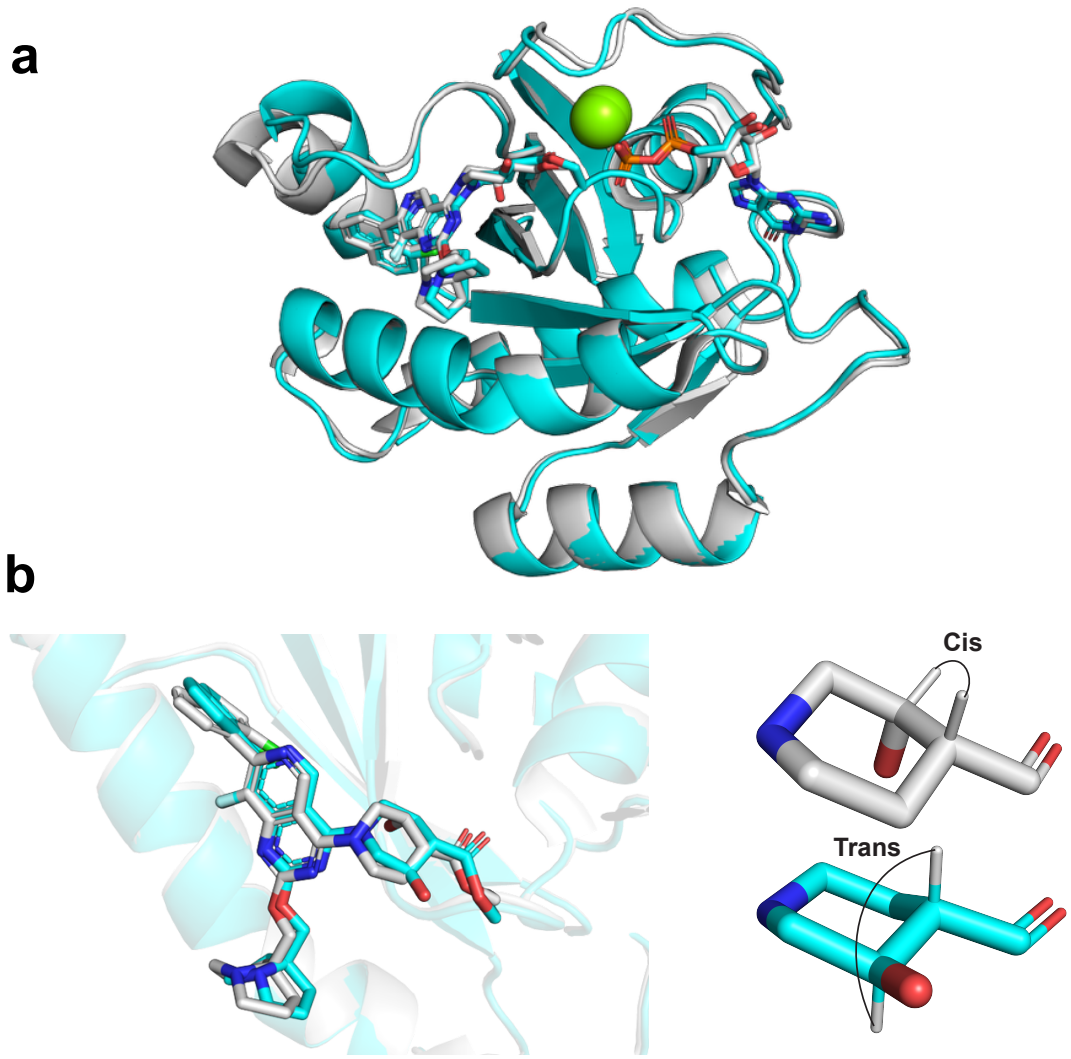


Figure 4 | *Effect of covalent restraint on K-Ras(G12S) complex.* (a) Overlay of the covalently restrained prediction (cyan) and crystallographic pose (grey) of the K-Ras(G12S)–G12Si-5 complex. Without the restraint (not shown), the desired covalent bond is not recapitulated in all five predictions. (b) Close-up overlay of G12Si-5 showing inversion of stereochemistry on the piperidine ring (crystal: cis; prediction: trans).

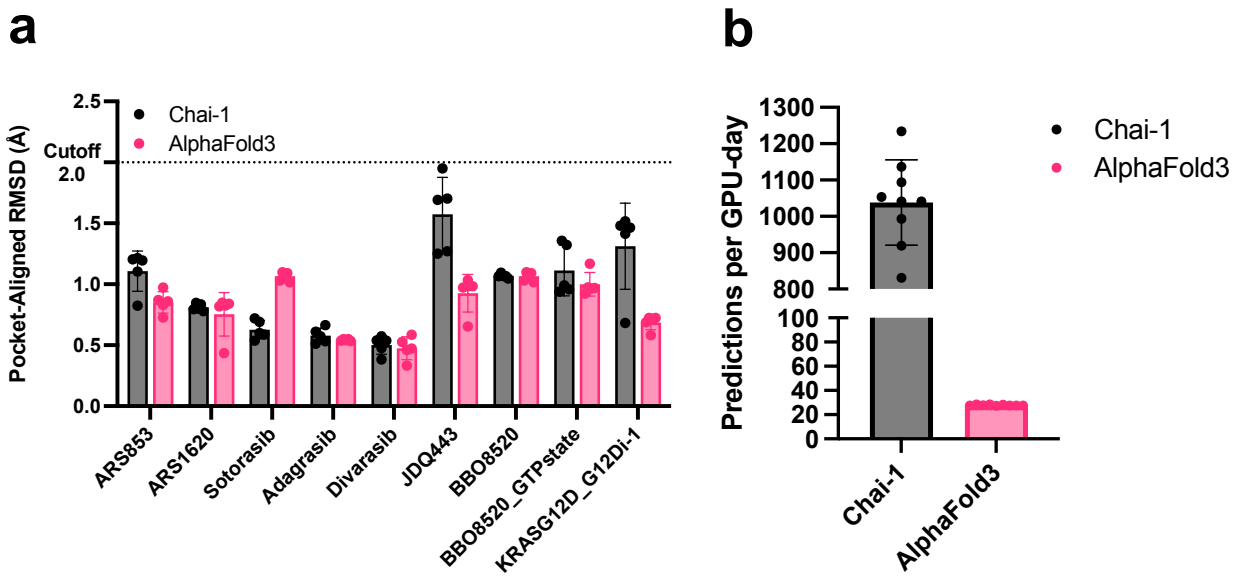
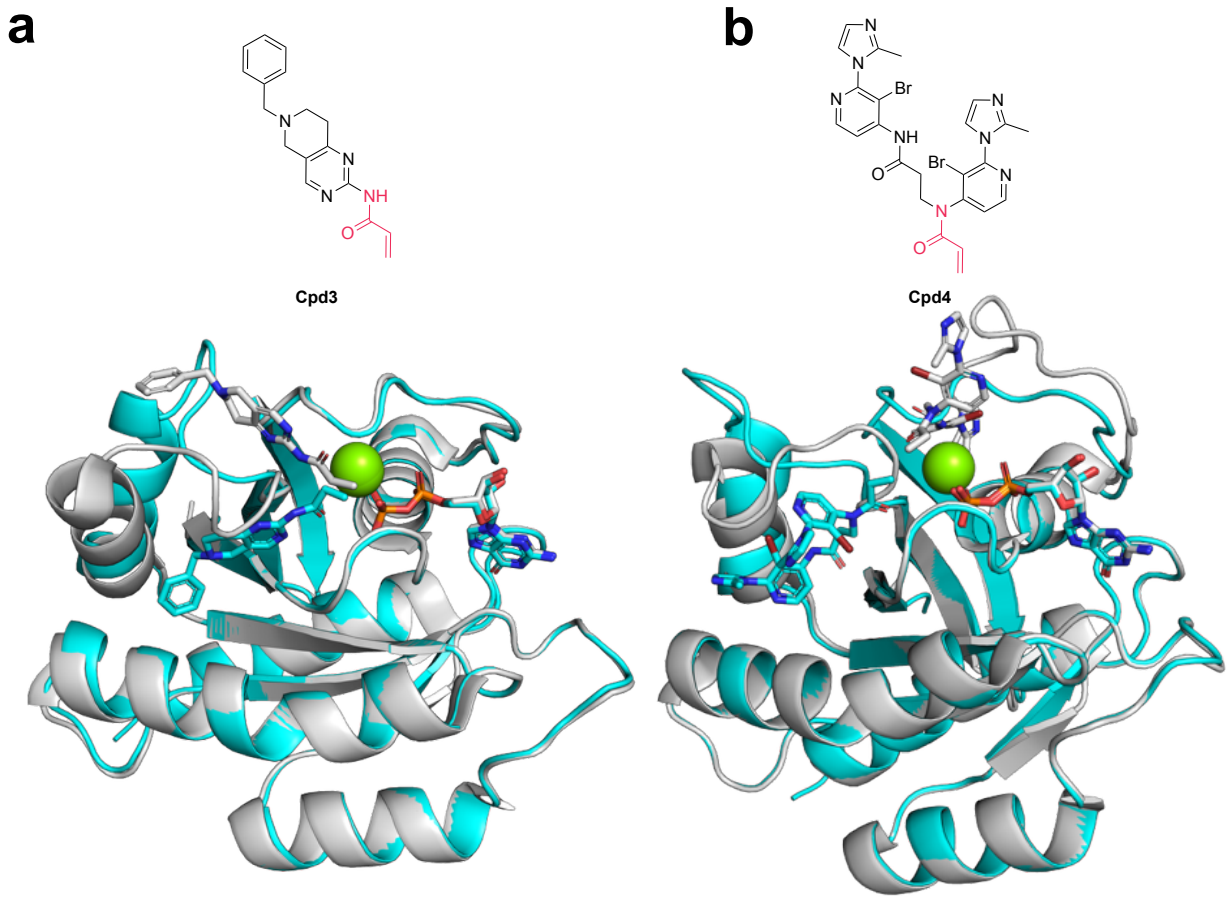


Figure 5 | ***Performance comparison between Chai-1 and AlphaFold3.*** (a) Pocket-aligned RMSDs for covalent K-Ras(G12C) and K-Ras(G12D) complexes with a covalent restraint. (b) Predicted complexes generated per 24 h on a single A40 GPU for Chai-1 in single-sequence mode versus AlphaFold3 with MSAs and templates.

EXTENDED DATA FIGURES



Extended Data Fig. 1 | *Lack of pose reproducibility for Switch I and Switch I/II-pocket fragments.* (a) Chemical structure of Cpd3 and overlay of the predicted (cyan) versus crystallographic (grey) KRAS (G12C)–Cpd3 complex. (b) Chemical structure of Cpd4 and overlay of the predicted versus crystallographic poses. In both cases, the model places the ligands in the Switch II pocket rather than the intended Switch I/II cleft (Cpd3) or Switch I (Cpd4), respectively.

REFERENCES

(1) Reita, D.; Pabst, L.; Pencreach, E.; Guérin, E.; Dano, L.; Rimelen, V.; Voegeli, A.-C.; Vallat, L.; Mascaux, C.; Beau-Faller, M. Direct Targeting KRAS Mutation in Non-Small Cell Lung Cancer: Focus on Resistance. *Cancers* **2022**, *14* (5), 1321. https://doi.org/10.3390/cancers14051321.

(2) Huang, L.; Guo, Z.; Wang, F.; Fu, L. KRAS Mutation: From Undruggable to Druggable in Cancer. *Signal Transduct. Target. Ther.* **2021**, *6* (1), 386. https://doi.org/10.1038/s41392-021-00780-4.

(3) Ostrem, J. M.; Peters, U.; Sos, M. L.; Wells, J. A.; Shokat, K. M. K-Ras(G12C) Inhibitors Allosterically Control GTP Affinity and Effector Interactions. *Nature* **2013**, *503* (7477), 548–551. https://doi.org/10.1038/nature12796.

(4) Canon, J.; Rex, K.; Saiki, A. Y.; Mohr, C.; Cooke, K.; Bagal, D.; Gaida, K.; Holt, T.; Knutson, C. G.; Koppada, N.; Lanman, B. A.; Werner, J.; Rapaport, A. S.; San Miguel, T.; Ortiz, R.; Osgood, T.; Sun, J.-R.; Zhu, X.; McCarter, J. D.; Volak, L. P.; Houk, B. E.; Fakih, M. G.; O'Neil, B. H.; Price, T. J.; Falchook, G. S.; Desai, J.; Kuo, J.; Govindan, R.; Hong, D. S.; Ouyang, W.; Henary, H.; Arvedson, T.; Cee, V. J.; Lipford, J. R. The Clinical KRAS(G12C) Inhibitor AMG 510 Drives Anti-Tumour Immunity. *Nature* **2019**, *575* (7781), 217–223. https://doi.org/10.1038/s41586-019-1694-1.

(5) Hallin, J.; Engstrom, L. D.; Hargis, L.; Calinisan, A.; Aranda, R.; Briere, D. M.; Sudhakar, N.; Bowcut, V.; Baer, B. R.; Ballard, J. A.; Burkard, M. R.; Fell, J. B.; Fischer, J. P.; Vigers, G. P.; Xue, Y.; Gatto, S.; Fernandez-Banet, J.; Pavlicek, A.; Velastagui, K.; Chao, R. C.; Barton, J.; Pierobon, M.; Baldelli, E.; Patricoin, E. F., III; Cassidy, D. P.; Marx, M. A.; Rybkin, I. I.; Johnson, M. L.; Ou, S.-H. I.; Lito, P.; Papadopoulos, K. P.; Jänne, P. A.; Olson, P.; Christensen, J. G. The KRASG12C Inhibitor MRTX849 Provides Insight toward Therapeutic Susceptibility of KRAS-Mutant Cancers in Mouse Models and Patients. *Cancer Discov.* **2020**, *10* (1), 54–71. https://doi.org/10.1158/2159-8290.CD-19-1167.

(6) Newby, Z. E. R.; O'Connel, J. D.; Gruswitz, F.; Hays, F. A.; Harries, W. E. C.; Harwood, I. M.; Ho, J. D.; Lee, J. K.; Savage, D. F.; Miercke, L. J. W.; Stroud, R. A General Protocol for the Crystallization of Membrane Proteins for X-Ray Structural Investigation. *Nat. Protoc.* **2009**, *4* (5), 619–637. https://doi.org/10.1038/nprot.2009.27.

(7) Blundell, T. L.; Jhoti, H.; Abell, C. High-Throughput Crystallography for Lead Discovery in Drug Design. *Nat. Rev. Drug Discov.* **2002**, *1* (1), 45–54. https://doi.org/10.1038/nrd706.

(8) Corso, G.; Stärk, H.; Jing, B.; Barzilay, R.; Jaakkola, T. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. arXiv February 11, 2023. https://doi.org/10.48550/arXiv.2210.01776.

(9) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions

with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.
(10) Shamir, Y.; London, N. State-of-the-Art Covalent Virtual Screening with AlphaFold3. bioRxiv March 19, 2025, p 2025.03.19.642201. https://doi.org/10.1101/2025.03.19.642201.
(11) Discovery, C.; Boitreaud, J.; Dent, J.; McPartlon, M.; Meier, J.; Reis, V.; Rogozhnikov, A.; Wu, K. Chai-1: Decoding the Molecular Interactions of Life. bioRxiv October 15, 2024, p 2024.10.10.615955. https://doi.org/10.1101/2024.10.10.615955.
(12) Zhang, Z.; Guiley, K. Z.; Shokat, K. M. Chemical Acylation of an Acquired Serine Suppresses Oncogenic Signaling of K-Ras(G12S). *Nat. Chem. Biol.* **2022**, *18* (11), 1177–1183. https://doi.org/10.1038/s41589-022-01065-9.
(13) Zhang, Z.; Morstein, J.; Ecker, A. K.; Guiley, K. Z.; Shokat, K. M. Chemoselective Covalent Modification of K-Ras(G12R) with a Small Molecule Electrophile. *J. Am. Chem. Soc.* **2022**, *144* (35), 15916–15921. https://doi.org/10.1021/jacs.2c05377.
(14) Zheng, Q.; Zhang, Z.; Guiley, K. Z.; Shokat, K. M. Strain-Release Alkylation of Asp12 Enables Mutant Selective Targeting of K-Ras-G12D. *Nat. Chem. Biol.* **2024**, *20* (9), 1114–1122. https://doi.org/10.1038/s41589-024-01565-w.
(15) Zheng, Q.; Shokat, K. M. Denitrogenative Alkylation of K-Ras(G12D) Inhibits Oncogenic Signaling in Cancer Cells. *J. Am. Chem. Soc.* **2025**. https://doi.org/10.1021/jacs.5c06745.
(16) Fernando, M. C.; Craven ,Gregory B.; and Shokat, K. M. The Structure of KRASG12C Bound to Divarasib Highlights Features of Potent Switch-II Pocket Engagement. *Small GTPases* **2024**, *15* (1), 1–7. https://doi.org/10.1080/21541248.2025.2505441.
(17) Pantsar, T. The Current Understanding of KRAS Protein Structure and Dynamics. *Comput. Struct. Biotechnol. J.* **2019**, *18*, 189–198. https://doi.org/10.1016/j.csbj.2019.12.004.
(18) Patricelli, M. P.; Janes, M. R.; Li, L.-S.; Hansen, R.; Peters, U.; Kessler, L. V.; Chen, Y.; Kucharski, J. M.; Feng, J.; Ely, T.; Chen, J. H.; Firdaus, S. J.; Babbar, A.; Ren, P.; Liu, Y. Selective Inhibition of Oncogenic KRAS Output with Small Molecules Targeting the Inactive State. *Cancer Discov.* **2016**, *6* (3), 316–329. https://doi.org/10.1158/2159-8290.CD-15-1105.
(19) Janes, M. R.; Zhang, J.; Li, L.-S.; Hansen, R.; Peters, U.; Guo, X.; Chen, Y.; Babbar, A.; Firdaus, S. J.; Darjania, L.; Feng, J.; Chen, J. H.; Li, S.; Li, S.; Long, Y. O.; Thach, C.; Liu, Y.; Zarieh, A.; Ely, T.; Kucharski, J. M.; Kessler, L. V.; Wu, T.; Yu, K.; Wang, Y.; Yao, Y.; Deng, X.; Zarrinkar, P. P.; Brehmer, D.; Dhanak, D.; Lorenzi, M. V.; Hu-Lowe, D.; Patricelli, M. P.; Ren, P.; Liu, Y. Targeting KRAS Mutant Cancers with a Covalent G12C-Specific Inhibitor. *Cell* **2018**, *172* (3), 578-589.e17. https://doi.org/10.1016/j.cell.2018.01.006.
(20) Lorthiois, E.; Gerspacher, M.; Beyer, K. S.; Vaupel, A.; Leblanc, C.; Stringer, R.; Weiss, A.; Wilcken, R.; Guthy, D. A.; Lingel, A.; Bomio-Confaglia, C.; Machauer, R.; Rigollier, P.; Ottl, J.; Arz, D.; Bernet, P.; Desjonqueres, G.; Dussauge, S.; Kazic-Legueux, M.; Lozac'h, M.-A.; Mura, C.; Sorge, M.; Todorov, M.; Warin, N.; Zink, F.; Voshol, H.; Zecri, F. J.; Sedrani, R. C.; Ostermann, N.; Brachmann, S. M.; Cotesta, S. JDQ443, a Structurally Novel, Pyrazole-Based, Covalent Inhibitor of KRASG12C for the Treatment

of Solid Tumors. *J. Med. Chem.* **2022**, *65* (24), 16173–16203. https://doi.org/10.1021/acs.jmedchem.2c01438.

(21) Maciag, A. E.; Stice, J. P.; Wang, B.; Sharma, A. K.; Chan, A. H.; Lin, K.; Singh, D.; Dyba, M.; Yang, Y.; Setoodeh, S.; Smith, B. P.; Ju, J. H.; Jeknic, S.; Rabara, D.; Zhang, Z.; Larsen, E. K.; Esposito, D.; Denson, J.-P.; Ranieri, M.; Meynardie, M.; Mehdizadeh, S.; Alexander, P. A.; Abreu Blanco, M.; Turner, D. M.; Xu, R.; Lightstone, F. C.; Wong, K.-K.; Stephen, A. G.; Wang, K.; Simanshu, D. K.; Sinkevicius, K. W.; Nissley, D. V.; Wallace, E.; McCormick, F.; Beltran, P. J. Discovery of BBO-8520, a First-In-Class Direct and Covalent Dual Inhibitor of GTP-Bound (ON) and GDP-Bound (OFF) KRASG12C. *Cancer Discov.* **2025**, *15* (3), 578–594. https://doi.org/10.1158/2159-8290.CD-24-0840.

(22) Zeissig, M. N.; Ashwood, L. M.; Kondrashova, O.; Sutherland, K. D. Next Batter up! Targeting Cancers with KRAS-G12D Mutations. *Trends Cancer* **2023**, *9* (11), 955–967. https://doi.org/10.1016/j.trecan.2023.07.010.

(23) Yu, Z.; He, X.; Wang, R.; Xu, X.; Zhang, Z.; Ding, K.; Zhang, Z.-M.; Tan, Y.; Li, Z. Simultaneous Covalent Modification of K-Ras(G12D) and K-Ras(G12C) with Tunable Oxirane Electrophiles. *J. Am. Chem. Soc.* **2023**, *145* (37), 20403–20411. https://doi.org/10.1021/jacs.3c05899.

(24) Weller, C.; Burnett, G. L.; Jiang, L.; Chakraborty, S.; Zhang, D.; Vita, N. A.; Dilly, J.; Kim, E.; Maldonato, B.; Seamon, K.; Eilerts, D. F.; Milin, A.; Marquez, A.; Spradlin, J.; Helland, C.; Gould, A.; Ziv, T. B.; Dinh, P.; Steele, S. L.; Wang, Z.; Mu, Y.; Chugh, S.; Feng, H.; Hennessey, C.; Wang, J.; Roth, J.; Rees, M.; Ronan, M.; Wolpin, B. M.; Hahn, W. C.; Holderfield, M.; Wang, Z.; Koltun, E. S.; Singh, M.; Gill, A. L.; Smith, J. A. M.; Aguirre, A. J.; Jiang, J.; Knox, J. E.; Wildes, D. A Neomorphic Protein Interface Catalyzes Covalent Inhibition of RASG12D Aspartic Acid in Tumors. *Science* **2025**, *389* (6758), eads0239. https://doi.org/10.1126/science.ads0239.

(25) Wen, C.; Yan, X.; Gu, Q.; Du, J.; Wu, D.; Lu, Y.; Zhou, H.; Xu, J. Systematic Studies on the Protocol and Criteria for Selecting a Covalent Docking Tool. *Molecules* **2019**, *24* (11), 2183. https://doi.org/10.3390/molecules24112183.

(26) Oyedele, A.-Q. K.; Ogunlana, A. T.; Boyenle, I. D.; Adeyemi, A. O.; Rita, T. O.; Adelusi, T. I.; Abdul-Hammed, M.; Elegbeleye, O. E.; Odunitan, T. T. Docking Covalent Targets for Drug Discovery: Stimulating the Computer-Aided Drug Design Community of Possible Pitfalls and Erroneous Practices. *Mol. Divers.* **2023**, *27* (4), 1879–1903. https://doi.org/10.1007/s11030-022-10523-4.

(27) Budai, B.; Vaupel, A.; Dickson, C. J.; Beyer, K. S.; Guthy, D. A.; Ostermann, N.; McGregor, L. M.; De Kanter, R.; Weiss, A.; Linder, M.; Sager, E.; Bomio-Confaglia, C.; Proctor, R.; Leblanc, C.; Yuan, J.; Cotesta, S.; Wilcken, R.; Danilack, A. D.; Garcia, F.; Rogemoser, P.; Kazic-Legueux, M.; Zhang, Z.; Zheng, Q.; Shokat, K. M.; Brachmann, S. M.; Ehmke, V. Promise and Challenge of β-Lactone Electrophiles to Target Aspartate 12 of Mutant KRASG12D. *J. Med. Chem.* **2025**, *68* (14), 15050–15064. https://doi.org/10.1021/acs.jmedchem.5c01214.

(28) Ishitani, R.; Moriwaki, Y. Improving Stereochemical Limitations in Protein-Ligand Complex Structure Prediction. bioRxiv July 30, 2025, p 2025.03.25.645362. https://doi.org/10.1101/2025.03.25.645362.

(29) Mathieu, M.; Steier ,Valérie; Fassy ,Florence; Delorme ,Cécile; Papin ,David; Genet ,Bruno; Duffieux ,Francis; Bertrand ,Thomas; Delarbre ,Laure; Le-Borgne ,Hélène; Parent ,Annick; Didier ,Patrick; Marquette ,Jean-Pierre; Lowinski ,Maryse; Houtmann ,Jacques; Lamberton ,Annabelle; Debussche ,Laurent; and Alexey, R. KRAS G12C Fragment Screening Renders New Binding Pockets. *Small GTPases* **2022**, *13* (1), 225–238. https://doi.org/10.1080/21541248.2021.1979360.

(30) Kazi, A.; Ranjan, A.; Kumar M.V., V.; Agianian, B.; Garcia Chavez, M.; Vudatha, V.; Wang, R.; Vangipurapu, R.; Chen, L.; Kennedy, P.; Subramanian, K.; Quirke, J. C. K.; Beato, F.; Underwood, P. W.; Fleming, J. B.; Trevino, J.; Hergenrother, P. J.; Gavathiotis, E.; Sebti, S. M. Discovery of KRB-456, a KRAS G12D Switch-I/II Allosteric Pocket Binder That Inhibits the Growth of Pancreatic Cancer Patient-Derived Tumors. *Cancer Res. Commun.* **2023**, *3* (12), 2623–2639. https://doi.org/10.1158/2767-9764.CRC-23-0222.

(31) Kessler, D.; Gmachl, M.; Mantoulidis, A.; Martin, L. J.; Zoephel, A.; Mayer, M.; Gollner, A.; Covini, D.; Fischer, S.; Gerstberger, T.; Gmaschitz, T.; Goodwin, C.; Greb, P.; Häring, D.; Hela, W.; Hoffmann, J.; Karolyi-Oezguer, J.; Knesl, P.; Kornigg, S.; Koegl, M.; Kousek, R.; Lamarre, L.; Moser, F.; Munico-Martinez, S.; Peinsipp, C.; Phan, J.; Rinnenthal, J.; Sai, J.; Salamon, C.; Scherbantin, Y.; Schipany, K.; Schnitzer, R.; Schrenk, A.; Sharps, B.; Siszler, G.; Sun, Q.; Waterson, A.; Wolkerstorfer, B.; Zeeb, M.; Pearson, M.; Fesik, S. W.; McConnell, D. B. Drugging an Undruggable Pocket on KRAS. *Proc. Natl. Acad. Sci.* **2019**, *116* (32), 15823–15829. https://doi.org/10.1073/pnas.1904529116.

(32) Das, R.; Baker, D. Macromolecular Modeling with Rosetta. *Annu. Rev. Biochem.* **2008**, *77* (Volume 77, 2008), 363–382. https://doi.org/10.1146/annurev.biochem.77.062906.171838.

(33) Wohlwend, J.; Corso, G.; Passaro, S.; Getz, N.; Reveiz, M.; Leidal, K.; Swiderski, W.; Atkinson, L.; Portnoi, T.; Chinn, I.; Silterra, J.; Jaakkola, T.; Barzilay, R. Boltz-1 Democratizing Biomolecular Interaction Modeling. bioRxiv May 6, 2025, p 2024.11.19.624167. https://doi.org/10.1101/2024.11.19.624167.

(34) Passaro, S.; Corso, G.; Wohlwend, J.; Reveiz, M.; Thaler, S.; Somnath, V. R.; Getz, N.; Portnoi, T.; Roy, J.; Stark, H.; Kwabi-Addo, D.; Beaini, D.; Jaakkola, T.; Barzilay, R. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. bioRxiv June 18, 2025, p 2025.06.14.659707. https://doi.org/10.1101/2025.06.14.659707.

(35) Herron, L.; Dakka, J.; Yao, K.; Shi, D.; Zhang, Y.; Jerome, S. V. DiffDock-Glide: A Hybrid Physics-Based and Data-Driven Approach to Molecular Docking. bioRxiv June 4, 2025, p 2025.06.02.657461. https://doi.org/10.1101/2025.06.02.657461.

(36) Shub, L.; Liu, W.; Skiniotis, G.; Keiser, M. J.; Robertson, M. J. MIC: A Deep Learning Tool for Assigning Ions and Waters in Cryo-EM and Crystal Structures. *Nat. Commun.* **2025**, *16* (1), 6182. https://doi.org/10.1038/s41467-025-61315-x.