JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

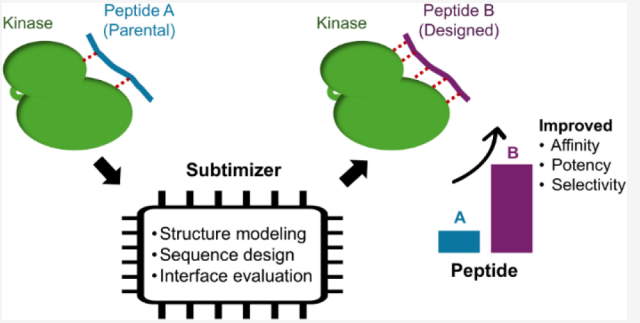
Subtimizer: Computational Workflow for Structure-Guided Design of Potent and Selective Kinase Peptide Substrates

Abeeb A. Yekeen, Cynthia J. Meyer, Melissa McCoy, Bruce Posner, and Kenneth D. Westover*

Cite This: J. Chem. Inf. Model. 2026, 66, 2041−2054 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Kinases are pivotal cell signaling regulators and prominent drug targets. Short peptide substrates are widely used in kinase activity assays essential for investigating kinase biology and drug discovery. However, designing substrates with high activity and specificity remains challenging. Here, we present Subtimizer (substrate optimizer), a streamlined computational pipeline for structure-guided kinase peptide substrate design using AlphaFold-Multimer for structure modeling, ProteinMPNN for sequence design, and AlphaFold2-based interface evaluation. Applied to five kinases, four showed substantially improved activity (up to 350%) with designed peptides. Kinetic analyses revealed >2-fold reductions in the Michaelis constant (K_m), indicating improved enzyme−substrate affinity. Designed peptides for MET and ROS1 exhibited reciprocal selectivity, with 4-fold and 11-fold preferences for their intended targets, respectively. This study demonstrates AI-driven structure-guided protein design as an effective approach for developing potent and selective kinase substrates, facilitating assay development for drug discovery and functional investigation of the kinome.



INTRODUCTION

Protein kinases, comprising over 500 members, are central regulators of cellular processes including metabolism, signal transduction, cell growth, and differentiation.[1−4] These enzymes catalyze the transfer of a phosphate group from ATP to specific serine, threonine, or tyrosine residues on their target protein substrates, thereby modulating protein function and signaling pathways.[4−6] The importance of kinases as a protein family is highlighted by the fact that they are dysregulated in numerous diseases, including cancer, inflammatory disorders, and metabolic syndromes, making kinases important targets for both fundamental research and therapeutic intervention.[2,7] Indeed, over 85% of the human kinome is implicated in human diseases.

Because of these disease connections, and because the active site of most kinases is amenable to competitive inhibition by small molecule inhibitors, kinases have been investigated extensively as drug targets.[4,8,9] This has led to the development and clinical success of numerous kinase inhibitors, with over 100 small-molecule inhibitors approved for clinical use.[6,8] However, these drugs target only about 10% of the human kinome, with the majority belonging to the tyrosine kinase family, leaving most kinases underexplored in clinical contexts.[8−11] A major barrier to inhibitor development has been the unavailability of robust kinase activity assays for a large portion of the kinome.[12,13] Indeed, more than 50% of known kinases do not have established high-throughput activity assays due to the lack of necessary tools, including optimal and assay-suitable substrates.[3,7,14−16] There is a critical need to discover kinase substrates that can generate sufficient activity to provide robust, unambiguous assay windows for inhibitor screening and structure-activity relationship studies.

In practice, researchers often use short synthetic peptide substrates to quantify kinase activity, enabling the study of kinase function and inhibitor development.[17−19] Compared to full-length protein substrates, peptides offer several advantages, including ease of synthesis, purification, and storage. They are also more cost-effective and readily adaptable to various assay formats, including high-throughput screening.[7,17−19] However, most existing kinase substrates are either promiscuous or nonselective, leading to high background and limited specificity. This hampers accurate activity measurement, especially for complex mixtures like cell lysates or tissue extracts.[7,18,20,21] Traditional substrate discovery and optimization methods, such as positional scanning peptide libraries or phosphoproteomic profiling, are costly, time-consuming and labor-intensive.[15,19,22,23] These methods often involve synthesizing libraries of peptides with systematic substitutions,

Received: October 9, 2025
Revised: January 23, 2026
Accepted: January 26, 2026
Published: February 7, 2026

followed by experimental determination of phosphorylation efficiency. While these approaches have been successful in identifying promiscuous substrates or minimal recognition motifs surrounding phosphosites, they are often limited by their inability to fully explore the vast sequence space beyond simple substitutions.[15,23] Thus, scalable, generalizable approaches to kinase substrate design and optimization are needed to address the growing need for tools to explore uncharacterized kinases, elucidate new mechanisms, and support kinase-targeted drug discovery.[9]

Existing computational efforts have primarily focused on identifying phosphorylation sites on protein substrates,[24,25] predicting kinase-protein substrate pairs,[26] or predicting kinases responsible for known phosphosites.[15,27] Despite the long-standing interest and practical importance of designing optimal peptide substrates for kinases,[28] computational design strategies have remained largely unexplored.[24,25] One notable existing design method is KINATEST-ID, developed to generate peptide substrates for use in chelation-enhanced fluorescence (CHEF) assays for tyrosine kinases.[17] However, KINATEST-ID's reliance on kinase compatibility with phosphorylation-dependent lanthanide ion coordination (specifically terbium, Tb^{3+}) limits its generalizability to diverse assay formats beyond Tb^{3+}-sensitization CHEF assays.[17]

Recent advances in AI-based protein modeling methods—propelled by the groundbreaking achievement of AlphaFold2 (AF2) for protein structure prediction[29] and followed by other related methods including RosettaFold,[30] ESMFold[31] and AF-Multimer[32]—have revolutionized structural biology and protein engineering.[33−35] Additionally, the field of protein sequence design (predicting amino acid sequences that fold into a given protein structure) has also witnessed the rapid emergence of innovative AI models with exceptional performance such as ProteinMPNN,[36] ESM-IF1,[37] and ABACUS-R.[38] The transformative impact of these advances was recognized with the 2024 Nobel Prize in Chemistry awarded to David Baker for computational protein design and jointly to Demis Hassabis and John Jumper for protein structure prediction.[39−41]

In this study, we evaluated the potential of AI-based protein design and structure prediction methods to generate optimized peptide substrates for kinases; we call our solution Subtimizer (substrate optimizer). We previously demonstrated that established protein design tools, such as the ABACUS2 statistical energy function[42,43] and the RosettaDesign physics-based method[44,45] can reprogram protease-substrate interactions.[46] We hypothesized that recent advances in AI-based protein modeling and design would overcome the limitations of physics-based and statistically learned empirical energy functions in selectively optimizing enzyme−substrate interactions.[46] To this end, we describe a streamlined structure-guided kinase substrate design workflow that utilizes AF-Multimer to predict the structure of a kinase in complex with a starting peptide substrate, ProteinMPNN for sequence optimization, and AF2-based interface prediction and evaluation.[47] As a proof of concept, we optimized known but suboptimal peptide substrates for a set of kinases, and the resulting designed peptides were experimentally evaluated.

METHODS

Kinase-Peptide Complex Prediction

The 3D structures of kinase-peptide complexes were predicted with AF-Multimer[32] using the ColabFold[48] implementation. The code for local installation of ColabFold was obtained from the repository https://github.com/sokrypton/ColabFold. The sequences of the kinase catalytic domain obtained from UniProt and the starting peptide substrate were used as input to AF-Multimer. For the proof-of-concept study, 25 kinases were each paired with one or more starting peptide substrates previously experimentally validated in-house, yielding a total of 45 kinase-peptide pairs. For each pair, five rounds of AF-Multimer predictions were run with different random seeds, generating five models per round. The number of AF-Multimer recycles and Amber relax cycles were set to 10 and 3, respectively. The top-ranking model from each round was selected, resulting in a total of 5 predictions. Only predictions with an ipTM > 0.75 were considered high-confidence and used for the downstream design step.

Peptide Sequence Design

The AF-Multimer models were passed to ProteinMPNN, which was used to design novel sequences on the backbone of the starting peptide while keeping the kinase residues fixed and preserving the amino acid identity of the phosphosite (Ser/Thr/Tyr) in the peptide. For the AKT2-gsk3tide complex where two crystal structures are available (PDB IDs 1O6K, 1O6L), these were also included as separate pairs (a total of 47 pairs) for use as input for ProteinMPNN. For each of the input kinase-peptide models, 480 sequences were designed (a total of 2400 for all five of the AF-Multimer models) with a sampling temperature of 0.1 and batch size of 32. The designed sequences were clustered using CD-Hit[49] at 100% sequence identity to eliminate redundant sequences. The ProteinMPNN code was obtained from the repository https://github.com/dauparas/ProteinMPNN.

Structure Prediction and Evaluation of Designed Sequences

The newly designed peptide sequences were paired with their target kinase sequences, and their complex structures were predicted again using AF-Multimer with minimal parameters (2 AF recycles, 4 prediction models, and no Amber relaxation). The top-ranking AF-Multimer model for each pair was used as an initial guess for the interface prediction and evaluation using a modified version of AF2 (AF2 with-initial-guess) as described by Bennett et al. (2023).[47] The code for AF2 with-initial-guess interface prediction was obtained from https://github.com/nrbennet/dl_binder_design. The ipAE, ipTM, pTM, and pLDDT scores were used to evaluate and rank the ProteinMPNN-generated sequences to identify a set of high-confidence designs for experimental validation. For the kinases (ALK, MET, ROS1, EGFR L858R, and SRC) selected for experimental validation, two to five designed peptides with low ipAE and high ipTM and pLDDT scores were chosen for experimental tests. Following experimental characterization, all designed and parental peptides for the five kinases were retrospectively evaluated using the recently introduced interface predicted Score from Aligned Errors (ipSAE) metric.[50] The ipSAE code, obtained from https://github.com/dunbracklab/IPSAE, processes the structure files and PAE matrices from AF-Multimer outputs. Both the ipSAE and the ipSAE_min (indicating, respectively, the maximum and minimum of the two directional predictions kinase→peptide and peptide→kinase) were computed for each peptide-kinase pair using default cutoffs of PAE < 15 Å and distance <15 Å.

Structural Analysis and Energy Evaluation

Structure refinement and quantitative assessment of predicted binding interactions was performed using the Rosetta FlexPepDock protocol.[51] The Rosetta suite source code was obtained from https://www.rosettacommons.org/software/license-and-download. AF-Multimer models of the kinase complexes of the most potent designed peptide and their parental counterparts were used as input

https://doi.org/10.1021/acs.jcim.5c02430
J. Chem. Inf. Model. 2026, 66, 2041−2054

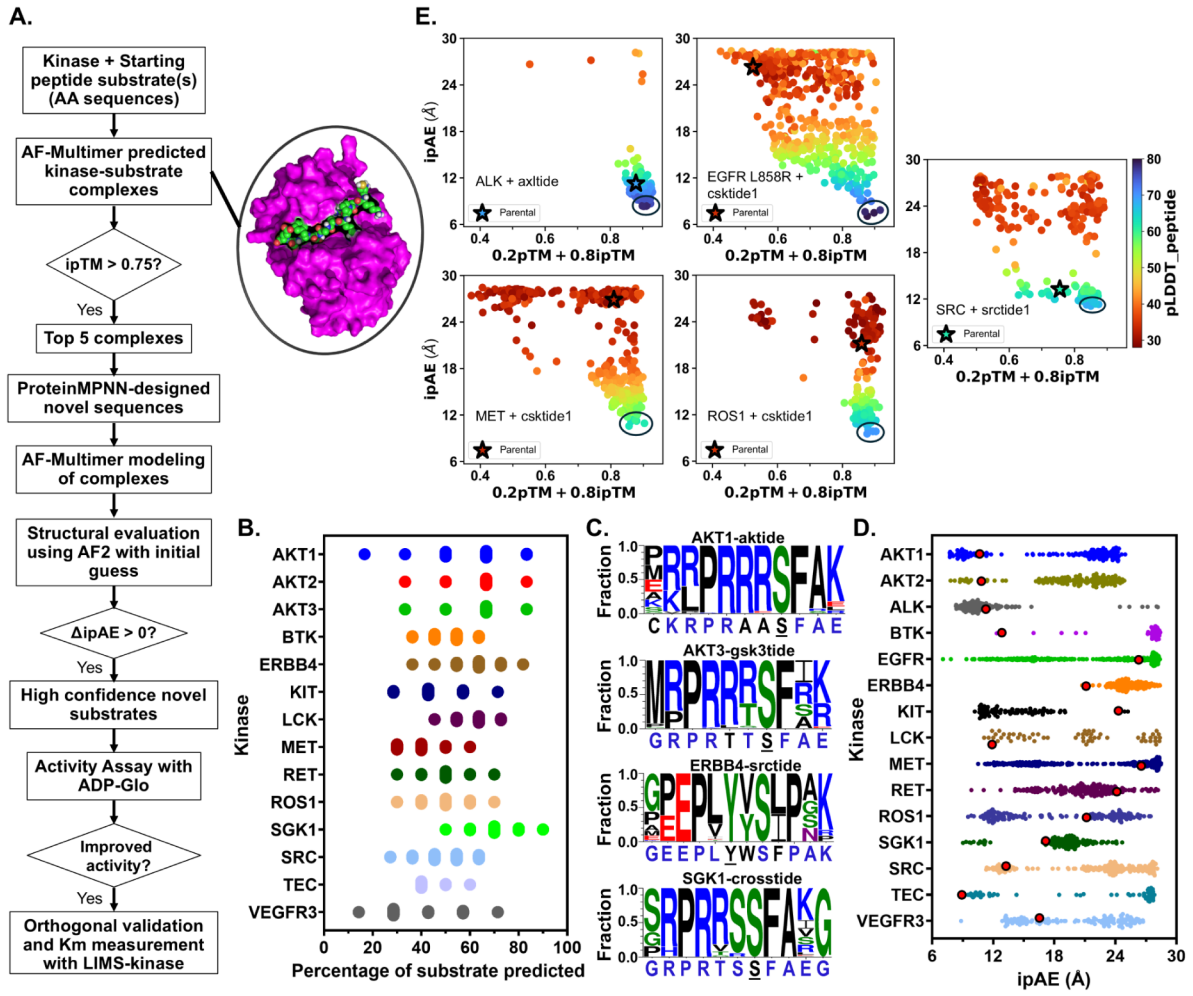


Figure 1. Computational workflow and validation of the Subtimizer pipeline for kinase peptide substrate design. (A) Schematic overview of the Subtimizer pipeline. The workflow proceeds from input sequences through AF-Multimer structure prediction, ProteinMPNN sequence design, and AF2-based evaluation to identify high-confidence substrate candidates. Inset: representative predicted structure of a kinase (magenta surface) bound to a peptide substrate (sphere representation). (B) Sequence recovery analysis showing the percentage of parental substrate residues that appear in the designed peptides for representative kinases. High recovery rates indicate that designed peptides retain known functional features. (C) Sequence logos comparing amino acid preferences between designed peptides (top logo in each panel) and parental substrates (bottom, colored sequence; the fixed phosphosite position is underlined). Letter height indicates frequency. (D) Distribution of interface predicted aligned error (ipAE) scores of designed and parental peptides for each of selected representative kinases. Each small colored dot represents a designed peptide; larger red dots indicate parental substrates. Lower ipAE values reflect higher confidence in the predicted interface. (E) Evaluation metrics for designed peptides across five kinases selected for experimental validation. Each dot represents a designed peptide, with position determined by ipAE (*y*-axis; lower is better) and a weighted confidence score (*x*-axis; 0.2 × pTM + 0.8 × ipTM; higher is better). Color gradient indicates peptide per-residue local confidence (pLDDT; green-dark blue indicates high confidence). Starred dots mark the parental peptides; circled clusters identify the top-scoring designed peptides selected for synthesis and testing. Note that MET, and ROS1 and EGFR L858R show clear separation between designed and parental peptides, while ALK and SRC show more overlap.

for FlexPepDock refinement and scoring using the Rosetta energy function. The refined structure with the lowest total Rosetta energy score was used for interaction analysis. Hydrogen bond interactions between kinase and peptide residues were calculated with HBPLUS 3.2[52] using angle and distance criteria of D-H-A > 125°, D-A < 3.45 Å. Predicted kinase-peptide complex structures and key interactions were visualized using PyMOL (Schrödinger, LLC). Plots were generated using Python or GraphPad Prism.

ADP-Glo Kinase Assays

ADP-Glo Kit (ADP-Glo™ Kinase Assay, Cat#V6930) was obtained from Promega. The kinases ALK (#08-518), MET (#08-151), ROS1 (#08-163), and EGFR L858R (#08-502) were purchased from Carna Biosciences. SRC was expressed and purified as previously reported.[53,54] Substrate peptides were obtained from Biomatik Corporation (Wilmington, DE, USA). All components were equilibrated to 25 °C prior to setting up reactions in 384-well microplates (White ProxiPlate 384-shallow well Plus). Final working concentrations were enzyme 1.25−10 nM, ATP 10 μM, peptide 1 μM. The assay buffer consisted of 100 mM HEPES pH 7.5, 0.003% Brij-35, 0.004% Tween 20, 10 mM MgCl2, and 2 mM DTT. All reagents were added manually and incubated for 90 min. Experiments were repeated twice with results expressed as mean ± standard deviation. Luminescence was detected on a Synergy Neo2 plate reader.

MRM-Kinase Assays: Enzyme Optimization

All components were equilibrated to 25 °C prior to setting up reactions in 384-well microplates (Costar 3657 round-bottom polypropylene). All reagents were added manually, and final working concentrations were enzyme 0.63−10 nM, ATP 100 μM, peptide 1−10 μM. Enzyme was first added to plate, followed by peptide/ATP mix in buffer (100 mM HEPES pH 7.5, 0.003% Brij-35, 0.004% Tween 20, 10 mM MgCl2, 2 mM DTT) to initiate reaction. Reactions

https://doi.org/10.1021/acs.jcim.5c02430
J. Chem. Inf. Model. 2026, 66, 2041−2054

were quenched at 20, 40, 60, 80, 100 min with formic acid at final concentration of 1%. Experiments were repeated twice and results expressed as mean ± standard deviation.

MRM-Kinase Assays: Kinetic Analysis

Final working concentrations of enzyme varied by kinase. ATP was at 100 μM, and peptide concentrations started at 300−400 μM followed by 2-fold or 3-fold dilutions. All components were equilibrated to 25 °C prior to setting up reactions. Peptide dilutions were added to 384-well microplates (Costar 3657 round-bottom polypropylene). All reagents were added manually. Buffer, enzyme, and ATP mixture were added to initiate reaction. Reactions were quenched at set time points with 1% formic acid (final). Activity was evaluated by measuring MS signal of ADP-specific fragment ion. Experiments were repeated twice, with results expressed as mean ± standard deviation.

Calculation of K_m

K_m was evaluated by linear regression of the first three time-points at each peptide concentration. The slope was used to represent initial velocity in a graph of initial velocity versus peptide concentration. Nonlinear regression using GraphPad Prism's Michaelis−Menten equation was used to estimate the K_m.

Peptide Selectivity Assays

For MET selectivity assay, reagent final working concentrations were MET 5 nM, peptides 134 μM (K_m for met-s1 with MET), and ATP 100 μM. For ROS1 selectivity assay, reagent final working concentrations were ROS1 0.63 nM, peptides 20.3 μM (K_m for ros1-s2 with ROS1), and ATP 100 μM. All components were equilibrated to 25 °C prior to setting up reactions. All reagents were added manually. Peptides were added to 384-well microplates (Costar 3657 round-bottom polypropylene). A premixed buffer, enzyme, and ATP solution was added to initiate the reaction. Reactions were quenched with 1% formic acid (final) following 30 min incubation at room temperature. Kinase activity was evaluated by measuring MS signal of the ADP-specific fragment ion. Experiments were repeated twice, with results expressed as mean ± standard deviation. Time 0 (background) values were subtracted from subsequent time-points.

RapidFire Chromatography and Mass Spectrometry

RapidFire liquid chromatography and mass spectrometry methods were performed as previously described with slight modifications.[16] Briefly, ADP production was used as a measure of kinase activity by directly monitoring the specific ADP ion fragment instead of the phosphorylated peptide ion fragment as in our previous report.[16]

RESULTS AND DISCUSSION

An AI-Driven Computational Pipeline for Designing Optimal Kinase Peptide Substrates

We hypothesized that AI-based protein structure prediction and sequence design tools could be combined to computationally optimize kinase peptide substrates. This hypothesis rests on two premises: first, that AF-Multimer can accurately model kinase-peptide complexes, capturing the structural determinants of substrate recognition; and second, that ProteinMPNN, trained to identify sequences compatible with protein structures, could redesign peptide sequences to strengthen kinase-substrate interactions. If correct, improved binding affinity should translate to enhanced catalytic efficiency, enabling the design of more potent substrates without extensive experimental screening. To test this hypothesis, we developed Subtimizer, a computational pipeline that integrates these AI tools in a streamlined workflow (Figure 1A). The pipeline takes a kinase sequence and a starting peptide substrate as input, then proceeds through four main stages: (1) structure prediction of the kinase-peptide complex, (2) sequence redesign of the peptide, (3) structural re-evaluation of the designed sequences, and (4) ranking to identify the most promising candidates for experimental testing.

In more detail, the pipeline (Figure 1A) begins with the amino acid sequences of a target kinase and a starting peptide substrate, which could be a known, literature-reported, or kinase family related substrate sourced from a phosphoproteomic database. AF-Multimer then generates multiple 3D structure models predicting how the peptide binds in the kinase active site. Given that the accuracy of the predicted complex structure is critical for downstream design, we implemented a filtering step based on the interface Predicted Template Modeling (ipTM) score, a confidence metric for protein−protein interfaces. Only complexes with ipTM > 0.75 (indicating high prediction confidence) were retained, with the top 5 models proceeding to the design stage. These structures then serve as templates for ProteinMPNN, which redesigns the peptide sequence while preserving the predicted backbone conformation. The amino acid identity of the phosphosite (Ser/Thr/Tyr) is fixed, as is the kinase sequence. The underlying rationale is that sequences optimized for tighter binding should also be better substrates, since productive catalysis requires the peptide to be properly positioned in the active site.

To assess whether the designed sequences would actually form the intended complexes, we performed a second round of structure prediction using AF-Multimer—this time with the new sequences rather than the parental peptides. This step tests whether the designed peptides are predicted to adopt a similar binding pose. The resulting models were further evaluated for interface quality using AF2 with initial guess, a method developed by Bennett et al. (2023) for assessing designed protein−protein interactions.[47] Designed peptides with average predicted aligned error of interchain residue pairs (ipAE) score $\leq$ 10 are considered high-confidence binders.[47] We considered a designed peptide to be computationally improved when ΔipAE > 0 relative to the parent peptide. Additional evaluation metrics utilized include peptide pLDDT (predicted local distance difference test)[29] and a weighted sum of pTM and interface pTM (0.2 × pTM + 0.8 × ipTM).[32]

To validate this workflow, we applied it to 47 kinase-substrate pairs experimentally validated in-house. This set comprised 25 unique kinases, each paired with one or more validated but suboptimal peptide substrates. We evaluated how well the Subtimizer workflow recovers residues present in the experimentally validated substrates. Figure 1B shows the percentage substrate recovery for a select set of kinases (see Figure S1 for all pairs). Figure 1C shows sequence logos depicting amino acid preferences for the designed peptides for a few representative kinase-substrate pairs (see Figure S2 for all pairs). Predictions for over half of the evaluated kinase-peptide pairs successfully recovered at least 70% (and up to 90% in some cases) of the residues in the validated substrates (Figures S1 and S2), demonstrating the pipeline's ability to generate sequences having features consistent with known substrates.

For each kinase-peptide pair, structural evaluation metrics, including ipTM and ipAE scores, were calculated using AF-Multimer and AF2 with-initial-guess.[47] The distribution of ipAE scores for representative kinase complexes of the designed and parental peptides is shown in Figure 1D (see Figure S3 for all kinase-peptide pairs). For all 47 kinase-peptide pairs except SGK1-1aktide, there are designed peptide complexes with ipAE lower than that of the parent peptide (Figures 1D and S3). In addition, most kinases have designed

https://doi.org/10.1021/acs.jcim.5c02430
J. Chem. Inf. Model. 2026, 66, 2041−2054

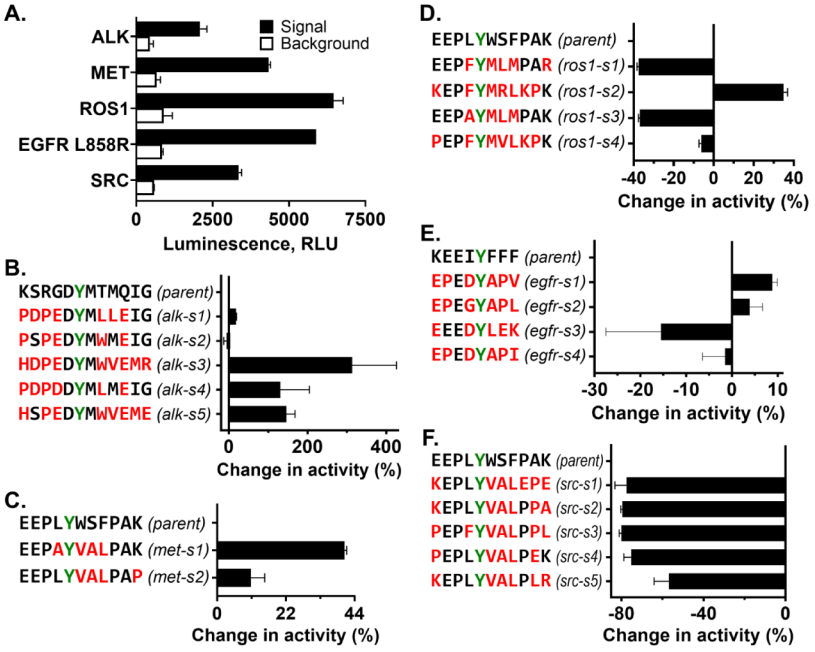


Figure 2. Experimental validation of designed peptides using the ADP-Glo kinase assay. (A) Baseline activity measurements comparing luminescence signal (proportional to ADP production) for parental substrates (black bars) versus no-enzyme background controls (white bars) across all five kinases. Error bars represent standard deviation (n = 2). (B–F) Percentage change in kinase activity for designed peptides relative to parental substrates: ALK (B), MET (C), ROS1 (D), EGFR L858R (E), and SRC (F). Peptide sequences are shown alongside each bar, with residues differing from the parent highlighted in red and the fixed phosphosite (Tyr) shown in green. Positive values indicate improved activity; negative values indicate reduced activity. ALK showed the largest improvements (up to >300%), while all SRC-designed peptides showed decreased activity.

peptides with ipAE < 10 for at least one substrate type. As shown in Figures 1E and S4, lower ipAE scores correlate with high ipTM and pLDDT scores, indicating that these metrics are suitable for evaluating prediction confidence. These structural evaluation metrics were used to filter the designed sequences down to a smaller set of high confidence designs to be prioritized for experimental testing.

Designed Peptide Substrates Showed Improved Potency In Vitro

Having established that our computational pipeline could generate peptide sequences with favorable predicted binding properties, we next asked whether these predictions would translate to improved catalytic activity in vitro. To test this, we selected five kinases (ALK, MET, ROS1, EGFR L858R, and SRC) for which recombinant protein was readily available. We note that designed peptides for three of the targets, EGFR, MET and ROS1, were far separated from the parental peptide ipAE scores whereas there was a narrow score separation for the SRC and ALK systems (Figure 1E). Using the ipAE, pLDDT, and ipTM scores, we ranked the Subtimizer-designed sequences, and for each kinase we selected two to five peptides for synthesis and experimental testing. We evaluated the activity of the peptides using the ADP-Glo kinase assay, a luminescence-based method that quantifies ADP production as a measure of kinase activity. First, we measured the ADP-Glo signals for the parental peptide and the baseline (no-enzyme background) for the five kinases. As shown in Figure 2A, we established that all five kinases are active under the assay conditions and that the signal generated from the phosphorylation reaction is significantly above background. The varying signal intensities across different kinases reflect differences in their specific activities and the efficiency of the parent peptide substrates.

Figure 2B–F shows, for ALK, MET, ROS1, EGFR, and SRC, respectively, the relative percentage change in kinase activity for the designed peptides compared to their respective parent peptides. As shown in Figure 2B, four out of five peptides designed for ALK showed improved activity compared to the parent peptide KSRGDYMTMQIG. Notably, peptide alk-s2 (PSPEDYMWMEIG) demonstrated a remarkable increase in activity, exceeding 300% relative to the parental axltide substrate. Alk-s5 (HSPEDYMWVEME) and alk-s4 (PDPDDYMLMEIG) also showed substantial improvements of 146% and 130%, respectively. For MET, two designed peptides met-s1 (EEPAYVALPAK) and met-s2 (EEPLYVALPAP) were tested as shown in Figure 2C. Compared to the parent peptide EEPLYWSFPAK, both met-s1 and met-s2 exhibited enhanced activity, showing relative improvements of approximately 41% and 11%, respectively.

For ROS1 (Figure 2D), only ros1-s2 (KEPFYMRLKPK) showed improved activity relative to the parent EEPLYWSFPAK, demonstrating an increase of about 35%. Evaluation of four designed substrates for EGFR L858R (Figure 2E) revealed two peptides, egfr-s1 (EPEDYAPV) and egfr-s2 (EPEGYAPL), with improved activity compared to the parent peptide KEEIYFFF. Egfr-s1 exhibited an increase in activity of about 10%, while egfr-s2 showed approximately 5% increase. Unlike ALK, MET, ROS1, and EGFR L858R, none of the five designed peptides tested for SRC activity (Figure 2F) showed better activity compared to the parent peptide. Instead, all designed SRC peptides showed reduced activity (57–80% decrease), suggesting that the computational predictions were less reliable for this kinase, consistent with the narrow score separation observed in Figure 1E. Overall, these results support

https://doi.org/10.1021/acs.jcim.5c02430
J. Chem. Inf. Model. 2026, 66, 2041−2054

our hypothesis that computationally optimized binding can translate to improved catalytic activity. Four of five kinases yielded at least one designed peptide with enhanced activity, with ALK showing the most dramatic improvement (>300% for alk-s2). The high success rate is notable: for most kinases, testing just two to five designed peptides was sufficient to identify improved substrates.

Improved Activity of Designed Peptides Correlates with Enhanced Binding Affinity

To further characterize the improved designed peptide substrates and gain insights into the mechanism underlying their enhanced activity, we determined the Michaelis constant (K_m) for the most potent designed peptides and their corresponding parent substrates using a mass spectrometry-based approach adapted from the LIMS-kinase assay previously developed in our group.[16] Figure 3A and B illustrate the schematic of a kinase phosphorylation reaction and the multiple reaction monitoring (MRM) detection system used to quantify ADP production. In this approach, a RapidFire liquid chromatography sampler (Agilent) is coupled to a triple quadrupole mass spectrometer (AB Sciex 6500). Rather than detecting the phosphorylated peptide product as in our original LIMS-kinase assay,[16] we monitored ADP production directly—this avoids potential variations in ionization efficiency across peptides with different sequences and enables unbiased comparison of activity across substrates (Figure 3B).

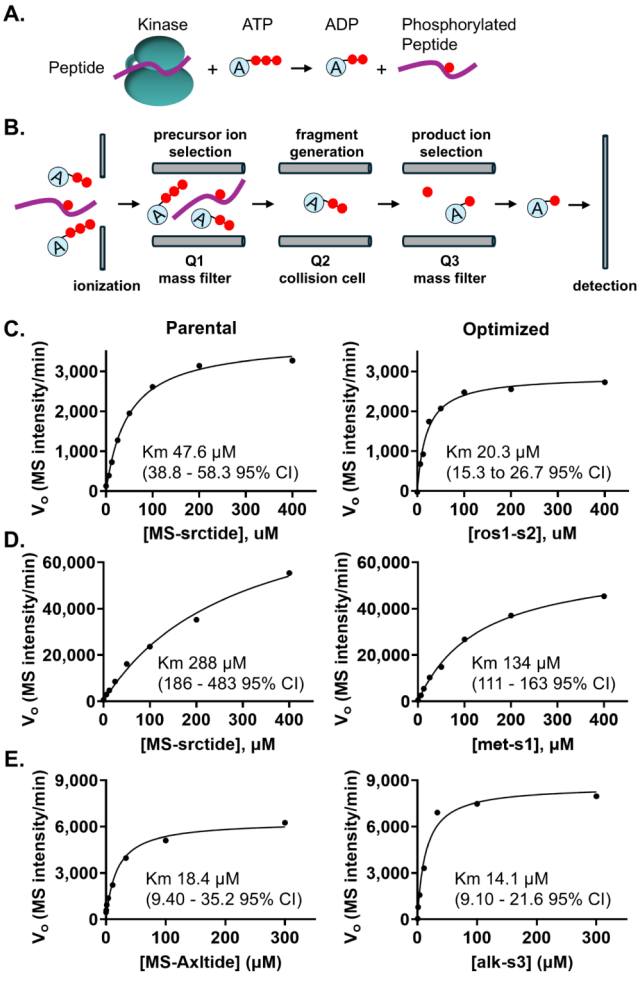


Figure 3. Kinetic characterization of designed peptides by ADP-based mass spectrometry. (A) Schematic of kinase phosphorylation reaction showing conversion of ATP to ADP. (B) Multiple reaction monitoring (MRM) detection scheme. ADP is ionized, filtered by mass (Q1), fragmented (Q2), and a specific product ion is monitored (Q3). This approach quantifies kinase activity independent of peptide sequence. (C–E) Michaelis–Menten analysis comparing parental (left) and optimized (right) peptides for ROS1 (C), MET (D), and ALK (E). Initial velocity (V_0) is plotted against peptide concentration; fitted K_m values and 95% confidence intervals. The designed peptides demonstrated lower K_m values, indicating improved apparent binding affinity.

Initial tests confirmed selective detection of ADP fragments in kinase/peptide reactions, with signal that was enzyme concentration- and time-dependent (Figure S5).[16] We optimized enzyme concentrations for ROS1, MET, ALK, and EGFR L858R, selecting conditions where activity was linear to 60 min with robust signal (Figure S5). We then performed kinetic analyses using a range of peptide concentrations. We plotted the initial reaction rate for each concentration to generate Michaelis–Menten curves. The curves were used to calculate K_m values for parental and most potent Subtimizer-designed peptides for ROS1, MET, and ALK, respectively (Figure 3C–E, Figures S6–S8). We could not run a successful K_m study with EGFR L858R due to the insolubility of the parental MS-csktide1 peptide at high concentrations, consistent with our previous observation.[16] For ROS1 (Figure 3C, Figure S6), the parent peptide (MS-srctide) exhibited a K_m of 47.6 μM while the optimized peptide ros1-s2 showed a significantly lower K_m of 20.3 μM. This more than 2-fold reduction in K_m indicates that the designed peptide has a substantially higher apparent affinity for ROS1 compared to the parent substrate.

While the parent peptide substrate MS-srctide had a K_m of 288 μM for MET (Figure 3D, Figure S7), the designed peptide met-s1 showed an improved activity with a K_m of 134 μM—an over 2-fold reduction in K_m for MET. For ALK (Figure 3E, Figure S8), the parent peptide MS-axltide exhibited a K_m of 18.4 μM while the designed peptide alk-s3 showed a lower K_m of 14.1 μM. This decrease in K_m suggests improved apparent affinity of the designed peptide for ALK, consistent with the enhanced activity observed in the ADP-Glo assay. The consistent observation of lower K_m values for the Subtimizer-designed peptides with improved activity across multiple kinases provides strong evidence that the computational pipeline optimizes substrates for kinase-substrate interaction, generating peptide sequences that bind more tightly to the kinase active site.

Subtimizer-Optimized Substrates Showed Improved Selectivity for Target Kinases

Beyond improving activity, we asked whether Subtimizer could also enhance substrate selectivity i.e., the preferential recognition of a peptide by its intended kinase over related kinases. This property is important for practical applications such as kinase activity assays in complex biological samples, where multiple kinases may be present. To test this, we took advantage of a natural experiment: MET and ROS1 are related receptor tyrosine kinases that both phosphorylate the MS-srctide peptide (EEPLYWSFPAK),[16] yet our pipeline produced distinct optimized sequences for each (met-s1 for MET, ros1-s2 for ROS1) even though we used the MS-srctide as the parent substrate for both (Figure 4A). If the design process successfully captured kinase-specific recognition features, each optimized peptide should show preferential

https://doi.org/10.1021/acs.jcim.5c02430
J. Chem. Inf. Model. 2026, 66, 2041–2054

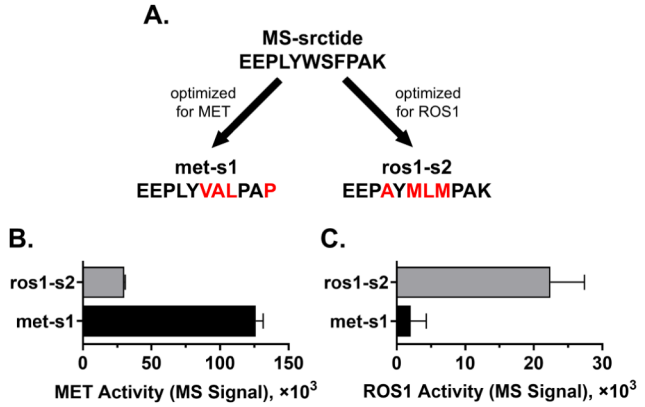


Figure 4. Subtimizer-designed peptides show reciprocal selectivity for their target kinases. (A) Experimental design: the parental MS-srctide peptide was independently optimized for MET (yielding met-s1) and ROS1 (yielding ros1-s2). Red letters indicate residues that differ from the parent sequence. (B) MET activity with both peptides measured by ADP-based mass spectrometry. Met-s1 (black bar) shows ∼4-fold higher activity than ros1-s2 (gray bar). (C) ROS1 activity with the same peptides shows the opposite pattern: ros1-s2 shows ∼11-fold higher activity than met-s1. Error bars represent standard deviation (n = 2). This reciprocal selectivity demonstrates that the optimization process captured kinase-specific recognition features.

activity with its intended target. This common parentage provides a direct context for evaluating whether the design process introduced kinase-specific selectivity.

Figure 4B shows the activity of MET kinase with each of the designed peptides ros1-s2 and met-s1, measured by ADP production using the mass spectrometry approach described above. MET showed high activity with its cognate designed peptide met-s1, consistent with the ADP-Glo results (Figure 2C). In contrast, MET kinase showed over 4-fold reduction in activity with the ros1-s2 peptide (designed for ROS1). This indicates that the modifications introduced during optimization of ros1-s2 for ROS1 reduced its efficiency as a substrate for MET, even though both peptides originated from the same MS-srctide parent.

Conversely, when the same two peptides ros1-s2 and met-s1 were incubated with ROS1 kinase (Figure 4C), ROS1 kinase showed an over 11-fold reduction in activity with the met-s1 peptide (designed for MET) compared to the ros1-s2 peptide. However, with the ros1-s2 peptide, ROS1 demonstrated high activity as anticipated from the ADP-Glo activity data (Figure 2D) and the K_m data (Figure 3C). This reciprocal pattern of activity demonstrates that the Subtimizer design process successfully introduced selectivity into the designed peptides. The MET-optimized peptide (met-s1) is preferentially phosphorylated by MET, while the ROS1-optimized peptide (ros1-s2) is preferred by ROS1, despite their common origin from the less selective MS-srctide peptide.

Structural and Computational Analyses of Optimized Kinase-Peptide Complexes

To gain deeper understanding of the molecular mechanisms underlying the observed improvements in activity, binding affinity (K_m), and selectivity of the designed peptide substrates, we asked whether the predicted structures could explain these improvements at the molecular level. We hypothesized that optimized peptides would show enhanced polar contacts with the kinase active site compared to parental peptides. To test this, we refined the AF-Multimer structural models using Rosetta FlexPepDock, a physics-based protocol for high-resolution peptide−protein docking.[51] We analyzed the kinase-peptide interactions observed in the refined structures (Figure 5). We also analyzed the computed Rosetta energy and interface scores of the refined structures to quantitatively assess the predicted binding interactions (Table 1).

Figure 5A−D presents structural representations and schematic diagrams highlighting key polar interactions between the kinases and their parental or optimized peptide substrates for ALK, ROS1, MET, and EGFR L858R. By comparing the hydrogen bonding patterns and overall predicted binding poses of the parental peptides with their optimized counterparts, we evaluated potential structural determinants that contribute to the enhanced functional properties. Although the parental peptide MS-axltide (green) forms several hydrogen bonds with residues of ALK (gray) in the structure model shown in Figure 5A, the modeled complex structure with the optimized peptide alk-s3 (yellow) forms slightly different hydrogen bonding patterns, including one additional hydrogen bond with ALK. While many of the hydrogen bonds are retained in the alk-s3, new interactions were made at the substituted peptide residues K1H, G4E, and G12R with kinase residues N151, M199, and N232. These structural changes, particularly the modifications in the hydrogen bonding pattern are predicted to lead to more favorable binding interaction, consistent with the enhanced activity and lower K_m observed for ALK-optimized peptides (Figures 2B and 3E).

In the case of ROS1 (Figure 5B), the two additional hydrogen bonds observed in the predicted model of ROS1 kinase with the optimized ros1-s2 substrate were formed by the unchanged residues E2 and K11. This indicates that the alterations in the sequences caused some changes in the binding mode of the peptide in a way that optimized the interactions of ROS1 with ros1-s2 for improved binding (Figure 3C) and higher activity (Figure 2D). As shown in Figure 5C, while the number of hydrogen bonds of the MET complexes of both the parental MS-srctide (green) and optimized met-s1 (yellow) peptide substrates are the same, the unchanged met-s1 peptide residues E2 and Y5 make new hydrogen bonds with the kinase residues T222 and D137, respectively. Furthermore, ros1-s2 residues P3, M8, and K11 form unique hydrogen bonds with ROS1 whereas no corresponding hydrogen bonds exist in met-s1 with MET. Similarly, met-s1 residue E1 makes a unique hydrogen bond with MET compared to the ROS1-ros1-s2 complex. These unique structural differences of the met-s1 and ros1-s2 complexes suggest that the design process introduced specific interactions tailored to the structural features of the binding pocket of the corresponding kinase. This indicates that these unique interactions are likely key determinants of the enhanced selectivity of met-s1 for MET over ROS1 (Figure 4B) and ros1-s2 for ROS1 over MET (Figure 4C).

For EGFR L858R, the unchanged residues E1 and E3 of the optimized egfr-s1 peptide make two new hydrogen bonds with the kinase residues R99 and K209 consistent with the improvement in activity (Figure 2E). Although one hydrogen bond between the parental MS-csktide1 peptide residue E2 and the kinase residue R99 was lost in egfr-s1, the lost bond was compensated for by the new bond between the egfr-s1 peptide residue E1 and the kinase residue K209. These changes in hydrogen bonding pattern suggest potential selectivity of the optimized egfr-s1 peptide for EGFR L858R as opposed to the promiscuous MS-csktide1, which is

https://doi.org/10.1021/acs.jcim.5c02430
J. Chem. Inf. Model. 2026, 66, 2041−2054

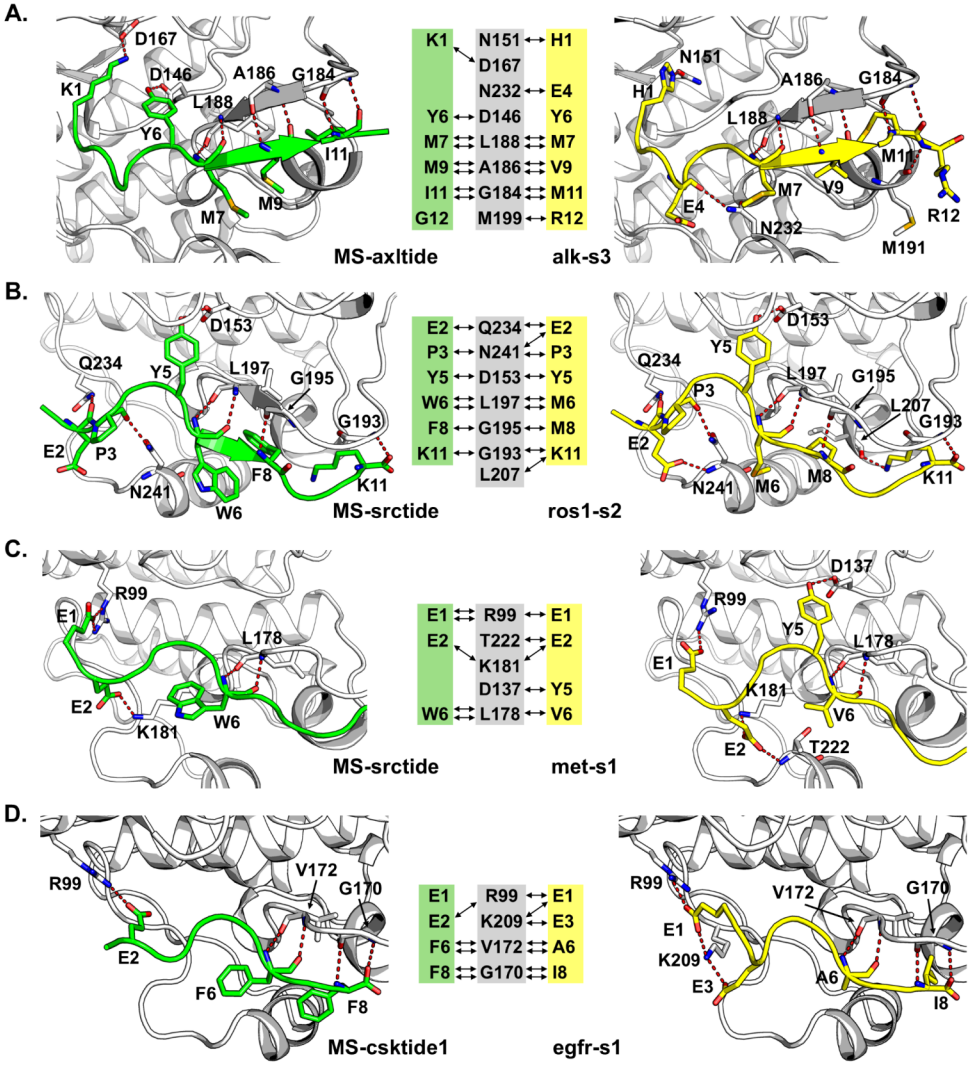


Figure 5. Structural comparison of parental and designed peptide-kinase complexes. Each panel shows the FlexPepDock-refined structures with parental peptide (green, left) and designed peptide (yellow, right), along with a central diagram mapping hydrogen bond contacts. Kinases are shown in gray ribbon representation; red dashed lines indicate hydrogen bonds. (A) ALK with MS-axltide versus alk-s3. The designed peptide forms one additional hydrogen bond through new residues at positions 1, 4, and 12. (B) ROS1 with MS-srctide versus ros1-s2. Two additional hydrogen bonds form despite sequence changes occurring at positions 6 and 8. (C) MET with MS-srctide versus met-s1. The same number of hydrogen bonds are maintained, but different residues participate. (D) EGFR L858R with MS-csktide1 versus egfr-s1. The designed peptide gains one hydrogen bond while losing another, with net improvement in computed binding energy.

phosphorylated by several other kinases, including MET, RET M918T, and ROS1.[16]

The Rosetta energy and interface scores calculated for each kinase-peptide complex offer a quantitative assessment of the predicted binding interactions of the parental and the most potent designed peptides (Table 1). For all kinases, the Rosetta scores "total_sc" (overall energy of the complex) and "reweighted_sc" (reweighted energy prioritizing interface and peptide residues) of the designed peptides were lower (better) than those of the parental peptides (Table 1), indicating that the designed peptides formed more energetically favorable and stable kinase-peptide interactions. Similarly, the designed alk-s3, ros1-s2, egfr-s2, and met-s1 peptides all have lower (better) Rosetta "pep_sc" (peptide score; overall peptide energy) than their corresponding parental peptides, indicating that the designed peptides have better stability and contribute more favorably to the total energy scores of their complexes.

The number of kinase-peptide interface hydrogen bonds (hb_I) for the designed peptides in complex with ALK, ROS1, and EGFR were higher than those of the parental peptides as shown in Table 1 and Figure 5. The increased numbers of polar interactions might play an important role in the improved activity of the designed alk-s3, ros1-s2, and egfr-s2 peptides. However, for MET, the number of interface hydrogen bonds is the same for both the parental and designed peptides. Although the hydrogen bonding pattern differs in the MET complexes of the parental and designed peptides, the improved activity of met-s1 over the parental MS-srctide might not be attributed to only polar interactions. The improvement might also be due to the overall stability of the met-s1 peptide and its complex with MET compared to the parental peptide as indicated by met-s1 having better Rosetta energy scores, including total_sc, reweighted_sc, pep_sc, fa_atr (LJ attractive energy), and fa_rep (LJ repulsive energy) (Table 1).

https://doi.org/10.1021/acs.jcim.5c02430
J. Chem. Inf. Model. 2026, 66, 2041−2054

Table 1. Rosetta FlexPepDock Analysis of Kinase-Peptide Complexes[a]

Kinase	Peptide	total_sc ↓	reweighted_sc ↓	hb_I ↑	I_sc ↓	hb_bb_sc ↓	hb_sc ↓	pep_sc ↓	fa_atr ↓	fa_elec ↓	fa_rep ↓
ALK	axltide (parental) KSRGDYMTMQIG	−647.6	−722.8	9	−53.5	−65.6	−24.09	−21.81	−1806.3	−507.8	273.7
	alk-s3 HDPEDYMWVEMR	−695.3	−776.3	10	−56.7	−68.6	−26.87	−24.21	−1824.0	−514.3	257.1
ROS1	srctide (parental) EEPLYWSFPAK	−778.7	−858.0	12	−58.2	−62.1	−23.16	−21.13	−1866.4	−530.3	219.2
	ros1-s2 EEPAYMLMPAK	−784.3	−879.0	15	−65.5	−64.1	−25.66	−29.15	−1848.6	−526.0	218.9
EGFR	csktide (parental) KEEIYFFF	−729.2	−784.6	8	−41.6	−42.7	−24.67	−13.81	−1732.2	−479.8	206.8
	egfr-s2 EPEDYAPI	−788.8	−849.1	10	−39.4	−46.3	−24.48	−20.88	−1743.3	−474.2	220.5
MET	srctide (parental) EEPLYWSFPAK	−494.8	−566.6	8	−51.2	−56.9	−19.76	−20.61	−1719.0	−446.8	308.1
	met-s1 EEPLYVALPAP	−559.6	−630.3	7	−46.6	−62.3	−17.03	−24.12	−1734.9	−433.4	277.9

[a]total_sc: Total score of the complex (REU: rosetta energy unit). reweighted_sc: Reweighted score of the complex (REU) with interface residues given double weight and peptide residues given triple weight. hb_I: Number of hbonds at the interface. I_sc: Interface score (sum over energy contributed by interface residues of both partners). hbond_bb_sc: sc-bb hbond energy. hbond_sc: sc-sc hbond energy. pep_sc: Peptide score (sum over energy contributed by the peptide to the total score; consists of internal peptide energy and interface energy). AF_hb: Number of hbonds in AF-multimer model used as input for Rosetta FlexPepDock. fa_atr: Lennard-Jones attractive between atoms in different residues. fa_rep: Lennard-Jones repulsive between atoms in different residues. fa_elec: Coulombic electrostatic potential with a distance-dependent dielectric.

Confidence Metrics Predict Optimization Potential and Guide Pipeline Refinement

The experimental results revealed an interesting contrast: while Subtimizer successfully improved substrates for ALK, MET, ROS1, and EGFR L858R, all designed peptides for SRC showed reduced activity (Figure 2F). We hypothesized that this failure might be predictable from computational confidence metrics. Across the representative kinase-substrate pairs, correlation analyses (Figure 1D-E, Figure S4) showed that the magnitude of confidence score separation between the parental substrates and their corresponding top-ranked Subtimizer designs varies substantially by kinase. For SRC, the parental substrate (MS-srctide) was located within the top-scoring region of the design landscape, with parental ipAE, ipTM and peptide pLDDT values close to those of the best-scoring designed peptides (Figure 1D and E). This contrasts with kinases such as MET and ROS1, where the parental peptide is separated from the best-scoring designs along the ipAE versus combined pTM-ipTM axis, indicating substantial room for computational optimization. We note that MS-srctide was derived from substrate specificity studies that identified optimal recognition motifs for Src-family kinases, and thus likely represents a high-affinity substrate with limited room for further improvement.[55,56]

Notably, the gap in interface confidence scores qualitatively mirrors the kinetic outcomes observed among the kinases tested. For ALK, where parental-versus-designed confidence separation is also narrow, the best designed peptide improved K_m only modestly (18.4 μM to 14.1 μM, a 1.3-fold improvement). In contrast, MET and ROS1 exhibited larger score deltas and their best-scoring designed peptides yielded correspondingly larger (>2-fold) improvements in affinity. Importantly, even though the designed peptides for ALK had a narrow score gap relative to the parental peptide (Figure 1D and E), they showed much higher confidence scores (ipAE < 10 and ipTM > 0.9) compared to the SRC designs (ipAE > 10 and ipTM < 0.9). These observations suggest that when designed peptides demonstrate both significant separation from the parental baseline and high absolute AF-derived interface confidence values, those improvements more reliably translate into measurable gains in apparent binding affinity (K_m). Conversely, smaller score separations, particularly when coupled with modest confidence values, may indicate limited capacity for functional improvement.

These observations raised the question of whether alternative metrics might better predict optimization success. We retrospectively scored all parental and designed peptides using the "interface predicted Score from Aligned Errors" (ipSAE) metric recently introduced by Dunbrack and coworkers.[57] The ipSAE is designed to correct the limitations of the ipTM score by focusing on high-confidence interface residue pairs.[57] Multiple recent benchmarking studies report that ipSAE improves binder/nonbinder discrimination and outperforms ipTM and ipAE for interaction prediction and affinity calibration.[50,51,58−60] From our retrospective evaluations, while ipSAE remained strongly correlated with ipTM (Pearson r > 0.9) across all kinases (as expected given that ipSAE is related to ipTM), a divergence emerged when comparing ipSAE to the ipAE and peptide pLDDT metrics (Figure S10). For the successful cases (ALK, EGFR, MET, ROS1), ipAE and peptide pLDDT correlated moderately well with ipSAE (*Pearson r* ranging from −0.32 to −0.64, with many of the designs having ipSAE > 0.6 (Figure S10A-D). However, for SRC, these correlations collapsed to −0.16 (ipAE vs ipSAE) and −0.19 (peptide pLDDT vs ipSAE) with all the designs having ipSAE < 0.6 (Figure S10E), which mirrors the low ipTM observed for SRC.

We also observed the same trend with ipSAE_min (Figure S11), defined as the minimum of the two asymmetric ipSAE scores (kinase→peptide and peptide→kinase), which has also been reported to be a good predictor of interaction quality.[58] The maximum ipSAE (simply represented as ipSAE) is, however, generally used for filtering. For all the evaluated kinases, we observed instances where low ipAE-ranked designs

https://doi.org/10.1021/acs.jcim.5c02430
J. Chem. Inf. Model. 2026, 66, 2041−2054

scored poorly on ipSAE and vice versa (Figure S12A−B). In addition, we observed that the ipSAE shows a wider gap between the parental peptide and the top ipSAE-ranked designs (Figure S12A). Taken together, these evaluations suggest that when the parental-versus-design ipAE score gap is narrow and especially coupled with weak or modest structure prediction scores (low ipTM or pTM), selection based solely on the ipAE ranking may not reliably predict functional binding or affinity improvements. Consequently, we have incorporated the ipSAE scoring into the Subtimizer pipeline and recommend its use as an additional filtering criterion. Specifically, we recommend prioritizing designs that show: (1) clear separation from the parental baseline in ipAE/ipTM, (2) high absolute confidence scores (ipTM above 0.9, ipAE below 10), and (3) ipSAE above 0.6. This multimetric approach should help identify cases where optimization is likely to succeed for the selection of candidate peptides for experimental testing.

DISCUSSION AND CONCLUSION

In this study, we present Subtimizer, a structure-guided computational workflow that approaches the design of potent and selective kinase peptide substrates as a protein design problem. The workflow integrates AF-Multimer for predicting kinase-peptide complex structures, ProteinMPNN for designing novel peptide sequences within this structural context, and structural metrics (ipTM, pTM, ipAE, and pLDDT) for evaluating the confidence and predicted quality of the designed peptides. Unlike previous studies on kinase-substrate relationships that mainly addressed the phosphorylation prediction task, the Subtimizer workflow combines both structure prediction and sequence design tasks to generate optimal peptide substrates for kinases.

We experimentally validated the Subtimizer pipeline by testing designed peptides for five kinases using a luminescence ADP-Glo kinase assay. For four out of the five kinases tested, we successfully identified designed peptides that demonstrated substantially improved kinase activity compared to their parental substrates, with the magnitude of improvement reaching over 300%. The most potent designed peptides for ALK, ROS1, and MET were further validated by the mass spectrometry-based MRM kinase assay, and kinetic characterization revealed that their improved activity was associated with lower K_m values (enhanced binding affinity). The improved binding affinity is consistent with the structure-guided design approach which aims to optimize kinase-peptide interactions within the substrate binding pocket. While we used the ADP-Glo and MRM-kinase assays for the experimental validation, we anticipate that these findings will translate to other kinase assay platforms. For example, the ADP-Glo assay has been repeatedly benchmarked against the ^{33}P-ATP-based radiometric assay and shown to give comparable results for various kinases and inhibitor profiling.[61,62]

We also demonstrated that the Subtimizer pipeline generates novel substrates with improved selectivity for the target kinase by showing that substrates designed for ROS1 and MET, while originating from the same parental peptide, exhibited 4-fold and 11-fold improvements in selectivity for their target kinase, respectively. Our selectivity analysis focused on these two kinases as proof-of-concept. High substrate affinity and selectivity are paramount for developing sensitive kinase assays suitable for low enzyme concentrations and complex biological samples like cell lysates and tissue extracts, as well as direct measurement in vivo.[7,18,20,21]

The computational confidence and quality evaluation steps of the workflow allowed for rapid prioritization of high-confidence designs for experimental validation, as only two to five designed peptides needed to be tested to identify peptides with improved activity. This significantly reduces the experimental burden compared to traditional screening, highlighting both the predictive power and practical efficiency of Subtimizer. Thus, the structure-guided protein design workflow offers a viable approach toward developing the necessary tools to enable comprehensive studies of the human kinome. The ability to generate high-affinity, selective substrates is crucial not only for fundamental research but also for accelerating drug discovery efforts, particularly for the large number of understudied kinases.[8,10,11,15,16]

Although the Subtimizer pipeline demonstrated high success rates by generating improved substrates for four (ALK, EGFR L858R, MET, and ROS1) of the five kinases tested, designed peptides tested for SRC showed reduced activity compared to the parental MS-srctide peptide. This highlights the potential limitations of the workflow in its current form and presents improvement opportunities that can be approached from three angles: by improving the accuracy of the (1) kinase-substrate structure models, (2) sequence design step, and (3) design filtering and selection.

First, the Subtimizer workflow could be improved by replacing AF-Multimer with the recently published AlphaFold 3, the latest iteration of AF that can predict structures of protein complexes with nucleic acids, small molecules, ions, and modified residues,[63] or comparable models like Boltz-2,[64] Protenix,[65] or Chai-1.[66] Second, LigandMPNN, the recently developed and ligand-aware version of ProteinMPNN that excels at designing residues near ligands and cofactors,[67] could be used in the sequence design step. Since kinases rely on metal cofactors and nucleotides for phosphorylation, implementing these ligand-aware modifications could improve the performance of the peptide design workflow. For example, such modifications could facilitate the optimization of not only substrate binding but also catalysis as design objectives, potentially overcoming the K_m-V_{max} trade-off seen with peptides designed for ROS1 and MET (Figure 3C−D). Our kinetic analyses revealed that while designed peptides consistently showed improved K_m (binding affinity), the magnitude of V_{max} improvements varied (Figure 3C−D). For example, ros1-s2 and met-s1 both achieved >2-fold K_m reductions, but their V_{max} values showed modest changes. This K_m-V_{max} trade-off likely reflects the fact that ProteinMPNN, which lacks explicit modeling of nucleotides and metals, optimizes only substrate binding rather than the complete catalytic geometry. Since kinases require precise positioning of ATP, metal cofactors, and the phosphoacceptor residue for efficient catalysis, incorporating these ligands during design is critical. LigandMPNN, which can explicitly model these ligands during sequence design, would enable optimization of both binding and catalysis, potentially yielding substrates with even greater overall activity improvements.

Third, our retrospective evaluation of the designed peptides using the recently introduced ipSAE metric (that has been reported to outperform the ipTM and ipAE scores used in filtering designs in the Subtimizer workflow) showed that incorporating ipSAE as an additional ranking criterion could improve the success of designs, especially for kinases with

https://doi.org/10.1021/acs.jcim.5c02430
J. Chem. Inf. Model. 2026, 66, 2041−2054

inadequate structure prediction metrics as we saw for SRC. Based on this retrospective analysis, we have incorporated the ipSAE ranking into the Subtimizer pipeline. Another recent metric that has improved on the ipTM is the actifpTM[68] which was introduced prior to ipSAE and has now been incorporated into the ColabFold version of AF2/AF-Multimer and a version of AF3 (https://github.com/Kuhlman-Lab/alphafold3). Although the actifpTM is also worthy of consideration, especially for protein-peptide complexes, we limited our retrospective evaluations to ipSAE because ipSAE is available as a standalone program that can be run on output structure and data files from already completed predictions.

Another potential limitation of the workflow is a case where the AF-Multimer structure prediction step fails to generate a complex structure that passes the threshold of the confidence metric. When this occurs, if using the improved structure prediction methods like AF3 or Boltz-2 does not solve the problem, it might be helpful to either incorporate an automated MD simulation tool for dynamics refinement,[69] or utilize methods such as AlphaRED[70] that significantly improved the success rate and accuracy of AF-Multimer on difficult cases by integrating ReplicaDock (a physics-based replica exchange docking algorithm). The combined implementation of ligand-aware sequence design with LigandMPNN, improved kinase-substrate-cofactor structure prediction with AF3 (or related methods), structural dynamics sampling with MD simulations, and improved filtering of designed peptides could be the key for challenging cases like SRC where initial design outcomes are poor.

In the current workflow, the solubility of the designed peptides was implicitly assumed given that sequences designed by ProteinMPNN have been shown to exhibit improved solubility compared to their natural counterparts.[36,71] We did in fact observe this in one of our validation experiments. We were unable to complete a K_m study for the EGFR L858R kinase because the parental peptide substrate was insoluble at high concentrations needed for kinetics study, whereas the designed peptides were fully soluble at such high concentrations. However, if solubility of designed peptides ever happens to be an issue, peptide solubility predictors like MahLool[72] or PeptideBERT[73] could be integrated into the workflow as explicit post-design solubility filtering.

In summary, this work demonstrates that AI-driven structure-guided protein design holds tremendous potential for scaling up kinase assay development by enabling rapid design of optimal kinase-specific peptide substrates for a larger set of kinases, including high-priority dark kinome members lacking validated substrates. More broadly, results from this study indicate that applications of recent advancements in AI-driven structure-guided protein engineering could generalize to other enzyme−substrate specificity engineering applications beyond kinases, such as proteases[46] and phosphatases.[74,75]

■ ASSOCIATED CONTENT

Data Availability Statement

The source code for Subtimizer is publicly available on GitHub (https://github.com/abeebyekeen/subtimizer). For ease of use, the pipeline is also distributed via the Python Package Index (PyPI) and can be installed using the "*pip install*" command. The outputs of AlphaFold, ProteinMPNN, AF2-with-initial-guess, and data analysis are available on Zenodo at 10.5281/zenodo.17309429.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.5c02430.

Figure S1: substrate sequence recovery analysis across kinase families; Figure S2: sequence logos of Subtimizer-designed peptides; Figure S3: distribution of interface predicted aligned error (ipAE) scores for all 47 kinase-substrate pairs; Figure S4: correlation between structural evaluation metrics; Figure S5: MRM-kinase assay optimization; Figure S6: kinetic analysis of ROS1; Figure S7: kinetic analysis of MET kinase; Figure S8: kinetic characterization of ALK kinase; Figure S9: structural comparison of AlphaFold-Multimer predicted kinase-bound complexes of parental substrates and Subtimizer-designed peptides; Figure S10: correlation analysis of the Interface Predicted Score from Aligned Errors (ipSAE) with standard interface evaluation metrics; Figure S11: correlation analysis of the ipSAE_min with standard interface evaluation metrics; Figure S12: scatter plots show the distribution of parental and designed peptides in the ipAE vs ipTM vs ipSAE confidence landscape (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Kenneth D. Westover − *Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9038, United States; Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9038, United States;* orcid.org/0000-0003-3653-5923; Email: Kenneth.Westover@UTSouthwestern.edu

Authors

Abeeb A. Yekeen − *Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9038, United States; Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9038, United States*

Cynthia J. Meyer − *Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9038, United States; Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9038, United States*

Melissa McCoy − *Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9038, United States*

Bruce Posner − *Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9038, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.5c02430

Author Contributions

A.A.Y. and K.D.W. conceived and designed the study. A.A.Y. developed and carried out computational investigations and wrote the first draft. C.J.M. carried out the wet experiments with help from A.A.Y. M.M. and B.P. coordinated the mass spectrometry (MS) data collection. C.J.M. and M.M. analyzed the MS data. K.D.W. edited and revised the draft, acquired funding, and supervised the study. All authors contributed to

the editing and revision of the manuscript. The final version of the manuscript was approved by all authors.

Funding

This work was supported by Welch Foundation (I-1829), NIH (UM1CA294119), (P50CA070907) and (P30CA142543).

Notes

The authors declare the following competing financial interest(s): K.D.W. has received consulting fees from Sanofi Oncology, Amgen, AstraZeneca, Boehringer Ingelheim and is a member of the SAB for Vibliome Therapeutics and Vellorum Pharmaceuticals. K.D.W. has received research funding from Revolution Medicines and Elekta. K.D.W. is a co-founder and equity holder in Stabilix LLC. K.D.W. declares that none of these relationships are directly or indirectly related to the content of this manuscript.

■ ACKNOWLEDGMENTS

We thank the University of Texas Southwestern Medical Center, Departments of Radiation Oncology and Biochemistry, and the Simmons Comprehensive Cancer Center for institutional support.

■ REFERENCES

(1) Soleymani, S.; Gravel, N.; Huang, L. C.; Yeung, W.; Bozorgi, E.; Bendzunas, N. G.; Kochut, K. J.; Kannan, N. Dark kinase annotation, mining, and visualization using the Protein Kinase Ontology. *PeerJ* **2023**, *11*, No. e16087.

(2) Berginski, M. E.; Moret, N.; Liu, C.; Goldfarb, D.; Sorger, P. K.; Gomez, S. M. The Dark Kinase Knowledgebase: an online compendium of knowledge and experimental results of understudied kinases. *Nucleic Acids Res.* **2021**, *49* (D1), D529−D535.

(3) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **2002**, *298* (5600), 1912−1934.

(4) Wilson, L. J.; Linley, A.; Hammond, D. E.; Hood, F. E.; Coulson, J. M.; MacEwan, D. J.; Ross, S. J.; Slupsky, J. R.; Smith, P. D.; Eyers, P. A.; Prior, I. A. New Perspectives, Opportunities, and Challenges in Exploring the Human Protein Kinome. *Cancer Res.* **2018**, *78* (1), 15−29.

(5) Hunter, T. Protein kinase classification. In *Methods in Enzymology*; Hunter, T.; Sefton, B. M., Eds.; Academic Press, 1991, Vol. *200*, pp. 3−37.

(6) Li, J.; Gong, C.; Zhou, H.; Liu, J.; Xia, X.; Ha, W.; Jiang, Y.; Liu, Q.; Xiong, H. Kinase Inhibitors and Kinase-Targeted Cancer Therapies: Recent Advances and Future Perspectives. *Int. J. Mol. Sci.* **2024**, *25* (10), 5489.

(7) Wu, D.; Sylvester, J. E.; Parker, L. L.; Zhou, G.; Kron, S. J. Peptide reporters of kinase activity in whole cell lysates. *Pept. Sci.* **2010**, *94* (4), 475−486.

(8) Attwood, M. M.; Fabbro, D.; Sokolov, A. V.; Knapp, S.; Schiöth, H. B. Trends in kinase drug discovery: targets, indications and inhibitor design. *Nat. Rev. Drug Discovery* **2021**, *20* (11), 839−861.

(9) Wu, P.; Nielsen, T. E.; Clausen, M. H. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol. Sci.* **2015**, *36* (7), 422−439.

(10) Anderson, B.; Rosston, P.; Ong, H. W.; Hossain, M. A.; Davis-Gilbert, Z. W.; Drewry, D. H. How many kinases are druggable? A review of our current understanding. *Biochem. J.* **2023**, *480* (16), 1331−1363.

(11) Fabbro, D. 25 Years of Small Molecular Weight Kinase Inhibitors: Potentials and Limitations. *Mol. Pharmacol.* **2015**, *87* (5), 766−775.

(12) Jafari, R.; Almqvist, H.; Axelsson, H.; Ignatushchenko, M.; Lundback, T.; Nordlund, P.; Martinez Molina, D. The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat. Protoc.* **2014**, *9* (9), 2100−2122.

(13) Warner, G.; Illy, C.; Pedro, L.; Roby, P.; Bosse, R. AlphaScreen kinase HTS platforms. *Curr. Med. Chem.* **2004**, *11* (6), 721−730.

(14) Perez, M.; Blankenhorn, J.; Murray, K. J.; Parker, L. L. High-throughput Identification of FLT3 Wild-type and Mutant Kinase Substrate Preferences and Application to Design of Sensitive In Vitro Kinase Assay Substrates. *Mol. Cell. Proteomics* **2019**, *18* (3), 477−489.

(15) Johnson, J. L.; Yaron, T. M.; Huntsman, E. M.; Kerelsky, A.; Song, J.; Regev, A.; Lin, T. Y.; Liberatore, K.; Cizin, D. M.; Cohen, B. M.; Vasan, N.; Ma, Y.; Krismer, K.; Robles, J. T.; van de Kooij, B.; van Vlimmeren, A. E.; Andree-Busch, N.; Kaufer, N. F.; Dorovkov, M. V.; Ryazanov, A. G.; Takagi, Y.; Kastenhuber, E. R.; Goncalves, M. D.; Hopkins, B. D.; Elemento, O.; Taatjes, D. J.; Maucuer, A.; Yamashita, A.; Degterev, A.; Uduman, M.; Lu, J.; Landry, S. D.; Zhang, B.; Cossentino, I.; Linding, R.; Blenis, J.; Hornbeck, P. V.; Turk, B. E.; Yaffe, M. B.; Cantley, L. C. An atlas of substrate specificities for the human serine/threonine kinome. *Nature* **2023**, *613* (7945), 759−766.

(16) Meyer, C.; McCoy, M.; Li, L.; Posner, B.; Westover, K. D. LIMS-Kinase provides sensitive and generalizable label-free in vitro measurement of kinase activity using mass spectrometry. *Cell Rep. Phys. Sci.* **2023**, *4* (10), 101599.

(17) Lipchik, A. M.; Perez, M.; Bolton, S.; Dumrongprechachan, V.; Ouellette, S. B.; Cui, W.; Parker, L. L. KINATEST-ID: a pipeline to develop phosphorylation-dependent terbium sensitizing kinase assays. *J. Am. Chem. Soc.* **2015**, *137* (7), 2484−2494.

(18) Fan, M. J.; Mehra, M.; Yang, K.; Chadha, R. S.; Anber, S.; Kovarik, M. L. Cross-Species Applications of Peptide Substrate Reporters to Quantitative Measurements of Kinase Activity. *ACS Meas. Sci. Au* **2024**, *4* (5), 546−555.

(19) Proctor, A.; Wang, Q.; Lawrence, D. S.; Allbritton, N. L. Development of a Peptidase-Resistant Substrate for Single-Cell Measurement of Protein Kinase B Activation. *Anal. Chem.* **2012**, *84* (16), 7195−7202.

(20) Shults, M. D.; Janes, K. A.; Lauffenburger, D. A.; Imperiali, B. A multiplexed homogeneous fluorescence-based assay for protein kinase activity in cell lysates. *Nat. Methods* **2005**, *2* (4), 277−284.

(21) Yeh, R.-H.; Yan, X.; Cammer, M.; Bresnick, A. R.; Lawrence, D. S. Real Time Visualization of Protein Kinase Activity in Living Cells*. *J. Biol. Chem.* **2002**, *277* (13), 11527−11532.

(22) Xiao, D.; Lin, M.; Liu, C.; Geddes, T. A.; Burchfield, J. G.; Parker, B. L.; Humphrey, S. J.; Yang, P. SnapKin: a snapshot deep learning ensemble for kinase-substrate prediction from phosphoproteomics data. *NAR:Genomics Bioinf.* **2023**, *5* (4), lqad099.

(23) Ljungdahl, T.; Veide-Vilg, J.; Wallner, F.; Tamás, M. J.; Grøtli, M. Positional Scanning Peptide Libraries for Kinase Substrate Specificity Determinations: Straightforward and Reproducible Synthesis Using Pentafluorophenyl Esters. *J. Comb. Chem.* **2010**, *12* (5), 733−742.

(24) Esmaili, F.; Pourmirzaei, M.; Ramazi, S.; Shojaeilangari, S.; Yavari, E. A Review of Machine Learning and Algorithmic Methods for Protein Phosphorylation Site Prediction. *Genomics, Proteomics Bioinf.* **2023**, *21* (6), 1266−1285.

(25) Audagnotto, M.; Dal Peraro, M. Protein post-translational modifications: In silico prediction tools and molecular modeling. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 307−319.

(26) Esmaili, F.; Qin, Y. F.; Wang, D. L.; Xu, D. Kinase-substrate prediction using an autoregressive model. *Comput. Struct. Biotechnol. J.* **2025**, *27*, 1103−1111.

(27) Ma, H.; Li, G.; Su, Z. KSP: an integrated method for predicting catalyzing kinases of phosphorylation sites in proteins. *BMC Genomics* **2020**, *21* (1), 537.

(28) Kemp, B. E.; Pearson, R. B. Design and use of peptide substrates for protein kinases. In *Methods in Enzymology*; Hunter, T.; Sefton, B. M., Eds.; Academic Press, 1991; Vol. *200*, pp. 121−134.

(29) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.;

https://doi.org/10.1021/acs.jcim.5c02430
J. Chem. Inf. Model. 2026, 66, 2041−2054

Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583−589.
(30) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millan, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373* (6557), 871−876.
(31) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **2023**, *379* (6637), 1123−1130.
(32) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J. et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*. 2022. .
(33) Read, R. J.; Baker, E. N.; Bond, C. S.; Garman, E. F.; van Raaij, M. J. AlphaFold and the future of structural biology. *Acta Crystallogr., Sect. D:Struct. Biol.* **2023**, *79* (7), 556−558.
(34) Varadi, M.; Bertoni, D.; Magana, P.; Paramval, U.; Pidruchna, I.; Radhakrishnan, M.; Tsenkov, M.; Nair, S.; Mirdita, M.; Yeo, J.; Kovalevskiy, O.; Tunyasuvunakool, K.; Laydon, A.; Zidek, A.; Tomlinson, H.; Hariharan, D.; Abrahamson, J.; Green, T.; Jumper, J.; Birney, E.; Steinegger, M.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **2024**, *52* (D1), D368−D375.
(35) Bertoline, L. M. F.; Lima, A. N.; Krieger, J. E.; Teixeira, S. K. Before and after AlphaFold2: An overview of protein structure prediction. *Front. Bioinform.* **2023**, *3*, 1120370.
(36) Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Courbet, A.; de Haas, R. J.; Bethel, N.; Leung, P. J. Y.; Huddy, T. F.; Pellock, S.; Tischer, D.; Chan, F.; Koepnick, B.; Nguyen, H.; Kang, A.; Sankaran, B.; Bera, A. K.; King, N. P.; Baker, D. Robust deep learning−based protein sequence design using ProteinMPNN. *Science* **2022**, *378* (6615), 49−56.
(37) Hsu, C.; Verkuil, R.; Liu, J.; Lin, Z.; Hie, B.; Sercu, T.; Lerer, A.; Rives, A. Learning inverse folding from millions of predicted structures. *bioRxiv*. 2022. .
(38) Liu, Y.; Zhang, L.; Wang, W.; Zhu, M.; Wang, C.; Li, F.; Zhang, J.; Li, H.; Chen, Q.; Liu, H. Rotamer-free protein sequence design based on deep learning and self-consistency. *Nat. Comput. Sci.* **2022**, *2* (7), 451−462.
(39) Prize, T. N. *Nobel Prize in Chemistry 2024*; 2024, https://www.nobelprize.org/prizes/chemistry/2024/press-release/.
(40) Abriata, L. A. The Nobel Prize in Chemistry: past, present, and future of AI in biology. *Commun. Biol.* **2024**, *7* (1), 1409.
(41) Lu, P.; Liu, L. Computational protein design and structure prediction—the 2024 Nobel prize in chemistry. *Sci. China Chem.* **2025**, *68* (3), 812−814.
(42) Xiong, P.; Hu, X.; Huang, B.; Zhang, J.; Chen, Q.; Liu, H. Increasing the efficiency and accuracy of the ABACUS protein sequence design method. *Bioinformatics* **2020**, *36* (1), 136−144.
(43) Xiong, P.; Wang, M.; Zhou, X.; Zhang, T.; Zhang, J.; Chen, Q.; Liu, H. Protein design with a comprehensive statistical energy function and boosted by experimental selection for foldability. *Nat. Commun.* **2014**, *5* (1), 5330.
(44) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K. W.; Renfrew, P. D.; Smith, C. A.; Sheffler, W., et al. Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. In *Methods in Enzymology*; Johnson, M. L.; Brand, L., Eds.; Academic Press, 2011; Vol. *487*, pp. 545−574.
(45) Leman, J. K.; Weitzner, B. D.; Lewis, S. M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R. F.; Aprahamian, M.; Baker, D.; Barlow, K. A.; Barth, P.; Basanta, B.; Bender, B. J.; Blacklock, K.; Bonet, J.; Boyken, S. E.; Bradley, P.; Bystroff, C.; Conway, P.; Cooper, S.; Correia, B. E.; Coventry, B.; Das, R.; De Jong, R. M.; DiMaio, F.; Dsilva, L.; Dunbrack, R.; Ford, A. S.; Frenz, B.; Fu, D. Y.; Geniesse, C.; Goldschmidt, L.; Gowthaman, R.; Gray, J. J.; Gront, D.; Guffy, S.; Horowitz, S.; Huang, P.-S.; Huber, T.; Jacobs, T. M.; Jeliazkov, J. R.; Johnson, D. K.; Kappel, K.; Karanicolas, J.; Khakzad, H.; Khar, K. R.; Khare, S. D.; Khatib, F.; Khramushin, A.; King, I. C.; Kleffner, R.; Koepnick, B.; Kortemme, T.; Kuenze, G.; Kuhlman, B.; Kuroda, D.; Labonte, J. W.; Lai, J. K.; Lapidoth, G.; Leaver-Fay, A.; Lindert, S.; Linsky, T.; London, N.; Lubin, J. H.; Lyskov, S.; Maguire, J.; Malmström, L.; Marcos, E.; Marcu, O.; Marze, N. A.; Meiler, J.; Moretti, R.; Mulligan, V. K.; Nerli, S.; Norn, C.; Ó'Conchúir, S.; Ollikainen, N.; Ovchinnikov, S.; Pacella, M. S.; Pan, X.; Park, H.; Pavlovicz, R. E.; Pethe, M.; Pierce, B. G.; Pilla, K. B.; Raveh, B.; Renfrew, P. D.; Burman, S. S. R.; Rubenstein, A.; Sauer, M. F.; Scheck, A.; Schief, W.; Schueler-Furman, O.; Sedan, Y.; Sevy, A. M.; Sgourakis, N. G.; Shi, L.; Siegel, J. B.; Silva, D.-A.; Smith, S.; Song, Y.; Stein, A.; Szegedy, M.; Teets, F. D.; Thyme, S. B.; Wang, R. Y.-R.; Watkins, A.; Zimmerman, L.; Bonneau, R. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods* **2020**, *17* (7), 665−680.
(46) Yekeen, A. A.; Zhang, L.; Liu, H.; Chen, Q. Simultaneous positive and negative selection of proteases in bacterium based on cell suicide and antibiotic resistance. *Biotechnol. J.* **2023**, *18* (6), 2200488.
(47) Bennett, N. R.; Coventry, B.; Goreshnik, I.; Huang, B.; Allen, A.; Vafeados, D.; Peng, Y. P.; Dauparas, J.; Baek, M.; Stewart, L.; et al. Improving de novo protein binder design with deep learning. *Nat. Commun.* **2023**, *14* (1), 2625.
(48) Mirdita, M.; Schutze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: making protein folding accessible to all. *Nat. Methods* **2022**, *19* (6), 679−682.
(49) Li, W.; Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22* (13), 1658−1659.
(50) Dunbrack, R. L., Jr. Rēs ipSAE loquuntur: What's wrong with AlphaFold's ipTM score and how to fix it. *bioRxiv*. 2025. .
(51) Raveh, B.; London, N.; Schueler-Furman, O. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins: Struct., Funct., Bioinf.* **2010**, *78* (9), 2029−2040.
(52) McDonald, I. K.; Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **1994**, *238* (5), 777−793.
(53) Du, G.; Rao, S.; Gurbani, D.; Henning, N. J.; Jiang, J.; Che, J.; Yang, A.; Ficarro, S. B.; Marto, J. A.; Aguirre, A. J.; Sorger, P. K.; Westover, K. D.; Zhang, T.; Gray, N. S. Structure-Based Design of a Potent and Selective Covalent Inhibitor for SRC Kinase That Targets a P-Loop Cysteine. *J. Med. Chem.* **2020**, *63* (4), 1624−1641.
(54) Gurbani, D.; Du, G.; Henning, N. J.; Rao, S.; Bera, A. K.; Zhang, T.; Gray, N. S.; Westover, K. D. Structure and Characterization of a Covalent Inhibitor of Src Kinase. *Front. Mol. Biosci.* **2020**, *7*, 2020.
(55) Schmitz, R.; Baumann, G.; Gram, H. Catalytic Specificity of Phosphotyrosine Kinases Blk, Lyn, c-Src and Syk as Assessed by Phage Display. *J. Mol. Biol.* **1996**, *260* (5), 664−677.
(56) Songyang, Z.; Carraway, K. L.; Eck, M. J.; Harrison, S. C.; Feldman, R. A.; Mohammadi, M.; Schlessinger, J.; Hubbard, S. R.; Smith, D. P.; Eng, C.; Lorenzo, M. J.; Ponder, B. A. J.; Mayer, B. J.; Cantley, L. C. Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. *Nature* **1995**, *373* (6514), 536−539.
(57) Eickhoff, J.; Choidas, A.Screening for Kinase Inhibitors: From Biochemical to Cellular Assays*Protein Kinases As Drug Targets*Wiley Online Library201145−67
(58) Overath, M. D.; Rygaard, A. S. H.; Jacobsen, C. P.; Brasas, V.; Morell, O.; Sormanni, P.; Jenkins, T. P. Predicting Experimental

https://doi.org/10.1021/acs.jcim.5c02430
J. Chem. Inf. Model. 2026, 66, 2041−2054

Success in De Novo Binder Design: A Meta-Analysis of 3,766 Experimentally Characterised Binders. *bioRxiv* **2025**.
(59) Ahad, A.; Leng, F.; Ichise, H.; Schrom, E.; So, J. Y.; Sellner, C.; Gu, Y.; Wang, W.; Kieu, C.; Park, W. Y. Immune Niche Formation in Engineered Mouse Models Reveals Mechanisms of Tumor Dormancy. *bioRxiv*. 2025. .
(60) Chow, A.; Chu, H.; Li, R.; Nalbant, B. N.; Dozic, A. V.; Kida, L. C.; Lareau, C. A. Sequence and structural determinants of efficacious de novo chimeric antigen receptors. *bioRxiv*. 2025.
(61) Sanghera, J.; Li, R.; Yan, J. Comparison of the Luminescent ADP-Glo Assay to a Standard Radiometric Assay for Measurement of Protein Kinase Activity. *Assay Drug Dev. Technol.* **2009**, *7* (6), 615−622.
(62) Takagi, T.; Shum, D.; Parisi, M.; Santos, R. E.; Radu, C.; Calder, P.; Rizvi, Z.; Frattini, M. G.; Djaballah, H. Comparison of Luminescence ADP Production Assay and Radiometric Scintillation Proximity Assay for Cdc7 Kinase. *Comb. Chem. High Throughput Screening* **2011**, *14* (8), 669−687.
(63) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493−500.
(64) Passaro, S.; Corso, G.; Wohlwend, J.; Reveiz, M.; Thaler, S.; Somnath, V. R.; Getz, N.; Portnoi, T.; Roy, J.; Stark, H.; et al. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. *bioRxiv* **2025**.
(65) ByteDance AML AI4Science TeamChen, X.; Zhang, Y.; Lu, C.; Ma, W.; Guan, J.; Gong, C.; Yang, J.; Zhang, H.; Zhang, K. et al. Protenix - Advancing Structure Prediction Through a Comprehensive AlphaFold3 Reproduction. *bioRxiv*. 2025. .
(66) Chai DiscoveryBoitreaud, J.; Dent, J.; McPartlon, M.; Meier, J.; Reis, V.; Rogozhnikov, A.; Wu, K. Chai-1: Decoding the molecular interactions of life. *bioRxiv*. 2024. .
(67) Dauparas, J.; Lee, G. R.; Pecoraro, R.; An, L.; Anishchenko, I.; Glasscock, C.; Baker, D. Atomic context-conditioned protein sequence design using LigandMPNN. *Nat. Methods* **2025**, *22* (4), 717−723.
(68) Varga, J. K.; Ovchinnikov, S.; Schueler-Furman, O. actifpTM: a refined confidence metric of AlphaFold2 predictions involving flexible regions. *Bioinformatics* **2025**, *41* (3), btaf107.
(69) Yekeen, A. A.; Durojaye, O. A.; Idris, M. O.; Muritala, H. F.; Arise, R. O. CHAPERONg: A tool for automated GROMACS-based molecular dynamics simulations and trajectory analyses. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 4849−4858.
(70) Harmalkar, A.; Lyskov, S.; Gray, J. J. Reliable protein-protein docking with AlphaFold, Rosetta, and replica-exchange. *eLife* **2025**, RP94029.
(71) Sumida, K. H.; Núñez-Franco, R.; Kalvet, I.; Pellock, S. J.; Wicky, B. I. M.; Milles, L. F.; Dauparas, J.; Wang, J.; Kipnis, Y.; Jameson, N.; Kang, A.; De La Cruz, J.; Sankaran, B.; Bera, A. K.; Jiménez-Osés, G.; Baker, D. Improving Protein Expression, Stability, and Function with ProteinMPNN. *J. Am. Chem. Soc.* **2024**, *146* (3), 2054−2061.
(72) Ansari, M.; White, A. D. Serverless Prediction of Peptide Properties with Recurrent Neural Networks. *bioRxiv* **2022**.
(73) Guntuboina, C.; Das, A.; Mollaei, P.; Kim, S.; Barati Farimani, A. PeptideBERT: A Language Model Based on Transformers for Peptide Property Prediction. *J. Phys. Chem. Lett.* **2023**, *14* (46), 10427−10434.
(74) Tian, C.; Yang, J.; Liu, C.; Chen, P.; Zhang, T.; Men, Y.; Ma, H.; Sun, Y.; Ma, Y. Engineering substrate specificity of HAD phosphatases and multienzyme systems development for the thermodynamic-driven manufacturing sugars. *Nat. Commun.* **2022**, *13* (1), 3582.
(75) Lu, J.; Lv, X.; Yu, W.; Zhang, J.; Lu, J.; Liu, Y.; Li, J.; Du, G.; Chen, J.; Liu, L. Reshaping Phosphatase Substrate Preference for Controlled Biosynthesis Using a "Design−Build−Test−Learn" Framework. *Adv. Sci.* **2024**, *11* (22), 2309852.

https://doi.org/10.1021/acs.jcim.5c02430
J. Chem. Inf. Model. 2026, 66, 2041−2054