BMC Methods

METHODOLOGY

Open Access

De Novo Rational design of peptide-based covalent inhibitors via mapping of complementary binding site residues – technical protocol and case study on KRAS^{G12C} and BTK^{481C}

Ernest Oduro-Kwateng[1], Musab Ali[1], Ibrahim Oluwatobi Kehinde[1], Li Rao[2] and Mahmoud E. S. Soliman[1*]

Abstract

Background Targeting undruggable proteins and challenging binding sites, such as protein–protein interaction (PPI) interfaces and allosteric pockets, using small-molecule inhibitors is often infeasible. Peptide-based irreversible inhibitors are promising emerging strategies for the treatment of such targets. However, there is currently no systematic in silico protocol for the rational design of peptide-based covalent inhibitors.

Methods We developed a streamlined computational framework for the de novo design of peptide-based irreversible inhibitors. Key considerations include peptide sequence optimization for binding, selection of electrophilic warheads, peptide folding, target specificity, and pharmacokinetic and toxicity profiles. Binding affinities were estimated using covalent molecular dynamics (MD^{cov}) simulations and thermodynamic binding free energy calculations.

Results Using KRAS^{G12C}, a strategic drug target traditionally considered undruggable, as a case study, the protocol identified top-hit peptide inhibitors (RVKDX, HVKXR, and XLKDH) with binding free energies (BFEs) of -48.84, -48.93, and -48.67 kcal/mol, respectively. These values are comparable to sotorasib (-50.63 kcal/mol) and lower than adagrasib (-71.73 kcal/mol), both FDA-approved KRAS^{G12C} inhibitors. Benchmarking against BTK^{481C} using zanubrutinib, an FDA-approved therapeutic agent for B-cell malignancies, further validated the protocol. Peptide inhibitors XDYMA, XDYVL, and QDWXL demonstrated BFEs of -83.40, -76.69, and -62.40 kcal/mol, outperforming zanubrutinib (-57.00 kcal/mol), acalabrutinib (-54.19 kcal/mol), and ibrutinib (-55.09 kcal/mol).

Discussion These findings underscore the robustness and adaptability of our protocol, offering a systematic, multifaceted approach that can be integrated into drug discovery workflows to design novel peptide-based irreversible inhibitors.

Keywords Peptide-based Covalent Inhibitors, Peptide Design, Binding Free Energy Calculations, KRAS^{G12C}, BTK^{481C}

*Correspondence:
Mahmoud E. S. Soliman
soliman@ukzn.ac.za
Full list of author information is available at the end of the article

Introduction

The history of peptide therapeutics in human diseases shows a progressive evolution from the utilization of natural peptides to the current era of sophisticated synthetic and engineered peptides [1]. While small-molecule inhibitors (SMIs) are commonly used against protein targets, their small and compact configuration, flatness, and lack of natural substrates make certain targets, like protein–protein interactions (PPIs), narrow allosteric pockets, transcription factors, and DNA-binding proteins, difficult to reach [2]. Peptide inhibitors (PIs) offer a larger interaction surface compared to SMIs, enabling them to engage in extensive interactions with target proteins and form multiple specific contacts, which enhances their binding affinity and selectivity. Moreover, the structural flexibility of peptides allows them to adopt various conformations, such as α-helices or β-sheets, to fit into the target binding site. Additionally, PIs can be tailored to interact with precise "hot spots" regions on target proteins, reducing off-target effects commonly associated with SMIs [3]. Peptide inhibitors have shown promise in various fields, including cancer therapy, infectious diseases, and autoimmune disorders to disrupt PPIs, inhibit enzymatic activity, or interfere with specific cellular pathways. Despite their potential, challenges such as stability, delivery, and cost-effectiveness remain hurdles to overcome. Advances in computational design and peptide engineering are continuously improving the development and optimization of PIs for improved cellular uptake and distribution, enhancing their bioavailability, permeability, stability, and therapeutic potential in vivo [4].

Covalent inhibition involves the formation of a covalent bond between the inhibitor and its target protein, leading to irreversible inhibition of the target (Fig. 1). This approach offers several advantages, including enhanced potency, prolonged target residence time, and the ability to target proteins that are traditionally considered undruggable [2]. There has been a resurgence of interest in covalent inhibitors, particularly bifunctional molecules, which consist of a binding scaffold decorated with a protein-reactive functional group. These targeted covalent inhibitors can selectively modify non-catalytic nucleophilic residues near the binding site, reducing off-target effects [5]. Peptide-based irreversible/covalent inhibition utilizes peptides as the scaffold for designing covalent inhibitors that irreversibly bind to target proteins. By incorporating covalent modifiers into the peptide sequence, these inhibitors can form stable covalent bonds with specific electrophilic residues on the target protein (Fig. 2). The design of peptide-based covalent inhibitors often involves a structure-based approach to identify suitable nucleophilic residues on the target protein and guide the incorporation of covalent modifiers [3]. Additionally, computational tools, like Rosetta [6] and Cov_DOX [7], can aid in the design and optimization of peptide-based covalent inhibitors, ensuring their efficacy and selectivity in disrupting critical protein interactions.

The de novo design of covalent-based peptide inhibitors presents a spectrum of challenges, spanning from selecting the reactive warhead and crafting the peptide sequence to achieving selectivity while limiting off-target effects [3]. Furthermore, considerations of peptide flexibility and folding are integral to this endeavour. Effectively addressing these challenges necessitates synergistic insights from computational tools, structural biology, and experimental validation [4, 7].

The choice of peptide sequence is crucial as it determines the binding affinity, specificity, and stability of the PI. One approach to addressing this challenge is to

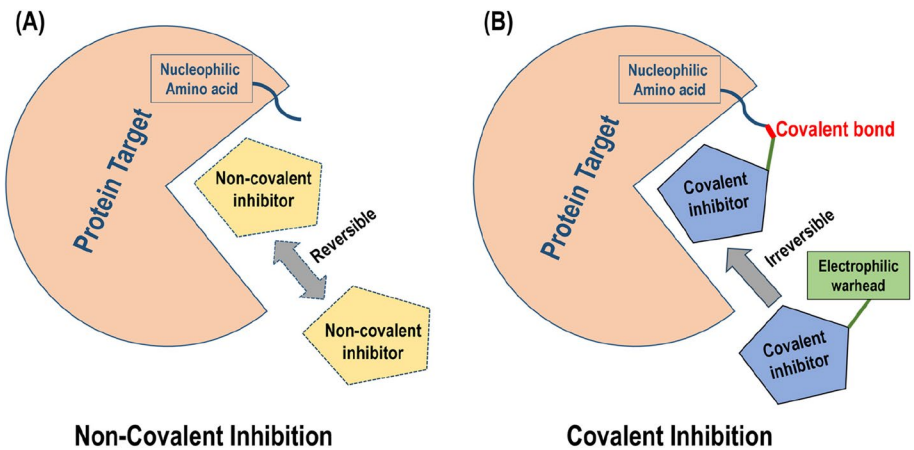


Fig. 1 Graphical illustration of non-covalent vs covalent inhibition in drug design

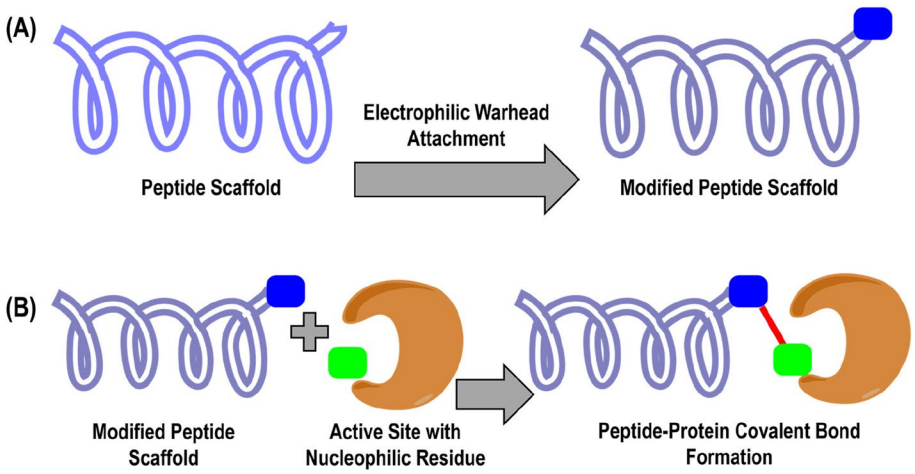


Fig. 2 Graphical illustration of the design of peptide-based covalent inhibitors. **(A)** Incorporation of a reactive warhead (blue) in the peptide scaffold (light purple) enables **(B)** the formation of covalent bond (red) between the modified peptide scaffold (dark purple) and the nucleophilic residue (green) near the active site (brown) of the protein target

leverage structural information on the target protein to identify key binding epitopes that can be mimicked by the peptide sequence [8]. Moreover, to further optimize the selection of the peptide sequence, sequence sampling strategies can be employed to produce diverse library of peptide variants for screening. The ideal peptide sequence may be amplified by predicting the various possibilities of combinations through the replacement of an amino acid residue of the same side chain biochemical property – acidic, basic, hydrophobic and aliphatic, aromatic, polar but uncharged, and unique (Fig. 3) [9]. Other strategies include random mutagenesis for exploring diverse mutations [10, 11], iterative design for continuous improvement [12], and focused libraries for screening peptide sequences [13], among others, allowing the systematic exploration of sequence space.

Furthermore, the choice of warhead influences the selectivity, potency, and reactivity of the PI with the target protein while minimizing off-target effects [5]. Unlike small molecule covalent inhibitors that have successfully bind numerous nucleophiles of targeted proteins, PIs have been shown to successfully bind only cysteine, lysine, and histidine residues [3]. Common examples of warheads include acrylamides, thioesters, alkynyl benzaldehydes, sulfanamides, and boronate-based reagents, among others (Fig. 4). One approach is through structure-based design, which involves identifying and characterizing the binding site of the target protein to guide the selection of an appropriate reactive warhead for the peptide inhibitor. Computational tools can also be employed to predict the reactivity and selectivity of different warheads [6]. Additionally, experimental screening of various

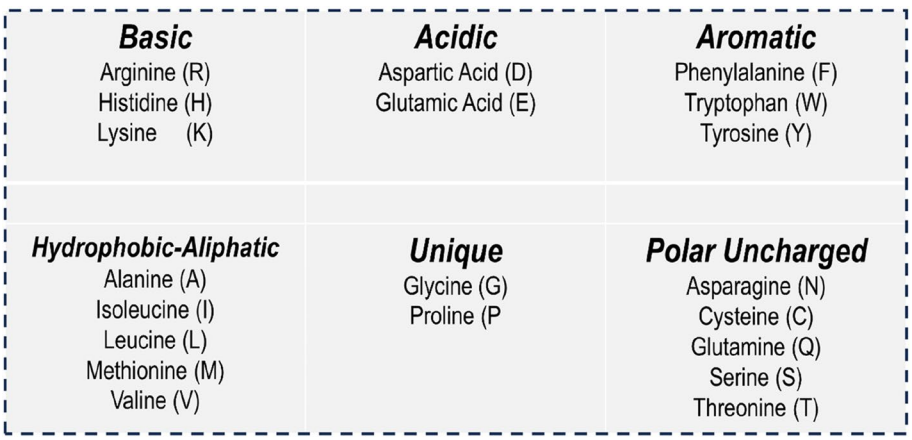


Fig. 3 Classification of the 20 amino acid residues based on shared physiochemical features of the side chain group

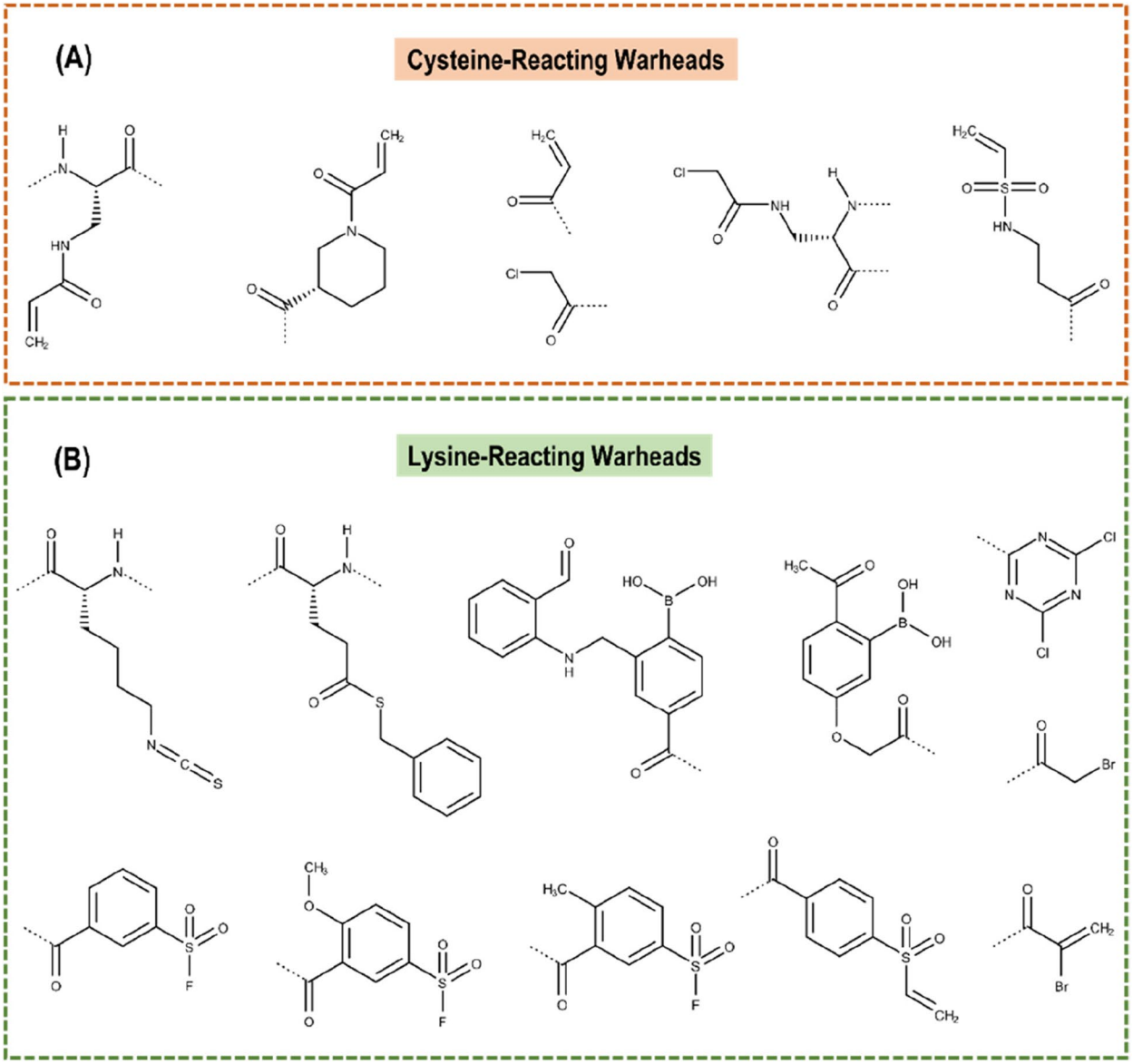


Fig. 4 2D Chemical structures of **(A)** cysteine-reacting and **(B)** lysine-reacting warheads commonly employed in the rational design of covalent-based peptide inhibitors. Dashed lines represent attachment bonds with the peptide tail

reactive warheads and modifications to the peptide scaffold can help fine-tune the reactivity profile of the inhibitor [14].

Peptides are inherently flexible and assume structural conformations, which may significantly impact their stability, conformational dynamics, and functional properties. Molecular dynamics and mechanics simulations can be used to predict the conformational dynamics and folding pathways of peptides [15]. Machine learning algorithms have been used to develop scoring functions that learn from large datasets of peptide structures and their associated properties. These algorithms can identify patterns and correlations in the data that can be used to predict the three-dimensional conformation and properties of peptide sequences [16]. Scoring functions have bene improved by incorporating advanced energy terms that consider various aspects of peptide stability, such as electrostatics, hydrogen bonding, van der Waals interactions, and solvation effects. These energy terms can be derived from experimental data or calculated using computational methods [17]. Secondly, scoring functions can be tailored to consider specific constraints relevant to peptide design, such as disulfide bond formation, backbone rigidity, and side-chain interactions [18].

To the best of our knowledge, there is currently no stand-alone computational protocol for the de novo design and optimization of covalent-based PIs. We applied this protocol to a case study aimed at optimizing a lead covalent-based PI against the $KRAS^{G12C}$ mutation by leveraging peptide sequence complementary to key interacting residues of $KRAS^{G12C}$ (PDB: 6OIM) [19]. The KRAS protein functions as a molecular switch in growth factor signaling pathways, regulating cell proliferation by cycling between an inactive GDP-bound form and an active GTP-bound form to engage downstream effector proteins. Mutations in the KRAS gene, such as the ubiquitous G12C mutation, disrupt this regulatory cycling, leading to persistent activation of pro-proliferative signaling pathways in cancer cells. Sotorasib, an FDA-approved covalent SMI, is used to treat various cancers, including non-small cell lung cancer (NSCLC), colorectal cancer, and other solid tumors mediated by the challenging and previously undruggable $KRAS^{G12C}$ mutation. The limitation of sotorasib against $KRAS^{G12C}$, lies in its potential for developing resistance over time due to the emergence of secondary mutations in the KRAS protein [20]. Thus, we propose PIs as alternative therapeutics against KRAS protein owing to their well-established benefits over SMIs.

Multiple peptide combination sequences were generated from the complementary peptide sequence and subjected through screening for their toxicity and physiochemical characteristics. Top-screened peptide sequences were investigated for proper folding and modelled for covalent attachment to C12 after warhead incorporation. Binding free energetics and thermodynamic profiles of the lead PI were analysed following MD^{cov} simulations. This approach ensures a systematic design, optimization, and evaluation of a PI with potential therapeutic efficacy against $KRAS^{G12C}$. We have also incorporated analyses on adagrasib (PDB: 6UT0) (21), a more selective and potent covalent inhibitor targeting $KRAS^{G12C}$, to provide a more comprehensive benchmarking of our designed peptide inhibitors.

To further confirm the versatility of the proposed protocol, we applied the workflow to Bruton's Tyrosine Kinase (BTK), an essential target for B-cell development and activation, using zanubrutinib (PDB: 6J6M) [22] as the primary reference covalent inhibitor. Zanubrutinib is highly potent and selective against BTK^{481C} as compared to previous analogues (acalabrutinib and ibrutinib), implicated in B-cell malignancies, including, Chronic Lymphocytic Leukemia (CLL), Mantle Cell Lymphoma (MCL), and Waldenström Macroglobulinemia (WM), and Diffuse Large B-Cell Lymphoma (DLBCL). The target sequence (T-E-Y-M-A) was selected based on the functional role of the critical residues, E475 and M477, which stabilize the kinase domain and facilitate ATP binding, and T474, the gatekeeper residue involved in inhibitor specificity. The peptides generated using this sequence exhibited superior binding free energies relative to zanubrutinib, acalabrutinib, and ibrutinib. Overall, our results highlight the therapeutic potential of the peptide candidates and emphasize the adaptability of our protocol while underscoring the need for experimental validation.

Methods

Our protocol provides a systemic guide for the computational de novo design of peptide-based covalent inhibitors (Fig. 5). Various in silico tools and algorithms have been referenced.

Mapping of Peptide Sequence Complementary to Hotspot Residues

The critical binding residues H95, Y96, and Q99 of $KRAS^{G12C}$ are crucial for interacting with sotorasib [20]. To maintain peptide connectivity, we included residues R97 and E98. Based on the target sequence H-Y-R-E-Q, we proposed a complementary sequence H-A-R-E-K [23]. Each residue in this complementary sequence was selected for its specific properties to enhance binding affinity and stability. 'H' mimics the histidine residue in the target sequence, facilitating specific interactions. 'A' provides a neutral residue that contributes to maintaining the flexibility of the peptide. 'R' corresponds to the arginine residue in the target sequence, capable of interacting through potential hydrogen bonds or electrostatic interactions. 'E' mimics the glutamic acid residue in the target sequence, ensuring proper charge complementarity. 'K' introduces a positively charged residue to complement the polar glutamine and negatively charged glutamic acid in the target sequence, enhancing electrostatic interactions.

For the BTK^{481C} case study, we applied the same systematic workflow used for $KRAS^{G12C}$, utilizing T-E-Y-M-A as the template to generate peptide inhibitors for BTK. The active-site residues T474, E475, and M477 were identified as critical for zanubrutinib binding, with Y476 and A478 included to maintain structural connectivity.

Generation of Peptide Combination Sequences

Each of the five residues in the template peptide, H-A-R-E-K, was alternately substituted with residues from the same group to generate a diverse set of peptide sequences. The groups included basic ('R', 'H', and 'K'), hydrophobic-aliphatic ('A', 'I', 'L', 'M', and 'V'), or acidic ('D' and 'E') residues. Similarly, each residue of T-E-Y-M-A was alternately substituted with a residue from the same group, namely polar uncharged ('N', 'C', 'Q', 'S', 'T'), acidic,

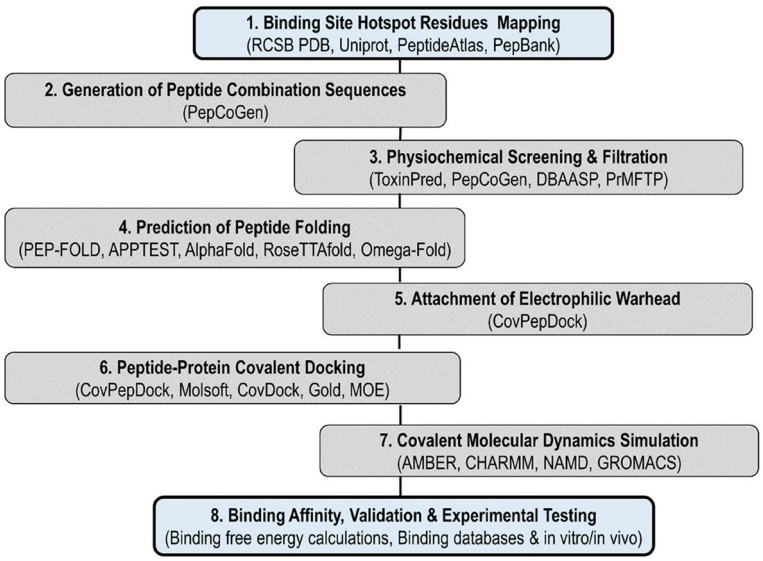


Fig. 5 Systemic computational workflow for the de novo design of peptide-based covalent inhibitors

aromatic ('F', 'W', 'Y'), and hydrophobic-aliphatic. The substitutions were carried out using the Peptide Combination Generator (PepCoGen) (https://pepcogen.bicfri.in/) [9].

Physicochemical Screening of Peptide Sequences

Toxicity and Physicochemical Characteristics Screening

The peptide sequences were screened for toxicity and physicochemical characteristics using ToxinPred (https://webs.iiitd.edu.in/raghava/toxinpred/), a tool that employs a Support Vector Machine (SVM) algorithm [10]. This tool is trained using diverse peptide sequences to predict peptide toxicity and physicochemical properties based on features such as amino acid composition (AAC), dipeptide composition (DPC), presence of motifs, and quantitative matrix (QM) models. Higher SVM scores suggest a higher likelihood of the peptide being a toxin, whereas lower scores indicate a lower likelihood. The SVM algorithm in ToxinPred encodes information regarding the physicochemical characteristics of peptide sequences, including hydrophobicity, amphipathicity, isoelectric point, and molecular weight [24]. The training data for the SVM models in ToxinPred consists of peptide sequences labeled as toxic or non-toxic, along with their corresponding physicochemical property values. These sequences are derived from experimental data with features such as AAC, DPC, motifs, and QM values, which are used as input variables for training the SVM model. SVM is used to build classification models that can distinguish between toxic and non-toxic peptides, and regression models that can predict continuous physicochemical properties to find the optimal hyperplane that best fits the training data. Performance metrics such as sensitivity, specificity, accuracy, Matthews correlation coefficient (MCC), and area under the receiver operating characteristic curve (AUC-ROC) have been used to validate the SVM model in ToxinPred [10].

Aggregation Propensity Screening

Peptide sequences were screened for in vivo aggregation propensity using the normalized a4v Sum Score (Na4vSS) score employed by AGGRESCAN (http://bioinf.uab.es/aggrescan/). This score is calculated using an aggregation propensity scale, identifying aggregation-prone protein segments based on validated in vivo experimental data [25]. In the AGGRESCAN algorithm, aggregation-prone segments within a peptide sequence are first identified using an aggregation propensity scale and validated against aggregation-promoting regions in experimental data. Subsequently, the evaluated aggregation-prone segments are ranked based on peak area or normalized peak area, providing a quantitative assessment of their contribution to peptide aggregation.

Modeling of Peptide Fold

For this study, PEP-FOLD3 (http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3), an advanced de novo peptide structure prediction tool was used for modeling the conformational folds of the peptide sequences [26]. PEP-FOLD3 utilizes a combination of innovative algorithms and procedures to model peptide structures accurately. Hidden Markov Model (HMM) for sub-optimal conformation sampling, enabling the prediction of various possible peptide conformations. Support Vector

Machine (SVM) Algorithm for sampling peptide fragments, ensuring the selection of fragments with the most probable conformations. Taboo Sampling Algorithm for constructing prototype conformations by avoiding previously explored conformations, enhancing the diversity of the predicted structures. Monte Carlo Procedure for refining the built conformations by exploring the conformational space more thoroughly, optimizing the peptide structures. Oscar-star for incorporating side chains into the peptide models, improving the accuracy of the spatial arrangement, and Gromacs 5 for rapid minimization of backbone geometry, ensuring the stability and realism of the predicted structures. The resulting peptide structures were then clustered based on the BCscore, which measures the structural similarity among peptides. The clusters are subsequently sorted using the sOPEP coarse-grained force field, which ranks the peptide conformations based on their potential energy, ensuring the identification of the most stable and functionally relevant structures [26]. The use of the sOPEP force field coarse-grained model can reduce the complexity of the modeling process, making it more computationally efficient and allowing for quicker exploration of the conformational space. Owing to the simplifications made in the force field, the coarse-grained model may not always be as precise as the atomistic model in predicting the detailed atomic interactions. This can potentially result in inaccuracies in capturing the fine structural features and interactions present in the atomistic models [27].

Modeling of Peptide-Based Covalent Inhibitor

To design irreversible covalent electrophilic peptide binders against $KRAS^{G12C}$ and BTK^{481C}, we utilized Rosetta CovPepDock (https://rosie.graylab.jhu.edu/cov_pep_dock). To the best of our knowledge, CovPecDock is the only computational tool that enables the modeling of peptide-based covalent inhibitors. This approach is tailored for designing peptides with electrophilic warheads, such as acrylamide and chloroacetamide, known for their mild reactivity that primarily enhances binding through recognition rather than reactivity, thus minimizing off-target effects in vitro [6].

Selection and Incorporation of Electrophilic Warhead

An electrophilic warhead suitable for peptide sequences was selected. This warhead was chosen to ensure effective covalent bond formation with the cysteine residue at position 12 (C12) and 481 (C481) of $KRAS^{G12C}$ and BTK^{481C}, respectively. The protocol identified peptide positions close to the reactive cysteine, typically with a Ca-Sg distance constraint of < 10 Å, crucial for modeling the covalent bond.

Covalent Peptide-Protein Docking

The designed electrophilic peptides were subjected to docking simulations using the Rosetta software suite. During these simulations, distance constraints were applied to guide the formation of the covalent bond between the electrophilic warhead and C12 and C481. Iterative cycles of rigid-body optimization and flexible backbone sampling were performed to explore the conformational space and optimize interactions with $KRAS^{G12C}$ and BTK^{481C}. The generated docking poses were scored based on energy minimization and validated against experimental data, typically from X-ray crystallography. Covalent constraint scores were utilized to assess the quality of the predicted covalent bonds. The model with the lowest covalent constraint score for each electrophilic peptide-protein complex was selected for further analysis. In addition, interface scores were evaluated to gauge the contributions of interface segments to the overall binding affinity.

Covalent Molecular Dynamics (MD) Simulation

Covalent MD (MD^{cov}) simulations were conducted using the AMBER 18 Particle Mesh Ewald Molecular Dynamics (PMEMD) Compute Unified Device Architecture (PMEMD.CUDA) central processing unit (CPU) package [28]. The systems were prepared using UCSF chimera and subjected to MD simulations for 200 ns with a 2-fs time step in an isothermal-isobaric (NPT) ensemble at 300 K and 1 bar pressure. Calculations of root mean square deviation (RMSD) and root mean square fluctuation (RMSF) were carried out using the CPPTRAJ and PTRAJ modules of AMBER 18 [29]. The detailed methodology of the (MD^{cov}) simulations can be referred to from our previous technical protocol [15].

Computation of Binding Free Energy (BFE)

The molecular mechanics/generalized Born surface area (MM/GBSA) method was applied to evaluate the binding free energy (BFE) between the protein targets and the covalent peptide inhibitors in comparison to the reference covalent inhibitors. This method is used to predict and analyze the binding energetics of molecular interactions in biological systems, providing a more detailed description of BFE than the scoring energetics used in docking studies. MM/GBSA integrates molecular mechanics calculations, which employ classical force fields to describe molecular interactions, with the generalized Born (GB) dielectric continuum solvent model and surface area (SA) terms to estimate BFE. Molecular mechanics calculations encompass internal energy, van der Waals interactions, electrostatic interactions, and other molecular forces. The GB model estimates polar

solvation free energy by considering the Born radii of the atoms and their pairwise interactions. The SA method quantifies changes in hydrophobic interactions upon binding by computing the buried surface area (BSA) during complex formation, correlating the nonpolar solvation energy with the peptide-protein–protein interface surface area using a water probe radius of 1.4 Å and surface tension constant (γ) set at 0.0072 kcal/mol Å^2 [30]. The BFE calculations were performed using 50,000 MD trajectory frames to capture detailed binding energetics. The formula for BFE (ΔG) is as follows:

$$\Delta G_{bind} = G_{complex} - G_{receptor} - G_{ligand}$$
$$\Delta G_{bind} = E_{gas} + G_{sol} - T\Delta S$$
$$E_{gas} = E_{int} + E_{vdw} + E_{ele}$$
$$G_{sol} = G_{GB} + G_{SA}$$
$$G_{SA} = \gamma SASA$$

where ΔG_{bind} represents the gas-phase summation, E_{gas} is the gas-phase energy, G_{sol} is the free solvation energy, TΔS is the total interaction entropy, E_{int} is the internal energy, E_{ele} is the Coulomb energy, and E_{vdw} is the van der Waals energy. E_{gas} was computed from the AMBER FF14SB force field, and G_{sol} was calculated from the energy contributions of polar and non-polar states.

Results

Computational Screening and Optimization of Complementary Peptide Sequence

A peptide library of 270 peptide sequences was generated from H-A-R-E-K by substituting one residue at each position for another belonging to the same physiochemical class. From this library, 60 peptide sequences were selected based on the criterion of not having the same recurring residue. Screening using the Support Vector Machine (SVM) algorithm on ToxinPred predicted that none of the 60 peptide sequences were toxic, with SVM scores ranging from -1.13 to -0.46 (Table S1). Peptide sequences with SVM scores < -0.8 were eliminated, resulting in 42 peptide sequences (Table S2). Screening based on amphipathicity retained 21 peptide sequences with higher indices of 1.51, while eliminating peptide sequences with lower indices of 1.77 (Table S3). Screening for hydrophobicity identified 10 peptide sequences with hydrophobic indices of -0.69. Peptide sequences with higher hydrophobic indices of -0.74 to -075 and those with lower hydrophobic indices of -0.65 were eliminated (Table S4). The top 10 peptide sequences had predicted isoelectric points (PI) of 9.1 and molecular weights ranging from 653.80 to 667.83 Da. For AGGRESCAN analysis, the peptide sequences had Normalized a4v Sum Score (Na4vSS) range of -72.4 to -79.2, and the 4 sequences with higher scores (> -77.5) were eliminated, resulting in top 6 peptide sequences (Table S5).

Refer to Tables S 7–S9 for the screening and optimization of peptide sequences from BTK^{481C}.

Peptide Structural Conformation Modeling

The sOPEP coarse-grained force field was used to predict the conformational folding of the 6 and 8 top-screened peptide sequences generated from KRAS^{G12C} and BTK^{481C}, respectively (Fig. 6). Low sOPEP energy values were observed (Table 1).

Covalent Peptide Inhibitor Modeling

Of the 6 peptide sequences, 3 of them were successfully modeled with electrophilic peptide binders targeting KRAS^{G12C}. The peptides demonstrated lower constraint and interface scores (Table 2).

Refer to Table S 10 for the characteristic features of the peptide sequences generated from BKT^{481C} after covalent modeling.

Protein–Ligand Interaction Framework Analysis

Comparative analysis of the protein–ligand interaction frameworks for the top-hit peptide inhibitors and the reference inhibitors targeting KRAS^{G12C} and BTK^{481C} revealed distinct interaction profiles that correlated with their binding affinities (Tables S6 and S11). For the binding landscape of KRAS^{G12C}, the peptide inhibitors engaged through a complex network of hydrogen bonds, electrostatic interactions, and pi/pi-alkyl contacts, supported by extensive van der Waals forces (Figs. 7A-C). In contrast, sotorasib predominantly relied on electrostatic pi-charge interactions with aromatic groups, complemented by mixed pi/alkyl hydrophobic interactions (Fig. 7D). Adagrasib, on the other hand, featured a unique fluorine-halogen interaction, alongside pi and mixed pi/alkyl hydrophobic contacts, largely attributed to its aromatic moieties (Fig. 7E). In parallel, the binding landscape of BTK^{481C} showed a diverse range of interactions, including hydrogen bonds, electrostatic forces, and hydrophobic contacts. Notably, the peptide inhibitors demonstrated a more robust hydrogen-bonding framework compared to zanubrutinib, acalabrutinib, and ibrutinib. The BTK^{481C} inhibitors engaged primarily via pi-alkyl bonds, especially for acalabrutinib (Fig. 8).

Structural and Flexible Conformational Changes of KRAS^{G12C} and BTK^{481C}

The post-molecular dynamics (MD) analysis focused on the evaluation of the structural stability and residue flexibility of KRAS^{G12C} and BTK^{481C} by calculating root mean square deviation (RMSD) and root mean square fluctuation (RMSF). The average RMSD for KRAS^{G12C} complexes with RVKDX, HVKXR, XLKDH, sotorasib, and adagrasib was 1.02 Å, 0.97 Å, 0.96 Å, 1.15 Å, and

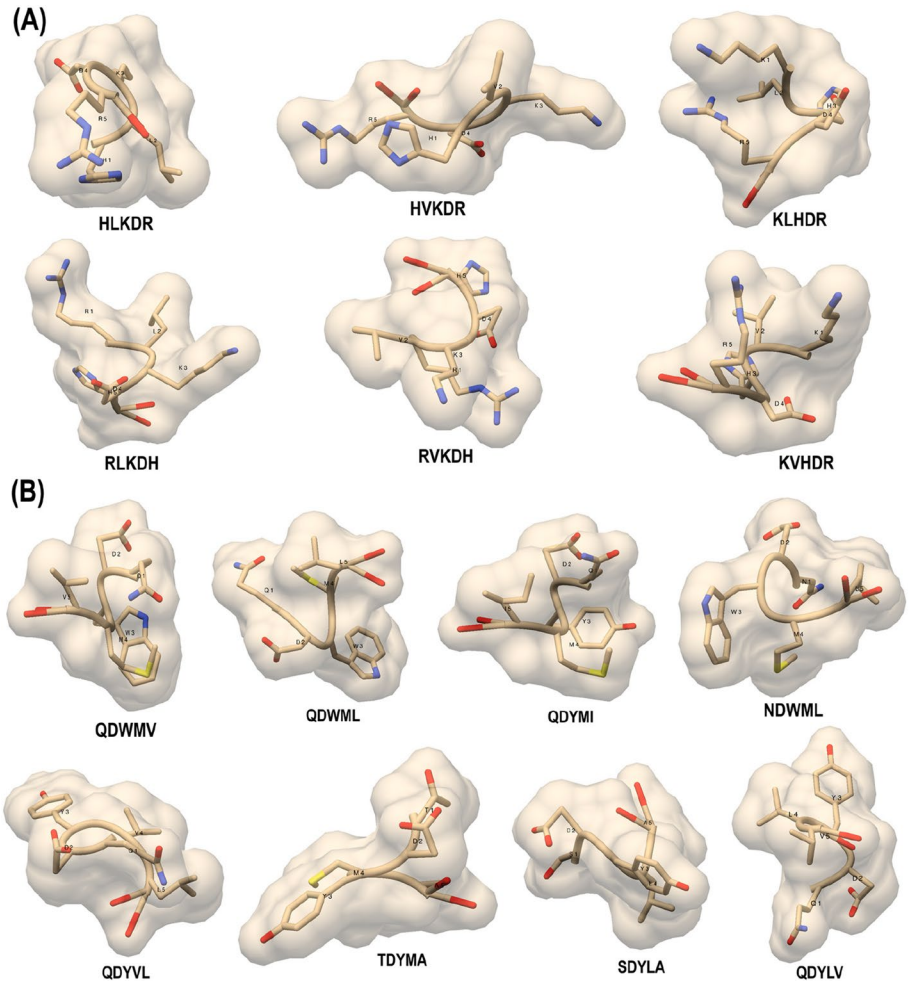


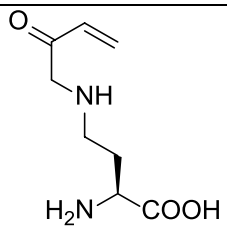
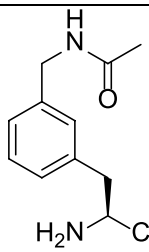
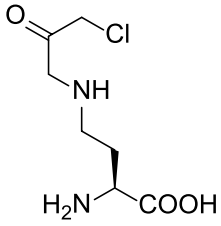
Fig. 6 3D solid structure of the conformational folds of the top screened peptide sequences generated from **(A)** $KRAS^{G12C}$ and **(B)** BTK^{481C}. Residues are colored by heteroatom

Table 1 Filtered peptide sequences generated from $KRAS^{G12C}$ and their predictive physiochemical characteristics, propensity of in vivo aggregation, and conformationally stable energy values

Peptide ID	Peptide Sequence	SVM Score	Toxicity prediction	Hydrophobicity	Amphipathicity	pI	Mol. Wt	Na4vSS	sOPEP Energy
18	RLKDH	-1.02	Non-Toxin	-0.69	1.51	9.10	667.83	-76.70	-1.91
51	KVHDR	-1.02	Non-Toxin	-0.69	1.51	9.10	653.80	-73.20	-2.17
40	HVKDR	-1.00	Non-Toxin	-0.69	1.51	9.10	653.80	-72.40	-1.54
17	KLHDR	-0.98	Non-Toxin	-0.69	1.51	9.10	667.83	-77.50	-2.19
52	HLKDR	-0.97	Non-Toxin	-0.69	1.51	9.10	667.83	-76.70	-1.55
13	RVKDH	-0.92	Non-Toxin	-0.69	1.51	9.10	653.80	-72.40	-1.78

0.87 Å, respectively (Fig. 9A). The corresponding average RMSF values were 0.66 Å, 0.68 Å, 0.63 Å, 0.74 Å, and 0.60 Å for $KRAS^{G12C}$ complexes with RVKDX, HVKXR, XLKDH, sotorasib, and adagrasib, respectively (Fig. 9B). For BTK^{481C} complexes, the average RMSD values for XDYMA, XDYVL, QDWXL, zanubrutinib, acalabrutinib, and ibrutinib were 1.06 Å, 1.05 Å, 1.14 Å, 1.00 Å, 1.28 Å, and 1.06 Å, respectively (Fig. 9C). The corresponding RMSF values were 0.75 Å, 0.79 Å, 0.74 Å, 0.70 Å, 0.78 Å, and 0.77 Å, respectively (Fig. 9D).

Table 2 Characteristic features of the peptide sequences generated from $KRAS^{G12C}$ after electrophilic warhead incorporation showing the position of modification, the chemical structure of warhead and constraint and restraint scores after covalent bond formation

Peptide Sequence	Modified Sequence	Electrophilic Warhead	Constraint Score	Interface Score
RVKDH	**RVKD[X]**		0.127	-10.232
HVKDR	**HVK[X]R**		0.045	-11.245
RLKDH	**[X]LKDH**		0.046	-17.472

X represents the peptide position mutated for warhead attachment.

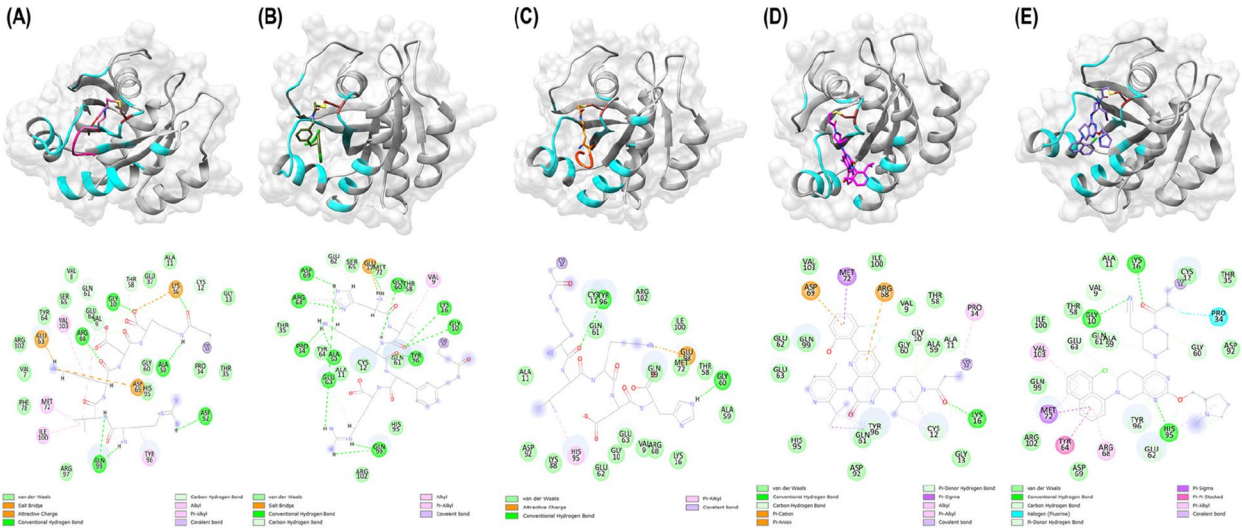


Fig. 7 3D solid structure of $KRAS^{G12C}$ (light gray) covalently bonded to **(A)** RVKDX (pink), **(B)** HVKXR (green), **(C)** XLKDH (orange), **(D)** sotorasib (magenta), and **(E)** adagrasib (purple) via the attached electrophilic warhead and at C12 (brown) showing interacting residues (cyan) at the cryptic binding pocket. This figure also depicts the 2D protein–ligand interaction frameworks of the complexes. Interaction types are shown

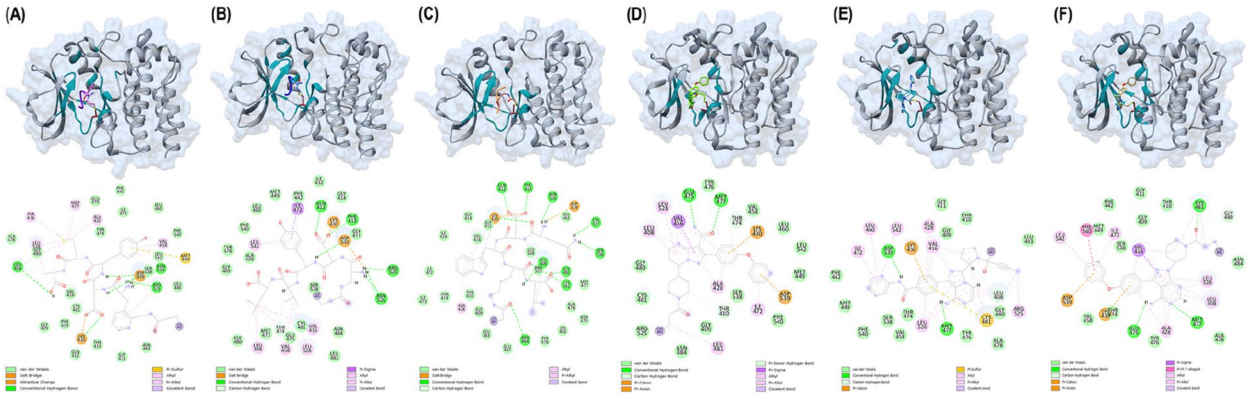


Fig. 8 3D solid structure of BTK^{481C} (light blue) covalently bonded to **(A)** XDYMA (orchid), **(B)** XDYVL (blue), **(C)** QDWXL (salmon), **(D)** zanubrutinib (chartreuse green), (E) acalabrutinib (sky blue), and (F) ibrutinib (dark khaki) via the attached electrophilic warhead and at C481 (brown) showing interacting residues (dark cyan) at the ATP binding pocket. This figure also depicts the 2D protein–ligand interaction frameworks of the complexes. Interaction types are shown

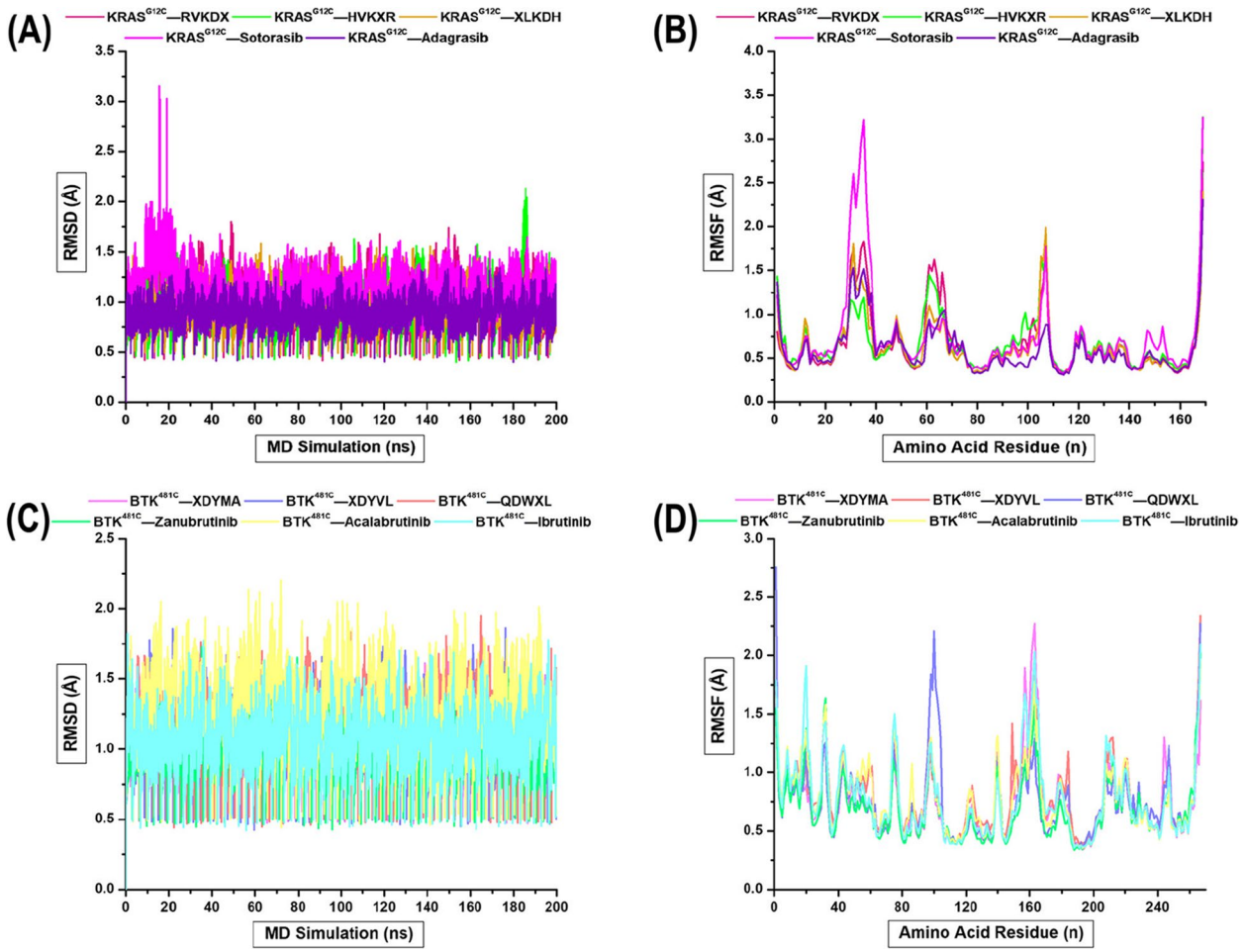


Fig. 9 Graphical plots of **(A)** RMSD and **(B)** RMSF of KRAS^{G12C}-RVKDX (pink), KRAS^{G12C}-HVKXR (green), KRAS^{G12C}-XLKDH (orange), KRAS^{G12C}-Sotorasib (magenta), and KRAS^{G12C}-Adagrasib (purple). Graphical plots of **(C)** RMSD and **(D)** RMSF of BTK^{481C}-XDYMA (orchid), BTK^{481C}-XDYVL (cornflower blue), BTK^{481C}-QDWXL (salmon), BTK^{481C}-Zanubrutinib (chartreuse green), BTK^{481C}-Acalabrutinib (yellow), and BTK^{481C}-Ibrutinib (cyan). Computations were made over 200 ns

Binding Free Energy (BFE) Calculation

The binding free energy (ΔG_{bind}) analysis for $KRAS^{G12C}$ inhibition shows that adagrasib demonstrated the strongest binding affinity at -71.73 ± 0.08 kcal/mol, with sotorasib following at -50.63 ± 0.07 kcal/mol. The peptide inhibitors RVKDX (-48.84 ± 0.13 kcal/mol), HVKXR (-48.93 ± 0.14 kcal/mol), and XLKDH (-48.67 ± 0.11 kcal/mol) showed comparable binding energies. Adagrasib and sotorasib exhibited more robust van der Waals interactions (ΔE_{vdW}), while HVKXR displayed a notably higher electrostatic energy contribution (ΔE_{elec}) (Table 3). For BTK^{481C} inhibition, peptide inhibitors demonstrated even greater binding affinities, with XDYMA achieving a remarkable ΔG_{bind} of -83.40 ± 0.15 kcal/mol, surpassing zanubrutinib (-57.00 ± 0.06 kcal/mol), acalabrutinib (-54.19 ± 0.07 kcal/mol), and ibrutinib (-55.09 ± 0.06 kcal/mol). XDYVL and QDWXL also showed strong binding free energies of -76.69 ± 0.14 kcal/mol and -62.40 ± 0.12 kcal/mol, respectively. All the peptide inhibitors exhibited highly favorable electrostatic (ΔE_{elec}) and van der Waals (ΔE_{vdW}) interactions (Table 4).

Discussion

The complementary sequence (H-A-R-E-K) was created to optimize interactions with the target residues (H-Y-R-E-Q) of $KRAS^{G12C}$. The generation of a peptide library from H-A-R-E-K ensured that the peptide sequences maintained their structural integrity and had specific drug-like characteristics, including amphipathicity, hydrophobicity, isoelectric point, and molecular weight (31). To balance structural stability, specificity, and computational feasibility, we opted for a 5-amino-acid peptide sequence. While longer peptides may enhance contact surface area, shorter peptides reduce the risk of proteolytic degradation, offer better pharmacokinetic properties, are easier to synthesize and modify, and maintain high binding affinity when optimally designed [32]. The Peptide Combination Generator (PepCoGen) offers a significant advantage in its capacity to generate up to 0.1 million peptide combination sequences tailored to specific physicochemical properties, enabling efficient

Table 3 Components of the free binding energetics between RVKDX, HVKXR, XLKDH, sotorasib, and adagrasib with $KRAS^{G12C}$ expressed in kcal/mol with standard error of means. Computations were made with 200,000 MD trajectory snapshots

Energy Component (kcal/mol)	$KRAS^{G12C}$-RVKDX	$KRAS^{G12C}$-HVKXR	$KRAS^{G12C}$-XLKDH	$KRAS^{G12C}$- Sotorasib	$KRAS^{G12C}$- Adagrasib
ΔE_{vdW}	-49.67 ± 0.09	-39.07 ± 0.11	-42.08 ± 0.09	-61.65 ± 0.06	-76.29 ± 0.06
ΔE_{elec}	-255.42 ± 0.62	-550.00 ± 0.85	-187.97 ± 0.57	-30.97 ± 0.09	-200.69 ± 0.25
ΔG_{GB}	264.39 ± 0.54	546.93 ± 0.75	187.04 ± 0.51	49.60 ± 0.09	214.01 ± 0.24
ΔG_{SA}	-8.15 ± 0.01	-6.79 ± 0.01	-5.66 ± 0.01	-7.61 ± 0.01	-8.77 ± 0.01
ΔG_{gas}	-305.09 ± 0.62	-589.07 ± 0.81	-230.05 ± 0.58	-92.61 ± 0.10	-276.98 ± 0.26
ΔG_{solv}	256.24 ± 0.54	540.14 ± 0.75	181.38 ± 0.50	41.99 ± 0.09	205.24 ± 0.24
ΔG_{bind}	-48.84 ± 0.13	-48.93 ± 0.14	-48.67 ± 0.11	-50.63 ± 0.07	-71.73 ± 0.08

ΔE_{elec} (electrostatic energy), ΔE_{vdW} (van der Waals energy), ΔG_{GB} (polar solvation energy), ΔG_{SA} (non-polar solvation energy), ΔG_{gas} (gas-phase energy), ΔG_{solv} (Total solvation free energy of polar and non-polar states), and ΔG_{bind} (total free binding energy)

Table 4 Components of the free binding energetics between XDYMA, XDYVL, QDWXL, zanubrutinib, acalabrutinib, and ibrutinib with BTK^{481C} expressed in kcal/mol with standard error of means. Computations were made with 200,000 MD trajectory snapshots

Energy Component (kcal/mol)	BTK^{481C}- XDYMA	BTK^{481C}- XDYVL	BTK^{481C}- QDWXL	BTK^{481C}- Zanubrutinib	BTK^{481C}- Acalabrutinib	BTK^{481C}- Ibrutinib
ΔE_{vdW}	-66.75 ± 0.08	-64.21 ± 0.07	-62.22 ± 0.07	-64.13 ± 0.05	-57.78 ± 0.06	-57.98 ± 0.05
ΔE_{elec}	-139.74 ± 0.32	-145.31 ± 0.28	-72.73 ± 0.41	-26.51 ± 0.06	-204.33 ± 0.30	-29.88 ± 0.05
ΔG_{GB}	133.27 ± 0.29	142.35 ± 0.24	81.25 ± 0.34	39.51 ± 0.05	215.43 ± 0.28	39.96 ± 0.04
ΔG_{SA}	-10.18 ± 0.01	-9.52 ± 0.01	-8.70 ± 0.01	-7.87 ± 0.01	-7.51 ± 0.01	-7.19 ± 0.01
ΔG_{gas}	-206.49 ± 0.29	-209.52 ± 0.28	-134.95 ± 0.42	-90.64 ± 0.08	-262.11 ± 0.29	-87.86 ± 0.07
ΔG_{solv}	123.09 ± 0.30	132.83 ± 0.24	72.55 ± 0.34	31.64 ± 0.05	207.92 ± 0.28	32.77 ± 0.04
ΔG_{bind}	-83.40 ± 0.15	-76.69 ± 0.14	-62.40 ± 0.12	-57.00 ± 0.06	-54.19 ± 0.07	-55.09 ± 0.06

ΔE_{elec} (electrostatic energy), ΔE_{vdW} (van der Waals energy), ΔG_{GB} (polar solvation energy), ΔG_{SA} (non-polar solvation energy), ΔG_{gas} (gas-phase energy), ΔG_{solv} (Total solvation free energy of polar and non-polar states), and ΔG_{bind} (total free binding energy)

exploration of vast sequence spaces. This feature proved indispensable for designing custom peptide libraries for optimization in this study, allowing for systematic and targeted peptide design. However, its utility suffers with longer peptide lengths, particularly owing to the exponential increase in computational cost. Despite this limitation, PepCoGen remains a robust and practical tool for the initial peptide design, particularly when working with shorter sequences ($\leq$21 residues) [9].

The Support Vector Machine (SVM) algorithm on ToxinPred was used for the screening process because of its ability to process complex data, model nonlinear relationships, and render accurate predictions of peptide characteristic features [11]. ToxinPred excels in predicting peptide toxicity by leveraging a hybrid approach that combines motif-based analysis with machine learning techniques. This methodology delivers exceptional sensitivity and specificity, achieving an impressive AUC of 0.98 on independent dataset. Such high accuracy establishes ToxinPred as a reliable resource for therapeutic peptide design. However, its utility is limited to short peptide sequences ($\leq$35 amino acids), which constrains its applicability to longer sequences. Nonetheless, the unique capability of ToxinPred to predict peptide physicochemical properties enhances its utility for the screening of peptide sequences with optimal therapeutic attributes as seen in this study. Moreover, the Support Vector Machine (SVM) algorithm integrated into ToxinPred has demonstrated exceptional efficacy in handling high-dimensional data and establishing robust decision boundaries that contribute to its high accuracy in classifying toxic and non-toxic peptides, and predicting physiochemical properties. The validation of the SVM algorithm through stratified k-fold cross-validation ensures a proportional distribution of samples across classes within each fold, enhancing its reliability. Its performance has been rigorously evaluated using metrics such as sensitivity, specificity, accuracy, Matthews correlation coefficient (MCC), and area under the receiver operating characteristic curve (AUC), offering a comprehensive assessment of its discriminative ability. However, the SVM model can be sensitive to the choice of kernel and hyperparameters, necessitating extensive tuning to optimize performance. Despite this challenge, SVM implementation in ToxinPred provides a robust and precise approach for predicting peptide toxicity and physicochemical properties [10].

We began the filtration process by screening for peptide sequences with optimal toxicity profiles, ensuring that potentially toxic peptide sequences were eliminated before screening for the other physiochemical characteristics [24]. A low SVM score indicates high toxicity, whereas a high score suggests low toxicity, with lower confidence of the peptide sequences exerting undesirable cytotoxicity, immunogenicity, and/or off-target effects in vivo. Following this, we screened the peptide sequences by prioritizing those with high amphipathicity. A high amphipathic index is important for peptide drugs because it influences membrane interactions, cellular uptake, and binding affinity [33]. We then screened for hydrophobicity by prioritizing peptide sequences with moderate indices. Peptides with higher or lower hydrophobicity may have reduced efficacy owing to poor solubility and nonspecific and/or poor interactions [34]. Furthermore, the predicted isoelectric points of the top 10 screened peptide sequences, close to physiological pH, are suggestive of optimum cellular reactivity and are suitable for peptide solubility and stability [35]. Their molecular weights within the ideal range ($\sim$ 500 – 5000 Da) is recommended for effective peptide therapeutics [36]. Peptides with moderate to high molecular weight offer a larger surface area for enhanced affinity and specificity and conformational stability for enhanced biological activity and half-life. The AGGRESCAN algorithm further validated the in vivo stability and reduced aggregation propensity of the peptide sequences, which are critical for their therapeutic viability [25]. AGGRESCAN demonstrates robust capability for high-throughput analysis of protein sequences, enabling precise identification of aggregation-prone regions or hot spots within polypeptides. Its predictive accuracy, validated at 84% and 75% on the Hex1421 and Hex31 datasets, respectively, underscores its strength in leveraging amino-acid-level features and in vivo experimental evidence of aggregation propensity. However, the tool's exclusive reliance on sequence data limits its ability to consider the extrinsic and intrinsic factors influencing protein aggregation under physiological conditions. Additionally, AGGRESCAN's moderate performance on more complex datasets, such as ALBase678, relative to advanced methodologies, such as ANuPP and APPNN, highlights its reduced versatility and suggests that its application may require supplementation with complementary approaches to ensure robust aggregation prediction across diverse scenarios [37].

The rigorous screening process, reducing the 270 peptide sequences to 6 final candidates, ensured that the final peptide sequences are safe, effective, stable, and suitable for potential therapeutic applications. In this present study, we have generated a comprehensive library of peptide sequences with therapeutic properties. Future research may focus on refining these peptide sequences, exploring additional candidates from the library, and optimizing their efficacy through structural modifications. Consequently, the peptide template sequence (T-E-Y-M-A) was used to build a library of

peptide sequences that were also rationally screened and optimized for their therapeutic efficacy against BTK^{481C}.

PEP-FOLD3 is highly effective for de novo peptide structure prediction from amino acid sequences, utilizing a hidden Markov model and coarse-grained force field. This methodology enables high-throughput modeling of peptides ranging from 5 to 50 amino acids, achieving commendable 95% accuracy for the tested targets. Its rapid and efficient prediction makes it an indispensable tool for preliminary peptide modeling, especially in scenarios demanding high-throughput analysis. However, its performance diminishes in capturing complex peptide conformations, with an accuracy of 50%. In such cases, detailed molecular dynamics simulations or experimental techniques offer superior precision for elucidating peptide behavior within biological contexts. Despite this drawback, PEP-FOLD3's accessibility and efficiency render it a valuable resource for early-stage peptide design, striking a practical balance between accuracy and computational demand [26]. The PEPFOLD3's sOPEP coarse-grained force field was particularly useful for studying peptide folding because it considers backbone hydrogen bonding ($E_{h\text{-}bond}$) and side chain interactions, namely local (E_{local}) and nonbonded ($E_{nonbonded}$) energetics. E_{local} accounted for the local interactions, including bond lengths, bond angles, and dihedral angles, whereas $E_{nonbonded}$ accounted for non-covalent interactions, such as van der Waals forces and electrostatic interactions [27]. The low sOPEP energy values observed for the top-screened peptides implies that they adopted the biologically correct and stable spatial arrangement necessary for their functional properties.

The application of Rosetta CovPepDock enabled the systematic design and validation of electrophilic peptide binders targeting $KRAS^{G12C}$ and BTK^{481C}. This approach ensured the design of peptides with optimal covalent bond geometries and binding affinities, laying the groundwork for potential therapeutic applications [6]. The lower constraint scores observed for the peptides indicate their closer adherence to the ideal covalent bond geometry, accurately representing the covalent interaction in the binding mode. In addition, the lower interface scores indicate stronger interactions between the peptides and the target proteins ($KRAS^{G12C}$ and BTK^{481C}), validating the therapeutic potential of the designed covalent-based peptide inhibitors. CovPepDock is a standalone tool for modeling covalent interactions between peptides and target proteins. The incorporation of electrophilic residues enables the design of peptides that form irreversible bonds with cysteine residues, effectively targeting challenging or undruggable proteins, such as $KRAS^{G12C}$. However, this capability requires higher computational resources and longer processing times than non-covalent docking tools.

Furthermore, the protein–ligand interaction analysis for $KRAS^{G12C}$ underscores stable and strong binding of the peptide inhibitors (RVKDX, HVKXR, and XLKDH). In contrast, the distinctive interaction profile of adagrasib likely contributed to its enhanced binding affinity compared to sotorasib. These observations highlight the potential for further optimization of the designed peptide inhibitors to enhance their efficacy against $KRAS^{G12C}$. For the BTK^{481C} interaction landmark, the robust hydrogen-bonding framework observed for the peptide inhibitors (XDYMA, XDYVL, and QDWXL) compared to the less extensive interaction frameworks of zanubrutinib, acalabrutinib, and ibrutinib indicates the sustained and potentially superior binding affinity of the peptide inhibitors over time. Thus, the peptide inhibitors may be considered viable alternatives to the conventional small-molecule drugs against BTK^{481C}. Collectively, these findings emphasize the importance of designing peptide inhibitors that leverage diverse and robust interaction frameworks to achieve high binding affinity and specificity.

On the case of conformational dynamics, the RMSD results indicate that adagrasib conferred the highest structural stability among the $KRAS^{G12C}$ complexes, with the lowest RMSD, while sotorasib exhibited the highest RMSD, indicating greater structural fluctuation. The peptide inhibitors (RVKDX, HVKXR, and XLKDH) demonstrated comparable stability, suggesting their potential for stable binding. Similarly, the RMSF values showed that adagrasib exhibited minimal residue fluctuation, while sotorasib displayed the highest flexibility. The intermediate degrees of flexibility conferred by the peptide inhibitors also suggest their ability to induce relatively stable binding conformations of $KRAS^{G12C}$. For BTK^{481C} complexes, the RMSD and RMSF results were consistent across the peptide inhibitors (XDYMA, XDYVL, and QDWXL) and the reference inhibitors (zanubrutinib, acalabrutinib, and ibrutinib), highlighting their reasonable structural stability and flexibility. The slight variations in RMSD and RMSF among the peptide inhibitors indicate their potential for stable interactions with BTK^{481C}, comparable to those of the reference inhibitors. Overall, the structural stability and flexibility profiles of the peptide inhibitors support their potential as competitive candidates for further therapeutic exploration.

Finally, the analysis of binding free energy (ΔG_{bind}) results highlights the superior binding affinity of adagrasib for $KRAS^{G12C}$, attributed to its strong van der Waals interactions, surpassing those of the peptide inhibitors and sotorasib. The comparable binding free energies of RVKDX, HVKXR, and XLKDH to sotorasib underscore their potential as competitive inhibitors.

Notably, the pronounced electrostatic energy contribution (ΔE_{elec}) of HVKXR emphasizes the significance of electrostatic interactions in its binding mechanism. For BTK^{481C} inhibition, XDYMA emerges as the strongest binder, achieving a ΔG_{bind} significantly better than those of the reference inhibitors (zanubrutinib, acalabrutinib, and ibrutinib). The remarkable binding energetics of XDYMA, XDYVL, and QDWXL, which are driven by favorable ΔE_{elec} and ΔE_{vdW} interactions, suggest their potential as highly effective inhibitors. Notably, the superior binding affinities of the peptide inhibitors relative to the reference inhibitors highlight their potential for therapeutic application against BTK^{481C}-mediated cancers, meriting further investigation and experimental validation. Although MM/GBSA provides reliable relative binding energy comparisons, its limitation in accurately predicting absolute binding free energies is acknowledged. However, it remains a valuable tool in drug discovery, as it has been benchmarked against experimental data, allowing for efficient screening and ranking of potential drug candidates [38, 39].

Covalent inhibition is governed by both thermodynamic and kinetic factors. Key kinetic parameters, such as the equilibrium inhibition constant (K_i) and inactivation rate constant (k_{inact}), are critical for determining the efficacy of covalent inhibitors [40]. Our current study focuses on the thermodynamics of binding but does not explicitly account for the kinetic aspects of covalent inhibition. To fully evaluate the effectiveness of the designed covalent peptide inhibitors, further experimental studies are needed to determine the kinetic properties of inhibitor binding and covalent bond formation using experimental enzyme assays.

Conclusions

The shift towards peptide-based covalent inhibitors addresses the limitations of noncovalent drugs, particularly in combating resistance mechanisms and achieving sustained target inhibition. Thus, the development of a systematic computational protocol for de novo design of peptide-based covalent inhibitors represents a significant advancement in drug discovery. By overcoming the challenges associated with the design of covalent-based peptide inhibitors, this approach offers promising avenues for developing next-generation therapeutics targeting undruggable proteins. This study also highlights the critical role of computational methodologies in expanding the scope and efficacy of peptide-based therapeutics, paving the way for new treatment modalities across diverse disease areas.

Furthermore, this study demonstrates the feasibility of rational peptide design targeting $KRAS^{G12C}$ and BTK^{481C} using computational methodologies. By systematically exploring peptide sequences, optimizing their physicochemical properties, and modeling covalent formation, we identified lead candidates with robust binding interactions and favourable structural stability in complex with $KRAS^{G12C}$ and BTK^{481C}. Molecular dynamics simulations and binding free energy calculations support their potential as therapeutic inhibitors, underscoring the effectiveness of our computational approach in peptide drug discovery.

Further experimental validation is warranted to advance their clinical applications, offering new insights into the treatment of $KRAS^{G12C}$-mutant and BTK^{481C} cancers. The integration of computational modeling with rational peptide design exemplifies a transformative approach in developing effective therapies against challenging oncogenic mutations. Although experimental validation was not within the scope of this study, we recommend that future studies build upon the computational pipeline established here to design and test the covalent peptide inhibitors.

Abbreviations

AMBER	Assisted Model Building with Energy Refinement
BFE	Binding Free Energy
BTK^{481C}	Bruton's Tyrosine Kinase (C481)
CPPTRAJ	Cartesian Coordinate Projection and Trajectory Analysis
FDA	Food and Drug Administration
$KRAS^{G12C}$	Kirsten Rat Sarcoma Viral Oncogene Homolog (G12C missense mutation)
MD	Molecular Dynamics
MD^{cov}	Covalent Molecular Dynamics
MM/GBSA	Molecular Mechanics/Generalized Born Surface Area
PI	Peptide Inhibitor
PPI	Protein–Protein Interaction
PTRAJ	Parallel Trajectory Analysis
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SMI	Small-Molecule Inhibitor

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s44330-025-00021-y.

Supplementary Material 1.

Acknowledgements

The authors express their deepest gratitude for the divine guidance of God Almighty, whose wisdom and insight have profoundly shaped their research journey. We also extend our heartfelt thanks to the Center for High-Performance Computing (CHPC) in Cape Town, South Africa (https://www.chpc.ac.za/), for their essential technical assistance, which was crucial to the success of this work.

Authors' contributions

Conceptualization: [L. Rao and M.E.S. Soliman]; Methodology: [M.E.S. Soliman]; Formal analysis and investigation: [E. Oduro-Kwateng and M. Ali]; Writing – original draft preparation: [E. Oduro-Kwateng, M. Ali, and I.O. KEHINDE]; Writing—review and editing: [E. Oduro-Kwateng and M.E.S. Soliman]; Supervision: [M.E.S. Soliman].

Funding

No funding was received for this work.

Data availability
All data generated or analysed during this study are included in this published article [and its supplementary information files].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Molecular Bio-Computation and Drug Design Research Group, School of Health Sciences, University of KwaZulu Natal, Westville Campus, Durban 4001, South Africa. [2]State Key Laboratory of Green Pesticide, College of Chemistry, Central China Normal University, Wuhan 430079, China.

Received: 19 July 2024 Accepted: 17 January 2025
Published online: 01 May 2025

References

1. Wang L, Wang N, Zhang W, et al. Therapeutic peptides: current applications and future directions. Signal Transduct Target Ther. 2022;7(1). https://doi.org/10.1038/s41392-022-00904-4
2. Kim H, Hwang YS, Kim M, Park SB. Recent advances in the development of covalent inhibitors. RSC Med Chem. 2021;12(7). https://doi.org/10.1039/d1md00068c
3. Paulussen FM, Grossmann TN. Peptide-based covalent inhibitors of protein–protein interactions. Journal of Peptide Science. 2023;29(1). https://doi.org/10.1002/psc.3457
4. Kazmirchuk TDD, Bradbury-Jost C, Withey TA, et al. Peptides of a Feather: How Computation Is Taking Peptide Therapeutics under Its Wing. Genes (Basel). 2023;14(6). https://doi.org/10.3390/genes14061194
5. Zheng L, Li Y, Wu D, et al. Development of covalent inhibitors: Principle, design, and application in cancer. MedComm – Oncology. 2023;2(4). https://doi.org/10.1002/mog2.56
6. Tivon B, Gabizon R, Somsen BA, Cossar PJ, Ottmann C, London N. Covalent flexible peptide docking in Rosetta. Chem Sci. 2021;12(32). https://doi.org/10.1039/d1sc02322e
7. Pan X, Kortemme T. Recent advances in de novo protein design: Principles, methods, and applications. Journal of Biological Chemistry. 2021;296. https://doi.org/10.1016/j.jbc.2021.100558
8. Pelay-Gimeno M, Glas A, Koch O, Grossmann TN. Structure-Based Design of Inhibitors of Protein-Protein Interactions: Mimicking Peptide Binding Epitopes. Angewandte Chemie - International Edition. 2015;54(31). https://doi.org/10.1002/anie.201412070
9. Ali N, Shamoon A, Yadav N, Sharma T. Peptide Combination Generator: A Tool for Generating Peptide Combinations. ACS Omega. 2020;5(11). https://doi.org/10.1021/acsomega.9b03848
10. Rathore AS, Choudhury S, Arora A, Tijare P, Raghava GPS. ToxinPred 3.0: An improved method for predicting the toxicity of peptides. Comput Biol Med. 2024;179. https://doi.org/10.1016/j.compbiomed.2024.108926
11. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS. In Silico Approach for Predicting Toxicity of Peptides and Proteins. PLoS One. 2013;8(9). https://doi.org/10.1371/journal.pone.0073957
12. Ochoa R, Cossio P, Fox T. Protocol for iterative optimization of modified peptides bound to protein targets. J Comput Aided Mol Des. 2022;36(11). https://doi.org/10.1007/s10822-022-00482-1
13. Vincenzi M, Mercurio FA, Leone M. Virtual Screening of Peptide Libraries: The Search for Peptide-Based Therapeutics Using Computational Tools. Int J Mol Sci. 2024;25(3). https://doi.org/10.3390/ijms25031798
14. Faucher F, Bennett JM, Bogyo M, Lovell S. Strategies for Tuning the Selectivity of Chemical Probes that Target Serine Hydrolases. Cell Chem Biol. 2020;27(8). https://doi.org/10.1016/j.chembiol.2020.07.008
15. Khan S, Bjij I, Olotu FA, Agoni C, Adeniji E, Soliman MES. Covalent simulations of covalent/irreversible enzyme inhibition in drug discovery: A reliable technical protocol. Future Med Chem. 2018;10(19). https://doi.org/10.4155/fmc-2017-0304
16. Kuhlman B, Bradley P. Advances in protein structure prediction and design. Nat Rev Mol Cell Biol. 2019;20(11). https://doi.org/10.1038/s41580-019-0163-x
17. Norn C, Wicky BIM, Juergens D, et al. Protein sequence design by conformational landscape optimization. Proc Natl Acad Sci U S A. 2021;118(11). https://doi.org/10.1073/PNAS.2017228118
18. Guedes IA, Pereira FSS, Dardenne LE. Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges. Front Pharmacol. 2018;9(SEP). https://doi.org/10.3389/fphar.2018.01089
19. Canon J, Rex K, Saiki AY, et al. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. Nature. 2019;575(7781). https://doi.org/10.1038/s41586-019-1694-1
20. Lanman BA, Allen JR, Allen JG, et al. Discovery of a Covalent Inhibitor of KRASG12C (AMG 510) for the Treatment of Solid Tumors. J Med Chem. 2020;63(1). https://doi.org/10.1021/acs.jmedchem.9b01180
21. Fell JB, Fischer JP, Baer BR, et al. Identification of the Clinical Development Candidate MRTX849, a Covalent KRASG12CInhibitor for the Treatment of Cancer. J Med Chem. 2020;63(13). https://doi.org/10.1021/acs.jmedchem.9b02052
22. Guo Y, Liu Y, Hu N, et al. Discovery of Zanubrutinib (BGB-3111), a Novel, Potent, and Selective Covalent Inhibitor of Bruton's Tyrosine Kinase. J Med Chem. 2019;62(17). https://doi.org/10.1021/acs.jmedchem.9b00687
23. Assaf L, Chi Ho N, Dana M, et al. Detection of peptide-binding sites on protein surfaces: The first step toward the modeling and targeting of peptide-mediated interactions. Proteins: Structure, Function, and Bioinformatics. 2014;82(7). https://doi.org/10.1002/prot.24579
24. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS. Peptide toxicity prediction. Methods in Molecular Biology. 2015;1268. https://doi.org/10.1007/978-1-4939-2285-7_7
25. Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X, Ventura S. AGGRESCAN: A server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics. 2007;8. https://doi.org/10.1186/1471-2105-8-65
26. Lamiable A, Thevenet P, Rey J, Vavrusa M, Derreumaux P, Tuffery P. PEP-FOLD3: faster denovo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res. 2016;44(1). https://doi.org/10.1093/nar/gkw329
27. Maupetit J, Tuffery P, Derreumaux P. A coarse-grained protein force field for folding and structure prediction. Proteins: Structure, Function and Genetics. 2007;69(2). https://doi.org/10.1002/prot.21505
28. Salmon-Ferrer R, Goetz AW, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER - Part II: Particle Mesh Ewald. *J* Chem Theory Comput. 2013;9(9).
29. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput. 2013;9(7). https://doi.org/10.1021/ct400341p
30. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. 2015;10(5). https://doi.org/10.1517/17460441.2015.1032936
31. Oduro-Kwateng E, Ali M, Kehinde IO, Zhang Z, Soliman MES. De Novo Rational Design of Peptide-Based Protein-Protein Inhibitors (Pep-PPIs) Approach by Mapping the Interaction Motifs of the PP Interface and Physicochemical Filtration: A Case on p25-Cdk5-Mediated Neurodegenerative Diseases. J Cell Biochem Published online. 2024. https://doi.org/10.1002/jcb.30633.
32. Hashemi ZS, Zarei M, Fath MK, et al. In silico Approaches for the Design and Optimization of Interfering Peptides Against Protein–Protein Interactions. Front Mol Biosci. 2021;8. https://doi.org/10.3389/fmolb.2021.669431
33. Edwards IA, Elliott AG, Kavanagh AM, Zuegg J, Blaskovich MAT, Cooper MA. Contribution of amphipathicity and hydrophobicity to the antimicrobial activity and cytotoxicity of β-hairpin peptides. ACS Infect Dis. 2016;2(6). https://doi.org/10.1021/acsinfecdis.6b00045

34. Yin LM, Edwards MA, Li J, Yip CM, Deber CM. Roles of hydrophobicity and charge distribution of cationic antimicrobial peptides in peptide-membrane interactions. Journal of Biological Chemistry. 2012;287(10). https://doi.org/10.1074/jbc.M111.303602
35. Frolov AI, Chankeshwara S V., Abdulkarim Z, Ghiandoni GM. pIChemiSt — Free Tool for the Calculation of Isoelectric Points of Modified Peptides. J Chem Inf Model. 2023;63(1). https://doi.org/10.1021/acs.jcim.2c01261
36. Al Musaimi O, Lombardi L, Williams DR, Albericio F. Strategies for Improving Peptide Stability and Delivery. Pharmaceuticals. 2022;15(10). https://doi.org/10.3390/ph15101283
37. Prabakaran R, Rawat P, Kumar S, Gromiha MM. Evaluation of in silico tools for the prediction of protein and peptide aggregation on diverse datasets. Brief Bioinform. 2021;22(6). https://doi.org/10.1093/bib/bbab240
38. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inf Model. 2011;51(1). https://doi.org/10.1021/ci100275a
39. Wang E, Sun H, Wang J, et al. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chem Rev. 2019;119(16). https://doi.org/10.1021/acs.chemrev.9b00055
40. Strelow JM. A Perspective on the Kinetics of Covalent and Irreversible Inhibition. J Biomol Screen. 2017;22(1). https://doi.org/10.1177/1087057116671509

Publisher's Note