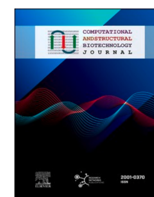
Computational and Structural Biotechnology Journal 23 (2024) 3155–3162

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj



Software/web server article

cPEPmatch Webserver: A comprehensive tool and database to aid rational design of cyclic peptides for drug discovery

Brianda L. Santini, Stephanie Wendel, Niklas Halbwedl, Asha Knipp, Martin Zacharias *

Center for Functional Protein Assemblies, Technical University of Munich, Ernst-Otto-Fischer-Straße 8, Garching, Germany

ARTICLE INFO

Keywords:
Cyclic peptides
Rational drug design
Molecular dynamics
Protein-protein complexes
Protein binding modulation

ABSTRACT

Cyclic peptides have emerged as versatile scaffolds in drug discovery due to their stability and specificity. Here, we present the cPEPmatch webserver (accessible at https://t38webservices.nat.tum.de/cpepmatch/), an easy-to-use interface for the rational design of cyclic peptides targeting protein-protein interactions combined with a semi-quantitative evaluation of binding stability. This platform also offers access to a comprehensive database of cyclic peptide crystal structures. We demonstrate the webserver's utility through a series of case studies involving medically relevant protein systems, highlighting its potential to significantly advance drug discovery efforts.

1. Introduction

Cyclic peptides are at the forefront of modern drug discovery research. These molecules, known for their cyclic structures, offer a unique blend of specificity, stability, and efficacy, distinguishing them from traditional linear peptides and small-molecule drugs [1–3]. Their unique properties have advanced the field by enabling the inhibition of key protein-protein interactions, marking a new frontier in therapeutic interventions [4–6]. Protein-protein interactions (PPIs) are essential to nearly all biological processes, including intercellular communication, cell signaling, metabolic and developmental control, and programmed cell death [7]. Distorted regulation of PPIs can lead to diseases such as cancers, immune disorders, and neurodegenerative conditions. Modulating protein-protein interactions is highly desirable, but targeting PPIs is challenging due to the nature of protein-protein interfaces, which typically feature large, flat surfaces devoid of pockets [8,9]. Despite this difficulty, cyclic peptides have shown promise in successfully modulating these interactions [10–13].

Advances in computational design tools have expanded the potential applications of cyclic peptides by helping to navigate the vast combinatorial space of potential peptide sequences to meet the specific requirements of effective PPI modulators. Techniques like virtual screening, pharmacophore matching, Molecular Dynamics (MD) simulations, and energy calculations predict the behavior of cyclic peptides in biological systems, providing insights into their binding modes and efficiencies [10,14–16]. Building on this foundation, we previously introduced the Cyclic Peptide Matching (cPEPmatch) method, which uses detailed experimental structures of cyclic peptides to identify those that mimic specific protein-protein interactions [17,18]. For a set of known cyclic peptide-protein complexes we could demonstrate that the cPEPmatch approach allowed the design of cyclic peptides that closely matched known cyclic peptide binders [17]. Our previous efforts also included the application to a large collection of known protein-protein complexes (> 170) and we could identify and suggest for the majority of complexes (> 75 %) cyclic peptides that mimic interface segments. The approach was also already successfully used to design and experimentally test cyclic peptides that interfere with the ICOS/ICOS-L interaction, an important interaction for immune reaction modulation [13]. The cPEPmatch webserver, which we present here, offers an easy-to-use interface for designing cyclic peptides targeting these interactions. This platform also includes a comprehensive database of cyclic peptide crystal structures. We detail the algorithm and webserver usage, and demonstrate its utility through case studies involving medically relevant protein systems, highlighting its potential to advance drug discovery efforts.

2. cPEPmatch Algorithm and Webserver

The cPEPmatch algorithm matches the backbone structures of short peptide segments at the protein-protein interface with those in a library of known cyclic peptide backbone structures. A cyclic peptide that aligns with the backbone structure of the segment serves as a template for a

* Corresponding author.
E-mail address: zacharias@tum.de (M. Zacharias).

https://doi.org/10.1016/j.csbj.2024.08.008
Received 26 June 2024; Received in revised form 9 August 2024; Accepted 9 August 2024
Available online 12 August 2024
2001-0370/© 2024 Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

binder, with its amino acid side chains modified to match those found in the target complex. The source code for cPEPmatch and its database are publicly available on GitHub and can be accessed via the following link: https://github.com/briandasantini/cPEPmatch. The webserver for cPEPmatch, hosted at the Technische Universität München (TUM) web services and accessible at https://t38webservices.nat.tum.de/cpepmatch/, presents a user-friendly interface for researchers to utilize the cPEPmatch tool. Developed using Django, a high-level Python web framework, the webserver ensures a robust and scalable platform with a queuing system that allows for multiple job submissions simultaneously. The core functionality lies in its job submission feature, where users can submit their protein structures and specify parameters for the cPEPmatch analysis. This section is designed to be intuitive, ensuring ease of use even for new users. To assist users, a quick guide is provided for rapid reference, and a comprehensive tutorial, available at https://t38webservices.nat.tum.de/tutorial/, includes a detailed example to demonstrate the process of using cPEPmatch.

At the time of publication, the foundational library of cPEPmatch comprises 432 unique cyclic peptide structures sourced from the Protein Data Bank. This database is continually updated on a yearly basis with new cyclic peptides as they are published. The accumulated database is a rich source of information on the sizes, secondary structures, and types of cyclization of these peptides (Fig. 1, A-D), and it also provides clean PDB files of the extracted cyclic peptides. The characteristic analysis reveals a diverse range of cyclic peptides, primarily 10–20 amino acids long, with head-to-tail cyclization being the most common. Peptide lengths vary with cyclization types: head-to-tail often correlates with shorter peptides, while head-to-tail with disulfide bonds or multiple disulfide bonds correlates with the largest peptides, indicating that more bonds within the cyclic peptide are necessary to stabilize larger structures. β-hairpin structures are more frequent, whereas α-helices are the least common. Within the cPEPmatch algorithm, the backbone structure of these cyclic peptides is characterized using CA carbon distance matrices, tailored according to user-defined parameters like motif size (ranging from 4 to 7 amino acids) and a consecutive or non-consecutive sequence search. A set of 432 cyclic peptide structures may still be considered as limited and of course a further extension is desirable (and as indicated above will be continuously performed). However, in a previous publication [17] we found that even such a limited set of cyclic peptides is sufficient to identify and design cyclic peptides that closely mimic interface segments in the great majority (>75 %) of a large set of tested known protein-protein complexes (>170). [17].

To ensure accurate matching, cPEPmatch also characterizes protein-protein interfaces by utilizing the same CA distance matrix method as applied to the cyclic peptides in the database The interface cutoff distance, measured in angstroms, determines which amino acids are

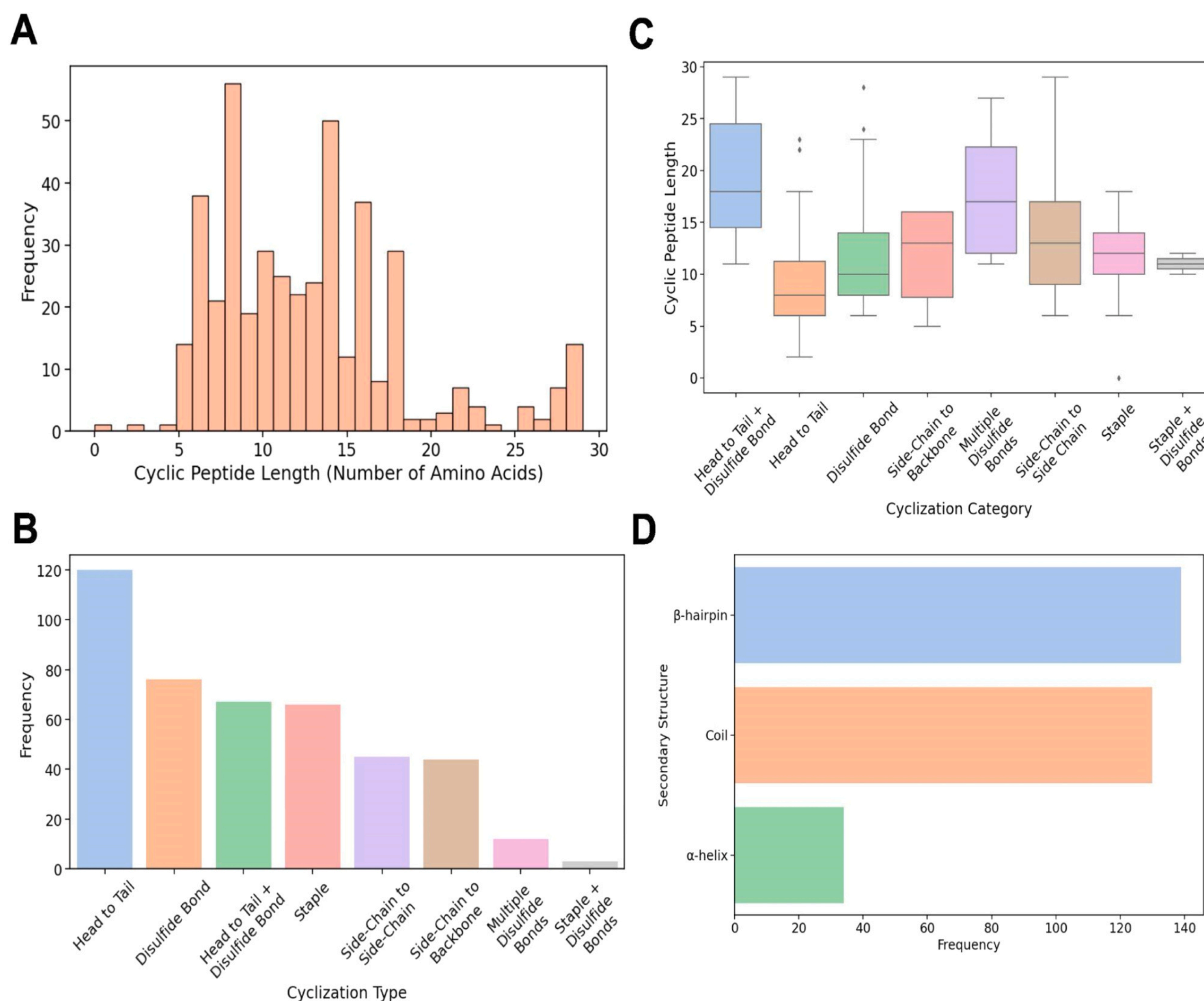


Fig. 1. : Overview of Cyclic Peptide Characteristics within the cPEPmatch Database. (A) Histogram of cyclic peptide lengths by amino acid count. (B) Bar chart of cyclization type frequencies. (C) Box plots correlating cyclization type with peptide length. (D) Stacked bar chart of secondary structure frequencies, including β-hairpin, α-helix, and coils.

considered part of the interface, thus influencing the matching process with cyclic peptides. Users have two primary options for defining the interface. (1) Proximity-Based Interface: Users can define the interface based on proximity. By selecting a distance cutoff in angstroms, cPEPmatch identifies amino acids in the vicinity of the receptor as part of the interface. This option is suitable for general interface identification where specific binding hotspots are not known or not be the primary focus. (2) Hot Spot-Based Interface: Alternatively, users can define the interface by specifying key amino acids known as hotspots. Selecting this option allows cPEPmatch to focus on these specific residues, enhancing the relevance and specificity of the cyclic peptide matches and reducing the output to more tailored matches. This dual approach in interface characterization ensures that cPEPmatch can be adapted to a wide range of research needs, from general exploratory studies to targeted investigations focusing on key interaction sites.

Following interface characterization, cPEPmatch executes a crucial step by matching the CA distance motifs between the characterized protein interfaces and the cyclic peptide database using a user-defined fit-RMSD (Root Mean Square Deviation) threshold in angstroms to ensure structural similarity. When dealing with many non-consecutive hot spots, we recommend clustering hot spots into groups of 5–7 nearest amino acids and running separate searches. This approach reduces the complexity of distance matrices and the number of combinations to consider, avoiding an extensive combinatorial search of irrelevant matches. Upon identifying suitable matches, cPEPmatch superimposes these onto the target protein structure and mutates the side chains of the cyclic peptide to mimic the protein, employing Modeller [19] software for this purpose. This process not only ensures functional mimicry of the target protein interface but also places these side chains accurately using homology modeling. Following this, a minimization step is carried out to ensure optimal positioning of the side chains, providing a more reliable starting structure for subsequent evaluations. This refinement is crucial for enhancing the accuracy and effectiveness of the matched cyclic peptides in potential binding and interaction studies. Note that Modeller cannot process non-standard amino acids, including methylated variants, during the mutation step of the cPEPmatch process. Consequently, when such matches are encountered, cPEPmatch outputs the matched structure without mutating it to resemble the target protein. In these cases, users must manually perform the mutations. Tools like pdb4amber, part of the AmberTools suite, can be used for this purpose. Additionally, when a matched residue falls on a disulfide-bonded amino acid of the cyclic peptide, it is not mutated to avoid disrupting its cyclization.

The output from cPEPmatch includes a detailed match list and all output structures—superimposed, mutated, and minimized. The output table provides comprehensive information, including the match number, PDB name, matched cyclic peptide residues, and fit-RMSD values. Tools such as PyMOL or VMD are recommended for the visual analysis of the matches. This visual inspection can be crucial for evaluating the steric fit of the matches and deciding which structures should be further evaluated using MD simulations and free energy calculations. A notable aspect of using cPEPmatch is its flexibility for parameter adjustments. Users might find that the initial recommended parameters yield too many or too few matches for their specific system. In such cases, adjusting parameters like Cutoff or Threshold is necessary to narrow down or increase the number of matches.

3. Application to medically relevant case studies

In our quest to develop novel therapeutic predictions, we employed cPEPmatch to identify cyclic peptides that could modulate critical protein-protein interactions. These interactions are medically significant due to their roles in various diseases such as viral infections, cancer, autoimmune disorders, and more. We targeted several PPI complexes, including (1) the interaction between the SARS-CoV-2 spike protein and the ACE2 receptor, crucial for viral entry into host cells and a critical target for COVID-19 therapeutic interventions [20]; (2) the N-terminal oligomerization domain within the Breakpoint Cluster Region (BCR)-Abelson (Abl) kinase fusion protein, significant in chronic myelogenous leukemia (CML) [21]; (3) the interaction between the Estrogen Receptor (ER) and its coactivators, a major regulator of disease development in breast cancer [22]; (4) the interaction between the gp120 protein of the Human Immunodeficiency Virus (HIV) and the CD4 receptor on host T-cells, critical for viral entry and infection [23]; (5) the binding of Interleukin-12 subunit beta (IL-12β) to its receptor subunit IL-12Rβ1, crucial for regulating inflammatory responses and thus critical in autoimmunity [24]; (6) the interaction between MDM2 and the tumor suppressor protein p53, central in cancer biology [25]; (7) the interaction between Programmed Death-1 (PD-1) on T cells and its ligand PD-L1, a key immune checkpoint relevant to cancer immunotherapy [26]; and (8) the interaction between S100P and RAGE, significant in tumor growth and metastasis [27]. These targeted interactions highlight the diverse therapeutic potentials addressed using the cPEPmatch tool.

3.1. Materials and methods

The cPEPmatch method requires structures of established protein-protein complexes as input. For this study, we utilized structures corresponding to the PDB codes listed in Table 1. These structures are critical, serving both as templates for simulation and analysis as well as hot spot identification, guiding residue selection within the cPEPmatch framework. Missing residues were added using MODELLER [19], except in the BCR system, for which we performed an AlphaFold2 [28] prediction.

For the initial assessment of our case studies binding interface hot spots, as well as the evaluation and ranking of protein-cyclic peptide complexes, we employed the Amber19 software package (extension of the Amber18 tools) [29]. We prepared the structures for energy minimization and molecular dynamics (MD-simulations using the tLeap module from Amber19. Protein parameters were sourced from the ff14SB force field [30]. To neutralize the complexes, we added either Na+ or Cl- ions and solvated them in an octahedral box, maintaining a minimum 15 Å distance to the box boundaries, using explicit TIP3P water molecules [31].

For all simulation systems, we initially performed energy minimization using the steepest descent method for 2000 steps via the Amber19 Sander module. For subsequent MD simulations, we utilized the pmemd.cuda module. We incrementally heated the systems to 310 K through

Table 1
Input parameters and configurations for cPEPmatch algorithm. This table enumerates the PDB codes from the Protein Data Bank representing the target and its corresponding ligand. Listed hotspots, derived from MMGBSA analysis of MD simulations, indicate key interface regions; a dash signifies utilization of the entire interface. Columns 'C', 'T' and 'M' represent the cutoff, threshold, and motif size, respectively.

Target	Ligand	PDB	Hot Spots	C	T	M
ACE2	Spike	6M0J	6 9 10 12 13 19 20 23 61 64 65 337	5;6	0.5;0.5	5
BCR-Abl	BCR-Abl	1K1F/AF2	198 201 204 218 222 223 226 259; 146 147 154 158 259 270 273 276 277	8;8	2;2	5
ER	Coactivator	5DZ1	242 243 244 245 246 247 248 249 250	8	1	5
GP120	CD4	1GC1	322 324 337 339 340 341 345 356 357	8	0.7	5
IL12b	IL12Rb1	6WDQ	17 19 21 62 86 88 196 197 198 199	8;8	1;2	5
MDM2	p53	4HFZ	-	6	0.5	5
PD1	PDL1	3BIK	2 37 39 49 51; 96 98 104 105 106 108	5;5	1;0.5	5
S100P	RAGE	2MJW	28 32 34 37 41 43 49 78 90 92	8	2	5

three stages (100 K, 200 K, 310 K), each lasting 100 ps, while applying positional restraints on all non-hydrogen atoms relative to their starting conformations. Following this, we systematically reduced the positional restraints from an initial value of 25 to 0.5 kcal·mol-1·Å-2 over the course of five successive 100 ps simulations at a constant pressure of 1 bar and a temperature of 310 K. The equilibrated structures then served as the starting points for the production runs, which were conducted without any restraints. We employed a 2 fs time step and constrained all bonds involving hydrogens to their optimal lengths using the Shake algorithm [32]. Data collection simulations were executed for 25 ns, and coordinates were saved every 30000 steps. These conditions were found optimal in a previous publication of the cPEPmatch approach with applications to more than 170 protein-protein complexes [17].

For the evaluation of hot spot residues and stable cyclic peptide binding, we examined the MD trajectories through visual inspection. Subsequently, we applied the Molecular Mechanics Generalized Born Surface Area (MMGBSA) methodology to quantify the mean interaction [33,34]. This analysis was conducted using the well-validated single trajectory approach [35], as facilitated by the MMPBSA.py module within Amber18. We utilized 1000 snapshots extracted from the last 5 ns of the production MD simulation for these calculations. The modified GB model (igb=5) with mbondi2 parameters was employed, along with specific α, β and γ values of 1.0, 0.8, and 4.85, respectively. Dielectric constants were set to 80 for the solvent and 5 for the solute. This analysis yielded the mean interaction energy between the cyclic peptide and its protein partner. Per-residue decomposition allowed for a detailed investigation of the contribution of interacting residues to identify the hot spots.

Furthermore, we performed Repulsive Scaling - Replica Exchange Molecular Dynamics (RS-REMD) [36] simulations for all our matches. Simulations were conducted in explicit TIP3P water at 300 K, together with neutralizing Na+ counterions in water and no additional salt. The prepared system structures were placed within periodic octahedral solvent boxes and a minimum distance between the peptide-protein complex and boundary of 15 Å. The systems were then minimized via steepest descent for 10000 steps, heated up step-wise to a target temperature of 300 K within 60 ps and equilibrated within NVT and NPT conditions for 1 ns each. The RS-REMD simulations were executed with 25 ns per replica and exchange attempts every picosecond. These simulations were carried out using 8 and 16 replicas, each exhibiting a progressive increase in the bias applied to the ligand and receptor Lennard-Jones interaction parameters as proposed by Siebenmorgen and Zacharias, 2020 [36]. The specific bias along the replicas leads to progressive destabilization of the ligand-receptor interactions and to partial or full dissociation of partners in the highest replicas. From the sampling overlap it is possible to estimate a free energy required to dissociate (or partially dissociate) the ligand from the receptor. The trajectories were analyzed using pytraj and the associated relative binding free energies were calculated with the Multistate Bennett Acceptance Ratio (MBAR) [36].

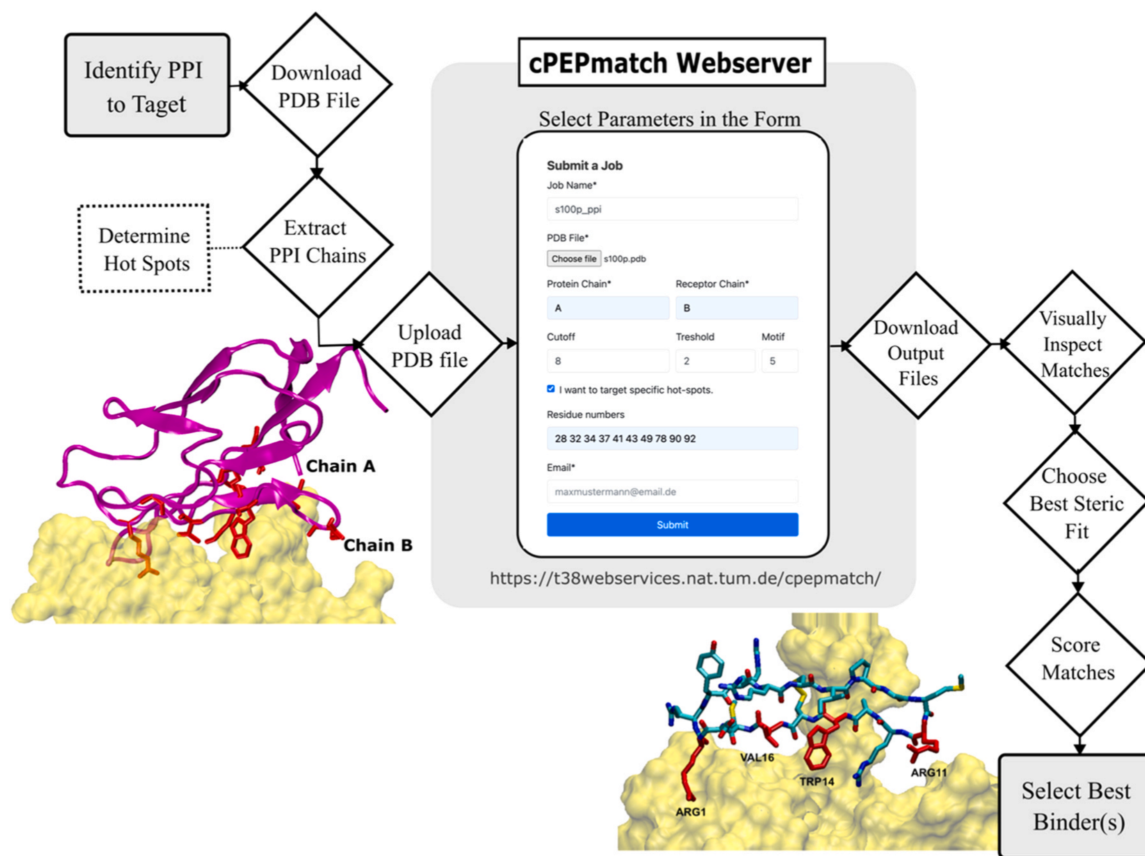


Fig. 2. : Flow chart illustrating the usage of the cPEPmatch web interface, using the S100P-RAGE system as an example target. (Left) Representation of the binding interface of the S100P-RAGE complex with Chain A, RAGE protein, in magenta and Chain B, S100P protein, in yellow, highlighting the interaction hotspots in red. (Middle) cPEPmatch job submission form where parameters such as cutoff, threshold, and motif are specified for the analysis, along with targeted hotspots for Chain A binding to Chain B. (Right) Visualization of the strongest cPEPmatch predicted cyclic peptide binder, Match 3 PDB2lwt, for the S100P system. The target protein S100P is rendered as a yellow surface, while the cyclic peptide binder is shown in licorice representation with atoms colored by element. Key hotspots on the cyclic peptide, corresponding to matched and mutated residues essential for binding, are highlighted in red.

3.2. Results and discussion

To illustrate the workflow of the cPEPmatch webserver, we described the highlighted system, S100P-RAGE, in detail as shown in Fig. 2. The first step is to download the crystal structure from the Protein Data Bank. The PDB file was inspected and prepared, including removing unnecessary chains and renumbering residues. Ensuring there are no missing residues in the binding site is crucial, and if missing residues are found outside the binding site, tools like Modeller can be used to fill them in for later simulations. An optional step is to identify the hotspots, specific residues that contribute the most to the overall binding free energy, to direct the search towards a more competitive cyclic peptide mimic, which has proven useful in our early cPEPmatch work [13,18]. For this, we ran MD and free energy calculations using MMGBSA. In the case of the S100P-RAGE system, we identified residues 28, 32, 34, 37, 41, 43, 49, 78, 90, and 92 in chain A, RAGE protein, as the most contributing to the binding to chain B, S100P protein. The next step is to choose the input parameters to run the cPEPmatch algorithm. The cutoff parameter defines the distance between the receptor and the interface, with residues within this cutoff distance considered part of the interface (Default: 6Å). The threshold parameter sets the fit-RMSD maximum for a match to fit into the targeted protein interface residues (Default: 0.7Å - Recommended: 0.3–1.5Å). The motif size refers to the length of the motif residues to target (Default: 5 amino acids) and should be adjusted based on the nature of the binding site. For our example, we selected an interface cutoff of 8Å, a threshold of 2Å, set the motif size to 5 residues, and chose the "target specific hot-spots" option to run a non-consecutive search to match the identified hotspot residues.

With the provided parameters, cPEPmatch identified 16 matches of cyclic peptides that mimic the RAGE PPI binding site. Candidates with steric clashes were eliminated via visual inspection using VMD. Five were selected for further simulations. To evaluate the binding affinity of the matches, we employed a three-criteria evaluation method: assessing stability during free MD simulations, calculating MMGBSA from the free MD trajectories, and employing MBAR analysis of RS-REMD simulations, from 8 and 16 replicas. It is important to note that in principle the MMGBSA approach only gives an estimate of interaction energy between cyclic peptide and receptor and neglects conformational entropy effects. Such effects are in principle included in the RS-REMD simulations (full flexibility of binding partners and inclusion of explicit solvent) but in this technique the dissociated state is only approximately defined. Hence, the calculated binding free energies need to be interpreted with care and can only be used as qualitative indicator for complex stability. Hence, our criteria for a stable match included sustained binding during the entire free MD trajectory, an MMGBSA interaction energy value less than −25 kcal/mol, and an RS-REMD derived MBAR free energy lower than −2 kcal/mol. These criteria were informed by previous studies: the MMGBSA threshold was based on our initial studies examining the binding affinity of small binders [17], and the RS-REMD free energy criterion was derived from an evaluation of our previously validated Match 41, a recognized inhibitor of the ICOS/ICOS-L PPI which presents values in a similar range, −2.6 + /- 0.14 kcal/mol [13]. All five selected matches for the S100P-RAGE complex demonstrated stability throughout the MD simulations and exceeded the established affinity criteria (Table 2). Detailed mutation information for these matches is available in Table 2. Notably, Match 3 exemplifies how effective mimicry of mutated hotspots can resemble those of the native binding site, as depicted in Fig. 2, right.

To refine our binding evaluation methods, we conducted RS-REMD simulations using 8 and 16 replicas across all systems. A Pearson Correlation coefficient of 0.85 was observed between these two setups (Fig. 3). Typically, MBAR free energy values from 16-replica simulations were lower compared to those from 8 replicas Table 2. A potential explanation for this discrepancy might lie in the sampling dynamics; 16-

Table 2
Comparative assessment of approximate binding free energies change (ΔG) in kcal/mol across tested systems, calculated using MBAR from RS-HREMD trajectories, with 16 and 8 replicas, and MMGBSA methods.

System	Match	16 Replicas ΔG (kcal/mol)	+ /-	8 Replicas ΔG (kcal/mol)	+ /-	MMGBSA ΔG (kcal/mol)
ACE2	match2_2ll5	−1.81	0.74	−1.93	0.48	−12.17
	match12_3uc7	−3.0	0.32	−3.31	0.4	−32.46
	match145_6ve9	−2.36	0.2	−2.41	0.63	−27.41
BCR	match1_1jdp	−4.81	0.9	−4.07	0.26	−33.69
	match3_4ttl	−5.87	0.77	−3.82	0.23	−30.71
ER	match1_3p72	−7.42	0.4	−3.99	0.12	−92.82
	match5_4ttl	−7.94	0.7	−4.65	0.35	−64.53
	match6_5wtt	−8.93	0.54	−5.95	0.35	−77.12
	match7_7m7x	−4.47	0.54	−3.11	0.21	−61.93
	match8_7m7x	−5.87	0.34	−5.37	0.74	−74.50
GP120	match5_1im7	−1.62	0.6	−3.99	0.61	−40.22
	match9_1znu	−3.31	0.2	−3.7	0.66	−21.96
	match18_3uc7	−2.4	0.91	−2.05	0.78	−38.09
	match29_4tto	−3.47	0.22	−3.63	0.54	−12.38
IL12	match1_1ebp	−0.64	0.38	−1.11	0.22	−32.92
	match2_2nb6	−1.2	0.85	−0.88	0.61	−24.32
	match3_1im7	−1.9	0.16	−2.12	0.24	−33.97
	match10_2ll5	−1.13	0.56	−0.94	0.47	−21.39
	match14_2ndm	−0.78	0.48	−2.21	0.52	−17.01
	match16_3pp4	−1.52	0.13	−1.22	0.24	−20.23
	match30_5otx	−2.85	0.49	−1.58	0.29	−5.36
	match31_6dhr	−2.05	0.24	−1.79	0.32	−38.19
MDM2	match6_3p72	−4.99	0.59	−4.32	0.46	−39.18
	match10_3pp4	−13.36	0.98	−7.0	0.43	−52.95
	match13_3uc7	−12.75	0.8	−6.02	0.24	−57.02
PD1	match9_2m7i	−0.17	0.18	−0.13	0.16	−30.10
	match12_2ns4	−5.68	0.37	−2.49	0.26	−52.40
	match13_2ns4	−3.57	0.4	−3.47	0.39	−43.12
	match25_6q1u	−2.07	0.88	−2.71	0.17	−48.08
S100P	match2_1znu	−3.31	0.41	−2.83	0.46	−33.93
	match3_2lwt	−8.65	0.7	−6.35	0.54	−66.98
	match5_2lwv	−2.76	0.56	−2.31	0.21	−43.25
	match7_2lyf	−9.33	0.78	−4.19	0.28	−57.66
	match15_7k7x	−8.59	0.79	−5.0	0.6	−41.60

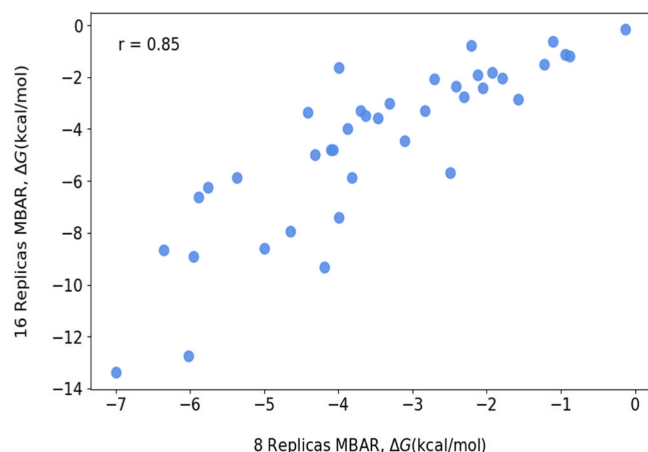


Fig. 3. : Correlation of RS-REMD derived Free Energy Calculations for 8 vs. 16 replicas. Comparison of 25 ns replica exchange simulations with repulsive scaling Hamiltonian using 8 vs. 16 replicas of RS-REMD simulation.

replica simulations are more effective at sampling a broader range of bound and unbound states between the cyclic peptide and its receptor, due to the stronger repulsion in the higher replicas. For instance, in the case of a strong binder like Match 3, the 16-replica simulations facilitated a comprehensive sampling of bound versus unbound states, while the 8-replica simulations failed to sample any completely unbound states, showing only the cyclic peptides in various conformations (Fig. 4). This observation was consistent across other systems.

A similar cPEPmatch search (Table 1) and evaluation criteria (Fig. 5) were applied to all other case study systems, resulting in the identification of stable matches with detailed mutation information presented in Table 3. Notably, while at least two stable matches were observed for all systems during MD simulations, IL12 and GP120 predominantly exhibited weak to moderate binders. These binders demonstrated lower binding MBAR scores compared to the ICOS/ICOS-L cyclic peptides, which are known experimentally validated binders or modulators. However, our findings in Abdel Rahman et al., 2024 suggest that a higher binding affinity does not invariably translate into more effective modulation, indicating potential viability for the predicted binders in this study [13]. In our systematic application of cPEPmatch, we identified a varying number of stable matches across systems: 2 for ACE2, 3 for BCR, 5 for ER, 2 for GP120, 2 for IL12, 3 for MDM2, and 3 for PD1 (Table 2). These results show the capability of cPEPmatch as a tool to identify potential lead molecules for therapeutic interventions, providing a solid foundation for further validation and optimization.

4. Conclusions

The cPEPmatch method is a computational approach developed to identify cyclic peptide templates for protein-protein interaction modulators, unique in its reliance on structural matching of protein interfaces. The method has already been tested successfully on known cyclic-peptide-protein complexes and has been used to identify cyclic peptides that mimic interface segments in a large fraction of known protein-protein complexes in previous publications. [13,17,18] Recently, it was also successfully used to generate cyclic peptides that mimic a binding site for glycosamin glycan molecules including experimental validation of specific binding [37]. Establishing the cPEPmatch webserver represents an advancement towards making our tool more user-friendly and accessible to a broader scientific community, thereby simplifying the process of designing cyclic peptides. Additionally, the growing cPEPmatch database, sourced from the Protein Data Bank, offers insights into the structural diversity and bioactivity of cyclic peptides, providing a valuable resource for future studies. Through case studies, we demonstrated its application in identifying promising cyclic peptides for

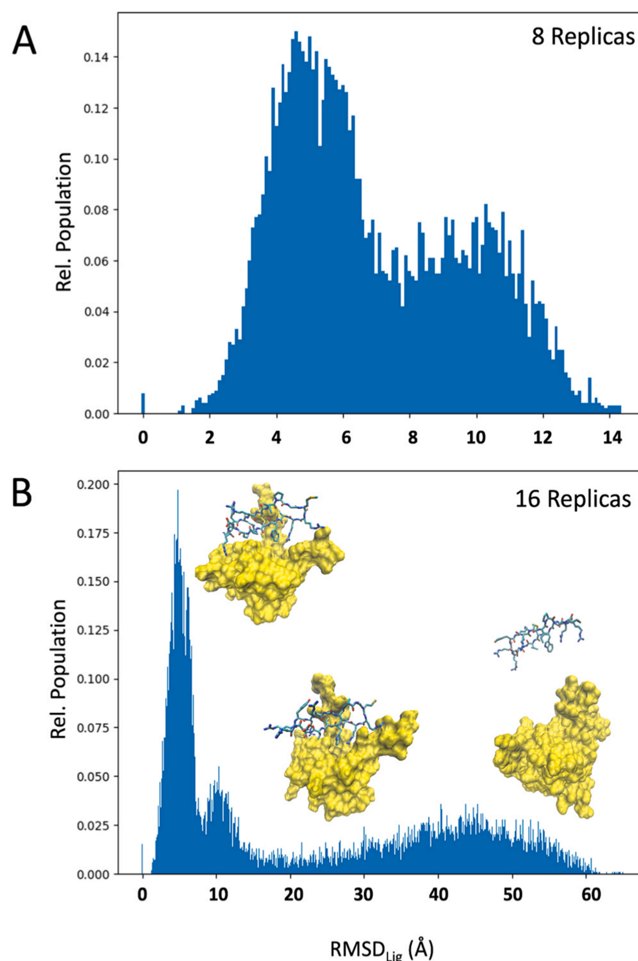


Fig. 4. Exemplary ligand RMSD Population Histograms for RS-REMD simulations of the S100P-RAGE system. Panel (A) represents the relative population distribution of ligand RMSD (RMSD of the ligand after best superposition of sampled structures on the receptor protein) for 8-replica simulations, indicating the diversity of conformations sampled during the simulation. However, complete dissociation is not achieved. Panel (B) indicates the corresponding distribution for 16-replica simulations alongside representative structures at key RMSD intervals, illustrating a wider range of sampling of bound and unbound states. The insets in Panel B depict the ligand in complex with the receptor (left), an intermediate state (middle), and a fully unbound state (right).

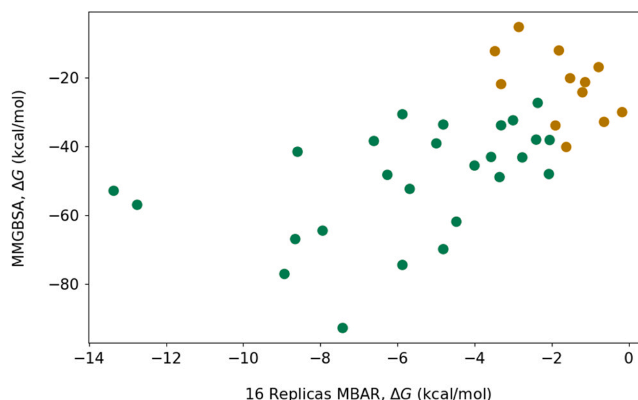


Fig. 5. : Selection criteria based on RS-REMD MBAR (x-axis) and MM-GBSA from regular MD (y-axis). Stable matches are identified as cases with an ΔG < −2 kcal/mol for the RS-REMD based binding free energy estimate and MM-GBSA based estimate of ΔG < −25 kcal/mol (green), with all other points considered non-matches (orange).

Table 3
Summary of stable matches identified through cPEPmatch and their corresponding point mutations. This table details the specific amino acid substitutions performed on the cyclic peptide templates to emulate the binding sites of their target proteins. Each row represents a unique match with its PDB code and a list of mutations that were introduced to achieve the desired structural and functional mimicry.

System	Match	PDB	Mutations
ACE2	Match 12	3uc7	6-THR; 7-PHE; 9-ASP; 10-LYS
ACE2	Match 145	6ve9	11-PHE; 12-ASN; 13-HIS; 14-GLU; 15-ALA
BCR	Match 1	1jdp	7-PHE; 11-GLN; 15-VAL
BCR	Match 3	4ttl	11-GLN; 17-VAL; 19-GLY; 20-PHE
ER	Match 1	3p72	4-HIS; 5-ARG; 6-LEU; 8-GLN
ER	Match 6	5wtt	3-ILE; 4-LEU; 6-ARG; 7-LEU
ER	Match 8	7m7x	4-ARG; 5-LEU; 7-GLN
ER	Match 5	4ttl	6-LEU; 7-HIS; 9-LEU; 10-LEU
ER	Match 7	7m7x	4-LEU; 5-HIS; 7-LEU
GP120	Match 5	1im7	7-GLN; 8-PHE; 12-ARG; 13-SER
GP120	Match 18	3uc7	9-HIS; 15-GLN; 17-PHE; 20-ARG
IL12	Match 31	6dhr	11-TRP; 15-PHE; 18-LYS; 20-LYS
IL12	Match 3	1im7	3-GLY; 11-THR
MDM2	Match 13	3uc7	1-PHE; 2-SER; 3-ASP; 4-LEU; 5-TRP
MDM2	Match 10	3pp4	15-PHE; 17-ASP; 18-LEU; 19-TRP
MDM2	Match 6	3p72	3-PHE; 4-SER; 5-ASP; 6-LEU; 7-TRP
PD1	Match 12	2ns4	2-ARG; 7-ALA; 8-ASP; 9-TYR
PD1	Match 25	6q1u	1-ARG; 3-MET; 10-ASP; 11-TYR; 13-ARG
PD1	Match 13	2ns4	2-ARG; 4-MET; 7-ALA; 8-ASP; 9-TYR
S100P	Match 3	2lwt	1-ARG; 11-ARG; 14-TRP; 16-VAL
S100P	Match 7	2lyf	1-ARG; 13-TRP; 15-VAL
S100P	Match 15	7k7x	4-LYS; 6-ASN; 27-TRP; 29-VAL
S100P	Match 5	2lwv	5-VAL; 8-GLY; 15-ARG; 17-ASN
S100P	Match 2	1znu	18-LYS; 20-TRP; 23-ARG; 25-LYS; 27-ASN

medically relevant systems. We used Molecular Dynamics simulations and free energy calculations, including MMGBSA and RS-REMD/MBAR, to evaluate the binding efficacies of the peptides. This combined approach has been tested already on cyclic peptide-protein complexes [17] and for evaluating docked protein-protein complexes. [36] Combining approaches allows potentially for a more rigorous assessment of their potential as binders. However, it should be emphasized that the *in silico* evaluation of the cyclic peptide binding properties is nevertheless approximate and the user is encouraged to employ alternative methods after downloading the designed complexes and may ultimately test the suggested cyclic peptide binders experimentally. Nevertheless, our findings indicate that cPEPmatch can effectively identify lead binders for competing with protein-protein interactions, highlighting its potential as a valuable tool in early-stage drug design research for addressing real-world health challenges such as viral infections, cancer, and autoimmune disorders.

Funding

This research was conducted within the Max Planck School Matter to Life supported by the German Federal Ministry of Education and Research (BMBF) in collaboration with the Max Planck Society. We acknowledge also support by the NHR@FAU supercomputer facility at Regionales Rechenzentrum Erlangen (RRZE), Germany.

CRediT authorship contribution statement

Asha Knipp: Writing – review & editing, Validation, Software, Investigation, Formal analysis, Data curation. **Niklas Halbwedl:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis. **Martin Zacharias:** Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Stephanie Wendel:** Writing – review & editing, Validation, Methodology, Investigation, Data curation. **Brianda L. Santini:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation.

Declaration of Competing Interest

None.

Acknowledgments

We thank Drs. Shu-Yu Chen and Patrick Quoika for helpful discussions, and Yasmin Saremi for her assistance with website development.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.08.008.

References

[1] Zorzi A, Deyle K, Heinis C. Cyclic peptide therapeutics: past, present and future. Curr Opin Chem Biol 2017;38:24–9.
[2] Naylor MR, Bockus AT, Blanco M-J, Lokey RS. Cyclic peptide natural products chart the frontier of oral bioavailability in the pursuit of undruggable targets. Curr Opin Chem Biol 2017;38:141–7.
[3] Ji X, Nielsen AL, Heinis C. Cyclic peptides for drug development. Angew Chem Int Ed 2024;63:e202308251.
[4] Driggers EM, Hale SP, Lee J, Terrett NK. The exploration of macrocycles for drug discovery — an underexploited structural class. Nat Rev Drug Discov 2008;7:608–24.
[5] Marsault E, Peterson ML. Macrocycles are great cycles: Applications, opportunities, and challenges of synthetic macrocycles in drug discovery. J Med Chem 2011;54:1961–2004.
[6] Mallinson, J. & Collins, I. *Macrocycles in New Drug Discovery*. (Future Science Ltd, London, 2012). doi:10.4155/fmc.12.93.
[7] Fontaine F, Overman J, François M. Pharmacological manipulation of transcription factor protein-protein interactions: opportunities and obstacles. Cell Regen 2015;4:2.
[8] Bahadur, R.P. & Zacharias, M. The interface of protein-protein complexes: Analysis of contacts and prediction of interactions. doi:10.1007/s00018–007-7451-x.
[9] Conte LL, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol Biol 1999;285:2177–98.
[10] Zhang Z, et al. Structure-Based Design and Synthesis of Potent Cyclic Peptides Inhibiting the YAP–TEAD Protein–Protein Interaction. ACS Med Chem Lett 2014;5:993–8.
[11] Nevola L, Giralt E. Modulating protein-protein interactions: The potential of peptides. Chem Commun 2015;51:3302–15.
[12] Siegert TR, Bird MJ, Makwana KM, Kritzer JA. Analysis of Loops that Mediate Protein-Protein Interactions and Translation into Submicromolar Inhibitors. J Am Chem Soc 2016;138:12876–84.
[13] Abdel-Rahman SA, Santini BL, Calvo-Barreiro L, Zacharias M, Gabr M. Design of cyclic peptides as novel inhibitors of ICOS/ICOSL interaction. Bioorg Med Chem Lett 2024;99:129599.
[14] Duffy FJ, et al. CycloPs: Generating virtual libraries of cyclized and constrained peptides including nonnatural amino acids. J Chem Inf Model 2011;51:829–36.
[15] Duffy FJ, et al. Virtual screening using combinatorial cyclic peptide libraries reveals protein interfaces readily targetable by cyclic peptides. J Chem Inf Model 2015;55:600–13.
[16] Qian Z, Dougherty PG, Pei D. Targeting intracellular protein–protein interactions with cell-permeable cyclic peptides. Curr Opin Chem Biol 2017;38:80–6.
[17] Santini BL, Zacharias M. Rapid in silico Design of Potential Cyclic Peptide Binders Targeting Protein-Protein Interfaces. Front Chem 2020;8.
[18] Santini BL, Zacharias M. Rapid Rational Design of Cyclic Peptides Mimicking Protein–Protein Interfaces. Computational Peptide Science, vol. 2405. New York, NY: Springer US,; 2022. p. 231–44.
[19] Fiser A, Šali A. Modeller: Generation and Refinement of Homology-Based Protein Structure Models. Macromolecular Crystallography, Part D, vol. 374. Academic Press,; 2003. p. 461–91.
[20] Lan J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 2020;581:215–20.
[21] Zhao X, Ghaffari S, Lodish H, Malashkevich VN, Kim PS. Structure of the Bcr-Abl oncoprotein oligomerization domain. Nat Struct Biol 2002;9:117–20.
[22] Nwachukwu JC, et al. Predictive features of ligand-specific signaling through the estrogen receptor. Mol Syst Biol 2016;12:864.
[23] Kwong PD, et al. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 1998;393:648–59.
[24] Glassman CR, et al. Structural basis for IL-12 and IL-23 receptor sharing reveals a gateway for shaping actions on T versus NK cells. Cell 2021;184:983–999.e24.
[25] Anil B, Riedinger C, Endicott JA, Noble ME. The structure of an MDM2-Nutlin-3a complex solved by the use of a validated MDM2 surface-entropy reduction mutant. Acta Crystallogr Biol Crystallogr 2013;69:1358–66.
[26] Lin DY, et al. The PD-1/PD-L1 complex resembles the antigen-binding Fv domains of antibodies and T cell receptors. Proc Natl Acad Sci U A 2008;105:3011–6.

[27] Penumutchu SR, Chou RH, Yu C. Structural insights into calcium-bound S100P and the V domain of the RAGE complex. PLoS One 2014;9:e103947.
[28] Jumper J, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9.
[29] Case DA, et al. The Amber biomolecular simulation programs. J Comput Chem 2005;26:1668–88.
[30] Maier JA, et al. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput 2015;11:3696–713.
[31] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79: 926–35.
[32] Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 1977;23:327–41.
[33] Wang H, Liu C, Deng L. Enhanced Prediction of Hot Spots at Protein-Protein Interfaces Using Extreme Gradient Boosting OPEN. Sci Rep 2018;8:14285.
[34] Wang J, Hou T, Xu X. Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. Curr Comput - Aided Drug Des 2006;2:287–306.
[35] Genheden S, Ryde U. Expert Opinion on Drug Discovery The MM / PBSA and MM / GBSA methods to estimate ligand-binding affinities The MM / PBSA and MM / GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov 2015;10.
[36] Siebenmorgen T, Zacharias M. Efficient Refinement and Free Energy Scoring of Predicted Protein–Protein Complexes Using Replica Exchange with Repulsive Scaling. J Chem Inf Model 2020;60:5552–62.
[37] Santini BL, Gaardløs M, Wyrzykowski D, Rothemund S, Penk A, Zacharias M, Samsonov SA. Rational design of glycosaminoglycan binding cyclic peptides using cPEPmatch. *Comput Struct Biotech J* 2024;23:2985–94.