Exploring a Structural Data Mining Approach to Design Linkers for Head-to-Tail Peptide Cyclization

Yasaman Karami, Samuel Murail, Julien Giribaldi, Benjamin Lefranc, Florian Defontaine, Olivier Lesouhaitier, Jérôme Leprince, Sjoerd de Vries, Pierre Tufféry

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

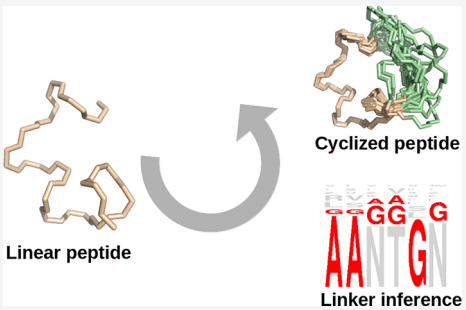
pubs.acs.org/jcim
Article

Exploring a Structural Data Mining Approach to Design Linkers for Head-to-Tail Peptide Cyclization

Yasaman Karami, Samuel Murail, Julien Giribaldi, Benjamin Lefranc, Florian Defontaine, Olivier Lesouhaitier, Jérôme Leprince, Sjoerd de Vries,* and Pierre Tufféry*

Cite This: J. Chem. Inf. Model. 2023, 63, 6436−6450
Read Online

ABSTRACT: Peptides have recently regained interest as therapeutic candidates, but their development remains confronted with several limitations including low bioavailability. Backbone head-to-tail cyclization, i.e., setting a covalent peptide bond linking the last amino acid with the first one, is one effective strategy of peptide-based drug design to stabilize the conformation of bioactive peptides while preserving peptide properties in terms of low toxicity, binding affinity, target selectivity, and preventing enzymatic degradation. Starting from an active peptide, it usually requires the design of a linker of a few amino acids to make it possible to cyclize the peptide, possibly preserving the conformation of the initial peptide and not affecting its activity. However, very little is known about the sequence−structure relationship requirements of designing linkers for peptide cyclization in a rational manner. Recently, we have shown that large-scale data-mining of available protein structures can lead to the precise identification of protein loop conformations, even from remote structural classes. Here, we transpose this approach to linkers, allowing head-to-tail peptide cyclization. First we show that given a linker sequence and the conformation of the linear peptide, it is possible to accurately predict the cyclized peptide conformation. Second, and more importantly, we show that it seems possible to elaborate on the information inferred from protein structures to propose effective candidate linker sequences constrained by length and amino acid composition, providing the first framework for the rational design of head-to-tail cyclization linkers. Finally, we illustrate this for two peptides using a limited set of amino-acids likely not to interfere with peptide function. For a linear peptide derived from Nrf2, the peptide cyclized starting from the experimental structure showed a 26-fold increase in the binding affinity. For urotensin II, a peptide already cyclized by a disulfide bond that exerts a broad array of biological activities, we were able, starting from models of the structure, to design a head-to-tail cyclized peptide, the first synthesized bicyclic 14-residue long urotensin II analogue, showing a retention of *in vitro* activity. Although preliminary, our results strongly suggest that such an approach has strong potential for cyclic peptide-based drug design.



INTRODUCTION

Several naturally occurring cyclic peptides constitute alternatives to antibiotics, and peptide backbone cyclization is frequently used in peptide-based drug design to convey druggable properties to linear bioactive sequences.[1,2] Peptides in general combine high affinity with high target selectivity and low toxicity and are a natural choice in the targeting of protein−protein interactions. While preserving these favorable properties, peptide cyclization additionally confers peptides with more rigid conformation and enhanced stability toward enzymatic proteolysis and improves the permeability through biological barriers.[3−9] Moreover, many natural-occurring cyclic peptides are known from different kingdoms of organisms, exhibiting diverse biological activities, including antitumor,[10,11] antimicrobial[12,13] and anthelmintic activities.[14−16] Together, this has caused a growing interest toward cyclic peptides, thus the number of designed cyclic peptide drugs is growing.[17,18] However, the design of cyclic peptides remains challenging, and particularly the compatibility with a desired structure.[9,19,20]

When designing new cyclic peptides, there are broadly two strategies that can be followed: (i) *de novo* design, or (ii) cyclization of an existing peptide. For the first strategy, a number of experimental techniques are available, such as SICLOPPS,[21] phage display, and mRNA display.[22] These are all based on libraries of random cyclic peptides that are subjected to an *in vitro* selection. They can be complemented with library-based computational approaches such as from Slough et al.,[23] CAESAR,[24] Omega,[25] and CycloPs[26] based on rdkit (https://github.com/rdkit/rdkit). Those approaches are conceptually similar to the molecular modeling of small ligands, with the corresponding strengths (arbitrary molecular topologies) and

Received: June 16, 2023
Published: October 12, 2023

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

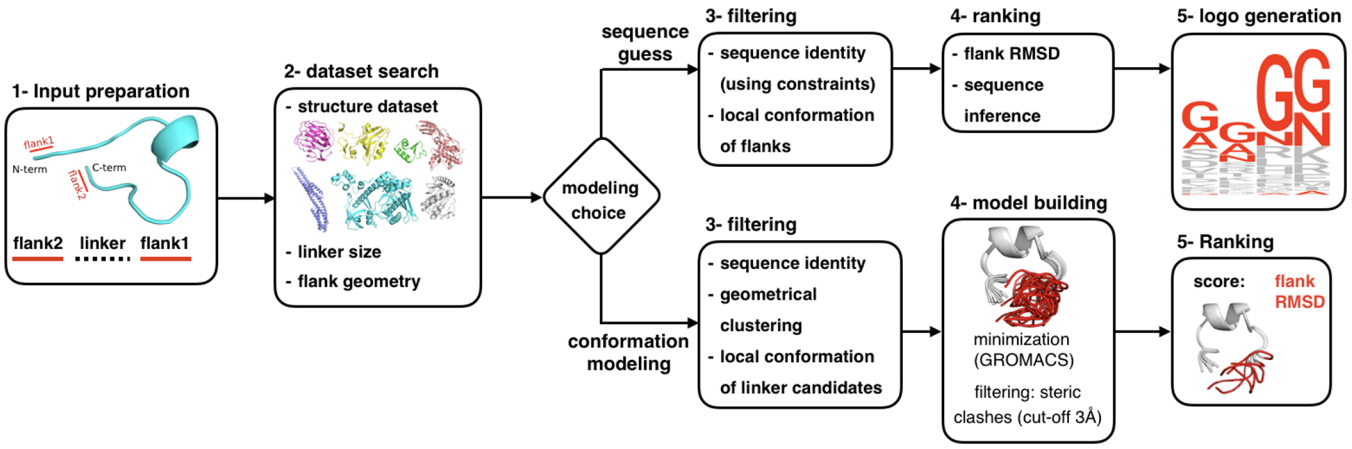


Figure 1. The workflow of PEP-Cyclizer. The workflow describes the main steps for peptide head-to-tail cyclization. The method provides two possibilities: proposing candidate sequences for the linker, or modeling the 3D conformation. The steps of the workflow are input preparation, linker candidate search, candidate filtering, model building, model selection, and logo generation in the case of sequence prediction. The inputs are a linear peptide and either the amino acid constraints for sequence prediction or the linker sequence for conformation modeling. In the final step, for conformation modeling, the 20 best models are returned as the final predictions. For sequence prediction a logo is generated and a forward-backtrack algorithm is used to sample the sequence space and assess the likelihood of the candidate linkers. Note that the sequence logo serves strictly as a global visualization of the ensemble of generated sequence candidates, and has no predictive power by itself.

weaknesses (limited number of flexible bonds). For computational *de novo design*, some recent approaches focusing on protein−protein interactions use hot spots identified at the interface to identify linker residues or template cyclic conformations matching the geometry of the hot spots.[27,28] They have not been experimentally validated so far. An alternative approach is to perform peptide structure prediction, using one of the many fragment-based methods that are available, such as PLOP,[29,30] Peplook,[31,32] PEPstrMOD,[33] or PEP-FOLD,[34,35] while imposing cyclization as a bond or distance restraint (see ref 36 for a review). Since these methods leverage the existing wealth of knowledge of protein and peptide structure, they can deal with larger peptides but have difficulties where this knowledge falls short, i.e., for unnatural amino acids.

For the second strategy, the starting point is an existing linear peptide of a known structure. It is well established that small linear peptides generally exist in solution in an interchangeable conformational equilibrium. This flexibility provides bioactive peptides the ability to interact with several types or subtypes of receptors, for instance. Stabilizing a bioactive conformation is a challenge that can be tackled by a variety of cyclization strategies. On the one hand, this can be as straightforward as mutating two spatially close residues into cysteins with the aim of introducing a disulfide bond. On the other hand, sophisticated chemical scaffolds or cyclotides can be used for the grafting or stitching of peptides or cyclotides into rigid bioactive conformations.[37,38] One particular successful strategy has been head-to-tail peptide backbone cyclization, i.e., the design of peptide in which a peptide bond connects the C-ter and N-ter amino acids.[39−48] Except for a few cases where the N- and C-terminal amino-acids are close enough to undergo direct cyclization, this involves the design of a sequence that links the N- and C-terminal extremities of the peptide. In principle, any amino acid can be part of the linker sequence, but Gly, Ala, and Pro residues are often favored because they are small and their side chains cannot form hydrogen bonds, which could potentially disrupt the bioactive conformation.

Head-to-tail cyclization leads to cyclic peptides with improved pharmacological properties (affinity, potency, efficiency, and selectivity) when compatible with target specificity (or bioactivity conservation). Whether the cyclic peptide is active or not, it is generally less sensitive to metabolic degradation. However, cyclization is often unsuccessful due to the imposed conformational restriction that is too strict and too far from the bioactive structure. In order to avoid this, it is necessary to understand the general sequence-structure requirements; in particular: what is the allowed sequence space of the linker and what will be the structure of the cyclized peptide? This is a challenging issue, and to the best of our knowledge, there is one computational protocol that has been successfully applied to head-to-tail cyclization linker design, namely the Rosetta protocol used by Bhardwaj et al.[6] In another recent study, the authors modified the AlphaFold framework for accurate design of cyclic peptides.[49] However, in both studies, the sequence and structure of the entire cyclic peptide were designed from scratch. At present, we do not know of any computational methods that can predict the sequence and structure of a head-to-tail cyclization linker while preserving the sequence and structure of the linear peptide that is being cyclized.

Recently, we have developed DaReUS-Loop,[50,51] a fast data-based approach that identifies loop candidates mining the complete set of available experimental protein structures. This is done by treating the loop as a gap in the structure and considering the flanking regions of the structure immediately before and after the gap. Loop candidates are then favored that (i) superimpose well onto the flanks, and (ii) have a compatible sequence. Recognizing the conceptual similarity, we developed PEP-Cyclizer, a method that extends the DaReUS-Loop approach and applies it to rational head-to-tail peptide cyclization. Given a conformation of the peptide to cyclize, it provides two complementary possibilities: (i) given a sequence for the linker, it can generate structural models for the cyclized peptide, leaving unaffected the original conformation of the linear peptide, (ii) it can propose candidate cyclization linker sequences, constrained by length and amino acid composition. For structure prediction, we have assessed the PEP-Cyclizer performance on a benchmark of five cyclic conotoxin structures for which a linear structure is available as well. With regard to the experimental structures, the predicted cyclized peptide models had a root-mean-square deviation (RMSD) of 2.0 Å (3.2 Å) for

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

Table 1. List of Real Cases for Head-to-Tail Cyclization[a]

uncyclized	no. NMR models	cyclized	no. NMR models	RMSD (Å)	uncyclized size	cyclized size	linker sequence
1m2c	14	2ajw	20	linear 1.22 ± 0.10	16	22	GGAAGG
1m2c	14	2ak0	20	1.09 ± 0.12	16	23	GAGGAAG
1mii	20	2ajw	20	1.26 ± 0.09	16	22	GGAAGG
1mii	20	2ak0	20	1.03 ± 0.12	16	23	GAGGAAG
2h8s	20	4ttl	1	0.40 ± 0.00	16	22	GGAAGG
1ixt	20	2mso	20	2.45 ± 0.07	27	30	GLP
2ew4	20	2j15	21	1.03 ± 0.40	13	15	AG
2ew4	20	-	-	-	13	16	RGD
1mxn (1mxp)	20	-	-	-	15	17, 19, 19	AG, AGGG, GGAA
2jut	20	-	-	-	13	19, 20	GGAAGG, GGAAGAG
1mvi	15	-	-	-	25	28	GLP

[a]The PDB code of the uncyclized and cyclized peptides (if available) are reported. With the exception of 4ttl, all the other structures are obtained using NMR and have several models. The average RMSD values are measured between all the models of the uncyclized and cyclized conformations. In some cases, more than one linker sequence exists, as reported in the last column of the table.

Table 2. Linker Modeling Quality Assessment[a]

			PEP-Cyclizer				Rosetta NGK			
name	*lsz*	N_{model}	$lRMSD^{*}_{20}$	$gRMSD^{*}_{20}$	$lRMSD^{*}_{1}$	$gRMSD^{*}_{1}$	$lRMSD^{*}_{20}$	$gRMSD^{*}_{20}$	$lRMSD^{*}_{1}$	$gRMSD^{*}_{1}$
2ew4	2	20	0.47 ± 0.13	2.03 ± 0.84	0.67 ± 0.18	3.57 ± 1.35	0.31 ± 0.22	2.53 ± 1.25	0.39 ± 0.26	3.10 ± 1.25
1ixt	3	20	0.46 ± 0.08	2.31 ± 0.25	1.43 ± 0.35	3.52 ± 1.50	0.37 ± 0.14	2.81 ± 0.54	0.43 ± 0.14	3.08 ± 0.66
1m2c	6	14	1.31 ± 0.15	1.99 ± 0.16	1.66 ± 0.12	2.58 ± 0.57	1.33 ± 0.17	4.58 ± 0.90	1.53 ± 0.16	6.08 ± 1.53
1mii[+]	6	20	1.35 ± 0.01	1.72 ± 0.01	1.75 ± 0.50	3.11 ± 1.74	1.56 ± 0.11	5.76 ± 0.55	1.72 ± 0.12	7.22 ± 0.45
2h8s	6	20	1.24 ± 0.12	2.12 ± 0.10	2.03 ± 0.29	4.05 ± 1.21	1.30 ± 0.19	3.01 ± 0.53	1.60 ± 0.21	3.81 ± 0.64
1m2c	7	14	1.59 ± 0.16	1.97 ± 0.25	2.09 ± 0.56	3.66 ± 1.56	1.84 ± 0.37	5.34 ± 1.30	2.12 ± 0.26	6.53 ± 1.29
1mii[+]	7	20	1.51 ± 0.08	1.89 ± 0.06	1.69 ± 0.01	2.30 ± 0.02	1.56 ± 0.03	5.51 ± 0.72	1.64 ± 0.06	7.80 ± 0.67
average[+]			1.01 ± 0.46	2.01 ± 0.43	1.49 ± 0.52	3.24 ± 1.26	0.95 ± 0.63	3.48 ± 1.39	1.13 ± 0.71	4.28 ± 1.77

[a]For each peptide, cyclization experiments were performed starting from each of the N_{model} experimental conformations of the linear peptide. For each, we calculated the local and global RMSD values (see Methods) and identified the top-1 and best out of top 20 values. The lRMSD measures the linker internal conformation deviation relative to the experimental one. The gRMSD measures the deviation of the positioning of the linker relative to that of the rest of the peptide. The RMSD values were calculated over the backbone atoms (N, C, Cα, and O). ($lRMSD^{*}_{1}$ and $gRMSD^{*}_{1}$) and ($lRMSD^{*}_{20}$ and $gRMSD^{*}_{20}$) correspond to the average and standard deviations over the respective N_{model} RMSD values. [+]: The structure of 1mii and 1m2c correspond to the same protein (α-Conotoxin MII), but cyclized structures are available for 2 different linker sizes (see Figure 2). We detail the results for 1m2c and 1mii separately, but for the average, to avoid redundancy, we considered the best predictions over 1m2c and 1mii.

the top 20 (top 1) models, an improvement of more than 1 Å over the Rosetta Next-generation KIC (NGK) protocol,[52] a high-resolution Rosetta protocol for the modeling of missing regions. For linker sequence prediction, PEP-Cyclizer was validated on the same benchmark, and as a result, experimental sequences were ranked significantly better than other sequences of the same length and composition.

As a functional validation, PEP-Cyclizer was used to design head-to-tail cyclized variants of the two peptides. The first one corresponds to the interaction of a 9-residue long fragment of Nrf2 in interaction with Keap1, as previously crystallized in linear form.[53] The second one corresponds to the human urotensin II (UII), that is an 11-residue long disulfide-bridged peptide.[54] UII exerts a broad array of biological activities, in particular, in the central nervous system, the cardiovascular system, and the kidney. It has been suggested that the cognate receptor of UII (UT), may emerge as a valuable and innovative therapeutic or diagnostic target.[55] Indeed, high affinity, potent and selective UT peptide ligands have been designed, from structure−activity relationship studies[56] to further elucidate the pharmacology and biology of UII toward new therapeutic opportunities, such as the treatment of sepsis-induced lung damage.[57] In this context, introduction of a main conformational restraint through head-to-tail cyclization has become a standard strategy to improve pharmacological profile of peptide ligands.[58]

RESULTS

The PEP-Cyclizer Protocol. PEP-Cyclizer considers all cyclization linker candidate structures that are compatible with the flanks of the uncyclized peptide structure. The sequences of these linker candidates, potentially filtered by *a priori* sequence constraints, are then used to build a linker sequence profile. This profile feeds a Hidden Markov Model from which it is possible to estimate the likelihood of candidate linkers using a forward-backtrack algorithm. Alternatively, if the linker candidates are restricted to one known linker sequence, they are clustered and superimposed onto the flanks, providing structural models of the cyclized peptide. Figure 1 depicts the workflow of the method. Note that when the linker size is not known, PEP-Cyclizer can estimate a minimum linker size required for head-to-tail cyclization based on the distance between the flanks of the uncyclized peptide (see below for the details).

Modeling Cyclic Peptides Given a Linker Sequence. As a positive control, PEP-Cyclizer was applied to 64 cyclic peptides from the CyBase database extracted on Oct. 28, 2018 (see Table S1 for a complete list of studied peptides). 1147 linear peptides were artificially generated by removing segments of 2−7 residues from the 64 cyclic peptides; details are reported in the Supporting Information section 'CyBase benchmark'. Unlike a real-world situation, where a peptide may undergo conformational changes upon cyclization, these artificial linear

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

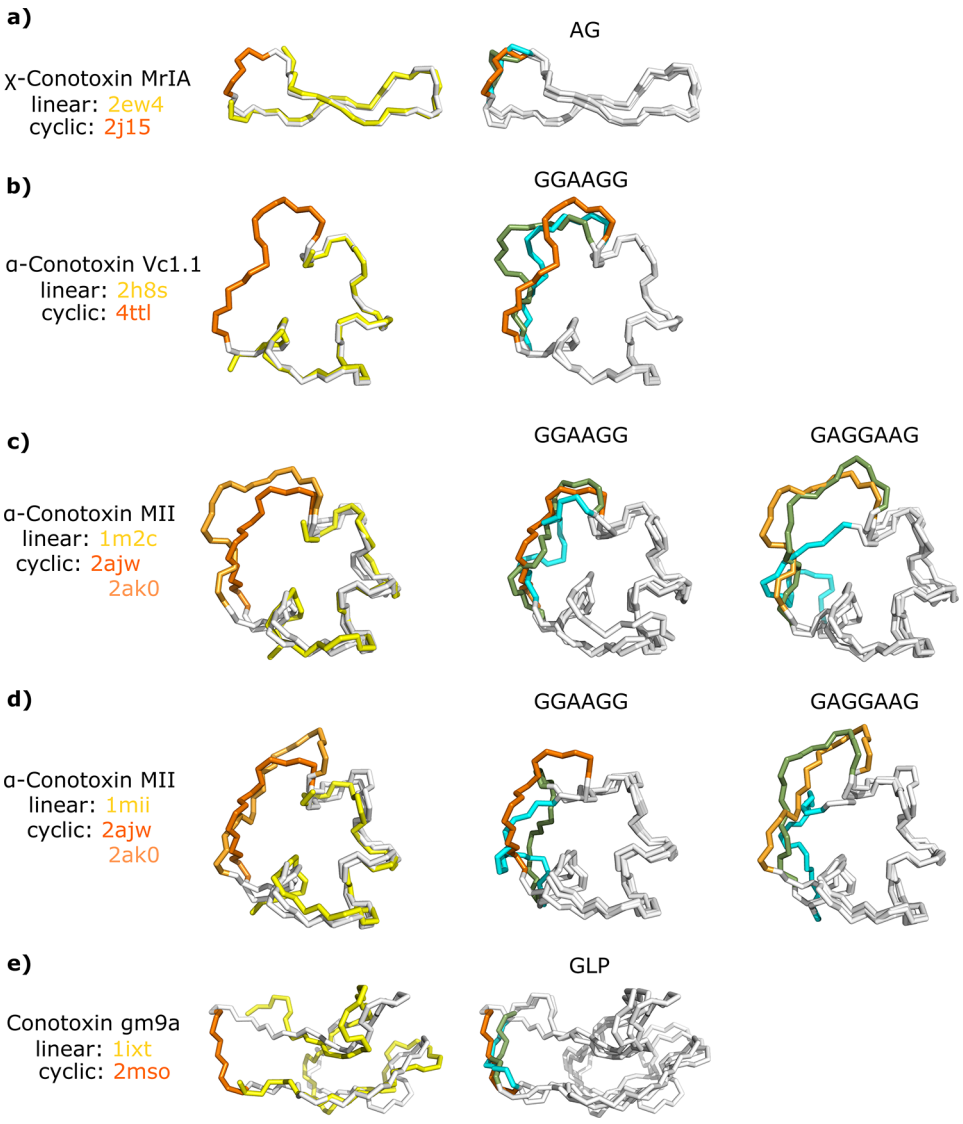


Figure 2. Structure of the studied linear conotoxins and their corresponding engineered cyclic peptides. The native linear and cyclic peptides are shown at the left column, colored yellow and orange, respectively. The structures on the middle and right columns represent the comparison between the native linker (in orange), linkers modeled by PEP-Cyclizer (in green), and Rosetta NGK (in cyan). See Table 2 for the corresponding $gRMSD_{20}$ (and $lRMSD_{20}$) values. The corresponding linker sequences are reported for every model at the top. (a): χ-Conotoxin MrIA, (b): α-Conotoxin Vc1.1, (c): α-Conotoxin MII, (d): α-Conotoxin MII, and (e): Conotocin gm9a.

peptides represent perfect conformations for modeling the removed linker conformation. For all linker sizes, PEP-Cyclizer was able to produce accurate models of the local linker conformation with an average accuracy of 1 Å or better. This is comparable (although not fully equal in accuracy) to models obtained for the same peptides using Rosetta NGK (comparisons reported per peptide and linker size in Tables S2 and S3, respectively).

PEP-Cyclizer was then applied to a small benchmark of real-world cases, in the form of several conotoxin peptides where both cyclized and noncyclized structures are available in the PDB. Seven distinct cyclized/noncyclized pairs of peptide structures were identified (Table 1 and Table S4). The range of backbone RMSD between the overlapping region of cyclized and uncyclized forms is 0.4−2.5 Å. Using the known linker sequence, only the noncyclized structure was used to model the linker. In this case, PEP-Cyclizer was able to return a model approximating the global structure of the linker at 2.01 Å on average (1.01 Å for the local linker conformation), as reported in Table 2. This is a considerable improvement over Rosetta NGK (3.48 Å), which suffers from structural imprecisions caused by conformational change upon cyclization. In contrast, our results show PEP-Cyclizer to be rather robust against such imprecisions. Figure 2 illustrates the results for the best predictions, out of the top 20, of PEP-Cyclizer (in green) and Rosetta NGK (in cyan), starting from the first NMR conformation of each uncyclized peptide. Overall, these results are compatible with those presented for DaReUS-Loop, where we found Rosetta NGK performing better when starting from an experimental backbone, and less accurate when starting from a somewhat divergent backbone. Finally, we have also considered the very recent evolution of AlphaFold for cyclized peptides.[49] For 2ew4, 1ixt, 1m2c/1mii (GGAAGG linker), 2h8s, and 1m2c/1mii (GAGGAAG linker)), we obtain $gRMSD^{*}_{20}$ values of 2.16, 1.49, 6.09, 3.32, and 3.50 Å, respectively, and $lRMSD^{*}_{20}$ of 0.42, 0.50, 1.99, 2.70 and 1.91, respectively. The corresponding averages are 2.91 and 1.50. As currently implemented, it seems AlphaFold is not as accurate as PEP-Cyclizer or RosettaNGK,

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

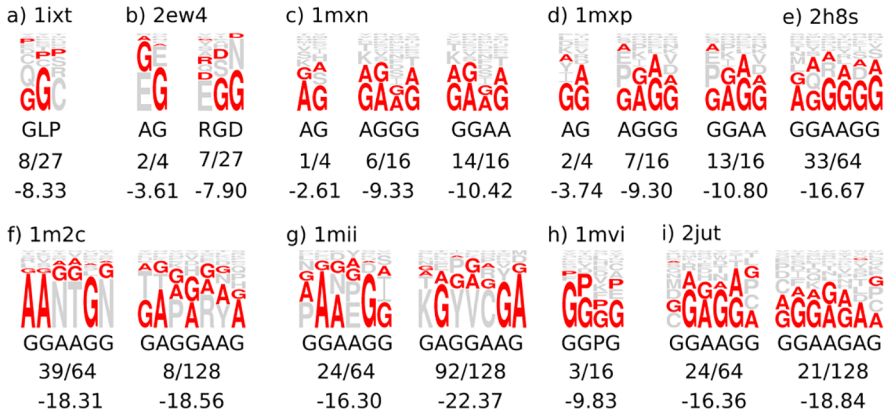


Figure 3. Sequence logo generated by PEP-Cyclizer for the studied cases. The pdb codes of the linear peptides used as input are reported for each case. Below every logo, the desired linker sequence, its rank, and score among the proposed sequences by the forward-backtrack algorithm are reported.

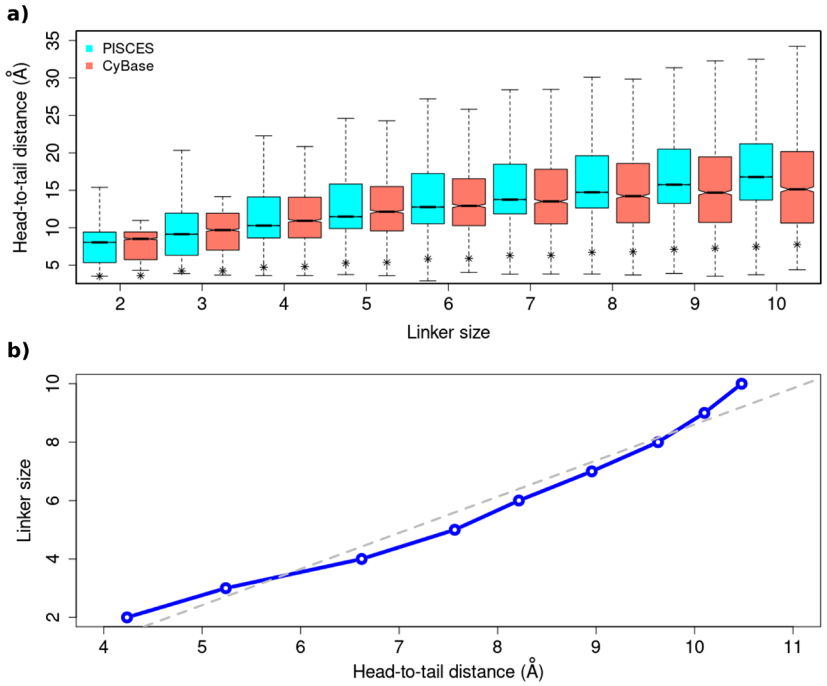


Figure 4. Linker size as a function of distance. (a) The distances between Cα atoms of every pair of nonconsecutive residues along each sequence with gap sizes ranging from 2 to 10 amino acids are reported for PISCES and CyBase data sets, in cyan and pink, respectively. The standard deviation values are shown with an asterisk for each box. (b) For every linker size, the *mean*−*sd* values calculated over both data sets are reported in blue. The fitted dashed line in gray corresponds to the final formula (eq 1) used here to predict a minimum linker size.

but it is important to note that AlphaFold performs the prediction for the complete peptide, whereas for PEP-Cyclizer and Rosetta NGK, we focused on the linker part.

Candidate Linker Sequence for a Given Linker Size. Next, the ability of PEP-Cyclizer to propose peptide linker sequences was tested. The same conotoxin benchmark was used, adding ten cyclic sequences with available structures for the uncyclized but not the cyclized peptide, for a total of 17 sequences. The details of the peptides are reported in Tables 1 and S4. As potential linker sequences, all combinations of all amino acids that are present in the experimental linker sequence (typically only Gly and Ala) were considered and ranked by the forward-backtrack algorithm. The results are shown in Figure 3. The experimental sequences were ranked significantly better (average percentile: 37.2, p = 0.024) than other potential sequences. It is important, however, to keep in mind that the structures of cyclized peptides with only a few linker sequences are available and that possible failures in either peptide synthesis or structure determination are not documented. Nevertheless, it is encouraging to observe that peptides cyclized with linkers well ranked by our approach could be successfully cyclized and adopt a conformation stable enough for structure determination.

Linker Size As a Function of Head-to-Tail Distance. Starting from a linear peptide, the size of the linker is not known and clues about likely linker sizes are needed. We focused on the estimation of a minimal linker size. Figure 4a shows the distribution of the inter-residue distance as a function of their separation in the sequence, for a separation from 2 to 10 residues, in a collection of experimental structures of the PDB and CyBase (see methods). As could be expected, one observes a rather linear relationship between the residue separation and the inter-residue distance. Figure 4b plots the relationship between the head and tail residue distance and the average linker

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

size reduced by one standard deviation (blue curve). The dashed line corresponds to a regression fit of the blue curve, and we find

$$L = \lfloor 1.24^{*}d - 3.77 \rfloor \quad (1)$$

, where d is the distance (in Å) between the flanks and L is the minimum linker size. It is in theory possible to make the relationship more or less stringent by varying the number of standard deviations, but probably considering a more complex model would be preferable. Considering Figure 4a, one observes that the minimum head-to-tail distance for all linker sizes is about 4−5 Å, which leads to the linker size of 2 using eq 1. Hence, the proposed formula is only a first approximation to estimate a potential minimal linker size for head-to-tail cyclization.

Application to the Design of Cyclized Nrf2 Peptides. PEP-Cyclizer was applied to predict a head-to-tail cyclization linker sequence for residues 76−85 of the Nrf2 protein. This peptide has been crystallized in interaction with the Keap1 protein.[53] The peptide conformation of the crystal structure was used as a starting point. For cyclization, a linker of size 3 was considered, according to the distance of 5.3 Å between the head and tail residues, accepting only alanine, proline, and glycines. Table 3 shows all 27 possible linkers ranked by PEP-Cyclizer. We selected two different linkers, namely the highly ranked AGG and the poorly ranked PAA.

Table 3. Likelihood of Each of the 27 Possible 27 Linker Sequences for Nrf2

linker	L
GGG	−5.71
AGG	−5.73
GAG	−6.32
PGG	−6.71
APG	−7.56
AAG	−7.65
GPG	−7.79
PAG	−7.84
PPG	−8.34
GGA	−8.71
AGA	−8.74
GGP	−9.66
AGP	−9.68
PGA	−9.71
GAP	−10.01
GAA	−10.63
PGP	−10.66
APP	−11.06
APA	−11.09
GPP	−11.30
GPA	−11.32
PPP	−11.85
PPA	−11.87
AAA	−11.97
PAA	−12.16

Three Nrf2-derived peptides were synthesized (see Methods), namely, the original linear $Nrf2_{(76-85)}$ peptide and the same peptide cyclized with the two linkers. The sequences of the peptides are listed in Figure 5.

Using microscale thermophoresis (MST), we then measured the binding affinities between his-tagged Keap1 protein and each of the synthesized peptides. The linear Nrf2 peptide (LV-5554) possessed an affinity for Keap1 with a K_d measured at 3.16 ± 0.64 μM Figure 6. Under the same conditions, LV-5565 (linker AGG) possessed a stronger affinity for the Keap1 protein with a K_d value of 0.12 ± 0.04 μM, 26 times better than that measured for linear Nrf2. Finally, the other cyclic peptide LV-5562 (linker sequence PAA) possessed an affinity slightly better than that of the linear peptide (K_d = 1.92 ± 0.44 μM). Therefore, the linker that is ranked highly by PEP-Cyclizer indeed showed a much higher affinity than the linker that is ranked last, although the latter still resulted in a cyclized peptide that had a binding affinity comparable to the linear peptide.

Application to Urotensin II. Finally, PEP-Cyclizer was applied to predict a head-to-tail cyclization linker sequence for UII. So far, only the structures of a fragment corresponding to the eight last amino acids of UII and its *N*-methylated tryptophan counterpart, $[(N\text{-Me})Trp^{7}]U\text{-}II_{4-11}$ in polar conditions (PDB entries 6HVB and 6HVC) have been solved by NMR. Since our goal was to obtain a head-to-tail cyclized version of UII, we decided to start from 3D models of the complete UII (11 amino acids) which includes one disulfide bond. Therefore, two ensembles of 8 and 5 conformations were generated using two distinct strategies: i) molecular dynamics simulations (MD) and ii) PEP-FOLD.[35] The models are highly structurally divergent (Supplementary Figure S2), with typical RMSD values in excess of 2 Å both within and between the ensembles (Supplementary Table S6). Consequently all of those models were used as the starting points for the cyclization (see Methods). The models of UII have head to tail distances ranging from approximately 4 up to 10 Å, dispatched in a bimodal manner, with 6 peptides having a distance of less than 6 Å, and the other having a distance of more than 7.8 Å. For 4 of the 6 peptides with a distance of less than 6 Å, a minimal size of 3 is suitable according to eq 1. Thus, we decided to consider a linker size of 3, accepting only alanine, proline, and glycines. Table 4 presents the results accumulated for each of the two ensembles of models. As can be observed, it is striking that despite the diversity of the conformations and the way they were obtained, those two independent ensembles of models resulted in a rather stable ranking of the predicted sequences. This is reflected by the fact that in both cases, the top 4 consist of the same four sequences as well as by the high overall correlation of the ranks (Spearman r = 0.98).

To test the significance of our approach, we analyzed the impact of top-, mid-, and bottom-ranked linkers on the agonist properties of the respective bicyclic UII analogues. Therefore, UII was cyclized with the sequences AGG, GAG, GGA, and GPA as displayed in Figure 5 leading to compounds LV-4131, LV-4130, LV-4132, and LV-4133, respectively (see Methods). Although obtained with low yield (<1%), each bicyclic analogue was highly pure. The pharmacological profile of these synthesis-challenging compounds was assessed by measuring their ability to increase intracellular calcium concentration ($[Ca^{2+}]i$) in human UT-transfected CHO cells (Eurofins-Cerep and Eurofins-screenFast) as previously described.[59] As shown in Figure 7, UII and LV-4130 induced a dose-dependent increase in $[Ca^{2+}]i$ with EC_{50} values of 0.7 and 46 nM, respectively. The other analogs were less potent than LV-4130 and exhibited EC_{50} values >6 μM. Noteworthy, LV-4130 is a first bicyclic UII analogue retaining a substantial ability to increase $[Ca^{2+}]i$ in UT-transfected CHO cells. Despite a shift in potency, the EC_{50} was less than 2 orders of magnitude lower, and LV-4130 is a nanomolar active UT agonist of peculiar interest. Indeed, its backbone cyclic structure may confer a lower susceptibility to

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

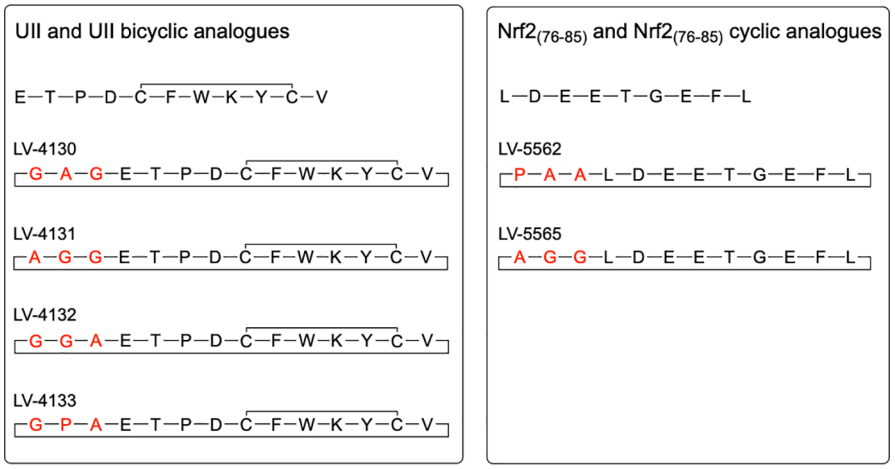


Figure 5. Sequence of human UII, human/mouse*Nrf*2$_{(76–85)}$ and head-to-tail cyclic analogues. Linker residues for ring closing are marked in red.

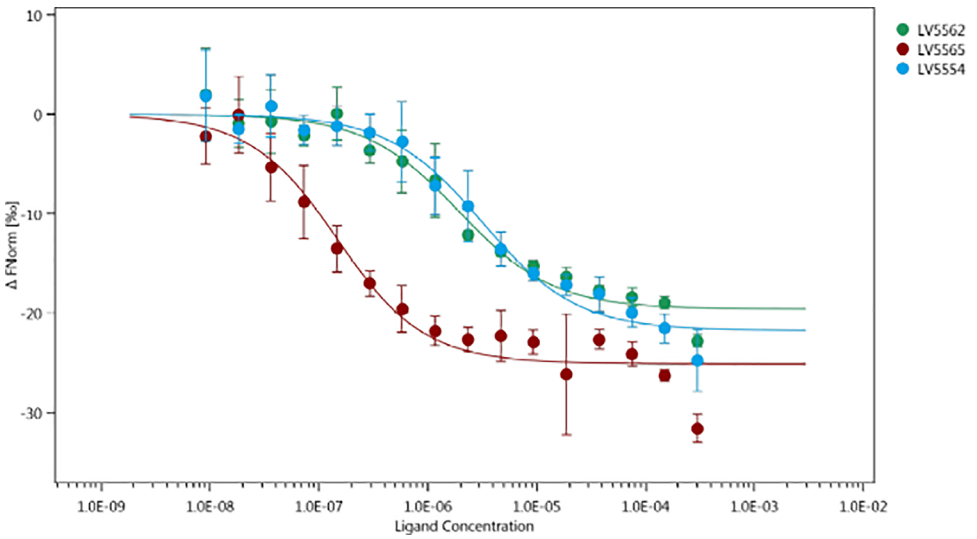


Figure 6. Dose response fit of MST for cyclized Nrf2 peptides. MST dose response curves for interaction between Keap1 labeled protein and the linear *Nrf*2$_{(76–85)}$ peptide (LV-5554, blue circles), the LV-5562 peptide (green circles) and the LV-5565 peptide (red circles). The data are the mean of three independent experiments for both cyclic (LV-5562 and LV-5565) peptides and of five independent experiments for the linear *Nrf*2$_{(76–85)}$ peptide (LV-5554). The thick lines correspond to the fit from the law of mass action. (Fl) fluorescently labeled; (FNorm) normalized fluorescence. ΔFnorm values of *Y*-axis in the dose response plot are calculated from the ratio of normalized fluorescence F0/F1, where F0 corresponds to the normalized fluorescence prior to MST activation. F1 is determined after an optimal MST power-dependent time interval, which yields the best signal-to-noise ratio.

metabolic degradation and a better selectivity for UT or a subset of UT's signaling cascade that deserves to be investigated: biased agonist concept.

DISCUSSION

Recently, we demonstrated that the current structural information available in the Protein Data Bank (PDB)[60] is sufficient to propose accurate protein loop candidates, in a manner that is robust for conformational inaccuracies. In the present study, we show that it is true for peptide cyclization linkers as well. We propose a computational framework to assist the design of head-to-tail cyclization of an existing peptide structure, a well-known strategy to enhance peptide resistance to enzymatic degradation and thus peptide bioavailability. The method addresses two complementary questions, namely, (i) proposing candidate sequences for the linker, a facility to assist medicinal chemists, and (ii) predicting the 3D conformation of the linker, for further peptide conformational stability analysis or peptide-receptor docking. Up until now, there has been an evident lack of computational methods to answer those questions. Existing methods[23−26,29−35] are oriented toward *de novo* design and do not perform head-to-tail cyclization of existing structures. We are aware of a single existing computational method, the Rosetta protocol by Bhardwaj et al.,[6] that is able to design head-to-tail cyclization linkers for pre-existing peptide structures. However, in that method, what is predefined is the complete structure of the entire cyclic peptide, including the linker; also, the sequence of the entire peptide is designed from scratch and not just that of the linker. In contrast, PEP-Cyclizer takes an existing structure of a linear peptide and predicts the sequence or structure of a cyclization linker, while leaving the rest of the peptide undisturbed. To the best of our knowledge, PEP-Cyclizer is the first computational method designed to do this. It has limitations though; particularly, the size of the linker is not presently guessed; it has to be specified. This is the subject of further investigations.

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

Table 4. Likelihood of Each of the Possible 27 Linker Sequences[a]

PEP-FOLD		MD	
linker	L	linker	L
AGG	−7.42	AGG	−7.21
APG	−7.65	GAG	−7.52
GAG	−7.67	PGG	−7.52
PGG	−7.73	APG	−7.61
AGA	−7.80	GGG	−7.64
PAG	−7.81	PAG	−7.69
GGG	−7.89	AAG	−7.78
AAG	−7.90	AGA	−7.80
PGA	−8.11	PGA	−8.10
GAP	−8.12	PPG	−8.11
PPG	−8.15	GAP	−8.21
AGP	−8.19	AGP	−8.22
PAP	−8.26	GGA	−8.23
GGA	−8.27	GPG	−8.36
APP	−8.29	PAP	−8.38
APA	−8.32	AAP	−8.47
AAP	−8.34	APA	−8.49
GPG	−8.44	APP	−8.50
PGP	−8.50	PGP	−8.53
GGP	−8.66	GGP	−8.65
GAA	−8.75	GAA	−8.79
PPP	−8.80	PAA	−8.96
PPA	−8.82	PPA	−8.99
PAA	−8.88	PPP	−9.00
AAA	−8.97	AAA	−9.06
GPP	−9.08	GPA	−9.24
GPA	−9.11	GPP	−9.25

[a]Two independent series of models (8 generated using MD and 5 using PEP-FOLD) were used as starting points.

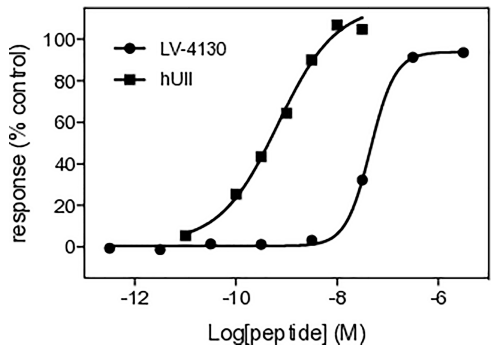


Figure 7. Concentration-dependent agonist-evoked Ca^{2+} responses on UT-transfected CHO cells. Agonist responses were expressed as a percent of the response observed with a maximally effective concentration of UII (100 nM). Data points represent the mean of duplicate.

The performance of the PEP-Cyclizer was validated on a set of conotoxins for which both linear and cyclic peptide structures are known. For comparison, we also evaluated a Rosetta protocol, not the one from Bhardwaj et al.,[6] but the Rosetta NGK protocol,[52] a state-of-the-art protocol for building missing loops in crystal structures. It must be mentioned that Rosetta NGK is not designed for peptide cyclization, and we had to modify the input data and convert the head-to-tail cyclization to loop modeling (i.e., dividing the peptides in two segments and switching them to generate a gapped structure). In all cases, the peptide linker models generated by PEP-Cyclizer had a significantly better global accuracy. This is especially evident for the two longest (7 amino acids) linkers, where the RMSD was <2 Å, while >5 Å for Rosetta NGK.

Finally, PEP-Cyclizer was applied to the design of two head-to-tail cyclized peptides, the first based on an Nrf2 fragment, the second on urotensin II. To avoid at maximum interference with peptide function, we considered linkers made of only a combination of 3 amino acids types: alanine, glycine, or proline. There are several significant differences between the two cases.

First, the Nrf2 peptide has been previously crystallized in linear form, in interaction with Keap1.[53] The known linear Keap1-bound peptide structure provides a straightforward starting point for PEP-Cyclizer in the design of a cyclic peptide. In contrast, for UII, the absence of a full-length structure made it necessary to perform molecular modeling in order to generate an ensemble of linear peptide structures.

Second, for Nrf2, its interaction with Keap1 provides an *in vitro* test, as it is possible to measure the affinity of binding of various peptide candidates using MST. In contrast, UII binds to a membrane receptor, making an *in vitro* test extremely challenging. However, a functional assay for UII is available in the form of an intracellular calcium concentration response to extracellular UII binding.

Third, for UII, the initial single cyclic peptide structure ensemble showed considerable conformational diversity, consistent with the knowledge that although the NMR structure of the disulfide-bridged core of UII is well-defined, the flanking linear extremities are very flexible.[61−64] Depending on the experimental environment (water or membrane mimetic micelles) and temperature, distinct conformations are stabilized within the disulfide-bridged core, involving different sets of intramolecular hydrogen bonds. This complicates the design of a head-to-tail cyclic peptide. Nevertheless, the fact that UII is a very potent molecule indicates that the loss of conformational entropy of the peptide upon binding is limited. In contrast, the Nrf2 peptide does not contain any disulfide bridges, opening up the possibility that head-to-tail cyclization may lead to rigidification, hence reduced loss of peptide conformational entropy upon binding, and hence improved affinity. Together with the known linear conformation, this makes Nrf2 a relatively straightforward case for PEP-Cyclizer; on the other hand, for UII, a cyclized peptide with demonstrated functional activity may have direct therapeutic relevance.

The differences between the two cases are reflected in the results. In both cases, using a linker predicted by PEP-Cyclizer, a head-to-tail cyclized peptide was synthesized, and its activity validated experimentally. For Nrf2, this resulted in an *in vitro* 26-fold increase of binding affinity, indicating that successful rigidification took place. In contrast, for UII, already rigidified through disulfide bond cyclization, head-to-tail cyclization with PEP-Cyclizer resulted in a loss of activity. However, as UII is very potent to begin with (0.7 nM), the cyclized UII is still a very strong UT ligand (46 nM). Peptide cyclization still has the expected benefit of increased stability against enzymatic degradation, paying a very modest price in terms of binding affinity.

PEP-Cyclizer is the extension of the DaReUS-Loop algorithm for the problem of head-to-tail peptide cyclization; details about the algorithm are reported in our previous study.[50] Essentially, the linker/loop is treated as a gap in the structure, and a structural database search is performed by using the flank regions on either side. Like DaReUS-Loop, PEP-Cyclizer is a

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

consensus method that considers both structural compatibility (i.e., good superposition of the linker candidate onto the flanks) and sequence compatibility. Therefore, when PEP-Cyclizer is used to predict linker conformations, it is essential to consider all 20 candidate structures. When the PEP-Cyclizer is forced to make a single prediction, the quality deteriorates considerably (from 2.0 to 3.2 Å). While a top-1 accuracy is naturally less favorable than a best-of-20 for any prediction method, it is specifically true for PEP-Cyclizer, as the effect is much weaker for Rosetta NGK (from 3.5 to 4.3 Å).

In contrast, the PEP-Cyclizer is shown to be very robust against conformational changes. For the conotoxin benchmark, the range of backbone RMSD between the overlapping region of cyclized and uncyclized forms is 0.4−2.5 Å. This is to be compared with the positive control, where the RMSD is zero. However, the global accuracy of the PEP-Cyclizer models is essentially the same between the two (2.0 vs 1.87 Å). This is a stark contrast to Rosetta NGK, which performs very well on the positive control (1.33 Å) but poorly on the real-world conotoxin benchmark (3.5 Å). This is an expected result, as Rosetta NGK is primarily designed to complete the missing regions in otherwise high-quality crystal structures. Note that as a high-resolution protocol Rosetta NGK does a good job in generating accurate local linker conformations; it is the global positioning of the linker onto the rest of the conotoxin structure where Rosetta NGK is outperformed by PEP-Cyclizer.

The robustness of PEP-Cyclizer for conformational change is also apparent for the prediction of linker sequences. For UII, sequence prediction was performed on two different structure ensembles that were of different origin and highly divergent with very similar results. Note that it is inherently complicated to evaluate linker sequence predictions, as we only have a few positive cases and no negatives, i.e., we normally do not know that a sequence does *not* cyclize. In addition, we must stress that PEP-Cyclizer proposes linker candidates based on likely sequence and structure only; in contrast, it cannot predict the effectiveness of the in vitro synthesis or the peptide conformational stability. Future research will focus on the prediction of the most likely length of the linker sequence, for which the current protocol does not show significant predictive power. Still, the result that experimental sequences were on average better ranked shows that PEP-Cyclizer has at least some predictive power. More importantly, the activity of the head-to-tail cyclic UII peptide LV-4130 demonstrates that PEP-Cyclizer has direct practical ability in cyclic peptide-based drug design.

MATERIALS AND METHODS

In this section we explain the details of PEP-Cyclizer, that is an extension of DaReUS-Loop, a data-based approach using remote or unrelated structures for loop modeling.[50,51] Starting from the geometry of flank residues, i.e., four residues before and four residues after the loop of interest, PEP-Cyclizer mines a structure database and identifies all possible candidates. It then integrates a filtering step and, in the end, ranks the candidates and proposes a final set of top models (structures or sequences). PEP-Cyclizer implements two complementary and new functionalities: (i) guessing the linker sequence and (ii) modeling the conformation of the linker. The details of those functionalities are depicted in Figure 1, and explained in the followings.

Structure Database. We employed two different structure databases. The first one is the database to search for linker candidates, which contains the entire set of protein structures available in the PDB. In March 2017, it consisted of 123,417 PDB entries, corresponding to 338,613 chains in total. The second database is the one to search for linker sequences and contains the entire set of protein structures available in the PDB70. For every database, each chain was split into segments that correspond to consecutive regions separated by gaps or nonstandard residues, but accepting seleno-methionines. This led to two databases with 758,143 and 172,693 protein segments, respectively.

Test Sets. To validate our approach, we have searched for cases for which both structures of the uncyclized and cyclized peptides are available. Backbone cyclization has been applied to few conotoxins, as reported in ref 65 and to the best of our knowledge, the structures (NMR/Xray) of only five engineered cyclic conotoxins for which the structure of the uncyclized form exists have been deposited in the PDB database.[60] For one of the cases, two structures of the open form have been deposited in the PDB (1m2c and 1mii), and their structures deviate by 1 Å,. We have included both structures in our test set. For 3 additional peptides, the structure of the uncyclized conformation and information about successful linkers are available. Table 1 reports the details of those studied cases. Of note, the structures of all the linear and cyclic peptides in this test set have been determined using NMR, with the exception of one case (4ttl) for which it has been solved by X-ray crystallography.

Since all the structures of the uncyclized forms of the peptides have been determined using NMR and have $N_{\text{uncyclized}}$ conformations, we have performed the head-to-tail cyclization starting from all $N_{\text{uncyclized}}$ NMR conformations. The final predictions for the cyclized forms of the peptides have been in turn compared with all of the N_{cyclized} conformations of the cyclized structures. Table 2 summarizes the average local and global $RMSD^{*}_{20}$ (best out of top 20) and $RMSD^{*}_{1}$ (top 1) values obtained for each linker (averaged over $N_{\text{uncyclized}}$ conformations).

Input Preparation and Candidate Search. We consider head-to-tail cyclization as a loop modeling problem, where the loop flanks are the first and the last four residues in the N-terminus and C-terminus, respectively. Accordingly, the minimum acceptable size for the input linear peptide is 8 residues. We then switch the flanks and search for linker candidates that match those flanks. We employ the method that was previously introduced to mine the database using a Binet-Cauchy (BC) kernel and a Rigidity score[66] (detail in Supporting Materials).

Predicting the Minimum Linker Size. To derive a minimal linker size as a function of the distance between head and tail residues, we have analyzed the distances between residues using a set of nonredundant protein chains from the PISCES server[67] (downloaded on July 20, 2023 from https://dunbrack.fccc.edu/pisces/download/) with 15% sequence identity, no chain breaks, resolutions better than 1 Å and determined by X-ray crystallography.[67] This collection was completed by all the natural cyclic peptides from the CyBase database (http://www.cybase.org.au/)[68,69] downloaded on Oct. 28, 2018. The inter-residue distances correspond to the distances between the Cα atoms of the residues.

Candidate Filtering. In most cases the number of candidates returned by BCLoopSearch is too large to be tractable, which implies limiting their number. Different filters were sequentially applied in our protocol for each mode of prediction:

Modeling the Conformation of the Linker.

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

- Sequence similarity: The sequence similarity of a linker candidate with the query linker sequence using BLOSUM62 score. Candidates with negative scores were discarded.
- Geometrical clustering: We used the python Numpy library to measure the pairwise distances (RMSD) between all the candidates.[70] In addition, we used the python Scipy package to perform hierarchical clustering.[71] A RMSD cutoff of 1 Å was used to group similar linker candidates. To consider memory constraints, we applied an iterative clustering over subsets of 25,000 candidates, until at most 25,000 clusters were obtained. Finally, one representative linker candidate with the highest sequence similarity to the query linker was selected for each cluster. The computational time of our clustering protocol is in the range of 1−5 min; however, it depends directly on the number of candidates detected by BCLoopSearch. In extreme cases, the needed time may increase up to 10−15 min.
- Local conformation: Previously, Shen et al. have shown that local conformation profiles predicted from sequence and profile−profile comparison can be employed to accurately distinguish similar structural fragments.[72] Consequently, we precomputed a collection of profiles for all the protein chains in the structure data set, and for all proteins of the test sets. For each linker candidate, it is thus possible to extract the subprofiles P and Q, corresponding to the query and candidate linker, and to measure the Jensen Shannon divergence ($JS(P, Q)$) between these profiles:

$$JS(P, Q) = \frac{1}{2}D_{KL}(P, M) + \frac{1}{2}D_{KL}(Q, M) \quad (2)$$

where M corresponds to $1/2(P + Q)$ and D_{KL} is the Kullback−Leibler divergence:

$$D_{KL}(P, Q) = \sum_{1\le i\le 27} P(i)\ln(P(i)/Q(i)) \quad (3)$$

$P(i)$ is the probability of SA letter i. Then we measured the average Jensen Shannon divergence (JSD) over the paired series of query and candidate profiles:

$$JSD(P, Q) = \sum_{1\le i\le n} JS(P_i, Q_i)/n \quad (4)$$

where P_i and Q_j are the two profiles corresponding to positions 1 to L on the query and candidate linker sequences. Note that a JSD of 0 indicates the perfect identity of the profiles. This procedure was applied on each linker candidate, and those with a $JSD > 0.40$ were discarded from the remaining set.
- Steric clash detection: After modeling the complete structure, models with steric clashes were discarded considering the Cα distance between linker residues and other residues of the protein, using a cutoff value of 3 Å.

Predicting the Linker Sequence.

- Sequence similarity: If sequence constraints are given, a subset of sequences that represent at least 50% sequence identity to any of the constraint amino acid types, regardless of their position, are kept.
- Local conformation: Measuring the local conformation of flanks (query and candidate flanks) and discarding candidates with flank $JSD > 0.40$.

Sequence Constraints. Throughout the study, linker sequences were predicted using the following sequence constraints. At each position of the linker, the set of amino acids of the entire experimental linker was considered, for instance, for the RGD linker of 2ew4, the amino acids Arg, Gly, and Asp were considered at all three positions; i.e., 3^3 different linker sequences are possible.

Model Building. Final energy minimization was conducted using Gromacs 2018,[73] the CHARMM36m force field[74] and the steepest descent algorithm for 1000 steps. All bonds were constrained by using the LINCS algorithm. The particle mesh Ewald algorithm was used to handle electrostatics with a 10 Å cutoff for the short-range part and a grid spacing of 1.2 Å for the long-range contribution in reciprocal space. The Verlet buffer scheme was used for nonbonded interactions, and the neighbor list was updated every 20 steps.

Model Selection. To rank the models, we considered the RMSD of the flanks. In the case of conformation modeling, our procedure returns a maximum of 20 models with the lowest *flank RMSD* score. And for sequence guessing, it returns a set of 30 sequences with the lowest *flank RMSD* score. From this set and considering the sequence constraints, we apply the sequence inference procedure (as explained below) to propose a final set of likely sequences for the linker.

Candidate Sequence Inference. To draw candidate sequences given the sequences of the candidate linkers identified, we used a forward-backtrack procedure. One advantage of such a procedure is to provide both sequences and their likelihood. The probabilities $p^l_{aa,linker}$ of observing each amino acid type aa at position l of the linker can be estimated from the amino acid sequences of the candidate linkers satisfying the condition of peptide cyclization. However, when a reduced number of amino acids are considered at a given position, these estimates can be performed on a rather low number of sequences. Consequently, we have estimated pseudofrequencies, with $p^l_{aa} = \alpha.p^l_{aa,linker} + (1 - \alpha).p^l_{aa,db}$ where α is a value between 0 and 1, and $p^l_{aa,db}$ is the frequency of amino acid type aa as observed in a large collection of sequences named db. For db, we have considered the sequences of the loops of 123,417 PDB entries (758,143 protein segments), identified using the procedure described in ref 50. Alternatively, we also considered db_s, which corresponds to the subset of db corresponding to a loop size of s. Transition probabilities have been estimated similarly. Pseudo transition probabilities $p(aa^l/aa^{l-1})$ were estimated as $p(aa^l/aa^{l-1}) = \beta.p(aa^l_{linker}/aa^{l-1}_{linker}) + (1 - \beta).p(aa^l_{db}/aa^{l-1}_{db})$, where β is a value between 0 and 1. Given estimates of p^l_{aa} and $p(aa^l/aa^{l-1})$ we have used the forward-backtrack algorithm to infer a series of amino acids that fit best the estimates. We prefer such procedure to for instance the $viterbi_{kbest}$ procedure that, in our experience,[75] usually returns less diverse sequences.

Linker Quality Assessment. To assess the quality of the final linker structures, we report the following RMSD values of the main chain heavy atoms (N, C, Cα, and O):

- Global RMSD ($gRMSD$): the modeled cyclic peptides are superimposed on the native structure excluding the linker region, then the RMSD is calculated over the linker.
- Local RMSD ($lRMSD$): RMSD is measured after performing the best fit superimposition of the linker region only.

Statistical Testing. To test the prediction of linker sequences of the conotoxin benchmark, the rank of the

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

experimental linker sequences was determined. To avoid pseudoreplication, five duplicate cyclic sequences were eliminated; using the remainder of the benchmark, the overall ranking of the experimental linker sequences was tested for statistical significance. With the total number of linker sequences varying from case to case and many instances of tied ranks, it was not feasible to compute an analytical *p*-value based on hypergeometric distributions. Instead, random ranks were simulated by sampling from flat rank distributions and converted to percentiles, and it was evaluated how often the overall mean percentile was better than the observed mean percentile (37.2) for the experimental linker sequences. This was the case in 2431/100000 random simulations, i.e., a *p*-value of 0.024.

Comparison with Other Approaches. In this work we compare the performance of our linker modeling protocol with the Rosetta NGK.[52] The Rosetta NGK runs were performed using the protocol provided by ref 52, and Rosetta energy values were employed for ranking the models. Considering the fact that Rosetta NGK is not designed for peptide cyclization, we converted the head-to-tail cyclization to loop modeling, by breaking every peptide into two segments and switching the two.

Experimental Tests. To test our procedure, we searched for linear peptides known to interact with a protein and for which some quantification of the impact of the cyclization can be done. Since the aim of our approach to design linkers is to perturb as little as possible the peptide conformation, while possibly rigidifying it, a straightforward possibility was to search the Protein Data Bank to identify peptide protein complexes solved at high resolution, for which the peptide does not contain unusual amino acids and for which the protein is available commercially so that it is possible to undergo affinity measurements between the different peptides and the protein. Two such systems were identified and correspond, respectively, to PDB entries 1x2r and 2qos. The first one corresponds to the interaction of a fragment of Nrf2 of 9 amino acids (sequence: LDEETGEFL) in an interaction with Keap1, an interaction spotted in the context of lung cancer. The second one corresponds to the C8 binding site, a peptide of 11 amino acids (sequence: LRYDSTAERLY) interacting with the complement protein C8. However, this peptide led to aggregation when tested for MST and did not lead to any exploitable results. A final system corresponds to urotensin II, a peptide of 11 amino acids (sequence: ETPDCFWKYCV) known to interact with a membrane receptor, for which functional tests of ligand-stimulated intracellular calcium response are available commercially.

Urotensin II Model Generation. Two sets of 3D models were used. The first one was generated using PEP-FOLD server,[35] a *de novo* approach to peptide structure prediction. Five independent runs of 3D generation (100 models) were run, and five models showing closed disulfide bonds in the PEP-FOLD coarse grained representation were then submitted to refinement using MD, with the aim of stabilizing the disulfide bond in the all atom representation. The model topology was created by using the Gromacs pdb 2gmx command, which did not include the disulfide bond. The topology was further modified to include the disulfide bond parameter using the gromacs_py library.[76] Simulations were performed using the CHARMM-36 force field[77] and the TIP3P model for water. The Gromacs 2018 software[73] was used to run the simulations. The five models were minimized two times for 10,000 steps with the steepest descent algorithm. During the first minimization, the bonds were not constraints, as in the second and following steps, all bonds were constrained using the LINCS algorithm. The five models were solvated in a water box and roughly 150 mM of NaCl. Systems were again minimized in two similar steps, then equilibrated in three successive steps: (i) 100 ps with position restraints of 1000 kJ mol^{-1} nm^{-2} applied on the peptide heavy atoms and an integration time step of 1 fs, (ii) 500 ps with position restraints of 1000 kJ mol^{-1} nm^{-2} applied on the peptide Cα atoms, the integration time step was fixed to 2 fs, (iii) 1 ns with position restraints of 100 kJ mol^{-1} nm^{-2} applied on the Cα atoms. Production runs were finally computed at 100 ns. The five 100 ns trajectories were then analyzed using MDAnalysis library.[78] PCA of backbone atom coordinates was computed, and the 15 first components were used to cluster the coordinates. The clustering DBSCAN algorithm[79] was used using a min_sample of 20, and sigma value of 5. A total of 13 clusters was identified, and the cluster centroids were chosen by taking the closest element in terms of RMSD to the average cluster structure. The conformations generated using this protocol are available as Supporting Information. All models underwent sequence guessing to cyclize the peptide.

Another set of models was kindly provided by Chatenet and co-workers at INRS Quebec, Canada. It consists of a set of 8 representative structures of UII displaying the heterogeneous conformational ensemble of this peptide. The three-dimensional structure of UII was generated from the sequence using the pdb utilities server https://spin.niddk.nih.gov/bax/nmrserver/pdbutil/ (performed in 2020). System preparation and MD simulations were performed using AMBER v16[80] and the ff14SB force field.[81] Simulations were performed at 300 K under constant energy (NVE) conditions using a 2 fs time step. The peptide was solvated using the SPC(E) water model in a rectangular box with periodic boundary conditions. The system was neutralized through the addition of counterions (Na^+). The preprocessing steps were followed by equilibration steps, as described previously.[82] All simulations were performed using the GPU-enabled version of the AMBER simulation engine pmemd. A Particle Mesh Ewald cutoff of 8 Å was used for the GPU-enabled simulations.[83] The peptide was simulated for a total of 100 ns. Representative structures were selected by clustering simulation ensembles obtained from the MD simulation trajectory. Clustering was performed using the hierarchical agglomerative approach with an epsilon cutoff of 3 Å, which represents the minimum distance between the clusters.

Peptide Synthesis and Functional Test. Linear peptide precursors of human UII (LV-4130, LV-4131, LV-4132, and LV-4133) and human/mouse $Nrf2_{(76-85)}$ (LV-5562 and LV-5565) cyclic analogues were synthesized by Fmoc solid phase methodology on a Liberty microwave-assisted automated peptide synthesizer (CEM, Saclay, France) using the standard manufacturer's procedure at 0.1 mmol scale on a preloaded Fmoc-Asp(Wang resin)-ODmab as previously described.[84] Reactive side chains were protected as follow: Thr, Tyr, *tert*-butyl (tBu) ether; Glu, *tert*-butyl (OtBu) ester; Lys, Trp, *tert*-butyloxycarbonyl (Boc) carbamate; Cys, trityl (Trt) thioether. After completion of the chain assembly, the C-terminal Dmab protective group was selectively removed by addition of a solution of 2% hydrazine in DMF for 3 min.[85] Treatment was repeated twice, and the resin was washed with DMF and DCM. Head-to-tail cyclization was performed on-resin by in situ activation of the free carboxyl group with BOP (1.1 equiv) and DIEA (10 equiv) in 10 mL of DMF at room temperature for about 72 h. The head-to-tail cyclic peptides were deprotected

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

and cleaved from the resin by adding 10 mL of the mixture TFA/TIS/H_2O (9.5:0.25:0.25) for 180 min at room temperature. After filtration, crude peptides were washed three times by precipitation in TBME followed by centrifugation (4500 rpm, 15 min). The synthetic peptides were purified by reversed-phase HPLC on a 21.2 mm × 250 mm Jupiter C18 (5 μm, 300 Å) column (Phenomenex, Le Pecq, France) using a linear gradient (10−50% over 45 min) of acetonitrile/TFA (99.9:0.1) at a flow rate of 10 mL/min. The disulfide bridge of UII analogues was then formed by treatment of the head-to-tail cyclic peptides with a mixture of *N*-chlorosuccinimide (1.05 equiv) in 20 mL of H_2O/CH_3CN (1:1) for 30 min at room temperature as previously described.[86] The resulting bicyclic UII analogues were purified as described above with a 20−60% linear gradient. The purified peptides were finally characterized by MALDI-TOF mass spectrometry on an ultrafleXtreme (Bruker, Strasbourg, France) in the reflector mode using α-cyano-4-hydroxycinnamic acid as a matrix. Analytical RP-HPLC, performed on a 4.6 mm × 250 mm Jupiter C18 (5 μm, 300 Å) column, indicated that the purity of all peptides was >95%. Purity analyses of 4 peptides are reported in Supplementary Figure S1.

Intracellular Calcium Assay. Ligand-stimulated intracellular calcium responses were measured at the human UT receptor expressed in transfected CHO cells using a fluorimetric detection method according to Eurofins-Cerep (catalog reference G099-1376) and EuroscreenFast (catalog reference FAST-0540A) standard assay protocols. The assays were performed in duplicate or triplicate. The results were expressed as a percent of human UII response at its EC_{100} concentration and plotted using Prism software (GraphPad, San Diego, CA).

Microscale Thermophoresis. Keap1 (His-tagged) was purchased from Tebu-Bio (Le Perray en Yvelines, France). This His-tagged protein was labeled using the Labeling Kit RED-tris-NTA second Generation Monolith (MO-L018, NanoTemper Technologies GmbH, Germany). According to the manufacturer's instructions we have first evaluated the affinity of the dye toward the His-Keap1, by preparing 16 dilution points of the his-Keap1 protein in PBS-T (consisting of PBS supplied with the labeling kit to which has been added 0.05% of tween 20) (from 4 μM to 0.12 nM; 10 μL each) and finally adding 10 μL of the RED-tris-NTA second generation dye (50 nM) (in PBS-T) in each tube. After 30 min of incubation at room temperature, the affinity (K_d) of the RED-tris-NTA second generation dye for Keap1 His-tag was measured using the Monolith NT.115Pico instrument. As recommended by the manufacturer, since we measured a K_d at 18 nM, we have chosen to label His-Keap1 as follows for all experiments dedicated to the affinity measurement of His-Keap1 with the different putative ligand. Briefly, we adjusted 90 μL of the Keap1 concentration at 362 nM and added 90 μL of RED-tris-NTA second generation dye (100 nM). After 30 min of incubation at room temperature, the sample was centrifugated (10 min; 4 °C; 15,000*g*) and then the supernatant was collected for binding assay. Concerning the binding assay experiments, the linear $Nrf2_{(76-85)}$ peptide, LV-5562 and LV-5565 were diluted in pure H_2O at 600 μM. Twenty μL of these stock solutions were placed in tube 1 and a series of 1:1 dilutions were prepared in PBS-T, in order to obtain a ligand concentration ranging from 600 μM to 18.32 nM (16 points). Afterward, 10 μL of the labeled protein (100 nM) was added in each 16 tubes containing the ligand at the 16 different concentration. Finally, for MST measurement the final concentration was 300 μM to 9.16 nM of ligand and 50 nM of His-Keap1 labeled protein. After this preparation, each solution was filled into Monolith NT standard capillaries. After loading the 16 capillaries into the Monolith NT.115Pico instrument, a scan of the fluorescence count for all capillaries was carried to check for consistent fluorescence while confirming the absence of ligand induced fluorescence changes or adsorption. MST was measured using a Monolith NT.115Pico instrument at an ambient temperature of 25 °C with 5 s/30 s/5 s laser off/on/off times, respectively. Instrument parameters were adjusted with 5% LED power and 40% MST power. Data of at least three independently pipetted measurements were analyzed (MO.Affinity Analysis software version 1.5.41, NanoTemper Technologies GmbH) using FHOT at 5 s. The data were fitted using the law of mass action from GraphPad Prism version 5, and MST figures were generated using MO.Affinity Analysis.

Availability of Materials and Data. All models generated for the conotoxins, Nrf2 and UII test cases, and the list of the template conformations identified from the subset of the Protein Data Bank are available as Supporting Information. Docker images and scripts making possible to execute PEP-Cyclizer are freely available at https://gitlab.rpbs.univ-paris-diderot.fr/tuffery/pep-cyclizer-dist.

■ ASSOCIATED CONTENT

Data Availability Statement

All models generated for the conotoxins, Nrf2 and UII test cases, as well as the list of the template conformations identified from the subset of the Protein Data Bank are available as Supporting Information. The source code to generate docker images and scripts to run them are available at: https://gitlab.rpbs.univ-paris-diderot.fr/tuffery/pep-cyclizer-dist.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.3c00865.

Database search, the CyBase benchmark, compound synthesis, and 3D models used for UII (PDF)

Conformations generated (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

Sjoerd de Vries − *Université Paris Cité, CNRS UMR 8251, INSERM ERL U1133, 75013 Paris, France*; Email: sjoerd.de-vries@inserm.fr

Pierre Tufféry − *Université Paris Cité, CNRS UMR 8251, INSERM ERL U1133, 75013 Paris, France*; orcid.org/0000-0003-1033-9895; Email: pierre.tuffery@u-paris.fr

Authors

Yasaman Karami − *Université Paris Cité, CNRS UMR 8251, INSERM ERL U1133, 75013 Paris, France*; orcid.org/0000-0001-8413-2665

Samuel Murail − *Université Paris Cité, CNRS UMR 8251, INSERM ERL U1133, 75013 Paris, France*; orcid.org/0000-0002-3842-5333

Julien Giribaldi − *Institut des Biomolécules Max Mousseron, UMR 5247, Université de Montpellier-CNRS, 34293 Montpellier, France*

Benjamin Lefranc − *Université de Rouen Normandie, INSERM U1239 NorDiC, Neuroendocrine, Endocrine and Germinal Differentiation and Communication, INSERM US51 HeRacLeS, F-76000 Rouen, France*

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

Florian Defontaine − *Université de Rouen Normandie, UR CBSA, Research Unit Bacterial Communication and Anti-infectious Strategies, 27000 Evreux, France*
Olivier Lesouhaitier − *Université de Rouen Normandie, UR CBSA, Research Unit Bacterial Communication and Anti-infectious Strategies, 27000 Evreux, France*
Jérôme Leprince − *Université de Rouen Normandie, INSERM U1239 NorDiC, Neuroendocrine, Endocrine and Germinal Differentiation and Communication, INSERM US51 HeRacLeS, F-76000 Rouen, France;* orcid.org/0000-0002-7814-9927

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.3c00865

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Funding from ANR-10-BINF-0003 (BipBip); ANR-14-2011-IFB; INSERM [U 1133]; Ressource Parisienne en Bioinformatique Structurale (RPBS) is acknowledged. The authors thank David Chatenet, Nicolas Doucet, and Chitra Narayanan, INRS, Quebec, Canada for communicating sample conformations of UII.

■ REFERENCES

(1) Zorzi, A.; Deyle, K.; Heinis, C. Cyclic peptide therapeutics: past, present and future. *Curr. Opin Chem. Biol.* **2017**, *38*, 24−29.
(2) Falanga, A.; Nigro, E.; De Biasi, M. G.; Daniele, A.; Morelli, G.; Galdiero, S.; Scudiero, O. Cyclic Peptides as Novel Therapeutic Microbicides: Engineering of Human Defensin Mimetics. *Molecules* **2017**, *22*, 1217.
(3) Daly, N. L.; Craik, D. J. Acyclic permutants of naturally occurring cyclic proteins. Characterization of cystine knot and beta-sheet formation in the macrocyclic polypeptide kalata B1. *J. Biol. Chem.* **2000**, *275*, 19068−19075.
(4) Barry, D. G.; Daly, N. L.; Clark, R. J.; Sando, L.; Craik, D. J. Linearization of a naturally occurring circular protein maintains structure but eliminates hemolytic activity. *Biochemistry* **2003**, *42*, 6688−6695.
(5) Zhang, Y.; Zhang, Q.; Wong, C. T. T.; Li, X. Chemoselective Peptide Cyclization and Bicyclization Directly on Unprotected Peptides. *J. Am. Chem. Soc.* **2019**, *141*, 12274.
(6) Bhardwaj, G.; Mulligan, V. K.; Bahl, C. D.; Gilmore, J. M.; Harvey, P. J.; Cheneval, O.; Buchko, G. W.; Pulavarti, S. V.; Kaas, Q.; Eletsky, A.; Huang, P. S.; Johnsen, W. A.; Greisen, P. J.; Rocklin, G. J.; Song, Y.; Linsky, T. W.; Watkins, A.; Rettie, S. A.; Xu, X.; Carter, L. P.; Bonneau, R.; Olson, J. M.; Coutsias, E.; Correnti, C. E.; Szyperski, T.; Craik, D. J.; Baker, D. Accurate de novo design of hyperstable constrained peptides. *Nature* **2016**, *538*, 329−335.
(7) Wang, C. K.; Craik, D. J. Designing macrocyclic disulfide-rich peptides for biotechnological applications. *Nat. Chem. Biol.* **2018**, *14*, 417−427.
(8) Buckton, L. K.; Rahimi, M. N.; McAlpine, S. R. Cyclic peptides as drugs for intracellular targets: The next frontier in peptide therapeutic development. *Chemistry−A European Journal* **2021**, *27*, 1487−1513.
(9) Huang, H.; Damjanovic, J.; Miao, J.; Lin, Y. S. Cyclic peptides: backbone rigidification and capability of mimicking motifs at protein-protein interfaces. *Phys. Chem. Chem. Phys.* **2021**, *23*, 607−616.
(10) Lindholm, P.; Göransson, U.; Johansson, S.; Claeson, P.; Gullbo, J.; Larsson, R.; Bohlin, L.; Backlund, A. Cyclotides: a novel type of cytotoxic agents. *Mol. Cancer Ther* **2002**, *1*, 365−369.
(11) Tang, J.; Wang, C. K.; Pan, X.; Yan, H.; Zeng, G.; Xu, W.; He, W.; Daly, N. L.; Craik, D. J.; Tan, N. Isolation and characterization of cytotoxic cyclotides from Viola tricolor. *Peptides* **2010**, *31*, 1434−1440.
(12) Tam, J. P.; Lu, Y. A.; Yang, J. L.; Chiu, K. W. An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 8913−8918.
(13) Gran, L.; Sletten, K.; Skjeldal, L. Cyclic peptides from Oldenlandia affinis DC. Molecular and biological properties. *Chem. Biodivers* **2008**, *5*, 2014−2022.
(14) Colgrave, M. L.; Kotze, A. C.; Ireland, D. C.; Wang, C. K.; Craik, D. J. The anthelmintic activity of the cyclotides: natural variants with enhanced activity. *Chembiochem* **2008**, *9*, 1939−1945.
(15) Colgrave, M. L.; Kotze, A. C.; Huang, Y. H.; O'Grady, J.; Simonsen, S. M.; Craik, D. J. Cyclotides: natural, circular plant peptides that possess significant activity against gastrointestinal nematode parasites of sheep. *Biochemistry* **2008**, *47*, 5581−5589.
(16) Dahiya, R.; Singh, S.; Sharma, A.; Chennupati, S. V.; Maharaj, S. First Total Synthesis and Biological Screening of a Proline-Rich Cyclopeptide from a Caribbean Marine Sponge. *Mar. Drugs* **2016**, *14*, 228.
(17) Jing, X.; Jin, K. A gold mine for drug discovery: Strategies to develop cyclic peptides into therapies. *Med. Res. Rev.* **2020**, *40*, 753−810.
(18) Zhang, H.; Chen, S. Cyclic peptide drugs approved in the last two decades (2001−2021). *RSC Chem. Biol.* **2022**, *3*, 18−31.
(19) Slough, D. P.; McHugh, S. M.; Lin, Y. S. Understanding and designing head-to-tail cyclic peptides. *Biopolymers* **2018**, *109*, No. e23113.
(20) Miao, J.; Descoteaux, M. L.; Lin, Y. S. Structure prediction of cyclic peptides by molecular dynamics + machine learning. *Chem. Sci.* **2021**, *12*, 14927−14936.
(21) Tavassoli, A. SICLOPPS cyclic peptide libraries in drug discovery. *Curr. Opin Chem. Biol.* **2017**, *38*, 30−35.
(22) Passioura, T.; Katoh, T.; Goto, Y.; Suga, H. Selection-based discovery of druglike macrocyclic peptides. *Annu. Rev. Biochem.* **2014**, *83*, 727−752.
(23) Slough, D. P.; McHugh, S. M.; Cummings, A. E.; Dai, P.; Pentelute, B. L.; Kritzer, J. A.; Lin, Y. S. Designing Well-Structured Cyclic Pentapeptides Based on Sequence-Structure Relationships. *J. Phys. Chem. B* **2018**, *122*, 3908−3919.
(24) Li, J.; Ehlers, T.; Sutter, J.; Varma-O'brien, S.; Kirchmair, J. CAESAR: a new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration. *J. Chem. Inf Model* **2007**, *47*, 1923−1932.
(25) Hawkins, P. C.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf Model* **2010**, *50*, 572−584.
(26) Duffy, F. J.; Verniere, M.; Devocelle, M.; Bernard, E.; Shields, D. C.; Chubb, A. J. CycloPs: generating virtual libraries of cyclized and constrained peptides including nonnatural amino acids. *J. Chem. Inf Model* **2011**, *51*, 829−836.
(27) Santini, B. L.; Zacharias, M. Rapid in silico design of potential cyclic peptide binders targeting protein-protein interfaces. *Frontiers in chemistry* **2020**, *8*, 573259.
(28) Delaunay, M.; Ha-Duong, T. Des3PI: a fragment-based approach to design cyclic peptides targeting protein−protein interactions. *Journal of Computer-Aided Molecular Design* **2022**, *36*, 605−621.
(29) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *Proteins* **2004**, *55*, 351−367.
(30) Mandell, D. J.; Coutsias, E. A.; Kortemme, T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods* **2009**, *6*, 551−552.
(31) Thomas, A.; Deshayes, S.; Decaffmeyer, M.; Van Eyck, M. H.; Charloteaux, B.; Brasseur, R. Prediction of peptide structure: how far are we? *Proteins* **2006**, *65*, 889−897.
(32) Beaufays, J.; Lins, L.; Thomas, A.; Brasseur, R. In silico predictions of 3D structures of linear and cyclic peptides with natural and non-proteinogenic residues. *J. Pept Sci.* **2012**, *18*, 17−24.

(33) Singh, S.; Singh, H.; Tuknait, A.; Chaudhary, K.; Singh, B.; Kumaran, S.; Raghava, G. P. PEPstrMOD: structure prediction of peptides containing natural, non-natural and modified residues. *Biol. Direct* **2015**, *10*, 73.
(34) Thévenet, P.; Shen, Y.; Maupetit, J.; Guyon, F.; Derreumaux, P.; Tufféry, P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res.* **2012**, *40*, W288−293.
(35) Shen, Y.; Maupetit, J.; Derreumaux, P.; Tufféry, P. Improved PEP-FOLD Approach for Peptide and Miniprotein Structure Prediction. *J. Chem. Theory Comput* **2014**, *10*, 4745−4758.
(36) McHugh, S. M.; Rogers, J. R.; Solomon, S. A.; Yu, H.; Lin, Y. S. Computational methods to design cyclic peptides. *Curr. Opin Chem. Biol.* **2016**, *34*, 95−102.
(37) Poth, A. G.; Chan, L. Y.; Craik, D. J. Cyclotides as grafting frameworks for protein engineering and drug design applications. *Biopolymers* **2013**, *100*, 480−491.
(38) Valeur, E.; Guéret, S. M.; Adihou, H.; Gopalakrishnan, R.; Lemurell, M.; Waldmann, H.; Grossmann, T. N.; Plowright, A. T. New Modalities for Challenging Targets in Drug Discovery. *Angew. Chem., Int. Ed. Engl.* **2017**, *56*, 10294−10323.
(39) Clark, R. J.; Fischer, H.; Dempster, L.; Daly, N. L.; Rosengren, K. J.; Nevin, S. T.; Meunier, F. A.; Adams, D. J.; Craik, D. J. Engineering stable peptide toxins by means of backbone cyclization: stabilization of the alpha-conotoxin MII. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13767−13772.
(40) Armishaw, C. J.; Dutton, J. L.; Craik, D. J.; Alewood, P. F. Establishing regiocontrol of disulfide bond isomers of alpha-conotoxin ImI via the synthesis of N-to-C cyclic analogs. *Biopolymers* **2010**, *94*, 307−313.
(41) Clark, R. J.; Jensen, J.; Nevin, S. T.; Callaghan, B. P.; Adams, D. J.; Craik, D. J. The engineering of an orally active conotoxin for the treatment of neuropathic pain. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 6545−6548.
(42) Halai, R.; Callaghan, B.; Daly, N. L.; Clark, R. J.; Adams, D. J.; Craik, D. J. *J. Med. Chem.* **2011**, *54*, 6984−6992.
(43) Armishaw, C. J.; Jensen, A. A.; Balle, L. D.; Scott, K. C.; Sérensen, L.; Strømgaard, K. Improving the stability of α-conotoxin AuIB through N-to-C cyclization: the effect of linker length on stability and activity at nicotinic acetylcholine receptors. *Antioxid. Redox Signal.* **2011**, *14*, 65−76.
(44) Lovelace, E. S.; Gunasekera, S.; Alvarmo, C.; Clark, R. J.; Nevin, S. T.; Grishin, A. A.; Adams, D. J.; Craik, D. J.; Daly, N. L. Stabilization of α-conotoxin AuIB: influences of disulfide connectivity and backbone cyclization. *Antioxid. Redox Signal.* **2011**, *14*, 87−95.
(45) Lovelace, E. S.; Armishaw, C. J.; Colgrave, M. L.; Wahlstrom, M. E.; Alewood, P. F.; Daly, N. L.; Craik, D. J. Cyclic MrIA: a stable and potent cyclic conotoxin with a novel topological fold that targets the norepinephrine transporter. *J. Med. Chem.* **2006**, *49*, 6561−6568.
(46) Dekan, Z.; Wang, C. I.; Andrews, R. K.; Lewis, R. J.; Alewood, P. F. Conotoxin engineering: dual pharmacophoric noradrenaline transport inhibitor/integrin binding peptide with improved stability. *Org. Biomol. Chem.* **2012**, *10*, 5791−5794.
(47) Hemu, X.; Taichi, M.; Qiu, Y.; Liu, D. X.; Tam, J. P. Biomimetic synthesis of cyclic peptides using novel thioester surrogates. *Biopolymers* **2013**, *100*, 492−501.
(48) Akcan, M.; Clark, R. J.; Daly, N. L.; Conibear, A. C.; de Faoite, A.; Heghinian, M. D.; Sahil, T.; Adams, D. J.; Marí, F.; Craik, D. J. Transforming conotoxins into cyclotides: Backbone cyclization of P-superfamily conotoxins. *Biopolymers* **2015**, *104*, 682−692.
(49) Rettie, S. A.; Campbell, K. V.; Bera, A. K.; Kang, A.; Kozlov, S.; De La Cruz, J.; Adebomi, V.; Zhou, G.; DiMaio, F.; Ovchinnikov, S.; Bhardwaj, G. Cyclic peptide structure prediction and design using AlphaFold. *bioRxiv* **2023**, 2023−02.
(50) Karami, Y.; Guyon, F.; De Vries, S.; Tufféry, P. DaReUS-Loop: accurate loop modeling using fragments from remote or unrelated proteins. *Sci. Rep* **2018**, *8*, 13673.
(51) Karami, Y.; Rey, J.; Postic, G.; Murail, S.; Tufféry, P.; de Vries, S. J. DaReUS-Loop: a web server to model multiple loops in homology models. *Nucleic Acids Res.* **2019**, *47*, W423−W428.
(52) Stein, A.; Kortemme, T. Improvements to robotics-inspired conformational sampling in rosetta. *PLoS One* **2013**, *8*, No. e63090.
(53) Fukutomi, T.; Takagi, K.; Mizushima, T.; Ohuchi, N.; Yamamoto, M. Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degron and Keap1. *Mol. Cell. Biol.* **2014**, *34*, 832−846.
(54) Coulouarn, Y.; Lihrmann, I.; Jegou, S.; Anouar, Y.; Tostivint, H.; Beauvillain, J. C.; Conlon, J. M.; Bern, H. A.; Vaudry, H. Cloning of the cDNA encoding the urotensin II precursor in frog and human reveals intense expression of the urotensin II gene in motoneurons of the spinal cord. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 15803−15808.
(55) Vaudry, H.; Leprince, J.; Chatenet, D.; Fournier, A.; Lambert, D. G.; Le Mevel, J. C.; Ohlstein, E. H.; Schwertani, A.; Tostivint, H.; Vaudry, D. International Union of Basic and Clinical Pharmacology. XCII. Urotensin II, urotensin II-related peptide, and their receptor: from structure to function. *Pharmacol Rev.* **2015**, *67*, 214−258.
(56) Leprince, J.; Chatenet, D.; Dubessy, C.; Fournier, A.; Pfeiffer, B.; Scalbert, E.; Renard, P.; Pacaud, P.; Oulyadi, H.; Segalas-Milazzo, I.; Guilhaudis, L.; Davoust, D.; Tonon, M. C.; Vaudry, H. Structure-activity relationships of urotensin II and URP. *Peptides* **2008**, *29*, 658−673.
(57) Cadirci, E.; Ugan, R. A.; Dincer, B.; Gundogdu, B.; Cinar, I.; Akpinar, E.; Halici, Z. Urotensin receptors as a new target for CLP induced septic lung injury in mice. *Naunyn Schmiedebergs Arch Pharmacol* **2019**, *392*, 135−145.
(58) Tapeinou, A.; Matsoukas, M. T.; Simal, C.; Tselios, T. Review cyclic peptides on a merry-go-round; towards drug design. *Biopolymers* **2015**, *104*, 453−461.
(59) Herold, C. L.; Behm, D. J.; Buckley, P. T.; Foley, J. J.; Wixted, W. E.; Sarau, H. M.; Douglas, S. A. The neuromedin B receptor antagonist, BIM-23127, is a potent antagonist at human and rat urotensin-II receptors. *Br. J. Pharmacol.* **2003**, *139*, 203−207.
(60) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(61) Bhaskaran, R.; Arunkumar, A. I.; Yu, C. NMR and dynamical simulated annealing studies on the solution conformation of urotensin II. *Biochim. Biophys. Acta* **1994**, *1199*, 115−122.
(62) Flohr, S.; Kurz, M.; Kostenis, E.; Brkovich, A.; Fournier, A.; Klabunde, T. Identification of nonpeptidic urotensin II receptor antagonists by virtual screening based on a pharmacophore model derived from structure-activity relationships and nuclear magnetic resonance studies on urotensin II. *J. Med. Chem.* **2002**, *45*, 1799−1805.
(63) Lescot, E.; Sopkova-de Oliveira Santos, J.; Dubessy, C.; Oulyadi, H.; Lesnard, A.; Vaudry, H.; Bureau, R.; Rault, S. Definition of new pharmacophores for nonpeptide antagonists of human urotensin-II. Comparison with the 3D-structure of human urotensin-II and URP. *J. Chem. Inf Model* **2007**, *47*, 602−612.
(64) Carotenuto, A.; Grieco, P.; Campiglia, P.; Novellino, E.; Rovero, P. Unraveling the active conformation of urotensin II. *J. Med. Chem.* **2004**, *47*, 1652−1661.
(65) Cemazar, M.; Kwon, S.; Mahatmanto, T.; Ravipati, A. S.; Craik, D. J. Discovery and applications of disulfide-rich cyclic peptides. *Curr. Top Med. Chem.* **2012**, *12*, 1534−1545.
(66) Guyon, F.; Martz, F.; Vavrusa, M.; Bécot, J.; Rey, J.; Tufféry, P. BCSearch: fast structural fragment mining over large collections of protein structures. *Nucleic Acids Res.* **2015**, *43*, W378−382.
(67) Wang, G.; Dunbrack, R. L. PISCES: a protein sequence culling server. *Bioinformatics* **2003**, *19*, 1589−1591.
(68) Mulvenna, J. P.; Wang, C.; Craik, D. J. CyBase: a database of cyclic protein sequence and structure. *Nucleic Acids Res.* **2006**, *34*, D192−194.
(69) Wang, C. K.; Kaas, Q.; Chiche, L.; Craik, D. J. CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. *Nucleic Acids Res.* **2007**, *36*, D206−210.

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450

(70) Oliphant, T. E. *A guide to NumPy*; Trelgol Publishing USA, 2006; Vol. *1*.
(71) Jones, E.; Oliphant, T.; Peterson, P. *SciPy: Open source scientific tools for Python*; SciPy Community, 2001. http://www.scipy.org/.
(72) Shen, Y.; Picord, G.; Guyon, F.; Tuffery, P. Detecting protein candidate fragments using a structural alphabet profile comparison approach. *PLoS One* **2013**, *8*, No. e80493.
(73) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19−25.
(74) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71−73.
(75) Lamiable, A.; Thevenet, P.; Tufféry, P. A critical assessment of hidden markov model sub-optimal sampling strategies applied to the generation of peptide 3D models. *J. Comput. Chem.* **2016**, *37*, 2006−2016.
(76) Murail, S. *gromacs_py: A gromacs python wrapper*, ver. 0.1; 2018; DOI: 10.5281/zenodo.1455734.
(77) Huang, J.; MacKerell, A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135−2145.
(78) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domanski, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. Proc. of the 15th Python in Science Conf. Austin, Texas, July 11, 2019 DOI: 10.25080/Majora-629e541a-00e.
(79) Mahesh Kumar, K.; Rama Mohan Reddy, A. A fast DBSCAN clustering algorithm by accelerating neighbor searching using Groups method. *Pattern Recognition* **2016**, *58*, 39−48.
(80) Case, D. A.; Berryman, J.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *AMBER*; University of California: California, US, 2015; *2015*.
(81) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput* **2015**, *11*, 3696−3713.
(82) Agarwal, P. K. Cis/trans isomerization in HIV-1 capsid protein catalyzed by cyclophilin A: insights from computational and theoretical studies. *Proteins* **2004**, *56*, 449−463.
(83) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput* **2013**, *9*, 3878−3888.
(84) Touchard, A.; Aili, S. R.; Téné, N.; Barassé, V.; Klopp, C.; Dejean, A.; Kini, R. M.; Mrinalini; Coquet, L.; Jouenne, T.; Lefranc, B.; Leprince, J.; Escoubas, P.; Nicholson, G. M.; Treilhou, M.; Bonnafé, E. Venom peptide repertoire of the european myrmicine ant manica rubida: identification of insecticidal toxins. *J. Proteome Res.* **2020**, *19*, 1800−1811.
(85) Cudic, M.; Wade, J. D.; Otvos, L., Jr Convenient synthesis of a head-to-tail cyclic peptide containing an expanded ring. *Tetrahedron Lett.* **2000**, *41*, 4527−4531.
(86) Postma, T. M.; Albericio, F. Immobilized N-chlorosuccinimide as a friendly peptide disulfide-forming reagent. *ACS Comb. Sci.* **2014**, *16*, 160−163.

https://doi.org/10.1021/acs.jcim.3c00865
J. Chem. Inf. Model. 2023, 63, 6436−6450