Combinatorial Discovery and Validation of Heptapeptides with UTP Binding Induced Structure

Daniela Kroiss,[a, c, d] James M. Aramini,[b] Sangitaa Ragoonath,[a, d] Rein V. Ulijn,[a, c, d] and Tell Tuttle*[e]

In biology, supramolecular recognition typically involves an ‘induced-fit’ mechanism, where structures rearrange upon complexation to accommodate binding ligands. Designing minimalistic compounds with such adaptability is challenging as they involve subtle conformational changes that are energetically similar. Here, we demonstrate the integration of combinatorial screening with molecular modelling to identify heptapeptides that form a stable loop upon recognition of uridine triphosphate (UTP). Peptide sequences selected using phage display were refined computationally and correlated with experimental K\textsubscript{D} values. This combined approach may serve as a method for the de novo selection and subsequent rationalization of the compositional and organizational principles that dictate chemical functionality in flexible structures with dynamic conformations.

The design and discovery of novel proteins and peptides with tailored folds and functions is of great interest due to their enormous potential for future applications in biotechnology and medicine.[1] One important characteristic of these bio receptors is their ability to accommodate ligands through conformational change, commonly referred to as ‘induced fit’. As proposed by Emil Fischer 125 years ago,[2] many protein-ligand interactions are initially weak, but increase in strength upon structural rearrangement of the receptor that favors complexation.[3]

While short peptides may not be able to match the performance of proteins, it is increasingly recognized that their simplicity and robustness makes them attractive candidates for incorporation in synthetic materials with life-like or life-interfacing functions.[4] Peptides are more amenable for the systematic investigation of the fundamentals underlying supramolecular recognition and induced structure formation. Moreover, short peptides do not have an inherently stable structure, due to their simplicity, but move along a shallow energy landscape, which makes them especially suited to study minimalistic induced-fit mechanisms.

Since the molecular principles that underlie these phenomena are still poorly understood, identification of new binders typically involves screening of large chemical search spaces using methods such as SELEX,[5] CoPhMoRe,[6] dynamic combinatorial libraries,[7] yeast[8] and phage[9] display. The latter has been used effectively to find peptides that bind to antibodies, cell surface proteins, carbohydrate antigens,[10] metal surfaces,[11] nanoparticles,[12] polymers[13] and macrocycles,[14] as well as for sequences that form dynamically interchanging complexes with ATP[15] and peptide catalysts.[16] In addition, molecular dynamics simulations have been used to predict the interaction of thereby identified sequences with gold[17] and silica[18] surfaces. Combining experimental screening with computation has been applied successfully to protein design and evolution[19] as well as directed discovery of peptides.[20] Thus, we reasoned that the work flow of phage display could be accelerated and improved through seamless integration of these complementary strategies.

The objective of the present work was to use computational chemistry to rationally direct the discovery of functional peptides from a random library displaying approximately 100 million different sequences to efficiently select for heptapeptides that interact with UTP. Selected sequences were exploited to elucidate the relationship between supramolecular complexation and structure formation and the obtained experimental results were found to be in good agreement with theoretical predictions. The thus identified lead sequence KAIHPMR forms a stable loop upon interaction with UTP, constituting the simplest example of induced fit in a peptide reported thus far.

A commercial library of M13 phages that display 5 copies of a heptapeptide on a terminal protein surface was mixed with biotinylated UTP (Scheme 1 and Figure S1) for post-selection.

---

[a] Dr. D. Kroiss, S. Ragoonath, Prof. Dr. R. V. Ulijn  
Nanoscience Initiative at Advanced Science Research Center (ASRC)  
The Graduate Center of the City University of New York  
85 St. Nicholas Terrace, New York, NY 10031 (USA)  
E-mail: Rein.Ulijn@asrc.cuny.edu

[b] Dr. J. M. Aramini  
Structural Biology Initiative at ASRC  
The Graduate Center of the City University of New York  
85 St. Nicholas Terrace, New York, NY 10031 (USA)

[c] Dr. D. Kroiss, Prof. Dr. R. V. Ulijn  
Ph.D. Programs in Biochemistry and Chemistry  
The Graduate Center of the City University of New York  
365 Fifth Avenue, New York, NY 10016 (USA)

[d] Dr. D. Kroiss, S. Ragoonath, Prof. Dr. R. V. Ulijn  
Department of Chemistry  
Hunter College, City University of New York  
695 Park Avenue, New York, NY 10065 (USA)

[e] Prof. Dr. T. Tuttle  
WestCHEM and Department of Pure and Applied Chemistry  
University of Strathclyde  
295 Cathedral Street, Glasgow G1 1XL (UK)  
E-mail: tell.tuttle@strath.ac.uk

[Supporting information for this article is available on the WWW under https://doi.org/10.1002/syst.202000025]

© 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
recovery using streptavidin-carrying paramagnetic agarose beads. The biotin-streptavidin decorated beads were separated into a pellet, thereby isolating phages that have bound to the target. The use of paramagnetic beads together with a magnetic rack facilitates selection compared to our previous centrifugation approach.\textsuperscript{[16,21]} After four rounds of panning and negative selection against streptavidin-agarose beads and soluble biotin, a total of 18 heptapeptide sequences were identified by DNA sequencing (Table S1).

Atomistic molecular dynamics simulations using the CHARMM force field were applied to investigate the interaction between these 18 heptapeptides (with free C-terminus) for their propensity to interact with UTP. The presence of H-bonds between peptides and the nucleotide was analyzed in VMD with a cut-off distance of 3.2 Å (see Supporting Information for details).

The results (Table S1) show that five of the tested sequences did not interact with UTP during the 50 ns simulation time. Eight sequences interacted with UTP only via a single hydrogen bond (often through the sequence-independent N-terminus interacting with the phosphates) and were therefore considered poor and non-selective binders. Consequently, these 13 peptide sequences were disregarded, leaving five peptide sequences for further investigation.

Since preliminary results suggested electrostatic interactions as the main driving force of peptide-nucleotide complexation, additional molecular dynamics simulations of the five selected sequences were carried out using C-terminally amidated versions of the peptides, to better reflect the phage surface conjugation. All simulations were run for 50 ns after the first peptide-UTP interaction was observed and the stability of the formed complexes was analyzed (Table 1). The results revealed the same trend that was observed for peptides with a free C-terminus, identifying KAIHPMR as the most promising candidate, as it remains bound to UTP 82% of the simulation time after initial contact. As expected, the net charge of +3 of this sequence favors electrostatic interactions with the nucleotide. The peptide ITLKGLT was observed in complex with UTP on

| Peptide  | Time to first peptide-UTP contact\textsuperscript{[a]} [ns] | % bound after initial contact\textsuperscript{[b]} | Net charge (amidated C-terminus) | Dissociation constant for UTP [mM] |
|----------|----------------------------------------------------------|---------------------------------|---------------------------------|-----------------------------------|
| KAIHPMR  | 5.63 ± 0.95                                             | 82.0 ± 1.0                      | +3                              | 0.74 ± 0.10                       |
| ITLKGLT  | 4.47 ± 1.45                                             | 51.1 ± 7.9                     | +2                              | 1.78 ± 0.17                       |
| SMRDGAV  | 8.24 ± 0.97                                             | 68.7 ± 3.0                     | +1                              | 3.31 ± 0.30                       |
| FPVVRTN  | 10.31 ± 5.01                                            | 67.4 ± 2.8                     | +2                              | 3.65 ± 0.47                       |
| TSVSLR   | 9.39 ± 0.67                                             | 74.6 ± 1.3                     | +2                              | 10.80 ± 1.56                      |
| KAIHPMR  | 4.31 ± 0.62                                             | 76.3 ± 1.1                     | +3                              | 0.83 ± 0.04                       |
| KAAPMR   | 3.77 ± 0.82                                             | 72.9 ± 4.2                     | +3                              | 3.45 ± 0.38                       |

\textsuperscript{[a]} Calculated using Equation (S1a). \textsuperscript{[b]} Calculated using Equation (S1b).
average 65% of the time. The peptides SMRDGAV and FPVVTRN were found to remain bound during 69% and 67% of the time, respectively. The peptide TSAVSLR formed a complex with the nucleotide during 75% of the time.

Solid-phase synthesis was used to produce the amidated versions of the five selected peptide sequences shown in Table 1. Following synthesis, all peptides were purified by preparative HPLC and extra care was taken to remove any residual trifluoroacetate using 10 mM HCl because of its propensity to form stable adducts with positively charged amino-groups, thereby impeding supramolecular interactions (see Supporting Information for details). $^{19}$F-NMR was performed to confirm that peptides were free from TFA (Figure S2).

$^1$H-NMR was applied to follow titrations of the five peptides with UTP and the changes in chemical shift perturbations of amino acid side chains were analyzed to calculate dissociation constants (Table 1 and Figures S3–12). The heptapeptide KAIHPMR showed the lowest dissociation constant with a $K_d$ of 0.74 mM (Figure 1A). The affinity of the sequence ITLKGLT was two-fold lower, with a $K_d$ of 1.78 mM. The peptides FPVVTRN and SMRDGAV show similar affinities, with $K_d$ of 3.65 mM and 3.31 mM, respectively. The lowest affinity was observed for the sequence TSAVSLR, which showed a $K_d$ of 10.80 mM, which is almost 15-times weaker than the best sequence KAIHPMR.

Comparison of experimental affinities and predicted stability of peptide-UTP complexes revealed that experimental and computational results are in good agreement for all peptides except TSAVSLR (Table 1, Figure 2). Closer analysis of this sequence revealed that the peptide interacts with UTP mainly via its C-terminal Arg-side chain, with the rest of the peptide pointing away from the nucleotide, thus not forming a complex.

The best performing sequence, KAIHPMR, was investigated computationally in more detail, revealing that the peptide interacts with UTP mainly via its Lys- and Arg-side chains as well as the N-terminal amino group (Figure 1B). Importantly, simulations of the peptide KAIHPMR in absence of UTP revealed that the free peptide remains in an extended conformation throughout a 50 ns simulation (Figure 1C,D) while the peptide forms a loop-structure upon complexation of the nucleotide (Figure 1E,F). Analysis of the distance between the Lys- and Arg-side chains shows that the loop, once formed, is stable over the course of the simulation.

More detailed investigation revealed that after 5 ns of equilibration, the Lys-residue forms the first contact to UTP via a stable hydrogen-bond to the phosphate-moiety of the nucleotide (Figure 1G). Subsequently, the Arg-residue engages in hydrogen-bonding with UTP as well, thereby inducing the loop structure (Figure 1H). To quantitatively analyze the stability of this UTP-binding induced structure, the simulation was extended further to yield 50 ns of simulation time after the initial peptide-UTP interaction. The results revealed that the loop-structure is retained during 91.7% of a 50 ns simulation after it was formed.

While no intermolecular H-bonds were observed between residue H4 and UTP, intramolecular H-bonds were detected between the side chains of residue H4 and R7 (18% of the time) and also between the backbone of these residues (23% of the time) after the loop has formed. This demonstrates that the formation of the loop is induced by binding to the nucleotide and, while it is a sampled conformation in the free peptide, it is not stable in the absence of UTP. Interaction with the nucleotide leads to a stabilization of the peptide structure, whereas the free peptide in solution remains dynamic throughout the simulation.

Unexpectedly, the computational results revealed that His, rather than the expected Pro residue, is important for stable complexation-induced loop formation. The loop structure is observed for the His-mutant KAIAPMR only during 45.4% of a 50 ns simulation after it was formed initially. Hence, the stability of the loop is decreased compared to the original peptide KAIHPMR (Figure 3A,B).
The side chain of residue K1 interacts with the phosphate moiety of the nucleotide 58% of the simulation time after initial contact. The side chain of residue R7, in contrast, forms a H-bond to UTP only 23% of the time (Figure 3C), the majority of which (20%) are with the phosphate-moiety. Experimental characterization of the peptide KAIAPMR binding to UTP by 1H-NMR revealed a significantly increased dissociation constant of 3.45 mM compared to 0.74 mM for KAIHPMR (Table 1 and Figures S13, 14).

Replacing the Pro-residue by Ala in peptide KAIHAMR does not affect the stability of the loop, as it is observed during 91.5% of a 50 ns simulation after formation. Intramolecular H-bonds between residue H4 and the N-terminus are observed 10% of the time and they are likely to contribute to the formation of the stable structure (Figure 3D–F).

Experimental characterization of the peptide KAIHAMR binding to UTP by 1H-NMR revealed a dissociation constant of 0.83 mM (Table 1 and Figures S15–16), which is comparable to that of the original sequence KAIHPMR. This confirms our hypothesis that the role of the H4-residue is to stabilize the intramolecular structure of the peptide and suggests that replacing it for residues that can perform that function too should give similar results.

The proposed model was further tested by additional molecular dynamics simulations on peptides KAIQPMR and KAIYPMR that were selected for their theoretical potential to form intramolecular H-bonds to residue R7. The analysis of these simulations revealed that the Q4-residue in KAIQPMR does not form a H-bond to R7, consequently, no stable loop structure was observed (Figure S17). In contrast, the formation of a stable loop structure was observed for KAIYPMR upon interaction with UTP, which coincides with the formation of a stable H-bond between residues Y4 and R7 (Figure S18). It is important to note that for these peptides H-bonds are not formed between the residue in position 4 and the nucleotide.

![Figure 2](image-url) **Figure 2.** Experimental affinities and computational predictions are in good agreement. Normalized fraction of peptide bound to 1 mM UTP determined using titrations followed by 1H-NMR (red) and normalized stability of peptide-UTP complex (blue) during a 50 ns simulation after initial interaction. Error bars show standard error of three independent titrations/simulations. [*] binds UTP only via C-terminal Arg-side chain.

![Figure 3](image-url) **Figure 3.** Histidine contributes to the formation of a stable loop-structure. Representation of His-mutant KAIAPMR (A) and Pro-mutant KAIHAMR (D) bound to UTP. Quantitative analysis of distance between Lys1 and Arg7 side chains of these peptide reveal that the loop is destabilized for KAIAPMR (B) but not affected by the PSA mutation (E). Quantitative analysis of H-bonds observed between KAIAPMR (C) and KAIHAMR (F) to UTP during a 50 ns simulation.
thus we conclude that stability of the induced loop depends on residue 4 to engage in intramolecular hydrogen bonding with residue R7. This exemplifies the difficulties encountered in rational protein/peptide design. The delicate balance between the strain induced in the structure and the enthalpic gain in binding the substrate is not simply a function of the interactions between the substrate and the receptor.

Rather, it includes subtle effects such as whether additional intramolecular interactions can be leveraged to mitigate the conformational strain and loss of entropy from the receptor.

Interestingly, analysis of the average distance $D$ between residues 1 and 7 of the investigated peptides over time revealed that complexation of the nucleotide decreases $D$ in peptide KAIHPMR by 9.4 Å, from 17.9 Å to 8.5 Å (Figure S19). A less significant change is observed for the peptides TSAVSLR and FPVVTRN, where the average $D$ is reduced by 5.3 Å and 4.5 Å, respectively. No reduction of $D$ is observed for the peptides ITLKGLT and SMRDGAV. Thus, we propose that the distance between the N- and C-terminal residues in short peptides can be used as a measure to predict ‘induced fit’ in short peptide sequences.

While functional proteins and peptides that have stable folds can increasingly be designed and rationally modified, there is only a limited understanding of the compositional and organizational principles that dictate chemical functionality in those with flexible structures and dynamic conformations.

In this work, we demonstrate the integration of experimental screening and computational refinement to identify novel minimalistic peptide sequences that take on a stable structure upon complexation with a single nucleotide. The simple binding measure used in this selection procedure was validated through the quantitative agreement shown with the experimentally determined binding constants. Moreover, molecular modelling allows for the rationalization of the interactions of selected sequences with the target.

Our approach resulted in the first example of an ‘induced fit’ mechanism in a heptapeptide and is therefore a minimalistic model of structural adaptation upon interaction with a ligand, a phenomenon which is widely observed in complex proteins but cannot yet be designed in simpler peptides. The findings contribute to our understanding of how the primary amino acid sequence affects supramolecular recognition and structure formation in minimalistic peptides, of relevance as components for adaptive peptide-based supramolecular systems with dynamically induced structures and functions [4].

Acknowledgements

The research leading to these results has received funding from the Airforce Office for Scientific Research (FA9550-19-1-0111) and the CUNI/Strathclyde partnership. Results were obtained using the ARCHIE-WeSt High Performance Computer (www.archie-west.ac.uk) based at the University of Strathclyde.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: induced fit · molecular dynamics simulations · peptides · supramolecular recognition

[1] P.-S. Huang, S. E. Boyken, D. Baker, Nature 2016, 537, 320–320.
[2] E. Fischer, Ber. Dtsch. Chem. Ges. 1894, 27, 2985–2993.
[3] a) D. E. Koshland, PNAS 1958, 44, 98 LP-104; b) D. E. Koshland Jr, Angew. Chem. Int. Ed. 1995, 33, 2375–2378.
[4] a) B. A. Gryzobowski, W. T. S. Huck, Nat. Nanotechnol. 2016, 11, 585–592; b) E. Martia, S. Otto, Nat. Nanotechnol. 2015, 10, 111–119; c) K. Sato, et al., Chem. Soc. Rev. 2018, 47, 7539–7551; d) I. R. Epstein, B. Xu, Nat. Nanotechnol. 2016, 11, 312–319; e) P. Makam, E. Gazit, Chem. Soc. Rev. 2018, 47, 3406–3420.
[5] a) A. R. Oliphant, C. J. Brandl, K. Struhl, Mol. Cell. Biol. 1989, 9, 2944–2949; b) C. Tuerk, L. Gold, Science 1990, 249, 505–510.
[6] J. Zhang, et al., Nat. Nanotechnol. 2013, 8, 959–959.
[7] a) P. T. Corbett, et al., J. Am. Chem. Soc. 2005, 127, 8902–8903; b) R. L. E. Furlan, et al., Tetrahedron 2002, 58, 771–778.
[8] E. T. Boden, K. D. Wittrup, Nat. Biotechnol. 1997, 15, 553–557.
[9] a) G. P. Smith, Science 1985, 228, 10–11; b) J. McCafferty, et al., Nature 1990, 348, 552–554.
[10] C.-H. Wu, et al., J. Biomed. Sci. 2016, 23, 1–14.
[11] S. Brown, Nat. Biotechnol. 1997, 15, 269–272.
[12] a) R. R. Naik, et al., Nat. Mater. 2002, 1, 169–172; b) S.-W. Lee, et al., Science 2002, 296, 892–895; c) S. V. Patwardhan, et al., J. Am. Chem. Soc. 2012, 134, 6244–6256.
[13] K. A. Günay, H.-A. Klok, Bioconjugate Chem. 2015, 26, 2002–2015.
[14] S. Sonzini, et al, Angew. Chem. Int. Ed. 2016, 55, 14000–14004; Angew. Chem. 2016, 128, 14206–14210.
[15] D. Kroiss, et al., ChemSystemsChem 2019, 1, DOI: 10.1002/syst.201900013.
[16] Y. Maeda, et al., J. Am. Chem. Soc. 2014, 136, 15893–15896.
[17] L. B. Wright, et al., Chem. Sci. 2015, 6, 5204–5214.
[18] a) V. Puddu, C. C. Perry, ACS Nano 2012, 6, 6356–6363; b) D. Oliver, et al., Phys. Chem. Chem. Phys. 2019, 21, 4663–4672.
[19] a) I. Cho, et al., Angew. Chem. Int. Ed. 2019, 58, 3138–3142; b) N. H. Joh, et al., Science 2014, 346, 1520 LP-1524; c) K. R. Mahendran, et al., Nat. Chem. 2017, 9, 411–419; d) G. Bhardwaj, et al., Nature 2016, 538, 329–329.
[20] A. Lampel, R. V. Ulijn, T. Tuttle, Chem. Soc. Rev. 2018, 47, 3737–3758.
[21] Z. Wei, Y. Maeda, H. Matsu, Angew. Chem. Int. Ed. 2011, 50, 10585–10588; Angew. Chem. 2011, 123, 10773–10776.
[22] V. V. Andrushchenko, H. J. Vogel, E. J. Prenner, J. Pept. Sci. 2007, 13, 37–43.
A good combination: We demonstrate integrated experimental screening with computational refinement to enable unbiased searching of the peptide sequence space for specific functions. This approach allows for the discovery of short peptides that undergo conformational adaption upon complexation with a single nucleotide (UTP), a feature that is ubiquitous in nature, but cannot be achieved through rational design.