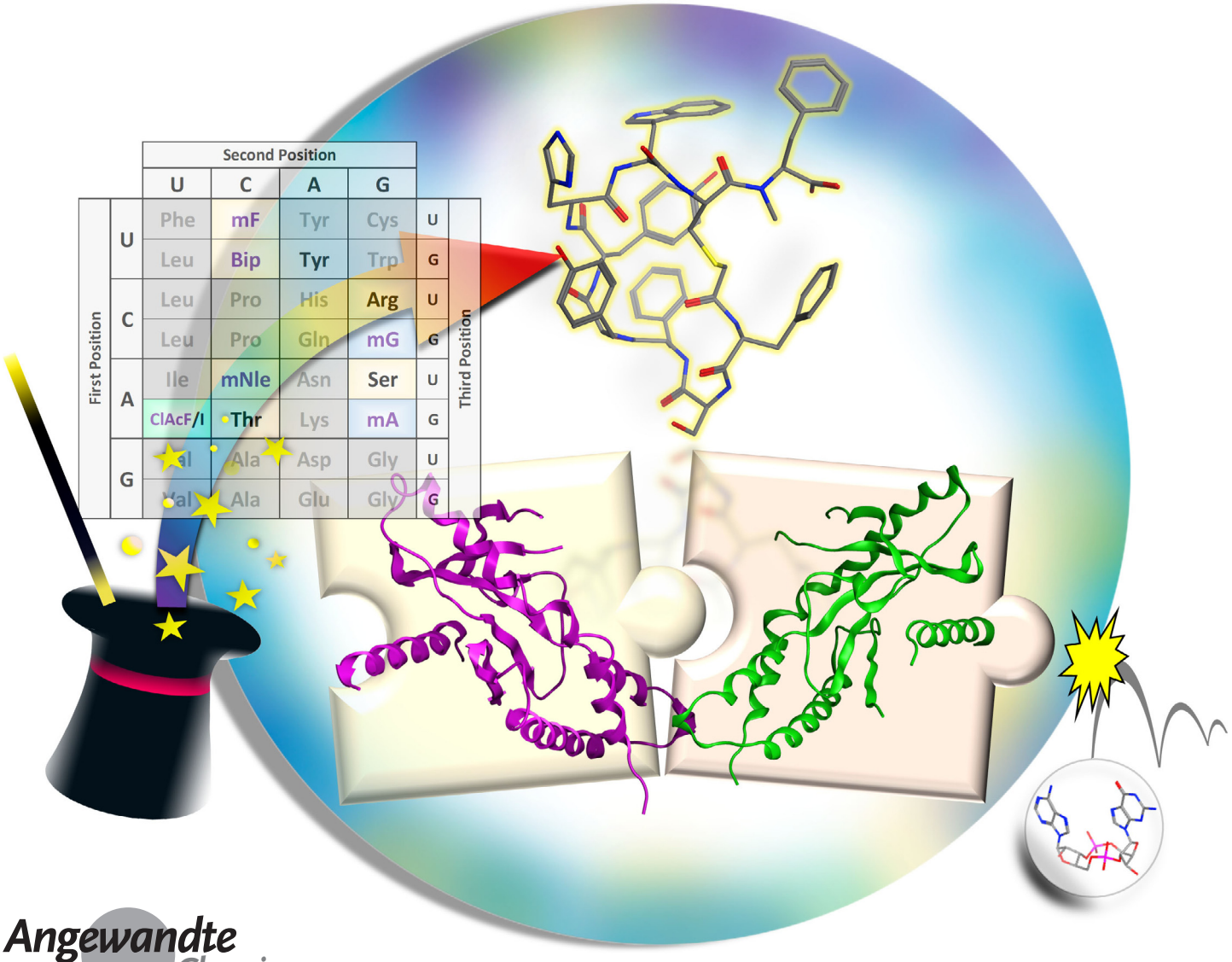
Peptides Hot Paper

How to cite: *Angew. Chem. Int. Ed.* **2021**, *60*, 22640–22645
International Edition: doi.org/10.1002/anie.202103043
German Edition: doi.org/10.1002/ange.202103043

A Selection of Macrocyclic Peptides That Bind STING From an mRNA-Display Library With Split Degenerate Codons

Chi-Wang Lin, Mary J. Harner, Andrew E. Douglas, Virginie Lafont, Fei Yu, Ving G. Lee, Michael A. Poss, Joanna F. Swain, Martin Wright, and Daša Lipovšek*



Abstract: *Recent improvements in mRNA display have enabled the selection of peptides that incorporate non-natural amino acids, thus expanding the chemical diversity of macrocycles beyond what is accessible in nature. Such libraries have incorporated non-natural amino acids at the expense of natural amino acids by reassigning their codons. Here we report an alternative approach to expanded amino-acid diversity that preserves all 19 natural amino acids (no methionine) and adds 6 non-natural amino acids, resulting in the highest sequence complexity reported to date. We have applied mRNA display to this 25-letter library to select functional macrocycles that bind human STING, a protein involved in immunoregulation. The resulting STING-binding peptides include a 9-mer macrocycle with a dissociation constant (K_D) of 3.4 nM, which blocks binding of cGAMP to STING and induces STING dimerization. This approach is generalizable to expanding the amino-acid alphabet in a library beyond 25 building blocks.*

Macrocyclic peptides result from chemical cross-linking between amino-acid residues separated by several peptide bonds. Their structures comprise at least one polypeptide ring and are both more constrained and more complex than the structure of linear peptides.[1] The molecular weight of macrocyclic peptides covers the range between small molecules and proteins, approximately 1 to 10 kDa. The preorganization and the semi-rigid character of macrocycles enable specific, high-affinity binding to protein targets, as well as inhibition of protein-protein interactions.[2–4]

Key to the discovery of novel macrocycles with desired properties are technologies that produce and screen a large number of diverse cyclic peptides. In vitro display methods, such as mRNA display[5,6] and RaPID,[7] are particularly powerful due to their libraries of up to 10^{12} unique peptides, each linked to an RNA tag that encodes it. These libraries are synthesized using purified components of the ribosomal translational machinery.[7,8] In addition to the standard, natural amino acids, this in vitro system can incorporate non-natural amino acids into peptide libraries, expanding the available chemical diversity, and, consequently, expanding the range of desirable properties surveyed. For example, *N*-methyl-, D-, and α, α′- di-substituted amino acids confer proteolytic stability of macrocycles,[1,5,9] and *N*-methyl amino acids hold the promise of improved oral bioavailability.[10]

To include non-natural amino acids in ribosomally translated peptide libraries, some codons need to be reassigned from encoding natural amino acids to encoding non-natural amino acids. Flexizymes, engineered ribozymes capable of charging a variety of non-natural amino acids onto a tRNA substrate, make possible such genetic reprogramming.[11,12] Traditionally, for each codon reassigned to a non-natural amino acid, the corresponding cognate natural amino acid and its aminoacyl-tRNA synthetase (aaRS) are removed from the system, and the resultant genetic blank is filled with a non-natural amino acid that is charged by a flexizyme onto the tRNA.[11] Due to this direct swap, the total number of different amino acids that can appear in a peptide library remains unchanged.

Alternatively, the size of the alphabet can be expanded by taking advantage of codon degeneracy. In this approach, some of the degenerate codons that encode the same natural amino acid are reassigned to multiple non-natural amino acids. Suga *et al* pioneered this approach by splitting the three degenerate codons for arginine (CGC, CGG, AGG) and assigning them to three different amino acids, resulting in an alphabet of 23 amino acids.[13–16] We have expanded this approach by simultaneously splitting three sets of degenerate codons, which encode Arg, Ser, and Thr in nature. We reassigned these codons to eight different amino acids, resulting in a library with an alphabet of 25 amino acids, comprised of 19 natural and six non-natural amino acids. To test the feasibility of extensive codon splitting, we designed two model peptides, each containing three degenerate Arg codons (CGU, CGG, AGG) or three degenerate Ser codons (UCU, AGU, UCG). We used two types of chemically synthesized tRNA (Table S2) to incorporate the chosen amino acids into these peptides during in vitro translation:

a) tRNA with primary nucleotide sequences identical to aaRS substrates, for charging natural amino acids (Gly, Ser, Arg), and
b) tRNA with engineered anti-codon loops, which nucleotides 32–38 were identical to those of the natural substrate tRNA for *E. coli* aaRS, for flexizyme-mediated acylation of reassigned amino acids.

The flexizyme charging efficiency, quantified by LC/MS of tRNA 3′- trinucleotide fragments, was found to be more than 10% for all amino acids tested (Figure S1). In vitro translation of the two mRNA templates yielded model peptides bearing one serine or arginine residue and two additional non-natural amino-acid residues (Figure 1b,c), all of which conformed to the design, confirming the intended genetic reprogramming (Figure S2). The in vitro translation took place with high fidelity, with no observable by-products due to mis-incorporation of unwanted amino acids.

Having confirmed the appropriate incorporation of reassigned amino acids at split-codon sites, we used this approach to make a complex library of macrocycles. The initiator codon (AUG) was reprogrammed to encode the non-natural amino

[*] Dr. C.-W. Lin, Dr. M. Wright, Dr. D. Lipovšek
Bristol Myers Squibb
100 Binney Street, Cambridge, MA 02142 (USA)
E-mail: chi-wang.lin@bms.com
Dr. M. J. Harner, Dr. A. E. Douglas, Dr. V. Lafont, F. Yu, V. G. Lee, Dr. M. A. Poss
Bristol Myers Squibb
Route 206 & Province Line Road, Lawrenceville, NJ 08543 (USA)
Dr. J. F. Swain
Repertoire Immune Medicines
Cambridge, MA (USA)
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202103043.

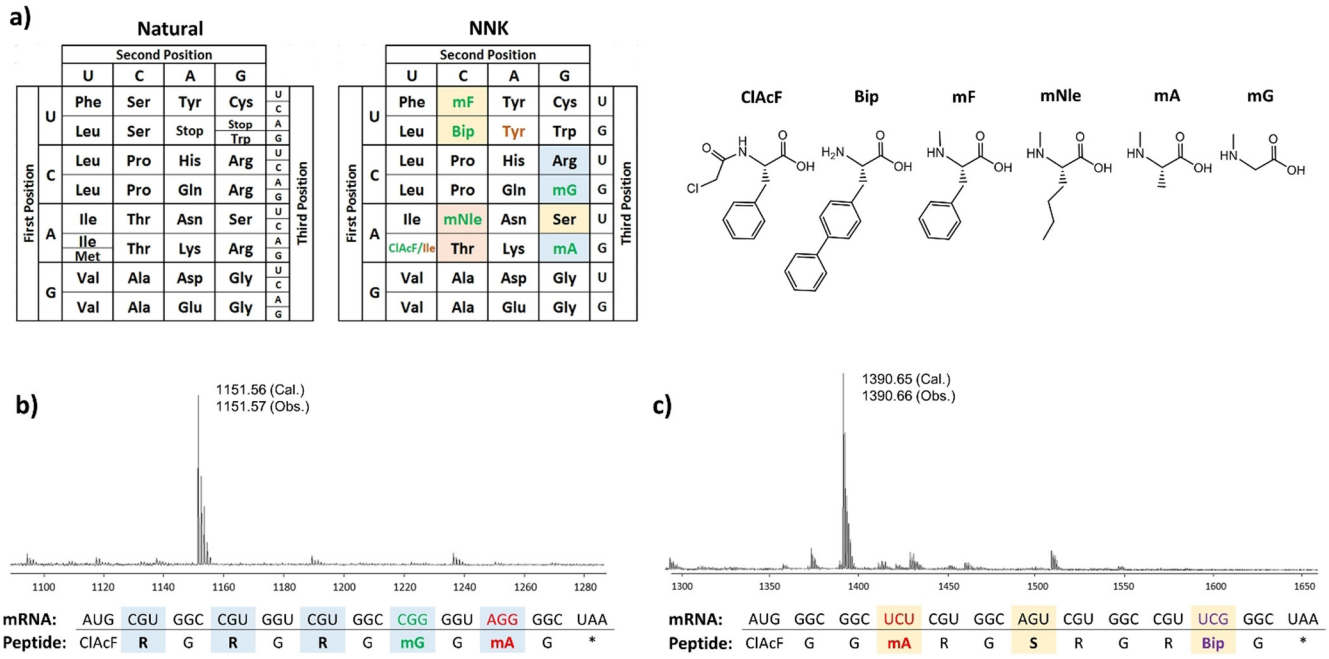


Figure 1. Splitting of degenerate codons to encode additional amino acids. a) Comparison of the natural codon table and the codon table designed for the NNK library. The degenerate codons of Arg, Ser, and Thr have been split and reassigned to different amino acids, shown in color. Non-natural amino acids are shown in green, and the reassigned natural amino acids are shown in brown. The chemical structures of the six non-natural amino acids are listed on the right. b) MALDI detection of a model peptide, in which R, mG, and mA were simultaneously assigned to three codons resulting from the split of the natural codons for Arg. c) MALDI detection of a model peptide, in which S, mA, and Bip were simultaneously assigned to codons resulting from the split of the natural codons for Ser.

acid ClAcF, which reacts with a cysteine further downstream in the sequence to cyclize each peptide.[17] The randomized amino-acid positions were encoded by NNK where N stands for any nucleotide and K stands for U or G. The redundant codons for Arg, Ser, and Thr were split simultaneously to sample a total of 25 amino acids, including six non-natural amino acids (ClAcF as an initiator, Bip, mA, mG, mF, and mNle) (Figure 1a). We retained 19 of the 20 natural amino acids, removing only methionine, due to its sensitivity to oxidation. Thus, our library composition is dramatically different from the library reported by Suga et al., who eliminated all hydrophilic natural amino acids, for the combination of 12 natural and 11 non-natural amino acids.[14,16] To demonstrate the utility of our NNK-based, 25-amino-acid mRNA display library, we used it to select macrocyclic peptides that bind human STING (Stimulator of Interferon Genes), an immune stimulator and a target of interest for cancer immunotherapy.[18] Much of the effort to discover novel STING activators has focused on cyclic dinucleotides, designed to imitate the natural STING ligand, cyclic GMP-AMP (cGAMP).[19,20] The recent discovery of a small-molecule class of STING activators, amidobenzimidazoles, demonstrates that this target can be activated by molecules with structure dramatically different from the natural ligand.[21]

We selected STING-binding macrocycles from libraries of thioether-linked monocyclic peptides with rings of 5 to 12 randomized amino-acid residues, followed by 0 to 2 randomized amino-acid residues, and by a C-terminal FLAG epitope (Figure 2a). Library DNA was transcribed in vitro into mRNA, which in turn was translated in vitro into peptides, yielding mRNA-peptide fusions associated with their complementary DNA. The fusion molecules were purified using an anti-FLAG antibody. The diversity of each purified library was approximately 5×10^{12}, as estimated by quantitative PCR.

We set up three parallel arms of selection to compare the performance of macrocycles of different size (10–14 mer vs. 7–9 mer) using two different forms of human STING (disulfide-linked protein that resembles the active signaling dimer vs. untethered STING monomer that can undergo dimerization; Figure 2b).

In each round of selection, mRNA-displayed peptide libraries were incubated with biotinylated human STING. The mRNA-displayed peptides that bound STING were captured on streptavidin beads. Next, the DNA encoding the captured peptide was recovered, amplified by PCR (Figure S4a), and used as template for the mRNA-peptide fusions to start the next round. By round four, the enriched populations showed detectable, STING-dependent binding (Figure 2b).

The enriched populations were tested for binding against additional proteins, including a human STING variant[22] (STING-AQ) and murine STING. The protein-bound fractions were recovered and analyzed by next-generation DNA sequencing (NGS) (Figure S4b). Individual sequences were triaged by their prevalence under different binding conditions and clustered into families based on sequence.

NGS of populations enriched during the three parallel arms of selections revealed peptide sequences with a wide range of length and amino-acid composition, some of which

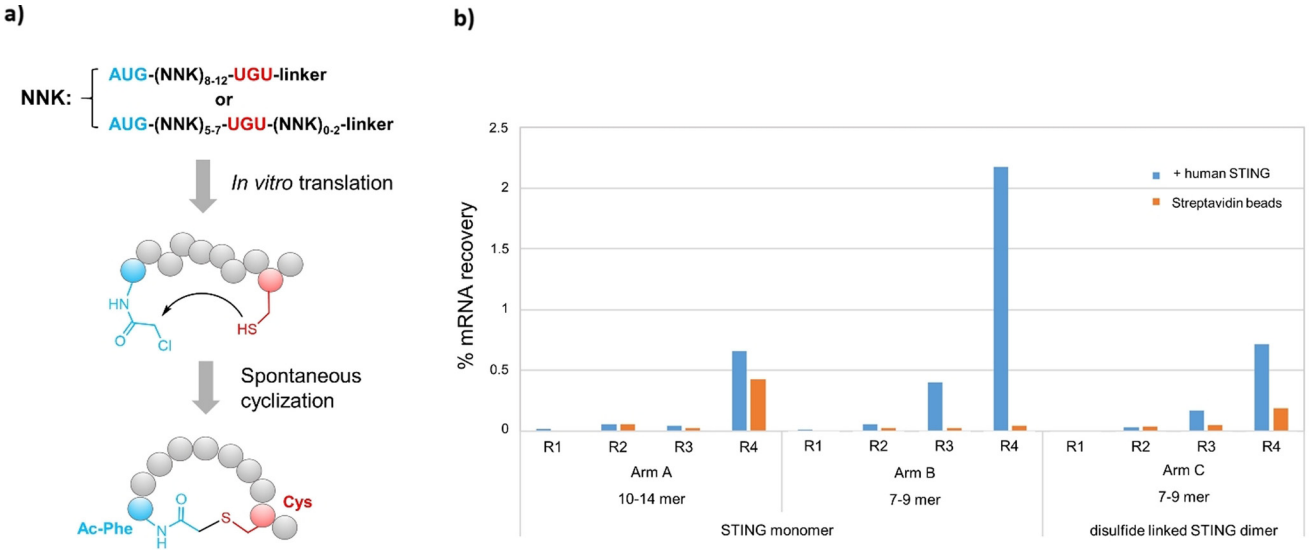


Figure 2. Selection from macrocyclic peptide libraries yielded peptides with high affinity for human STING. a) Structure of the macrocyclic peptide libraries, showing the covalent cross-link between the initiator residue (ClAcF) and the C-terminal Cys. b) The selection profile of the three parallel selection arms. Arm A and Arm B were selected for binding to human STING monomer, and Arm C was selected for binding to disulfide-linked STING dimer. All three arms showed enriched target binding by round 4. Streptavidin-only binding was monitored as a negative control.

clustered in highly homologous families (Figure S4c). For example, the library of small macrocycles (7–9 mer) yielded a family of related 9-mer peptides, with members in populations that were selected against each of the two forms of STING (Arm B and Arm C). The peptides in this series, named NNK-1 to NNK-7 (Figure 3, Table 1, Figure S4c), consisted of an 8-mer macrocycle and a one-amino-acid overhang. In addition, peptides in this family shared a stretch of hydrophobic residues between positions 3 and 5, and one or two charged residues (H, R, K, D) near the C-terminal cysteine. In contrast, the longer peptides enriched from the 10–14 mer library, e.g., NNK-8 to NNK-13, shared no obvious sequence motifs with the peptides selected from the library of 9-mers.

We chose thirteen individual sequences, which included the most highly enriched sequences, as well as several moderately enriched peptides representing different lengths, for chemical synthesis and detailed characterization. The peptides were synthesized by standard Fmoc chemistry and mass-confirmed using LC-MS (Table S3). The synthetic peptides were screened using two different biochemical assays: displacement of cGAMP, an endogenous ligand of STING, and induction of STING dimerization, which is critical for STING signaling.[23] The ten peptides with the highest potency in the cGAMP-displacement assay (NNK-1 to NNK-10, Table 1) were characterized further by a thermal shift assay, and surface plasmon resonance (SPR). We found that the binding affinity of these peptides correlated closely with their potency in the two assays (Figures S6,7). For example, NNK-2, which binds to human STING with the highest affinity (SPR $K_D = 3.4$ nM), was also the most potent in both the cGAMP- displacement assay ($IC_{50} = 7.5$ nM) and in the dimerization assay ($EC_{50} = 105$ nM) (Figure 3).

The other members of the 9-mer family with high sequence homology to NNK-2 (peptides NNK-1 and NNK-3 to NNK-7) had higher affinity for STING than unrelated

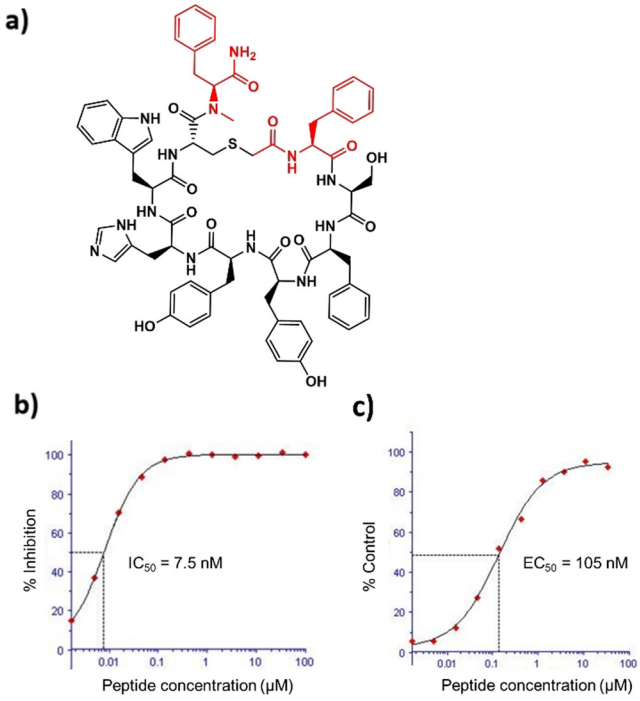


Figure 3. Characterizations of the synthetic peptide, NNK-2. a) The molecular structure with the non-natural amino acids highlighted in red. b) HTRF assay showing NNK-2 in competition of cGAMP ligand binding to human STING. c) NNK-2 induced dimerization of human STING, using cGMAP as the control.

peptides, with K_D between 14 and 250 nM (Table 1 and Figures S8,9). In addition, when bound to human STING, peptides NNK-1 to NNK-7 raised the STING melting temperature by more than 5 °C (Figure S10). In contrast, the larger macrocycles, NNK-8 to NNK-13, showed lower activity in the ligand displacement assay ($IC_{50} > 300$ nM) and no

Table 1: Sequence and function of the chemically synthesized peptides NNK-1 to NNK-13. *: thioether linkage between the initiator and the cysteine. Non-natural amino acids are shown in bold text. Since NNK-3 showed non-1:1 binding by SPR, the affinity of this peptide for STING was determined by ITC.

ID	Position															SPR Kinetics			Ligand Displacement	Dimerization
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	k_a (1/Ms)	k_d (1/s)	K_D (nM)	IC_{50} (nM)	EC_{50} (nM)
NNK-1	**AcF***	S	F	L	Y	H	W	**C***	**mF**							2.3E+06	3.0E-02	14	15	240
NNK-2	**AcF***	S	F	Y	Y	H	W	**C***	**mF**							1.4E+06	4.7E-03	3.4	7.5	105
NNK-3	**AcF***	R	F	Y	Y	W	R	**C***	D							K_D = 16 nM (by ITC)			20	390
NNK-4	**AcF***	V	F	Y	I	W	N	**C***	H							7.8E+05	1.4E-02	17	20	390
NNK-5	**AcF***	F	F	Y	I	W	N	**C***	H							4.3E+05	1.5E-02	36	25	690
NNK-6	**AcF***	P	V	F	**mF**	K	W	**C***	Y							7.1E+04	1.7E-02	250	180	3,700
NNK-7	**AcF***	L	V	**mF**	W	I	F	**C***	N							1.4E+05	2.3E-02	180	150	1,900
NNK-8	**AcF***	L	F	V	Y	N	N	D	F	Y	F	V	Y	**C***	G	No binding			320	>100,000
NNK-9	**AcF***	L	V	F	A	**mF**	R	F	D	V	Y	F	V	**C***	G	No binding			20,000	>100,000
NNK-10	**AcF***	L	V	Y	**mA**	**mF**	G	Y	A	W	I	**C***	G			No binding			61,000	>100,000
NNK-11	**AcF***	L	V	Y	P	**mF**	N	Y	V	Y	L	**C***	G			Not tested			38,000	>100,000
NNK-12	**AcF***	F	V	F	**mA**	**mF**	G	Y	V	Y	V	**C***	G			Not tested			87,000	>100,000
NNK-13	**AcF***	L	V	F	**mF**	**mNle**	G	Y	V	K	F	**C***	G			Not tested			>100,000	>100,000

observable activity in the dimerization assay (Figure 2). For NNK-8 to NNK-13, direct binding to STING could not be observed by SPR or in temperature shift assay (Figures S5a, S7, S10).

The 9-mer peptide family showed similar potency against the two forms of human STING (STING and STING-AQ, 98.9% sequence identity) in both assays (Figure S5a). In contrast, these peptides exhibited lower potency against murine STING (81.8% sequence identity with human STING; Figures S3, S5a).

At first glance, it is surprising that the macrocycles selected from the 9-mer library outperformed those selected from the 10–14 mer library, since a peptide with a larger loop presents a larger surface area that may be available for binding to the target. This may be due to a more thorough sampling of sequence space by the library with the fewer randomized positions. Alternatively, the shorter peptides might bind the target with higher affinity due to a smaller loss of conformational entropy upon binding.

The sequences of the macrocycles with the highest affinity for STING validate the utility of degenerate-codon splitting in large libraries. In particular, NNK-1 and NNK-2 each contain two amino-acid residues that were encoded by codons split from the degenerate codons for serine, i.e., serine, in position 2 (encoded by AGU), and mF, in position 9 (encoded by UCU). The presence of these two amino-acid residues in the STING-binding peptides was first inferred from the sequence of the DNA encoding the peptides, and then confirmed when the peptides with the inferred sequence were synthesized chemically and shown to have high affinity for STING.

We used an NGS positional scan[24] to compare STING binding between 184 single-substitution variants of NNK-2 and the wild-type. The experiment shows that positions 2 and 6 allow a wide range of substitutions without impacting STING binding. By contrast, mF in position 9 could be replaced by only two structurally similar amino acids (Bip and F), demonstrating the importance of the mF side chain for binding to STING (Figure S11).

In summary, the high affinity, potency, and specificity of the 9-mer cyclic peptides selected in this study demonstrate that a 25-amino-acid alphabet (19 natural and 6 non-natural amino acids) provides sufficient chemical and structural diversity for selected macrocycles to disrupt protein-protein interactions. These nonapeptides are the first reported peptidic binders to STING. The dissociation constant between NNK-2 and STING, 3.4 nM, is comparable to that of the natural ligand cGMAP (3.79 nM),[25] despite their drastically different structures. To the best of our knowledge, the 3.4 nM K_D for binding of NNK-2 to human STING is the highest reported affinity between a protein and a 9-mer (or smaller) cyclic peptide. Such high affinity binding had been reported for in vitro selected macrocycles with rings in the range of 13 to 16 amino-acid residues.[1,26]

Whereas the libraries described here split three degenerate codons to introduce six non-natural amino acids, this approach can be expanded. Splitting all available degenerate codons in the 32-codon NNK codon table would generate a maximal alphabet complexity of 32 amino acids (in addition to the initiator amino acid, ClAcF), enabling in vitro libraries of macrocyclic peptides with an even broader diversity of structure and function.

Acknowledgements

We thank R. Rampulla for facilitating synthesis of amino acids for flexizyme charging. We thank X. Xiao, R. Kandasamy, C. Terragni, and D. Kirdeeva for MiSeq sequencing. We thank Y. Zhang for performing quantitation of tRNA acylation. We thank B. Claus, M. Davis, and S. Johnson for assistance with NGS data analysis. We are grateful to P. Zhang and M. Witmer for expression and purification of protein constructs. We thank Y.-K. Wang for help with ligand displacement and STING dimerization assays. We also thank B. Warrack, J. Haugner, C. Stutz, J. Juneja, J. Sack, B. Fink, G. Schieven, and S. Cload for useful discussions.

Conflict of Interest

All authors are or were employees of Bristol Myers Squibb, which develop medicines and research models for profit.

Keywords: codon expansion · genetic reprograming · mRNA display · peptides · selection

[1] A. A. Vinogradov, Y. Yin, H. Suga, *J. Am. Chem. Soc.* **2019**, *141*, 4167 – 4181.
[2] E. M. Driggers, S. P. Hale, J. Lee, N. K. Terrett, *Nat. Rev. Drug Discovery* **2008**, *7*, 608 – 624.
[3] A. K. Yudin, *Chem. Sci.* **2015**, *6*, 30 – 49.
[4] Y. Huang, M. M. Wiedmann, H. Suga, *Chem. Rev.* **2019**, *119*, 10360 – 10391.
[5] Y. V. Guillen Schlippe, M. C. Hartman, K. Josephson, J. W. Szostak, *J. Am. Chem. Soc.* **2012**, *134*, 10469 – 10477.
[6] K. Josephson, A. Ricardo, J. W. Szostak, *Drug Discovery Today* **2014**, *19*, 388 – 399.
[7] C. J. Hipolito, H. Suga, *Curr. Opin. Chem. Biol.* **2012**, *16*, 196 – 203.
[8] F. P. Seebeck, J. W. Szostak, *J. Am. Chem. Soc.* **2006**, *128*, 7150 – 7151.
[9] D. Schwarzer, R. Finking, M. A. Marahiel, *Nat. Prod. Rep.* **2003**, *20*, 275 – 287.
[10] A. F. B. Räder, F. Reichart, M. Weinmuller, H. Kessler, *Bioorg. Med. Chem.* **2018**, *26*, 2766 – 2773.
[11] M. Ohuchi, H. Murakami, H. Suga, *Curr. Opin. Chem. Biol.* **2007**, *11*, 537 – 542.
[12] J. Lee, K. E. Schwieter, A. M. Watkins, D. S. Kim, H. Yu, K. J. Schwarz, J. Lim, J. Coronado, M. Byrom, E. V. Anslyn, A. D. Ellington, J. S. Moore, M. C. Jewett, *Nat. Commun.* **2019**, *10*, 5097.
[13] K. B. Lee, C. Y. Hou, C. E. Kim, D. M. Kim, H. Suga, T. J. Kang, *ChemBioChem* **2016**, *17*, 1198 – 1201.
[14] C. Nitsche, T. Passioura, P. Varava, M. C. Mahawaththa, M. M. Leuthold, C. D. Klein, H. Suga, G. Otting, *ACS Med. Chem. Lett.* **2019**, *10*, 168 – 174.
[15] Y. Iwane, A. Hitomi, H. Murakami, T. Katoh, Y. Goto, H. Suga, *Nat. Chem.* **2016**, *8*, 317 – 325.
[16] T. Passioura, W. Liu, D. Dunkelmann, T. Higuchi, H. Suga, *J. Am. Chem. Soc.* **2018**, *140*, 11551 – 11555.
[17] T. Kawakami, H. Murakami, H. Suga, *Chem. Biol.* **2008**, *15*, 32 – 42.
[18] L. Corrales, L. H. Glickman, S. M. McWhirter, D. B. Kanne, K. E. Sivick, G. E. Katibah, S. R. Woo, E. Lemmens, T. Banda, J. J. Leong, K. Metchette, T. W. Dubensky, Jr., T. F. Gajewski, *Cell Rep.* **2015**, *11*, 1018 – 1030.
[19] B. J. Francica, A. Ghasemzadeh, A. L. Desbien, D. Theodros, K. E. Sivick, G. L. Reiner, L. Hix Glickman, A. E. Marciscano, A. B. Sharabi, M. L. Leong, S. M. McWhirter, T. W. Dubensky, Jr., D. M. Pardoll, C. G. Drake, *Cancer Immunol. Res.* **2018**, *6*, 422 – 433.
[20] J. Wu, L. Sun, X. Chen, F. Du, H. Shi, C. Chen, Z. J. Chen, *Science* **2013**, *339*, 826 – 830.
[21] J. M. Ramanjulu, G. S. Pesiridis, J. Yang, N. Concha, R. Singhaus, S. Y. Zhang, J. L. Tran, P. Moore, S. Lehmann, H. C. Eberl, M. Muelbaier, J. L. Schneck, J. Clemens, M. Adam, J. Mehlmann, J. Romano, A. Morales, J. Kang, L. Leister, T. L. Graybill, A. K. Charnley, G. Ye, N. Nevins, K. Behnia, A. I. Wolf, V. Kasparcova, K. Nurse, L. Wang, Y. Li, M. Klein, C. B. Hopson, J. Guss, M. Bantscheff, G. Bergamini, M. A. Reilly, Y. Lian, K. J. Duffy, J. Adams, K. P. Foley, P. J. Gough, R. W. Marquis, J. Smothers, A. Hoos, J. Bertin, *Nature* **2018**, *564*, 439 – 443.
[22] L. Jin, L. G. Xu, I. V. Yang, E. J. Davidson, D. A. Schwartz, M. M. Wurfel, J. C. Cambier, *Genes Immun.* **2011**, *12*, 263 – 269.
[23] W. Sun, Y. Li, L. Chen, H. Chen, F. You, X. Zhou, Y. Zhou, Z. Zhai, D. Chen, Z. Jiang, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 8653 – 8658.
[24] J. M. Rogers, T. Passioura, H. Suga, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 10959 – 10964.
[25] X. Zhang, H. Shi, J. Wu, X. Zhang, L. Sun, C. Chen, Z. J. Chen, *Mol. Cell* **2013**, *51*, 226 – 235.
[26] K. Deyle, X. D. Kong, C. Heinis, *Acc. Chem. Res.* **2017**, *50*, 1866 – 1874.

Manuscript received: March 1, 2021
Revised manuscript received: July 11, 2021
Accepted manuscript online: August 12, 2021
Version of record online: September 6, 2021

