Translation of DNA into a Library of 13,000 Synthetic Small-Molecule Macrocycles Suitable for In Vitro Selection

Brian N. Tse, Thomas M. Snyder, Yinghua Shen, David R. Liu*

Supporting Information

Amino acids were purchased from Novabiochem, unless otherwise noted. All other solvents and reagents were purchased from Sigma-Aldrich, unless otherwise noted. All reactions were performed at 25 °C, unless otherwise noted.

Preparation of 5'-Scaffolds With a Side-Chain Amine Tartaramide. The oligonucleotide-linked CPG beads were transferred to a 1.7 mL tube. The 5' MMT group of the 5'-Amino-5 group was removed using three washes (1 mL each) of 3% trichloroacetic acid in DCM, or until the supernatant was colorless. The relevant diamino acid (Fmoc-Lys(MMT)-OH, Fmoc-Orn(MTT)-OH, Fmoc-Dab(MTT)-OH, or Fmoc-Dpr-(MTT)-OH) was installed by addition of a solution of the amino acid (100 μmol), HOBt (100 μmol, Novabiochem), HBTU (90 μmol, Novabiochem), and DIPEA (200 μmol) in 600 μL dry DMF, with agitation for 2 h. The resin was then washed with DMF and acetonitrile. The side-chain protecting group (MMT or MTT) was removed using three washes (1 mL each) of 3% (mol %) trichloroacetic acid in DCM, or until the supernatant was colorless. The O,O'-diacetyl-L-tartaric acid monobenzylamide group (see synthesis below) was installed using a solution of the tartaramide (32.3 mg, 100 μmol), HOBt (100 μmol), HBTU (90 μmol), and DIPEA (200 μmol) in 600 μL dry DMF, with agitation for 2 h. The resin was then washed with DMF and acetonitrile. The product was cleaved and fully deprotected from the resin with AMA at 65 °C for 10 min, dried *in vacuo*, and purified by reverse-phase HPLC using a gradient of acetonitrile (8% to 80%) in 100 mM triethylammonium acetate (TEAA), pH 7.0

Preparation of 5' Scaffolds With an α-Amine Tartaramide. The relevant diamino acid (Fmoc-Lys(MMT)-OH, Fmoc-Orn(MTT)-OH, Fmoc-Dab(MTT)-OH, or Fmoc-Dpr-(MTT)-OH, Novabiochem) was installed on the oligonucleotide as previously described with the side-chain analogs. The α-amine FMOC group was removed using three consecutive washes with 20% piperidine/80% DMF (5 min agitation per wash, monitored by UV). The resin was washed with DMF and acetonitrile. The O,O'-diacetyl-L-tartaric acid monobenzylamide group (see synthesis below) was installed using a solution of the tartaramide (32.3 mg, 100 μmol), HOBt (100 μmol), HBTU (90 μmol), and DIPEA (200 μmol) in 600 μL dry DMF, with agitation for 2 h. The resin was washed with DMF and acetonitrile.

For the Lys, Orn, and Dab scaffolds, the side-chain protecting group (MMT or MTT) was removed using three washes (1 mL each) of 3% (mol %) trichloroacetic acid in DCM, or until the supernatant was colorless. The product was cleaved and fully deprotected from the resin with AMA at 65 °C for 10 min, dried *in vacuo*, and purified by reverse-phase HPLC as described above.

For the Dpr scaffold, positional exchange between the α amine and side-chain amine was observed for the tartaramide group after on-bead acid treatment (as determined by LC/MS); to avoid this problem, the product was cleaved using AMA at 65 °C with the MTT group intact. The oligonucleotide was then purified by reverse-phase HPLC as previously described, lyophilized, and resuspended in 300 μL H_2O. The MTT group was then cleaved by addition of TFA to 3.5% (v/v) for 15 min. The sample was treated with 50 μL 3M NaOAc (pH 5.0) with 20 μg of glycogen and precipitated with ethanol to recover the pure product.

Preparation of Scaffold for Template 5a. The oligonucleotide CPG resin bearing the 5'-Fmoc-Lys(tartaramide) was prepared as previously described. The FMOC group was removed from the lysine α-amine using three consecutive washes with 20% piperidine/80% DMF (5 min agitation per wash). The resin was washed with DMF and acetonitrile. The (D)-phenylalanine was installed by addition of a solution of Fmoc-(D)-Phe-OH (100 μmol), HOBt (100 μmol), HBTU (90 μmol), and DIPEA (200 μmol) in 600 μL dry DMF, with agitation for 2 h. The resin

was then washed with DMF and acetonitrile. The desired material was cleaved from resin and purified by reverse-phase HPLC as previously described.

Preparation of Scaffold for Template 6a. The oligonucleotide CPG resin bearing the 5'- Fmoc-(D)-Phe-Lys(tartaramide) was prepared as previously described (synthesis of **5a**). The FMOC group was removed from the (D)-phenylalanine α-amine using three consecutive washes with 20% piperidine/80% DMF (5 min agitation per wash). The resin was washed with DMF and acetonitrile. The 2-furylalanine was installed by addition of a solution of Fmoc-2-furylalanine-OH (100 μmol, Peptech), HOBt (100 μmol), HBTU (90 μmol), and DIPEA (200 μmol) in 600 μL dry DMF, with agitation for 2 h. The FMOC group was removed from the 2-furylalanine α-amine using three consecutive washes with 20% piperidine/80% DMF (5 min agitation per wash, monitored by UV). The resin was washed with DMF and acetonitrile. The desired material was cleaved from resin and purified by reverse-phase HPLC as previously described.

Synthesis of the O,O'-Diacetyl-L-tartaric acid monobenzylamide. 2.00 g (9.25 mmol) of (+)-diacetyl-L-tartaric anhydride was dissolved in 25 mL dry DCM in a flame-dried 100 mL roundbottom with a magnetic stir bar. To this, 991 mg (1.01 mL, 9.25 mmol) of benzylamine was added dropwise while stirring; the reaction was stirred under nitrogen for 10 h, producing a solid white precipitate. This precipitate was collected by filtration, washed three times with ~5 mL cold DCM, and dried under vacuum to generate 2.86 g of a white powder (95.6% yield). ESI-MS analysis (positive mode): theoretical = 324.11, observed = 324.1059±0.3. ^{1}H-NMR (500 MHz, $CDCl_3$): δ 2.06 (3H, s), 2.18 (3H, d, J ~ 2.5 Hz), 4.45 (1 H, dd, J ~ 10.0, 5.0), 4.68 (1 H, dd, J ~ 6.5, 15.5), 5.68 (1 H, d, J ~ 3.0), 5.86 (1 H, d, J ~ 3.0), 6.49 (1 H, t, J ~ 6.0), 7.29 (2 H, m), 7.34, (1 H, m), 7.37 (2H, m).

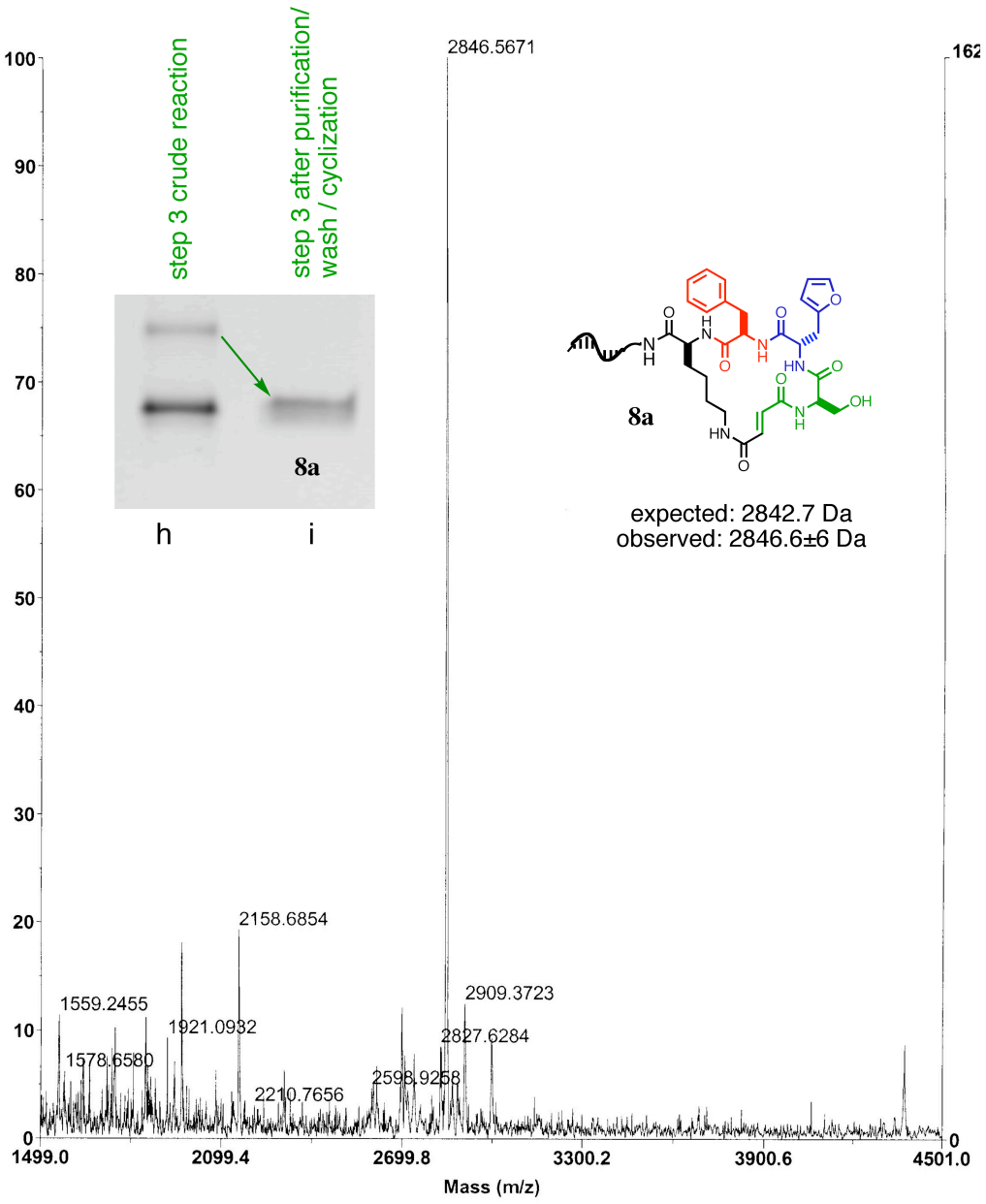

Figure S1. Denaturing PAGE and MALDI-TOF analysis of macrocycle **8a**, as synthesized using the capping-based approach (see main text, Figure 3).

scaffold	codon
Lys-s	5'- AAC
Orn-s	5'- CTA
Dab-s	5'- AGA
Dpr-s	5'- TAC
Lys-α	5'- CAA
Orn-α	5'- TGA
Dab-α	5'- ACA
Dpr-α	5'- GAA

Figure S2. Scaffold codons for template libraries **24**, **26**, and **28**.

scaffold	expected (M-H)	observed (±0.1 Da)	LC retention time (min)
1b (Lys-s)	742.29	742.29	8.78
14 (Orn-s)	728.27	728.27	8.64
15 (Dab-s)	714.26	714.25	8.58
16 (Dpr-s)	700.24	700.24	8.54
17 (Lys-α)	742.29	742.27	8.75
18 (Orn-α)	728.27	728.26	8.75
19 (Dab-α)	714.26	714.25	8.82
20 (Dpr-α)	700.24	700.24	8.64

Figure S3. LC/MS data for the eight template-linked scaffolds after S1 nuclease digestion.

Masses of Reagents for the Building-Block Compatibility Study (see main text, Figure 7)

"gly": glycine (Novabiochem)

"ach": 1-amino-1-cyclohexyl carboxylic acid (Sigma-Aldrich)

"pro": proline (Novabiochem)

"sta": statine (Bachem)

"tml": *N',N',N'*-trimethyl-lysine (Bachem)

"bdp": (D)-4-benzoylphenylalanine (Chem-Impex International)

amino acid	R_1 expected (M-H)	R_1 observed (±8 Da)	R_2 expected (M-H)	R_2 observed (±9 Da)	R_3 expected (M-H)	R_3 observed (±9 Da)
gly	3669.9	3674.3	4220.0	4226.3	4740.0	4747.7
ach	3579.9	3584.2	4140.0	4144.9	4710.1	4717.1
pro	3647.9	3654.1	4208.0	4213.9	4778.1	4785.1
sta	3619.9	3625.9	4180.0	4187.3	4750.1	4757.0
tml	3679.9	3685.0	4240.0	4245.4	4810.1	4817.3
bdp	3694.0	3699.8	4254.1	4260.6	4824.2	4830.3
control	3773.9	3781.1	4334.0	4342.3	4904.1	4909.8

Figure S4. MALDI-TOF characterization of the building-block compatibility study DNA-linked reagents.

codon	amino acid	commercial source	expected (M-H)	observed (±11 Da)
1A	(D)-phenylalanine	Sigma-Aldrich	5227.2	5226.1
1B	glycine	Novabiochem	5128.1	5126.4
1C	5-hydroxytryptophan	Novabiochem	5295.1	5296.8
1C (**25**)	asparagine	Novabiochem	5207.1	5209.5
1D	norvaline	Novabiochem	5152.1	5151.6
1E	isoleucine	Novabiochem	5184.1	5186.7
1F	isoglutamine	Novabiochem	5190.1	5190.7
1G	3-(2-thienyl)-alanine	Sigma-Aldrich	5215.0	5221.3
1H	statine	Bachem	5228.1	5229.8
1I	(O_2)-methionine	Bachem	5216.0	5216.2
1J	4-nitrophenylalanine	Bachem	5263.1	5267.5
1K	*N'*-pyrazinylcarbonyl-ornithine	Bachem	5282.1	5282.3
1L	(D)-4-benzoylphenylalanine	Chem-Impex-Intl.	5322.1	5321.6

Figure S5. MALDI-TOF characterization for the step 1 DNA-linked reagents.

codon	amino acid	commercial source	expected (M-H)	observed (±11 Da)
2A	2-furylalanine	Peptech	5168.1	5170.7
2B	β-alanine	Sigma-Aldrich	5102.1	5101.7
2B	proline	Novabiochem	5137.2	5137.9
2D	threonine	Novabiochem	5141.1	5140.9
2E	4-*trans*-hydroxyproline	Novabiochem	5153.2	5156.2
2F	β-(3-pyridyl)-alanine	Bachem	5179.2	5180.0
2G	*N*-methyl-phenylalanine	Sigma-Aldrich	5192.1	5193.7
2G (**25**)	phenylglycine	Novabiochem	5165.1	5169.8
2H	β-cyclohexylalanine	Bachem	5193.2	5196.6
2I	citrulline	Novabiochem	5179.1	5178.9
2J	*N',N',N'*-trimethyllysine	Bachem	5220.1	5223.2
2K	(D)-tryptophan	Sigma-Aldrich	5217.1	5219.6
2L	(*R*)-3-amino-3-(2,3-dimethoxy phenyl)-propionic acid	CSPS	5238.1	5239.7

Figure S6. MALDI-TOF characterization of the step 2 DNA-linked reagents.

codon	amino acid	commercial source	expected (M-H)	observed (±11 Da)
3A	(D)-serine	Sigma-Aldrich	5742.1	5745.0
3B	sarcosine	Novabiochem	5668.2	5665.4
3C	2-amino-4-pentenoic acid	Sigma-Aldrich	5654.2	5658.1
3D	4-amino-2-hydroxybutyric acid	Bachem	5667.2	5665.6
3E	β-cyclopropyl alanine	Bachem	5708.2	5711.3
3F	glutamine	Sigma-Aldrich	5734.2	5734.6
3G	*threo-tert*-butyl serine	Acros	5658.2	5657.4
3G (**32**)	homoserine	Novabiochem	5700.2	5702.5
3H	arginine	Novabiochem	5713.2	5715.6
3I	β-phenylalanine	CSPS	5758.2	5757.5
3J	styrylalanine	Peptech	5810.2	5810.0
3K	cyclohexylstatine	Bachem	5803.3	5804.6
3L	4-phosphonomethylphenylalanine	CSPS	5845.2	5847.6

Figure S7. MALDI-TOF characterization of the step 3 DNA-linked reagents.

codon	amino acid	expected (M-H)	observed (±0.1 Da)
1A	(D)-phenylalanine	972.35	972.36
1B	glycine	882.30	882.30
1C	5-hydroxytryptophan	1027.36	not observed
1D	norvaline	924.35	924.35
1E	isoleucine	938.38	938.38
1F	isoglutamine	953.34	953.34
1G	3-(2-thienyl)-alanine	978.31	978.32
1H	statine	982.39	982.40
1I	(O_2)-methionine	988.31	988.31
1J	4-nitrophenylalanine	1017.34	1017.35
1K	*N'*-pyrazinylcarbonyl-ornithine	1045.38	1045.36
1L	(D)-4-benzoylphenylalanine	1076.38	1076.40

Figure S8. LC/MS data for macrocycle library **21** (12x1x1).

codon	amino acid	expected (M-H)	observed (±0.1 Da)
2A	2-furylalanine	972.35	972.36
2B	β-alanine	906.34	906.36
2C	proline	932.36	932.36
2D	threonine	936.35	936.37
2E	4-*trans*-hydroxyproline	948.35	948.38
2F	β-(3-pyridyl)-alanine	983.37	983.40
2G	*N*-methyl-phenylalanine	996.39	not observed
2H	β-cyclohexylalanine	988.42	988.43
2I	citrulline	992.39	992.40
2J	*N',N',N'*-trimethyllysine	1005.44	1005.46
2K	(D)-tryptophan	1021.38	1021.38
2L	(*R*)-3-amino-3-(2,3-dimethoxy phenyl)-propionic acid	1042.39	1042.39

Figure S9. LC/MS data for macrocycle library **22** (1x12x1).

codon	amino acid	expected (M-H)	observed (±0.1 Da)
3A	(D)-serine	972.35	972.35
3B	sarcosine	956.35	956.36
3C	2-amino-4-pentenoic acid	982.37	982.37
3D	4-amino-2-hydroxybutyric acid	986.37	986.37
3E	β-cyclopropyl alanine	996.39	996.40
3F	glutamine	1013.38	1013.40
3G	*threo-tert*-butyl serine	1028.41	1028.41
3H	arginine	1041.42	1041.45
3I	β-phenylalanine	1046.40	1046.45
3J	styrylalanine	1058.40	1058.41
3K	cyclohexylstatine	1082.46	1082.47
3L	4-phosphonomethylphenylalanine	1126.37	1126.40

Figure S10. LC/MS data for macrocycle library **23** (1x1x12).

Complete Ref. 60:

(60) Woodward, R. B.; Logusch, E.; Nambiar, K. P.; Sakan, K.; Ward, D. E.; Au-Yeung, B,-W.; Balaram, P.; Browne, L. J.; Card, P. J.; Chen, C. H.; Chevert, R. B.; Fliri, A.; Frobel, K.; Gais, H.-J.; Garratt, D. G.; Hayakawa, K.; Heggie, W.; Hesson, D. P.; Hoppe, D.; Hoppe, I.; Hyatt, J. A.; Ikeda, D.; Jacobi, P. A.; Kim, K. S.; Kobuke, Y.; Kojima, K.; Krowicki, K.; Lee, V. J.; Leutert, T.; Malchenko, S.; Martens, J.; Matthews, R. S.; Ong, B. S.; Press, J. B.; Rajan Babu, T. V.; Rousseau, G.; Sauter, H. M.; Suzuki, M.; Tatsuta, K.; Tolbert, L. M.; Truesdale, E. A.; Uchida, I.; Ueda, Y.; Uyehara, T.; Vasella, A. T.; Vladuchick, W. C.; Wade, P. A.; Williams, R. M; Wong, H. N.-C. *J Am Chem Soc* **1981**, *103*, 3213-3215.