Supplementary information

Fragment expansion with NUDELs - poised DNA-encoded libraries

Catherine L. A. Salvini,[a] Benoit Darlot,[b] Jack Davison,[a] Mathew P. Martin,[c] Sue Tudhope,[c] Shannon Turberville,[c] Akane Kawamura,[b] Martin E. M. Noble,[c] Stephen R. Wedge,[c] James Crawford,[d] Michael J. Waring[a]*

[a] Cancer Research Horizons Newcastle Drug Discovery Group, Chemistry, School of Natural and Environmental Sciences, Bedson Building, Newcastle University, NE1 7RU, UK.

[b] Chemistry, School of Natural and Environmental Sciences, Bedson Building, Newcastle University, NE1 7RU, UK.

[c] Cancer Research Horizons Newcastle Drug Discovery Group, Paul O'Gorman Building, Newcastle University, NE2 2AH, UK.

[d] Genentech Inc. 1 DNA Way, South San Francisco, California, 94080, USA.

Contents

Reagents and Equipment

A. **Solvents and Reagents**

Chemicals were purchased from Fluorochem and Sigma-Aldrich (Merck); they were used without further purification unless otherwise indicated. Fmoc-NH-PEG_4-CO_2H linker was purchased from key organics. TPGS-750-M was purchased from Sigma-Aldrich (Merck). All water used with DNA substrates was nuclease-free water purchased from ThermoFisher. DNA was purchased from Sigma-Aldrich (Merck) or IDT either attached to solid support, or as single stranded product, used without further purification unless otherwise specified. Deuterated solvents for NMR spectroscopy were purchased from Sigma-Aldrich. Anhydrous solvents using SureSeal™ or Acroseal™ were purchased from either Sigma-Aldrich or Acros, respectively.

B. **Analysis and Equipment**

FTIR spectra were measured using an Agilent Cary 630 FTIR as a neat sample. LC-MS analyses were conducted using a Waters Acquity UPLC system with PDA and ELSD. When a 2 min gradient was used, the sample was eluted on Acquity UPLC BEH C^{18}, 1.7 µm, 2.1 x 50 mm, with a flow rate of 0.6 µL/min

using 5-95% 0.1% HCOOH in MeCN. HRMS analyses were conducted using an Agilent 6550 iFunnel QTOF LC-MS with an Agilent 1260 Infinity UPLC system. The sample was eluted on Acquity UPLC BEH C^{18} (1.7 µm, 2.1 x 50mm) with a flow rate of 0.7 $mLmin^{-1}$, and run at a gradient of 1.2 min 5-95% 0.1% HCOOH in MeCN with 0.1% aq. HCOOH. Calculated exact masses were quoted from ChemDraw Professional 20.0. All final compounds are >95% purity by HPLC.

DNA mass spectra were measured on an Agilent 6550 QTOF in negative mode, using standard 3200 m/z maximum and 2 GHz extended dynamic range. Drying gas temperature was at 260 °C at 12 L min^{-1}, sheath gas temperature was 400 °C at 12 L min^{-1}, nebuliser at 45 psig, VCap voltage of 4000 V and nozzle voltage of 2000 V.

The LC was carried out on an Agilent 1260 Infinity 2 using either an Agilent Advancedbio oligonucleotides column, 2.1x150 mm or 2.1x100 mm using methods A and B, respectively. Method A: the gradient was run at 0.4 mL min^{-1} from 10% MeOH to 40% MeOH over 8 mins against 800 mM HFIP:8 mM DIPEA buffer solution. A 3 min flush at 95% MeOH ended each run. Method B: the gradient was run at 0.8 mL min^{-1} from 10% MeOH to 50% MeOH over 3 mins against 200 mM HFIP:15 mM DIPEA buffer solution. A 1 min flush at 95% MeOH ended each run. Analysis of data was carried out using Agilent Qualitative Analysis version 7.0.

DNA transformations including phosphorylations, ligations, and PCR amplifications carried out using a Techne Prime thermal cycler, 96 x 0.2ml thermal cycler. Precise conditions described below. qPCR analysis carried out using either a CFX Opus 96 Real-Time PCR system, or a CFX96 Touch Deep Well Real-Time PCR Detection System. Reagents purchased from ThermoFisher or Sigma Aldrich (Merck). Next generation sequencing (NGS) carried out by GENEWIZ (South Plainfield, NJ) to undergo their Amplicon-EZ service on an Illumina platform.

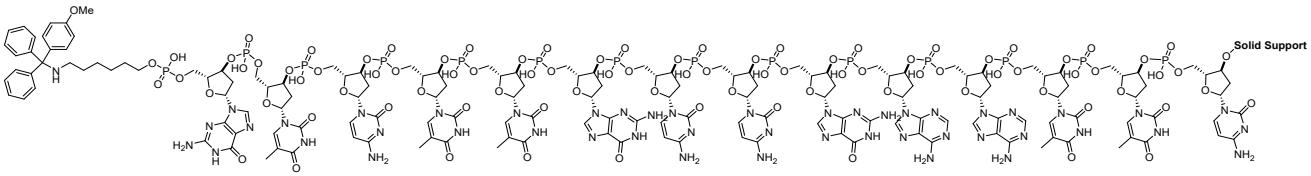
C. **Chromatography Techniques**

Normal phase column chromatography purifications were carried out using Biotage SP4 and Isolera automated flash system with UV monitoring at 278 nm and collection at 254 nm. Grace Resolve pre-packed flash cartridges were used for normal phase separations.

Preparative HPLC purification was carried on an Agilent 1260 infinity system using a Phenomenex Clarity 5 µM Oligo-RP column, 10x150 mm. The gradient was run at 5 mL/min from 10% MeOH to 90% MeOH over 22 mins against an 800 mM HFIP:8 mM TEA buffer solution. Fractions were analysed at 210 and 260 nm wavelengths.

D. **DNA Headpiece Material**

DNA headpiece materials. Two types of DNA were used to construct the DNA headpiece **HP-01** were both purchased from Sigma-Aldrich (Merck) either attached to solid support, or as single stranded product. DNA headpiece modifiable strand (5'- /5Phos/GTCTTGCCGAATTC-3', Figure 22 this is purchased attached to a polymer support through the 3' hydroxyl group with an MMT protection of the modified 5' phosphate. DNA headpiece complementary strand, comp. strand, (5'-/5OH/CAGAACGGCTTA -3', Figure 3) was purified by HPLC as described above prior to use. DNA components are received as granular solids.

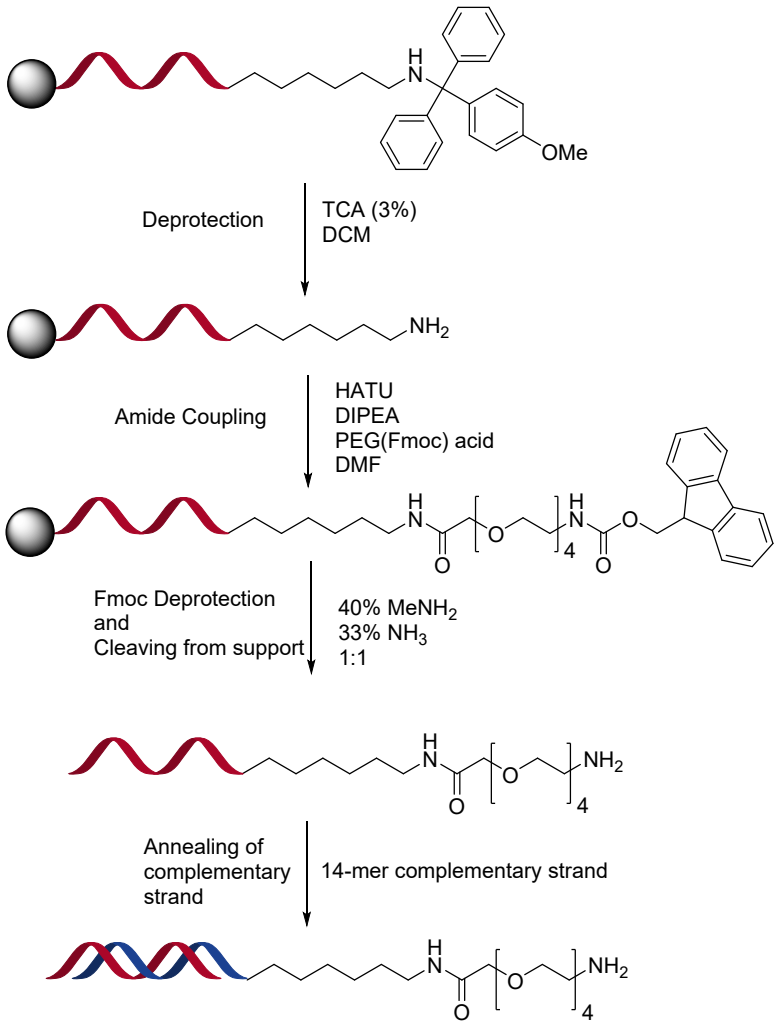


DNA headpiece as purchased modifiable strand with chemical spacer.

Chemistry Procedures On-DNA

A. On DNA Generic Procedures

A.I. Synthesis of PEG ds-14mer



To cleave the MMT protecting group ps-ss-14mer(MMT) (100 mg, 2.0 µmol) was washed with 3% TCA in DCM (15 mL), the filtrate was a strong yellow colour. This was continued until the filtrate ran colourless. The ps-DNA was then washed with DCM and allowed to air dry[24].

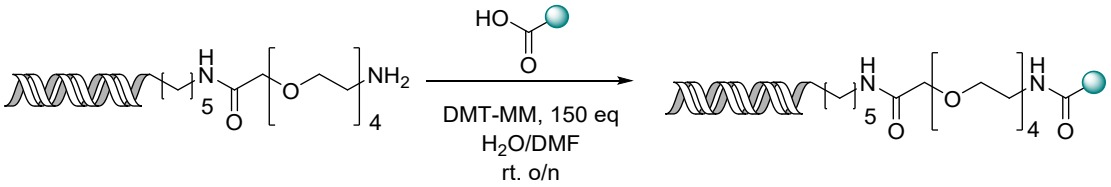
An Eppendorf was charged with 1-(9H-fluoren-9-yl)-3-oxo-2,7,10,13,16-pentaoxa-4-azaoctadecan-18-oic acid (19 mg, 40 µmol), DMF (1.0 mL), DIPEA (17 µL, 100 µmol), and HATU (17 mg, 44 µmol). The mixture was vortexed for 20 minutes before the dry ps-ss-14mer was added and the reaction was shaken for 16 hours at room temperature. The ps-DNA was washed with DMF (1500 µL), MeCN (1500 µL), MeOH (1500 µL), and DCM (1500 µL) and allowed to air dry before being suspended in a 1:1 mixture of NH_3 aq. (40%, 500 µL) and $MeNH_2$ aq. (33%, 500 µL). The suspension was then shaken for

16 hours at room temperature. The DNA product was collected by filtration and purified by HPLC. The gradient was run at 5 mL/min from 10% MeOH to 90% MeOH over 22 mins against an 800 mM HFIP:8 mM TEA buffer solution.

A.II. **General Ethanol Precipitation Procedure**

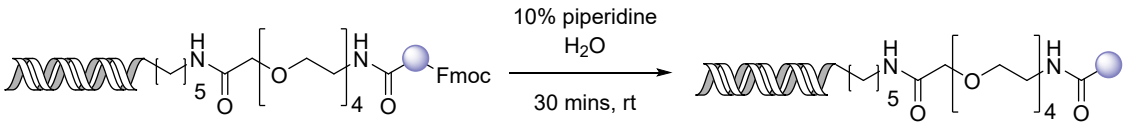
Aqueous sodium chloride (10% volume, 4 M) added to aqueous DNA solution followed by cold absolute ethanol to give an 80% ethanol solution. The mixture was incubated at - 78 °C for 1 hour. The mixture was then centrifuged, and the ethanol layer removed. Aqueous ethanol solution (70% v/v) was added, and the process repeated. The pellet of DNA then dissolved in water to give a 1 mM solution.[48]

A.III. **Amide Coupling with DMT-MM in DMF**



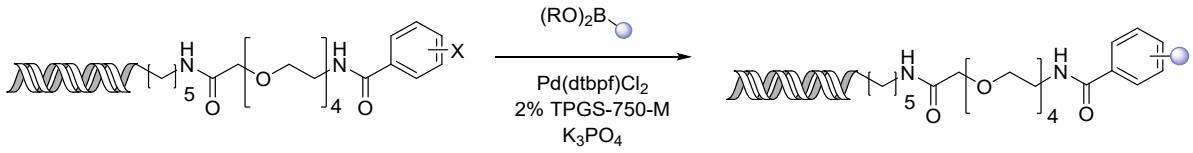
A solution of DNA **6** (1.0 mM) was pipetted into a PCR tube, to which was added pH 9.4 borate buffer (150 mM, 50 µL), amino acid in DMF (150 mM, 12.6 µL), and freshly prepared DMT-MM in water (250 mM, 7.6 µL). Each well was vortexed for 10 seconds and left to shake for 6 hours. A second addition of DMT-MM solution (250 mM, 7.6 µL) was carried out and the mixture allowed to shake for a further 16 hours at RT. The reaction was worked up by ethanol precipitation according to procedure A.II.[6]

A.IV. **Fmoc Deprotection Conditions**



Fmoc-protected DNA-construct solution (1.0 mM) was added to an Eppendorf followed by piperidine to give a 10% aqueous solution (v/v). The mixture was incubated at room temperature for 1 hour. The reaction is worked up by ethanol precipitation according to procedure A.II.[48]

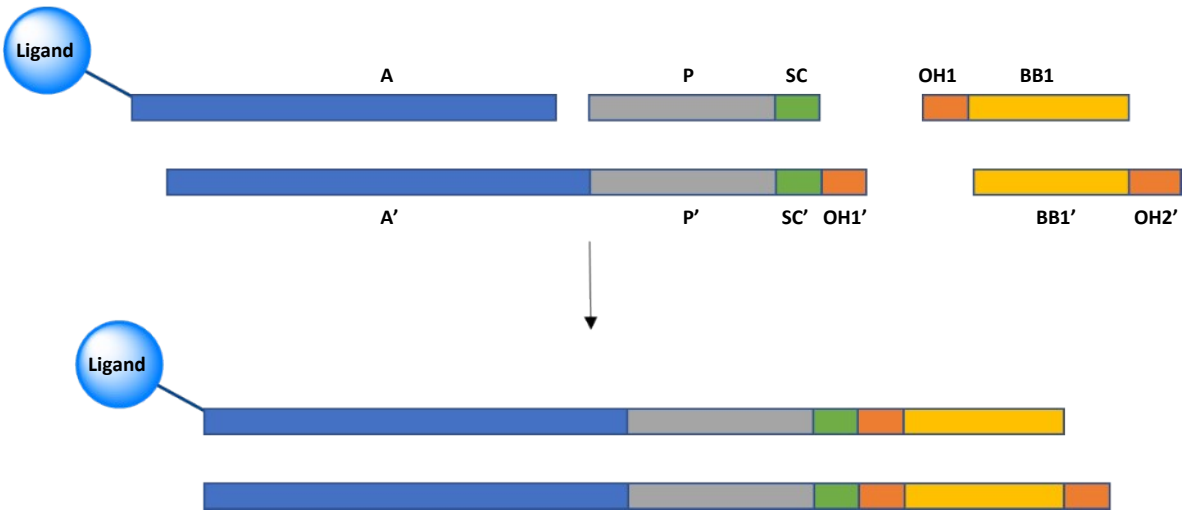
A.V. **Suzuki Cross Coupling Conditions**



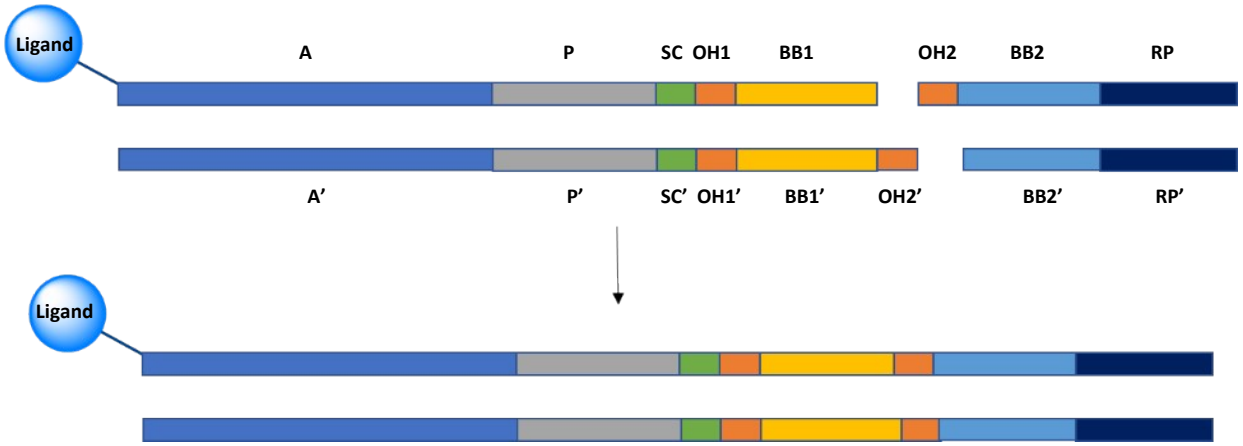
Boronic acid/ester solution (20 µl, 0.75 M in DMF) was weighed into a 50 µl glass insert for a para-dox 96-well micro-Para-dox™ photoredox/optimisation plate. To the insert was added; a solution of 5% TPGS-750-M in water (4.0 µl), potassium phosphate (6.0 µl of a solution of 113 mg in 200 µl water) and the halogenated headpiece (20 µl of 1 mM in 2% TPGS-750-M in water). The vials were subsequently vortexed for 30 seconds each. $Pd(dtbpf)Cl_2$ (4.5 µl of 6.37mg in 200 µl THF) was added and the samples vortexed again for 10 seconds each. The mixtures were heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 60 °C for 5 hours. QTOF mass spectrometry was used to analyse reactions. Samples prepared by diluting reaction mixture with water (300 µl) DCM (300 µl) was added to each vial and vortexed. The organics were removed, and aliquot taken for mass spec analysis. The reaction was worked up by ethanol precipitation according to procedure A.II.[66]

B. **2D Library Label Synthesis**

First Ligation



Second Ligation



Experimental Table 4 - Codes, functions, and corresponding sequence for each DNA section

Code	Function	Sequences 5'-3'
A	Adapter – 5' aminolinked head piece	GTCTTGCCGAATTC
A'	Complimentary adapter	GAATTCGGCAAGAC
P	Primer	AGGTCGGTGTGAACGGATTTG
P'	Complementary primer	CAAATCCGTTCACACCGACCT
S	Scaffold code	CATGTAA
S'	Complementary scaffold code	
OH1	Ligation overhang 1	GTAT
OH1'	Complementary OH1	ATAC
BB1	Building block 1	xxxxxxxx
BB1'	Complementary BB1	xxxxxxxx
OH2	Ligation overhang 2	CCTA
OH2'	Complementary OH2	TAGG
BB2	Building block 2	xxxxxxxx
BB2'	Complementary BB2	xxxxxxxx
P2	Complementary to P2'	TGACCTCAACTACATGGTCTACA
P2'	Primer (reverse)	TGTAGACCATGTAGTTGAGGTCA

B.I. **Phosphorylation**

Prior to ligation, the 5' terminus of each strand was phosphorylated in separate reactions. DNA strands (450 µM, 9000 pmol in overall reaction media of 20 µl) were added PNK reaction buffer (2 µl, 500mM Tris-HCl [pH 7.6 at 25°C], 100mM MgCl2, 50 mM DTT, 1mM spermidine), ATP (2 µl, 10 mM, Thermo Scientific), T4 Polynucleotide Kinase (1 µl, 10U/µl, Thermo Scientific) and nuclease free water (up to 20 µl). The reaction was carried out at 37 °C for 1 hour, followed by heating to 75 °C for 10 mins. DNA was used in the ligation steps without purification or precipitation.

B.II. **Ligation**

Ligations contained DNA (100 µM, 9000 pmol in overall reaction media of 90 µl), phosphorylated DNA strands (20 µl, 9000 pmol), 10X T4 DNA ligase buffer (9 µl, 400 mM Tris-HCl, 100 mM MgCl2, 100 mM DTT, 5 mM ATP9), water (up to 90 µl) and T4 DNA Ligase (3 µl, 30 Weiss U/µL). The ligations were carried out at 25°C for 16 hours, followed by heating to 75°C for 10 mins. Each ligation was purified by ethanol precipitation prior to the subsequent organic reaction taking place.

B.III. **PCR Amplification**

Forward and reverse primers were designed to amplify the DEL library, flanked by 5' Illumina adapter sequences to enable downstream sequence analysis and differentiation from target sequence. Each PCR was performed in a 50µl reaction mixture containing AmpliTaq Gold® 360 Master Mix (Thermo Fisher) and 200 ng of 1x1 prototype library (at 4.2 µM). The final concentration of each primer used was 10 µM. PCR amplification carried out using a Techne Prime thermal cycler, 96 x 0.2ml. The thermal cycling conditions were as follows: 10 min at 95 °C, followed by 40 cycles of 30 s at 95 °C, 30 s at 55°C, and 1 min at 72°C, with a final extension time of 420 s at 72°C. A negative control (distilled water in place of primers) was included in each run. Following the PCR reactions, the samples were run on a 4% Agarose E-gel.

Table 1 Forward and reverse primers for PCR. Primer sequence (blue) and NGS Illumina elongation sequence (purple).

Primer	**Sequence 5'-3'**	**Length**
Long Forward primer (LFP)	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGTAG ACCATGTAGTTGAGGTCA	56
Long Reverse primer (LRP)	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGTC GGTGTGAACGGATTTG	53

Selection Procedures

C. Selection Against BRD4

C.I. Control Selection with Positive Control

Product	Comments
His6-BRD4	20uL aliquot (76uM)
Dynabeads™ His-Tag Isolation and Pulldown	/
Buffer*	HEPES 40 mM, NaCl 300 mM, 0.1 mg/mL BSA, 0.1 mg/mL Salmon Sperm DNA, 0.05% TWEEN20 (pH 7.8)
SYBR™ Green PCR Master Mix	/
Hard-Shell® Low-Profile, Thin-Wall, Skirted 96-Well PCR Plates, White Shell/White Well	/
Microseal 'B' PCR Plate Sealing Film, adhesive, optical	/
Protein LoBind Tubes, Protein LoBind®, 0.5 mL, PCR clean, colorless, 1 bag × 500 tubes	/

Forward Primer (FW)	AGGTCGGTGTGAACGGATTTG
Reverse Primer (RV)	TGTAGACCATGTAGTTGAGGTCA

JQ1-DNA (JQ1 attached to 5' written 5'→3')	GTCTTGCCGAATTCAGGTCGGTGTGAACGGATTTGCATGTAAGT ATACTGGCTACCTGGCGTACATTGACCTCAACTACATGGTCTACA

An appropriate amount of His Trap Dynabead was washed 3 times with 100 µL of buffer and then re-suspended in the original amount of slurry. After a bead loading experiment, it was found that the loading capacity of the His-trap Dynabeads was 16 pmol BRD4/uL of slurry. Positive beads (POS): 20 pmol of BRD4 was incubated with 1.25 µL of washed Dynabead slurry and then buffer was added up to 20 µL in Low binding tube. Negative beads 1 (NEG1): 20 pmol of BRD4 was incubated with 1.25 µL of washed Dynabead slurry and then buffer was added up to 20 µL in Low binding tube. Negative beads 2 (NEG2): 1.25 µL of washed Dynabead slurry was diluted with buffer up to 20 µL in Low binding tube. POS, NEG1 and NEG2 were incubated at 4°C for 30 min on a rotating wheel. POS, NEG1 and

NEG2 were then washed 3 times with 100 µL buffer and finally left dry. POS: 2 µMol of JQ1-DNA in 20 µL of buffer was added to the beads and left to incubate at 25°C for 1h. NEG1: 2 µMol of DNA (without JQ1) in 20 µL of buffer was added to the beads and left to incubate at 25°C for 1h. NEG2: 2 µMol of JQ1-DNA in 20 µL of buffer was added to the beads and left to incubate at 25°C for 1h. This need to be quick, but precise to avoid losing binding molecules in the washes. Each sample supernatant is then removed, and beads are quickly resuspended in 20 µL ice cold buffer and then transferred in a fresh low binding tube. This step is repeated twice more. The resulting beads are then suspended in 20 µL of fresh buffer. All beads are then heated at 95°C for 5min. The supernatant is then immediately removed and stored in a new fresh low binding tube.

qPCR mix	Volume (uL)
SYBR Green (2X)	330
FW primer (100 µM)	1.65
RV primer (100 µM)	1.65
H_2O	293.7
Total for 10 reactions	627

qPCR was set up as follows with 1uL of sample mixed with 19uL of qPCR mix.

	1	2	3	4	5	6	7	8	9
A	std 10^{10}	std 10^{8}	std 10^{6}	std 10^{4}	H_2O	input	POS	NEG1	NEG2
B	std 10^{10}	std 10^{8}	std 10^{6}	std 10^{4}	H_2O	input	POS	NEG1	NEG2
C	std 10^{10}	std 10^{8}	std 10^{6}	std 10^{4}	H_2O	input	POS	NEG1	NEG2

qPCR program: 95°C for 3min, (95°C for 40s, 61°C for 40s, 72°C for 40s, imaging)*40. Results were analysed using the CFX manager software and label

C.II. **BRD4 Selection with Library**

Product	Comments
His6-BRD4	20uL aliquot (76uM)
Dynabeads™ His-Tag Isolation and Pulldown	/
Buffer	HEPES 40 mM, NaCl 300 mM, 0.1 mg/mL BSA, 0.1 mg/mL Salmon Sperm DNA, 0.05% TWEEN20 (pH 7.8)
SYBR™ Green PCR Master Mix	/

Hard-Shell® Low-Profile, Thin-Wall, Skirted 96-Well PCR Plates, White Shell/White Well	/
Microseal 'B' PCR Plate Sealing Film, adhesive, optical	/
Protein LoBind Tubes, Protein LoBind®, 0.5 mL, PCR clean, colorless, 1 bag × 500 tubes	/

Forward Primer (FW)	AGGTCGGTGTGAACGGATTTG
Reverse Primer (RV)	TGTAGACCATGTAGTTGAGGTCA

JQ1-DNA (JQ1 attached to 5')	GTCTTGCCGAATTCAGGTCGGTGTGAACGGATTTGCATGTAAGT ATACTGGCTACCTGGCGTACATTGACCTCAACTACATGGTCTACA

An appropriate amount of His Trap Dynabead was washed 3 times with 100 µL of buffer and then re-suspended in the original amount of slurry. After a bead loading experiment it was found that the loading capacity of the His-trap Dynabeads was **16 pmol BRD4/uL of slurry.** Positive beads (POS_JQ1_sl): 20 pmol of BRD4 was incubated with 1.25 µL of washed Dynabead slurry and then buffer was added up to 20 µL in Low binding tube. Negative beads (NEG_JQ1_sl): 1.25 µL of washed Dynabead slurry was diluted with buffer up to 20 µL in Low binding tube. Positive beads (POS_csDEL_1): 20 pmol of BRD4 was incubated with 1.25 µL of washed Dynabead slurry and then buffer was added up to 20 µL in Low binding tube. Negative beads (NEG_csDEL_1): 1.25 µL of washed Dynabead slurry was diluted with buffer up to 20 µL in Low binding tube. POS and NEG were incubated at 4°C for 30 min on a rotating wheel. POS and NEG were then washed 3 times with 100 µL buffer and finally left dry. POS_JQ1_sl: 2 nmol of JQ1_sl-DNA in 20 µL of buffer was added to the beads and left to incubate at 25°C for 1h. NEG_JQ1_sl: 2 nmol of JQ1_sl-DNA in 20 µL of buffer was added to the beads and left to incubate at 25°C for 1h. POS_csDEL_1: 2 nmol of csDEL-DNA in 20 µL of buffer was added to the beads and left to incubate at 25°C for 1H (containing 0.052nmol of JQ1-DNA (PEG linker) same quantity as each of the 42 members of this library). NEG_csDEL_1: 2 nmol of JQ1_sl-DNA in 20 µL of buffer was added to the beads and left to incubate at 25°C for **30 min** (containing 0.052nmol of JQ1-DNA (PEG linker) same quantity as each of the 42 members of this library). POS_csDEL_1: supernatant from NEG_csDEL_1 was transferred on POS_csDEL_1 beads and further incubated at 25°C for

1h. **IMPORTANT STEP: (need to be quick, but precise to avoid losing binding molecules in the washes).** Each sample supernatant is then removed, and beads are **quickly** resuspended in 20 µL ice cold buffer and then transferred in a fresh low binding tube. This step is repeated twice more. The resulting beads are then suspended in 20 µL of fresh buffer. All beads are then heated at 95°C for 5min. The supernatant is then immediately removed and stored in a new fresh low binding tube.

qPCR mix	Volume (uL)
SYBR Green (2X)	330
FW primer (100 µM)	1.65
RV primer (100 µM)	1.65
H_2O	293.7
Total for 10 reactions	627

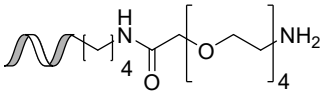
qPCR was set up as follows with 1uL of sample mixed with 19uL of qPCR mix.

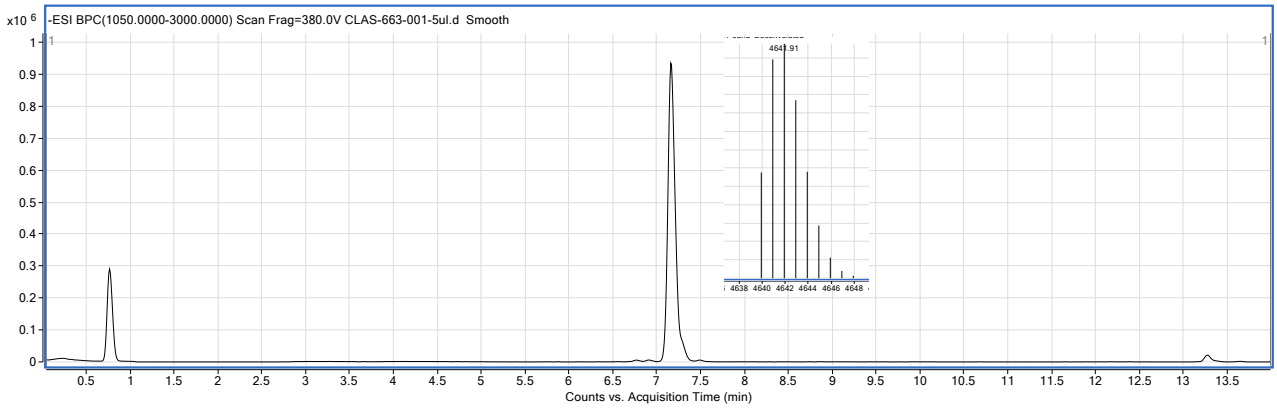
qPCR program: 95°C for 3min, (95°C for 40s, 61°C for 40s, 72°C for 40s, imaging)*40. Results were analysed using the CFX manager software and label

Synthesised Constructs and Chromatograms

D. **Chromatograms**

D.I. **Headpiece, 6**

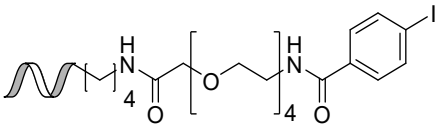


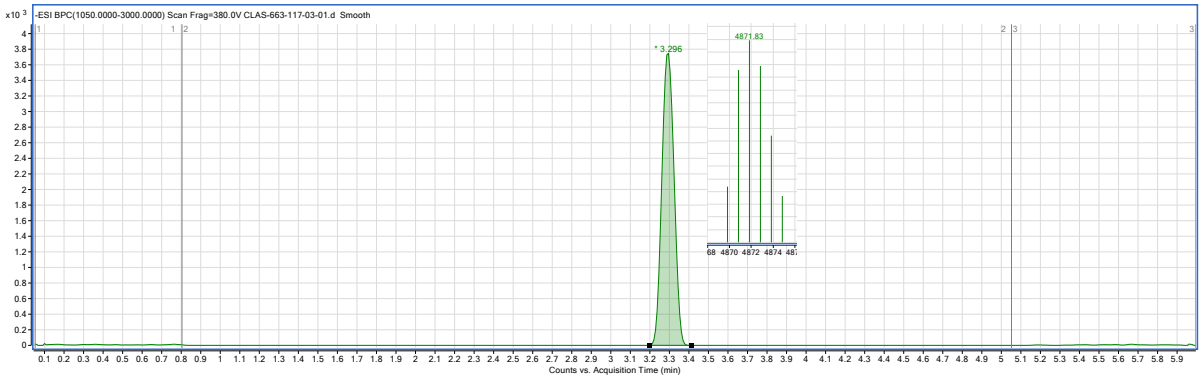


Mass calculated: 4639.93
Mass recorded: 4639.91
6 was synthesised according to procedure A.I (Synthesis of PEG ds-14mer aka. Headpiece-01) and analysed by mass spectrometry.

Example of Validation of Warhead Coupling Aryl Acid to 6

D.II. **13**

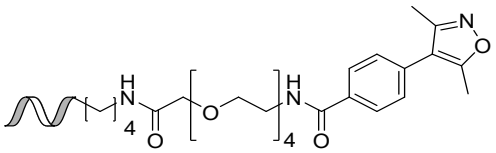


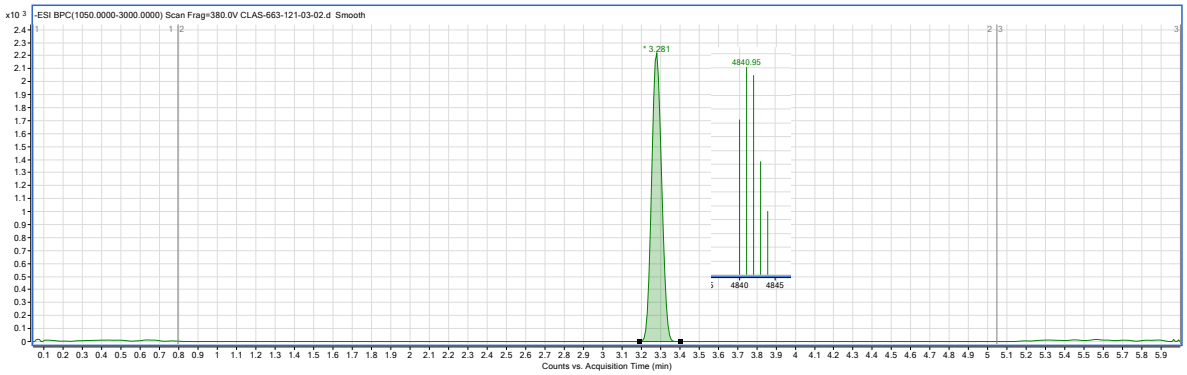


Mass calculated: 4869.85

Mass recorded: 4869.83
Conversion: 100%.
Synthesised from **6** according to procedure E.III.i. and analysed by mass spectrometry.

D.III. **Isoxazole Aryl-1 HP, 13-SI**

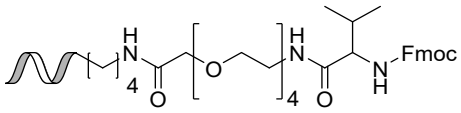


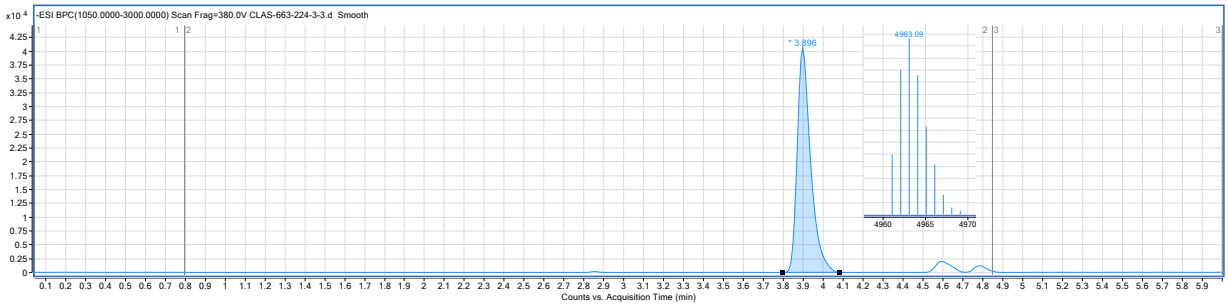


Mass calculated: 4838.99
Mass found: 4839.94
Conversion: 100%.
Synthesised from **13** according to procedure E.V and analysed by mass spectrometry.

1x1 Exemplar Library Member

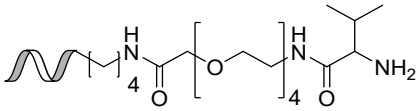
D.IV. **Fmoc-Valine-Headpiece 1B-SI**

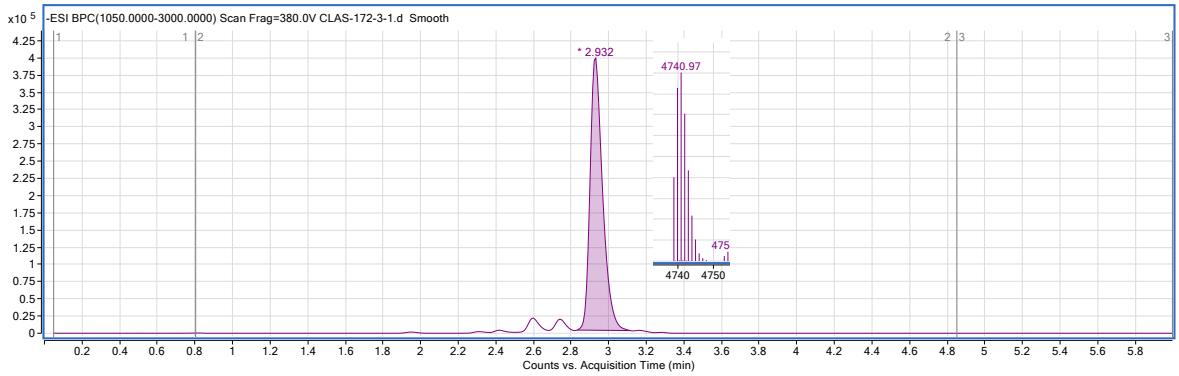




Mass calculated: 4961.07
Mass found: 4961.09
Conversion: 100%.
Synthesised from **6** according to procedure E.V and analysed by mass spectrometry.

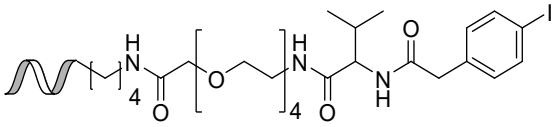
D.V. **Valine-Headpiece 2B-SI**

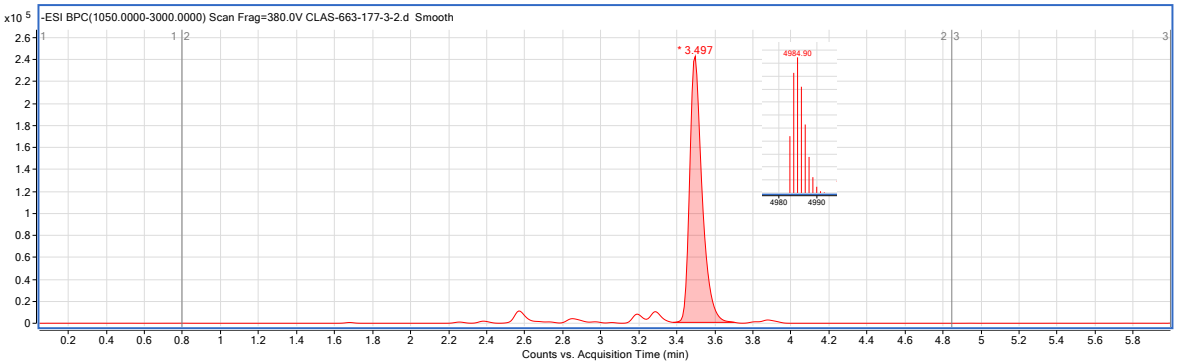




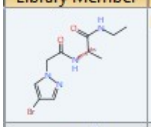
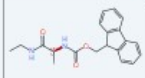
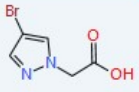
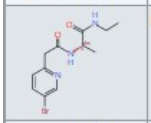
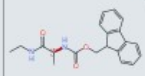
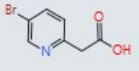
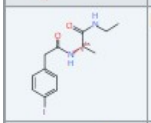
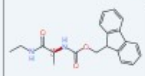
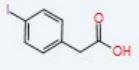
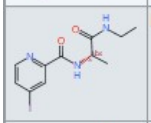
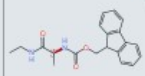
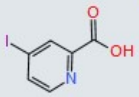
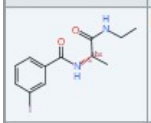
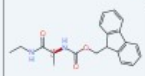
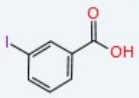
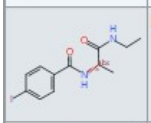
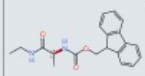
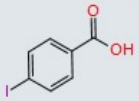
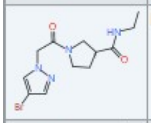
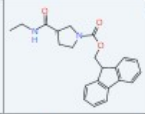
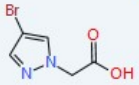
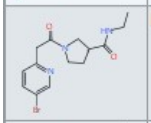
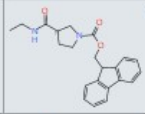
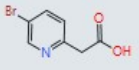
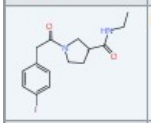
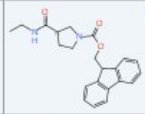
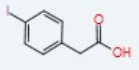
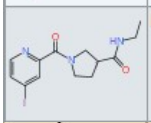
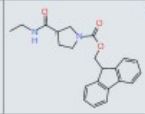
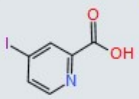
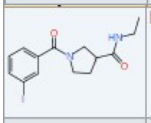
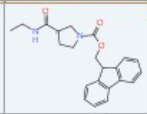
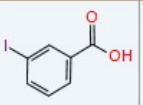
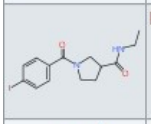
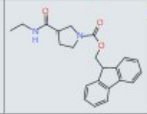
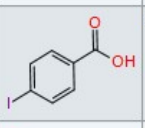
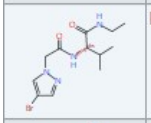
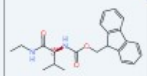
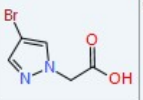
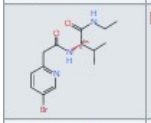
Mass calculated: 4739.00
Mass recorded: 4738.97
Conversion: 100%.
Synthesised from **27** according to procedure D.V. and analysed by mass spectrometry.

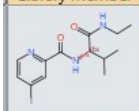
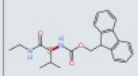
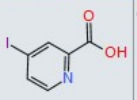
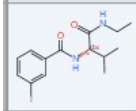
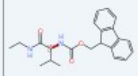
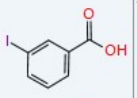
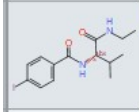
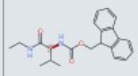
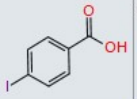
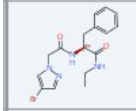
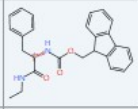
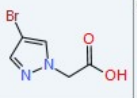
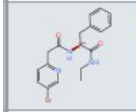
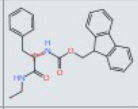
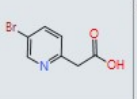
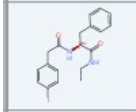
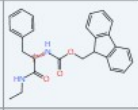
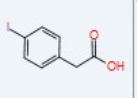
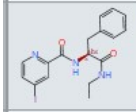
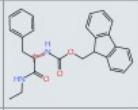
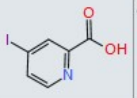
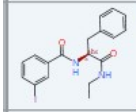
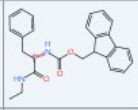
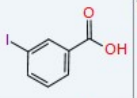
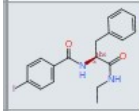
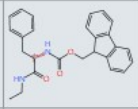
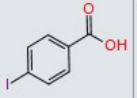
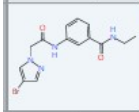
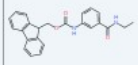
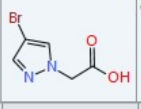
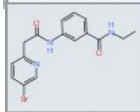
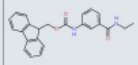
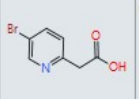
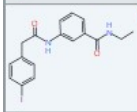
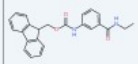
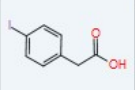
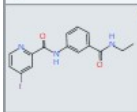
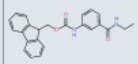
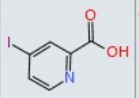
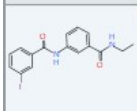
D.VI. **3-iodobenzoyl-Valine-Headpiece (3B-SI)**

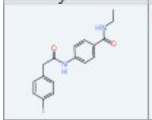
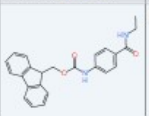
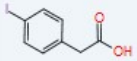
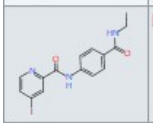
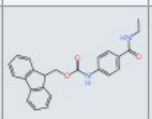
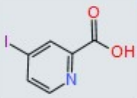
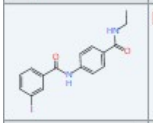
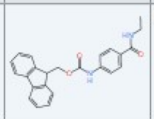
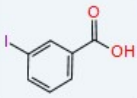
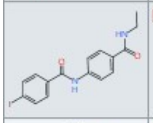
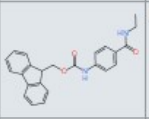
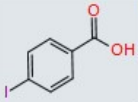
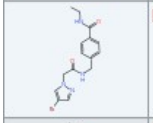
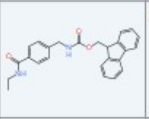
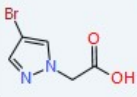
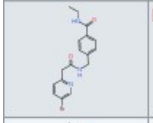
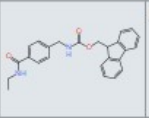
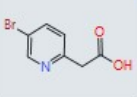
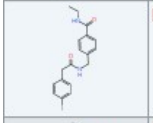
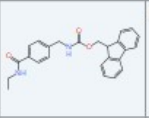
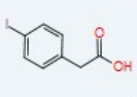
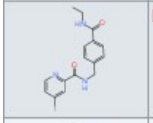
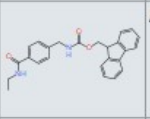
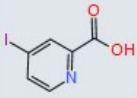
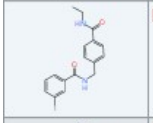
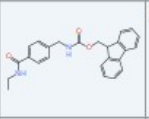
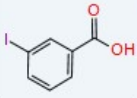
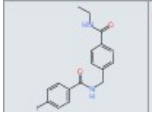
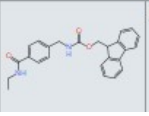
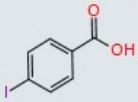




Mass calculated: 4982.94
Mass recorded: 4982.89
Conversion: 100%.
Synthesised from **36** according to procedure D.V. and analysed by mass spectrometry.

Library Member	BB1 Number	BB1	BB1 Codon	BB1 Complementary Codon	BB2 Number	BB2	BB2 Codon	BB2 Complementary Codon
	BB1a		ACTCTGGA	TCCAGAGT	BB2a		CTTAGAGC	GCTCTAAG
	BB1a		ACTCTGGA	TCCAGAGT	BB2b		GATCGACT	AGTCGATC
	BB1a		ACTCTGGA	TCCAGAGT	BB2c		TCTGGAAC	GTTCCAGA
	BB1a		ACTCTGGA	TCCAGAGT	BB2d		ATTGACCG	CGGTCAAT
	BB1a		ACTCTGGA	TCCAGAGT	BB2e		TGTCACGA	TCGTGACA
	BB1a		ACTCTGGA	TCCAGAGT	BB2f		GCTAACTG	CAGTTAGC
	BB1b		GGTGTTAC	GTAACACC	BB2a		CTTAGAGC	GCTCTAAG
	BB1b		GGTGTTAC	GTAACACC	BB2b		GATCGACT	AGTCGATC
	BB1b		GGTGTTAC	GTAACACC	BB2c		TCTGGAAC	GTTCCAGA
	BB1b		GGTGTTAC	GTAACACC	BB2d		ATTGACCG	CGGTCAAT
	BB1b		GGTGTTAC	GTAACACC	BB2e		TGTCACGA	TCGTGACA
	BB1b		GGTGTTAC	GTAACACC	BB2f		GCTAACTG	CAGTTAGC
	BB1c		AGTGTCGT	ACGACACT	BB2a		CTTAGAGC	GCTCTAAG
	BB1c		AGTGTCGT	ACGACACT	BB2b		GATCGACT	AGTCGATC
	BB1c		AGTGTCGT	ACGACACT	BB2c		TCTGGAAC	GTTCCAGA

Library Member	BB1 Number	BB1	BB1 Codon	BB1 Complementary Codon	BB2 Number	BB2	BB2 Codon	BB2 Complementary Codon
	BB1c		AGTGTCGT	ACGACACT	BB2d		ATTGACCG	CGGTCAAT
	BB1c		AGTGTCGT	ACGACACT	BB2e		TGTCACGA	TCGTGACA
	BB1c		AGTGTCGT	ACGACACT	BB2f		GCTAACTG	CAGTTAGC
	BB1d		GCTCTAAG	CTTAGAGC	BB2a		CTTAGAGC	GCTCTAAG
	BB1d		GCTCTAAG	CTTAGAGC	BB2b		GATCGACT	AGTCGATC
	BB1d		GCTCTAAG	CTTAGAGC	BB2c		TCTGGAAC	GTTCCAGA
	BB1d		GCTCTAAG	CTTAGAGC	BB2d		ATTGACCG	CGGTCAAT
	BB1d		GCTCTAAG	CTTAGAGC	BB2e		TGTCACGA	TCGTGACA
	BB1d		GCTCTAAG	CTTAGAGC	BB2f		GCTAACTG	CAGTTAGC
	BB1e		CATCGTGA	TCACGATG	BB2a		CTTAGAGC	GCTCTAAG
	BB1e		CATCGTGA	TCACGATG	BB2b		GATCGACT	AGTCGATC
	BB1e		CATCGTGA	TCACGATG	BB2c		TCTGGAAC	GTTCCAGA
	BB1e		CATCGTGA	TCACGATG	BB2d		ATTGACCG	CGGTCAAT
	BB1e		CATCGTGA	TCACGATG	BB2e		TGTCACGA	TCGTGACA
	BB1e		CATCGTGA	TCACGATG	BB2f		GCTAACTG	CAGTTAGC
	BB1f		AATCGGTC	GACCGATT	BB2a		CTTAGAGC	GCTCTAAG
	BB1f		AATCGGTC	GACCGATT	BB2b		GATCGACT	AGTCGATC
	BB1f		AATCGGTC	GACCGATT	BB2c		TCTGGAAC	GTTCCAGA

Library Member	BB1 Number	BB1	BB1 Codon	BB1 Complementary Codon	BB2 Number	BB2	BB2 Codon	BB2 Complementary Codon
	BB1f		AATCGGTC	GACCGATT	BB2c		TCTGGAAC	GTTCCAGA
	BB1f		AATCGGTC	GACCGATT	BB2d		ATTGACCG	CGGTCAAT
	BB1f		AATCGGTC	GACCGATT	BB2e		TGTCACGA	TCGTGACA
	BB1f		AATCGGTC	GACCGATT	BB2f		GCTAACTG	CAGTTAGC
	BB1g		ACTGGCTA	TAGCCAGT	BB2a		CTTAGAGC	GCTCTAAG
	BB1g		ACTGGCTA	TAGCCAGT	BB2b		GATCGACT	AGTCGATC
	BB1g		ACTGGCTA	TAGCCAGT	BB2c		TCTGGAAC	GTTCCAGA
	BB1g		ACTGGCTA	TAGCCAGT	BB2d		ATTGACCG	CGGTCAAT
	BB1g		ACTGGCTA	TAGCCAGT	BB2e		TGTCACGA	TCGTGACA
	BB1g		ACTGGCTA	TAGCCAGT	BB2f		GCTAACTG	CAGTTAGC

Selection data

E. **Pre-Selection Library Composition:**

Read 1

Total sequence count = 173603

Sequence	Count
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGTCACGATGATACTTACATGCAAATCCGTTC ACACCGACCT	5465
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGGTAACACCATACTTACATGCAAATCCGTTC ACACCGACCT	5164
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGTCACGATGATACTTACATGCAAATCCGTTC ACACCGACCT	5122
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGGTAACACCATACTTACATGCAAATCCGTTC ACACCGACCT	5090
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGGTAACACCATACTTACATGCAAATCCGTTC ACACCGACCT	4975
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGGACCGATTATACTTACATGCAAATCCGTTC ACACCGACCT	4738
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGACGACACTATACTTACATGCAAATCCGTTC ACACCGACCT	4708
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGACGACACTATACTTACATGCAAATCCGTTC ACACCGACCT	4645
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGGACCGATTATACTTACATGCAAATCCGTTC ACACCGACCT	4643
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGTCCAGAGTATACTTACATGCAAATCCGTTC ACACCGACCT	4592
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGTCCAGAGTATACTTACATGCAAATCCGTTC ACACCGACCT	4589
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGTCACGATGATACTTACATGCAAATCCGTTC ACACCGACCT	4530
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGTCCAGAGTATACTTACATGCAAATCCGTTC ACACCGACCT	4526
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGACGACACTATACTTACATGCAAATCCGTTC ACACCGACCT	4501
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGGACCGATTATACTTACATGCAAATCCGTTC ACACCGACCT	4488

TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGGTAACACCATACTTACATGCAAATCCGTTC ACACCGACCT	4281
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGTCCAGAGTATACTTACATGCAAATCCGTTC ACACCGACCT	3926
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGACGACACTATACTTACATGCAAATCCGTTC ACACCGACCT	3870
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGGACCGATTATACTTACATGCAAATCCGTTC ACACCGACCT	3779
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGTCACGATGATACTTACATGCAAATCCGTTC ACACCGACCT	3703
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGGTAACACCATACTTACATGCAAATCCGTTC ACACCGACCT	3686
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGGTAACACCATACTTACATGCAAATCCGTTC ACACCGACCT	3670
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGTCACGATGATACTTACATGCAAATCCGTTC ACACCGACCT	3449
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGTCCAGAGTATACTTACATGCAAATCCGTTC ACACCGACCT	3382
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGGACCGATTATACTTACATGCAAATCCGTTC ACACCGACCT	3348
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGGACCGATTATACTTACATGCAAATCCGTTC ACACCGACCT	3259
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGACGACACTATACTTACATGCAAATCCGTTC ACACCGACCT	3240
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGTCCAGAGTATACTTACATGCAAATCCGTTC ACACCGACCT	3192
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGACGACACTATACTTACATGCAAATCCGTTC ACACCGACCT	3072
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGCTTAGAGCATACTTACATGCAAATCCGTTC ACACCGACCT	1931
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGCTTAGAGCATACTTACATGCAAATCCGTTC ACACCGACCT	1906
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGTCACGATGATACTTACATGCAAATCCGTTC ACACCGACCT	1894
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGCTTAGAGCATACTTACATGCAAATCCGTTC ACACCGACCT	1879
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGCTTAGAGCATACTTACATGCAAATCCGTTC	1615

ACACCGACCT	
TGTAGACCATGTAGTTGAGGTCAATGTACGCCAGGTAGCCAGTATACTTACATGCAAATCCGTTC ACACCGACCT	1349
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGCTTAGAGCATACTTACATGCAAATCCGTTC ACACCGACCT	1208
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGTAGCCAGTATACTTACATGCAAATCCGTTC ACACCGACCT	991
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGTAGCCAGTATACTTACATGCAAATCCGTTC ACACCGACCT	951
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGTAGCCAGTATACTTACATGCAAATCCGTTC ACACCGACCT	925
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGCTTAGAGCATACTTACATGCAAATCCGTTC ACACCGACCT	786
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGTAGCCAGTATACTTACATGCAAATCCGTTC ACACCGACCT	752
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGTAGCCAGTATACTTACATGCAAATCCGTTC ACACCGACCT	699
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGTAGCCAGTATACTTACATGCAAATCCGTTC ACACCGACCT	670

Read 2

Total sequence count = 173603

Sequence	Count
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGTCACGATGATACTTACATGCAAATCCGTTC ACACCGACCT	5527
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGGTAACACCATACTTACATGCAAATCCGTTC ACACCGACCT	5230
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGTCACGATGATACTTACATGCAAATCCGTTC ACACCGACCT	5156
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGGTAACACCATACTTACATGCAAATCCGTTC ACACCGACCT	5145
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGGTAACACCATACTTACATGCAAATCCGTTC ACACCGACCT	5036
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGGACCGATTATACTTACATGCAAATCCGTTC	4780

ACACCGACCT	
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGACGACACTATACTTACATGCAAATCCGTTC ACACCGACCT	4731
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGGACCGATTATACTTACATGCAAATCCGTTC ACACCGACCT	4696
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGACGACACTATACTTACATGCAAATCCGTTC ACACCGACCT	4685
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGTCCAGAGTATACTTACATGCAAATCCGTTC ACACCGACCT	4670
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGTCCAGAGTATACTTACATGCAAATCCGTTC ACACCGACCT	4669
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGTCCAGAGTATACTTACATGCAAATCCGTTC ACACCGACCT	4593
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGTCACGATGATACTTACATGCAAATCCGTTC ACACCGACCT	4575
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGACGACACTATACTTACATGCAAATCCGTTC ACACCGACCT	4545
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGGACCGATTATACTTACATGCAAATCCGTTC ACACCGACCT	4537
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGGTAACACCATACTTACATGCAAATCCGTTC ACACCGACCT	4337
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGTCCAGAGTATACTTACATGCAAATCCGTTC ACACCGACCT	3986
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGACGACACTATACTTACATGCAAATCCGTTC ACACCGACCT	3924
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGGACCGATTATACTTACATGCAAATCCGTTC ACACCGACCT	3835
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGTCACGATGATACTTACATGCAAATCCGTTC ACACCGACCT	3747
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGGTAACACCATACTTACATGCAAATCCGTTC ACACCGACCT	3732
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGGTAACACCATACTTACATGCAAATCCGTTC ACACCGACCT	3724
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGTCACGATGATACTTACATGCAAATCCGTTC	3534

ACACCGACCT	
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGTCCAGAGTATACTTACATGCAAATCCGTTC ACACCGACCT	3414
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGGACCGATTATACTTACATGCAAATCCGTTC ACACCGACCT	3395
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGGACCGATTATACTTACATGCAAATCCGTTC ACACCGACCT	3322
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGTCCAGAGTATACTTACATGCAAATCCGTTC ACACCGACCT	3285
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGACGACACTATACTTACATGCAAATCCGTTC ACACCGACCT	3279
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGACGACACTATACTTACATGCAAATCCGTTC ACACCGACCT	3121
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGCTTAGAGCATACTTACATGCAAATCCGTTC ACACCGACCT	1947
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGTCACGATGATACTTACATGCAAATCCGTTC ACACCGACCT	1941
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGCTTAGAGCATACTTACATGCAAATCCGTTC ACACCGACCT	1929
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGCTTAGAGCATACTTACATGCAAATCCGTTC ACACCGACCT	1898
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGCTTAGAGCATACTTACATGCAAATCCGTTC ACACCGACCT	1641
TGTAGACCATGTAGTTGAGGTCAATGTACGCCAGGTAGCCAGTATACTTACATGCAAATCCGTTC ACACCGACCT	1369
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGCTTAGAGCATACTTACATGCAAATCCGTTC ACACCGACCT	1247
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGTAGCCAGTATACTTACATGCAAATCCGTTC ACACCGACCT	962
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGTAGCCAGTATACTTACATGCAAATCCGTTC ACACCGACCT	939
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGCTTAGAGCATACTTACATGCAAATCCGTTC ACACCGACCT	789
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGTAGCCAGTATACTTACATGCAAATCCGTTC	768

ACACCGACCT	
CAAAGACCCCAACGAGAAGAGCACACGTCTGAACTCCAGTCATGTGGATACGCTGCTT	761
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGTAGCCAGTATACTTACATGCAAATCCGTTC ACACCGACCT	710
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGTAGCCAGTATACTTACATGCAAATCCGTTC ACACCGACCT	690

E.I. **Selection Output Composition:**

Read 1 Overall (anything with >100 reads) Read 1 Total sequence count = 153396

Sequence	Count	Length	BB 2		BB 1	
			BB 2 Sequence	BB2 Count	BB 1 Sequence	BB 1 Count
TGTAGACCATGTAGTTGAGGTCAATGTACGCTAGGTAGCCAGTA TACTTACATGCAAATCCGTTCACACCGACCT	28553	75	ATGTACGC	42401	TAGCCAGT	44666
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGTCCAGAGTA TACTTACATGCAAATCCGTTCACACCGACCT	25824	75	GCTCTAAG	36871	TCCAGAGT	39620
TGTAGACCATGTAGTTGAGGTCAATGTACGCCAGGTAGCCAGTA TACTTACATGCAAATCCGTTCACACCGACCT	9466	75	GTTCCAGA	15879	GACCGATT	16446
TGTAGACCATGTAGTTGAGGTCAGGTGCCGTCGTATGTGCGCGT AGGAGTCGATCATACAGCCAAATCCGTTCACACCGACCT	6219	83	CGGTCAAT	14693	GTAACACC	14499
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGGACCGATTA TACTTACATGCAAATCCGTTCACACCGACCT	2905	75	CAGTTAGC	14619	TCACGATG	14046
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGGACCGATTA TACTTACATGCAAATCCGTTCACACCGACCT	2829	75	TCGTGACA	13746	ACGACACT	13080
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGTCACGATGA TACTTACATGCAAATCCGTTCACACCGACCT	2796	75	AGTCGATC	10645	CTTAGAGC	5519
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGGACCGATTA TACTTACATGCAAATCCGTTCACACCGACCT	2667	75	ATGTACGT	336	TAGCTAGT	385
TGTAGACCATGTAGTTGAGGTCATATGTGTACGTAGACGCTATTA GGTCTTCTCCATACCAAATCCGTTCACACCGACCT	2633	80	ATGTATGC	243	TAGCCCGT	226

TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGGACCGATTA TACTTACATGCAAATCCGTTCACACCGACCT	2593	75	ATGTACAC	142	TAGTCAGT	224
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGGTAACACCA TACTTACATGCAAATCCGTTCACACCGACCT	2533	75	TGGTCAAT	114	TAGTTAGT	172
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGTCACGATGA TACTTACATGCAAATCCGTTCACACCGACCT	2459	75	GTGTACGC	103	CCCAGAGT	143
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGGTAACACCA TACTTACATGCAAATCCGTTCACACCGACCT	2452	75	GCCCTAAG	100	TAGCAAGT	138
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGTCACGATGA TACTTACATGCAAATCCGTTCACACCGACCT	2435	75	GTTCCAAA	93	TCCAGGGT	132
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGTCCAGAGTA TACTTACATGCAAATCCGTTCACACCGACCT	2401	75	ACGTACGC	92	TAACCAGT	107
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGTCCAGAGTA TACTTACATGCAAATCCGTTCACACCGACCT	2364	75	GCTCTAGG	89	TCCAAAGT	86
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGGTAACACCA TACTTACATGCAAATCCGTTCACACCGACCT	2358	75	ATGTACTC	88	TAGCCAGC	84
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGTCCAGAGTA TACTTACATGCAAATCCGTTCACACCGACCT	2353	75	GCTCTAAA	78	TAGCCAAT	84
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGACGACACTA TACTTACATGCAAATCCGTTCACACCGACCT	2319	75	GCTCCAGA	77	TGGCCAGT	83
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGGTAACACCA TACTTACATGCAAATCCGTTCACACCGACCT	2280	75	GTTCTAAG	67	TCCAGAGC	83
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGTCCAGAGTA TACTTACATGCAAATCCGTTCACACCGACCT	2250	75	ATGCACGC	65	TTCAGAGT	82

TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGACGACACTA TACTTACATGCAAATCCGTTCACACCGACCT	2171	75	ACTCTAAG	64	TCTAGAGT	81
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGACGACACTA TACTTACATGCAAATCCGTTCACACCGACCT	2158	75	CTGTCAAT	59	TTACGATG	80
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGACGACACTA TACTTACATGCAAATCCGTTCACACCGACCT	2137	75	GTTCCAGG	58	TCCGGAGT	79
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGGACCGATTA TACTTACATGCAAATCCGTTCACACCGACCT	2077	75	GTTACAGA	58	TCATGATG	77
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGGACCGATTA TACTTACATGCAAATCCGTTCACACCGACCT	1889	75	CAGTTAAC	55	GTAACACT	75
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGTCACGATGA TACTTACATGCAAATCCGTTCACACCGACCT	1863	75	ATGTGCGC	55	TCACAATG	73
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGGTAACACCA TACTTACATGCAAATCCGTTCACACCGACCT	1839	75	TCGTGATA	54	TCACGATT	73
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGTCACGATGA TACTTACATGCAAATCCGTTCACACCGACCT	1787	75	GCTTTAAG	54	ACGACATT	71
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGGTAACACCA TACTTACATGCAAATCCGTTCACACCGACCT	1745	75	CAGTTAGT	53	AACCGATT	71
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGACGACACTA TACTTACATGCAAATCCGTTCACACCGACCT	1623	75	CAGTCAAT	53	ATAACACC	71
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGTCCAGAGTA TACTTACATGCAAATCCGTTCACACCGACCT	1602	75	CGGCCAAT	52	CAGCCAGT	70
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGACGACACTA TACTTACATGCAAATCCGTTCACACCGACCT	1550	75	GCTATAAG	51	TAGCCGGT	69

TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGTCACGATGA TACTTACATGCAAATCCGTTCACACCGACCT	1410	75	ATATACGC	50	ACAACACT	63
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGCTTAGAGCA TACTTACATGCAAATCCGTTCACACCGACCT	1031	75	GCTCCAAG	46	GACTGATT	61
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGCTTAGAGCA TACTTACATGCAAATCCGTTCACACCGACCT	929	75	TCGTAACA	46	TCAAGAGT	59
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGCTTAGAGCA TACTTACATGCAAATCCGTTCACACCGACCT	926	75	GGTCGATC	45	TATCCAGT	59
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGCTTAGAGCA TACTTACATGCAAATCCGTTCACACCGACCT	901	75	GTTCTAGA	44	GTAATACC	59
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGCTTAGAGCA TACTTACATGCAAATCCGTTCACACCGACCT	630	75	TTGTGACA	44	TAGCCACT	56
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGCTTAGAGCA TACTTACATGCAAATCCGTTCACACCGACCT	578	75	ATGTATGT	43	GGCCGATT	54
TGTAGACCATGTAGTTGAGGTCACAGTTAGCTAGGTAGCCAGTA TACTTACATGCAAATCCGTTCACACCGACCT	495	75	AGTCGATT	42	TCCAGAAT	53
TGTAGACCATGTAGTTGAGGTCACGGTCAATTAGGTAGCCAGTA TACTTACATGCAAATCCGTTCACACCGACCT	460	75	AGTTGATC	42	GTAACACA	53
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGTAGCCAGTA TACTTACATGCAAATCCGTTCACACCGACCT	447	75	TCATGACA	41	ATGACACT	52
TGTAGACCATGTAGTTGAGGTCATCGTGACATAGGTAGCCAGTA TACTTACATGCAAATCCGTTCACACCGACCT	440	75	CCGTGACA	41	TCCATAGT	50
TGTAGACCATGTAGTTGAGGTCAATGTACGCTATGTAGCCAGTAT ACTTACATGCAAATCCGTTCACACCGACCT	399	75	GTTTCAGA	40	ACGATACT	48

TATCGCTGGATGTGTCTGCGGCGTTTTATCATCTTCCTCTTCATCC TGCTGCTATGCCTCATCTTCTTGTTGGTTCTTCTGGACTATCAAGG TATGTTGCCCGTTTGTCCTCTAATTCCAGGATCCTCAACCACCAGC ACGGGACCATGCCGAACCTGCATGACTACTGCTCAAGGAACCTC TATGTATCCCTCCTGTTGCTGTACCAAACCTTCGGAC	329	219	GTTCAAGA	38	GATCGATT	47
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGTAGCCAGTA TACTTACATGCAAATCCGTTCACACCGACCT	315	75	GTTCCGGA	37	ACGACACC	46
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGTAGCCAGTA TACTTACATGCAAATCCGTTCACACCGACCT	301	75	CGGTAAAT	37	GCGACACT	45
TGTAGACCATGTAGTTGAGGTCAATGTACGCTAGGTAGCTAGTA TACTTACATGCAAATCCGTTCACACCGACCT	290	75	ATTTACGC	36	CTTAAAGC	44
TGTAGACCATGTAGTTGAGGTCAATGTACGTTAGGTAGCCAGTA TACTTACATGCAAATCCGTTCACACCGACCT	258	75	CAGTTGGC	36	CCACGATG	44
TGTAGACCATGTAGTTGAGGTCAATGTATGCTAGGTAGCCAGTA TACTTACATGCAAATCCGTTCACACCGACCT	194	75	GCTCTGAG	36	CTTAGAGT	43
TGTAGACCATGTAGTTGAGGTCAATGTACGCTAGGTAGTCAGTA TACTTACATGCAAATCCGTTCACACCGACCT	171	75	CGGTTAAT	35	GACAGATT	42
TGTAGACCATGTAGTTGAGGTCAATGTACGCTAGGTAGCCAGTA TATTACATGCAAATCCGTTCACACCGACCT	142	74	TCGTGGCA	35	TACAGAGT	42
TGTAGACCATGTAGTTGAGGTCAATGTACGCTAGGTAGCCCGTA TACTTACATGCAAATCCGTTCACACCGACCT	134	75	TAGTTAGC	34	TAGCCAGA	41
TGTAGACCATGTAGTTGAGGTCAGCAATATTCGTAGTACATGCTA GGGCGTGTGCATACCAAATCCGTTCACACCGACCT	131	80	AGTAGATC	33	TCACGATA	41
TGTAGACCATGTAGTTGAGGTCAATGTACGCTAGGTAGTTAGTA	127	75	ATTCCAGA	33	GACCAATT	40

TACTTACATGCAAATCCGTTCACACCGACCT						
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGCCCAGAGTA TACTTACATGCAAATCCGTTCACACCGACCT	103	75	ATGTACGA	32	TCCCGAGT	38

E.II. **Enrichments by Selection**

Read 2

Selection Input Library		Selection Output		Enrichment	Frequency in selection
Sequence	Count	BB2 Sequence	BB2 Count		
AGTCGATC	19357	AGTCGATC	10645	0.5356	7
ATGTACGC	1792	ATGTACGC	42401	23.0437	1
CAGTTAGC	27917	CAGTTAGC	14619	0.5100	5
CGGTCAAT	23637	CGGTCAAT	14693	0.6054	4
GCTCTAAG	19566	GCTCTAAG	36871	1.8353	2
GTTCCAGA	28371	GTTCCAGA	15879	0.5451	3
TCGTGACA	24329	TCGTGACA	13746	0.5503	6

Read 1

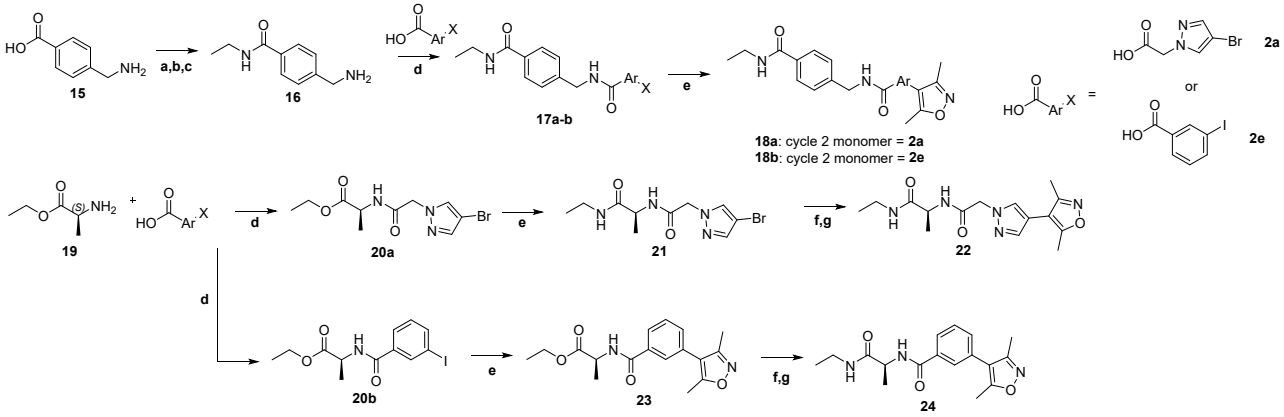
Selection Input Library		Selection Output		Enrichment	Frequency in selection
Sequence	Count	BB1 sequence	BB1 Count		
ACGACACT	24867	ACGACACT	13080	0.5137	6
CTTAGAGC	9650	CTTAGAGC	5519	0.5586	7
GACCGATT	25108	GACCGATT	16446	0.6398	3
GTAACACC	27766	GTAACACC	14499	0.5100	4
TAGCCAGT	6946	TAGCCAGT	44666	6.2807	1
TCACGATG	25012	TCACGATG	14046	0.5485	5
TCCAGAGT	25083	TCCAGAGT	39620	1.5428	2

Read count minus JQ1 from selection	115377
Number of reads for **22**	25824
Representation of **22** in selection output	0.2238

Read count for clean library	173603
Read count for **22** in clean library	3382
Representation in clean library	0.0195

Enrichment of **22** by selection	11.5

Off-DNA Synthesis



Scheme S 1 a) TFAA; b) $SOCl_2$; c) NH_2Et in THF 2M; d) HATU, DIPEA, DCM e) $Pd(dtbpf)Cl_2$, Cs_2CO_3, Diox./H_2O 10:1, 3,5-dimethyl-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)isoxazole; f) LiOH, THF/H_2O 1:1; g) HATU, DIPEA, DCM, NH_2Et in THF 2M.

F. **Chemistry**

General Procedure A:

Flask charged with acid starting material (1 eqv.), HATU (1 eqv.), DIPEA (2.5 eqv.) and then dissolved in dry DCM (0.35 M) under N_2. Preactivated for 10 minutes at room temperature, before the amine reactant was added to the solution. Stirred at room temperature for 16 hrs. Dried under vacuum and purified by reverse phase flash chromatography 5→95% ACN in HCO_2H 0.1% (aq).

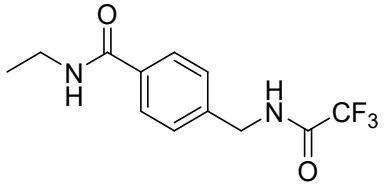
General Procedure B:

Aryl halide (1 eqv.), 3,5-dimethyl-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)isoxazole (1.1 eqv.), Cs_2CO_3 (2 eqv.), and $Pd(dtbpf)Cl_2$ (10 mol%) were dissolved in a degassed dioxane water mixture 10:1 under N_2. The reaction was heated to 100°C for 3 hr. Reaction mixture filtered through celite, and the filtrate dried under vacuum and purified by reverse phase flash chromatography 5→95% ACN in HCO_2H 0.1% (aq).

General Procedure C:

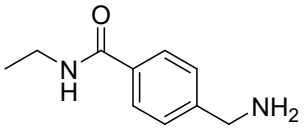
Ester (1 eqv.) dissolved in 1:1 mixture of H_2O/THF (2 mL), $LiOH.H_2O$ (1 eqv.) added. Reaction stirred vigorously for 1 hour at room temperature. Reaction dried under vacuum, before resuspending the residue in dry DCM (0.15 M). The solution was degassed, and the atmosphere purged with N_2 to the solution was added HATU (2.0 eqv.), DIPEA (2.5 eqv.), and $EtNH_2$ 2M in THF (5 eqv.). Dried under vacuum and purified by reverse phase flash chromatography 5→95% ACN in HCO_2H 0.1% (aq).

F.I. N-ethyl-4-((2,2,2-trifluoroacetamido)methyl)benzamide **16-SI**



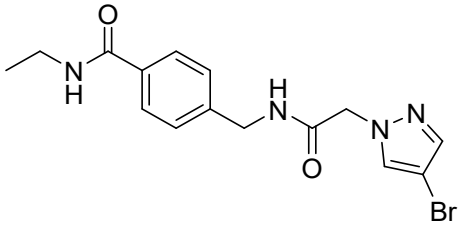
Aminomethylbenzoic acid **16** (13.2 mmol, 2 g) was dissolved in DCM (30mL) and pyridine (4 mmol, 4.4 mL) under N_2 and cooled to 0°C. Trifluoroacetic acid (26.4 mmol, 3.75 mL) was added slowly and allowed to warm to room temperature and stir for 2 hours. The solvent was dried under vacuum. The residue was resuspended in thionyl chloride (132 mmol, 10 mL) at 0°C under N_2. Allowed to warm to room temperature and stir for 3 hours. The liquid was dried under vacuum then the residue was resuspended in DCM (30 mL) dry under N_2. To the solution was added $EtNH_2$.HCl (132 mmol, 10.8 g) and pyridine (12.4 mL), dropwise. The mixture stirred overnight at room temperature. The reaction was worked up by diluting with additional DCM, washing with water x3 and sat. $NaHCO_3$. The product, N-ethyl-4-((2,2,2-trifluoroacetamido)methyl)benzamide, crashed out into the organic layer and was collected by filtration. White solid (1.57 g, 43%).^{1}H NMR (500 MHz, Methanol-d_4) δ 9.39 – 9.33 (m, 2H), 8.98 – 8.92 (m, 2H), 6.47 (s, 7H), 6.07 (s, 2H), 4.96 (q, *J* = 7.2 Hz, 2H), 2.78 (t, *J* = 7.3 Hz, 3H). ^{13}C NMR (126 MHz, Methanol-d_4) δ 168.22, 157.51, 140.78, 133.76, 127.30, 127.22, 117.33, 115.05, 48.11, 47.94, 47.77, 47.60, 47.43, 47.26, 47.09, 42.44, 34.43, 13.49. Calculated for $C_{12}H_{13}F_3N_2O_2$ $[MH]^+$ 275.1, LCMS $[MH]^+$ 275.3.

F.II. 4-(aminomethyl)-N-ethylbenzamide **16**



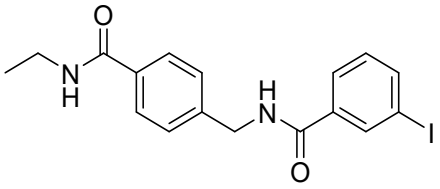
N-ethyl-4-((2,2,2-trifluoroacetamido)methyl)benzamide (5.73 mmol, 1.57 g) was dissolved in MeOH/NH_4OH 1:1 and heated to 60°C 8 hrs. Solvent removed under vacuum. Purified by flash chromatography, amine column 0→10% MeOH in DCM. White solid (0.85 g, 84%). ^{1}H NMR (500 MHz, Methanol-d_4) δ 7.92 – 7.86 (m, 2H), 7.58 – 7.52 (m, 2H), 4.18 (s, 2H), 3.42 (q, *J* = 7.3 Hz, 2H), 1.23 (t, *J* = 7.3 Hz, 3H). ^{13}C NMR (126 MHz, Methanol-d4) δ 167.73, 136.45, 135.10, 128.64, 127.62, 48.11, 47.93, 47.76, 47.59, 47.42, 47.25, 47.08, 42.44, 34.49, 13.46. Calculated for $C_{10}H_{15}N_2O$ $[MH]^+$ 179.2, LCMS $[MH]^+$ 179.1.

F.III. 4-((2-(4-bromo-1H-pyrazol-1-yl)acetamido)methyl)-N-ethylbenzamide **17a**



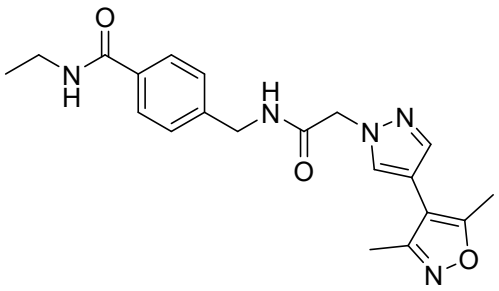
Compound **17a** was synthesised from 2-(4-bromo-1H-pyrazol-1-yl)acetic acid (3.54 mmol, 0.73 g) and **16** (0.42 g, 2.36 mmol) by general procedure A. White solid (0.86 g, 73%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.81 (d, J = 0.7 Hz, 1H), 7.79 – 7.76 (m, 2H), 7.53 (d, J = 0.6 Hz, 1H), 7.40 – 7.35 (m, 2H), 4.91 (s, 9H), 4.45 (s, 2H), 3.40 (q, J = 7.2 Hz, 2H), 1.22 (t, J = 7.2 Hz, 3H). ^{13}C NMR (126 MHz, Methanol-d4) δ 169.70, 169.06, 143.27, 141.62, 134.75, 133.28, 128.55, 128.50, 94.43, 55.50, 49.51, 49.34, 49.17, 49.00, 48.83, 48.66, 48.49, 43.83, 35.82, 14.91. Calculated for $C_{15}H_{18}{}^{79}BrN_4O_2$ $[MH]^+$ 364.1, LCMS $[MH]^+$ 364.0.

F.IV. N-(4-(ethylcarbamoyl)benzyl)-3-iodobenzamide **17b**



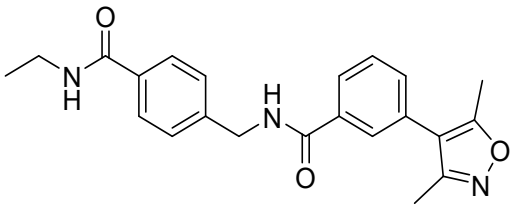
Compound **17b** was synthesised from 3-iodobenzoic acid (2.36 mmol, 0.59 g) and **16** (2.36 mmol, 0.42 g) by general procedure A. The product was a white solid (0.963 g, 76 %). ^{1}H NMR (500 MHz, Methanol-d4) δ 8.22 (t, J = 1.7 Hz, 1H), 7.91 (dt, J = 7.9, 1.4 Hz, 1H), 7.86 (dt, J = 7.9, 1.4 Hz, 1H), 7.82 – 7.76 (m, 2H), 7.46 – 7.40 (m, 2H), 7.26 (t, J = 7.9 Hz, 1H), 4.61 (s, 2H), 3.40 (q, J = 7.2 Hz, 2H), 1.22 (t, J = 7.3 Hz, 3H). ^{13}C NMR (126 MHz, Methanol-d4) δ 169.80, 143.82, 141.77, 137.46, 134.69, 131.47, 128.52, 128.51, 127.62, 124.46, 94.78, 49.63, 49.51, 49.34, 49.17, 49.00, 48.83, 48.66, 48.49, 44.22, 35.83, 14.91. Calculated for $C_{17}H_{18}IN_2O_2$ $[MH]^+$ 409.0, LCMS $[MH]^+$ 409.1.

F.V. 4-((2-(4-(3,5-dimethylisoxazol-4-yl)-1H-pyrazol-1-yl)acetamido)methyl)-N-ethylbenzamide **18a**



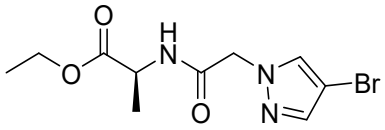
Synthesised from **17a** (1.73 mmol, 0.63 g) via general procedure B. Product was a white solid (0.40 g, 61%). ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.71 (t, *J* = 6.0 Hz, 1H), 8.45 (t, *J* = 5.5 Hz, 1H), 8.00 (d, *J* = 0.8 Hz, 1H), 7.83 – 7.77 (m, 2H), 7.69 (d, *J* = 0.8 Hz, 1H), 7.37 – 7.31 (m, 2H), 4.92 (s, 2H), 4.36 (d, *J* = 5.9 Hz, 2H), 3.27 (qd, *J* = 7.2, 5.5 Hz, 2H), 2.43 (s, 3H), 2.27 (s, 3H), 1.11 (t, *J* = 7.2 Hz, 3H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 166.69, 165.63, 163.95, 158.15, 142.05, 137.74, 133.33, 129.79, 127.18, 126.99, 109.68, 107.81, 54.12, 41.97, 40.02, 39.85, 39.69, 39.52, 39.35, 39.19, 39.02, 34.02, 14.86, 11.60, 10.85. Calculated for $C_{20}H_{23}N_5O_3$ $[MH]^+$ 382.1874, LCMS $[MH]^+$ found 382.3, HRMS $[MH]^+$ found 382.1865.

F.VI. 3-(3,5-dimethylisoxazol-4-yl)-N-(4-(ethylcarbamoyl)benzyl)benzamide**18b**



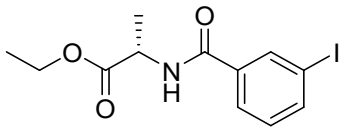
Synthesised from **17b** (1.79 mmol, 0.73 g) via general procedure B. Product was a white solid (0.23 g, 34%). ^{1}H NMR (500 MHz, DMSO-d6) δ 9.17 (t, J = 6.0 Hz, 1H), 8.42 (t, J = 5.6 Hz, 1H), 7.92 (dt, J = 7.4, 1.7 Hz, 1H), 7.87 (d, J = 1.8 Hz, 1H), 7.82 – 7.77 (m, 2H), 7.63 – 7.53 (m, 2H), 7.39 (d, J = 8.2 Hz, 2H), 4.54 (d, J = 5.9 Hz, 2H), 3.27 (qd, J = 7.2, 5.5 Hz, 2H), 2.42 (s, 3H), 2.24 (s, 3H), 1.11 (t, J = 7.2 Hz, 3H). ^{13}C NMR (126 MHz, DMSO-d6) δ 165.79, 165.67, 165.42, 158.10, 142.62, 134.75, 133.23, 131.79, 130.07, 129.02, 127.44, 127.15, 126.90, 126.50, 115.50, 42.44, 40.02, 39.85, 39.69, 39.52, 39.35, 39.19, 39.02, 33.97, 14.82, 11.32, 10.41. Calculated for $C_{22}H_{24}N_3O_3$ $[MH]^+$ 378.1813, LCMS $[MH]^+$ found 378.3, HRMS $[MH]^+$ found 378.1802.

F.VII. ethyl (2-(4-bromo-1H-pyrazol-1-yl)acetyl)-L-alaninate **20a**



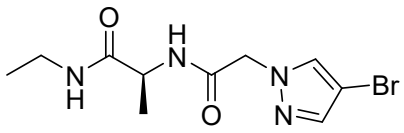
Synthesised from 2-(4-bromo-1H-pyrazol-1-yl)acetic acid (2.15 mmol, 0.44 g and ethylalaninoate hydrochloride (1.95 mmol, 0.30 g) general procedure B. White solid (0.414 g, 72%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.78 (d, J = 0.7 Hz, 1H), 7.51 (d, J = 0.7 Hz, 1H), 4.90 (s, 7H), 4.40 (q, J = 7.3 Hz, 1H), 4.17 (qd, J = 7.1, 0.6 Hz, 2H), 1.40 (d, J = 7.3 Hz, 3H), 1.26 (t, J = 7.1 Hz, 3H). ^{13}C NMR (126 MHz, Methanol-d4) δ 173.84, 168.68, 141.40, 133.18, 94.38, 62.46, 55.19, 49.74, 49.51, 49.34, 49.17, 49.00, 48.83, 48.66, 48.49, 17.44, 14.42. Calculated for $C_{10}H_{15}{}^{79}BrN_3O_3$ $[MH]^+$ 304.0, LCMS $[MH]^+$ 304.0

F.VIII. Ethyl (3-iodobenzoyl)alaninate **20b**



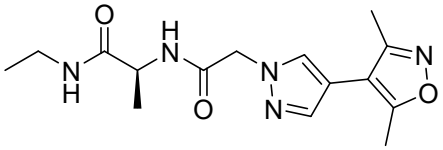
Synthesised from 3-iodobenzoic acid (6.55 mmol, 1.62 g) and ethyl alaninoate hydrochloride (6.55 mmol, 1.0 g) via general procedure A. Product was an oily solid (1.52 g, 71%). ^{1}H NMR (500 MHz, Methanol-d4) δ 8.21 (t, J = 1.7 Hz, 1H), 7.86 (dddd, J = 23.1, 7.8, 1.8, 1.1 Hz, 2H), 7.24 (t, J = 7.9 Hz, 1H), 4.56 (q, J = 7.3 Hz, 1H), 4.19 (q, J = 7.1 Hz, 2H), 1.49 (d, J = 7.4 Hz, 3H), 1.27 (t, J = 7.1 Hz, 3H). ^{13}C NMR (126 MHz, Methanol-d4) δ 174.20, 168.34, 141.74, 137.51, 137.11, 131.32, 127.77, 94.67, 62.35, 50.26, 49.51, 49.34, 49.17, 49.00, 48.83, 48.66, 48.49, 17.13, 14.48. Calculated for $C_{12}H_{14}INO_3$ $[MH]^+$ 348.0, LCMS $[MH]^+$ 348.1.

F.IX. (S)-2-(2-(4-bromo-1H-pyrazol-1-yl)acetamido)-N-ethylpropanamide **21**



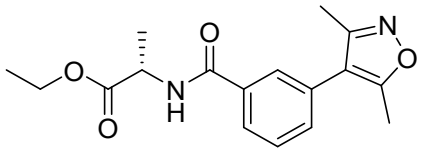
Synthesised from **20a** (1.37 mmol, 0.41 g) general procedure C. White solid (10 mg, 7.2%). ^{1}H NMR (500 MHz, Methanol-d_4) δ 7.78 (d, *J* = 0.7 Hz, 1H), 7.52 (d, *J* = 0.8 Hz, 1H), 4.88 (d, *J* = 0.9 Hz, 2H), 4.32 (q, *J* = 7.2 Hz, 1H), 3.20 (q, *J* = 7.3 Hz, 2H), 1.35 (d, *J* = 7.2 Hz, 3H), 1.11 (t, *J* = 7.3 Hz, 3H). ^{13}C NMR (126 MHz, Methanol-d_4) δ 174.31, 168.58, 141.48, 133.27, 94.37, 55.32, 50.60, 49.51, 49.34, 49.17, 49.00, 48.83, 48.66, 48.49, 35.31, 18.31, 14.73. Calculated for $C_{10}H_{16}{}^{79}BrN_4O_2$ $[MH]^+$ 303.0, LCMS $[MH]^+$ 303.1.

F.X. (S)-2-(2-(4-(3,5-dimethylisoxazol-4-yl)-1H-pyrazol-1-yl)acetamido)-N-ethylpropanamide **22**



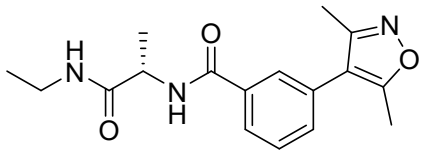
Synthesised from **21** (0.10 mmol, 30 mg) general procedure B. White solid (10 mg, 31%). ^{1}H NMR (500 MHz, MeOD) δ 7.88 (s, 1H), 7.69 (s, 1H), 4.96 (s, 2H), 4.34 (q, *J* = 7.1 Hz, 1H), 3.20 (q, *J* = 7.3 Hz, 2H), 2.45 (s, 3H), 2.30 (s, 3H), 1.37 (d, *J* = 7.1 Hz, 3H), 1.11 (t, *J* = 7.3 Hz, 3H). ^{13}C NMR (126 MHz, MeOD) δ 174.30, 168.86, 166.50, 160.14, 139.86, 131.62, 112.11, 109.13, 55.03, 54.81, 50.60, 49.51, 49.34, 49.17, 49.00, 48.83, 48.66, 48.49, 35.30, 18.39, 14.75, 11.63, 10.95. Calculated for $C_{15}H_{22}N_5O_3$ $[MH]^+$ 320.1718, LCMS $[MH]^+$ found 320.3, HRMS $[MH]^+$ found 320.1715 (HPLC purity >98%).

F.XI. ethyl (3-(3,5-dimethylisoxazol-4-yl)benzoyl)alaninate **23**



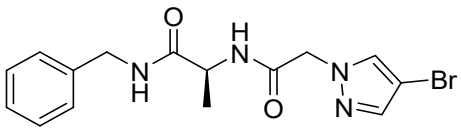
Synthesised from **20b** via general method B giving **23** which was an oily off white solid (0.76 g, 64%). ^{1}H NMR (500 MHz, MeOD) δ 7.89 (ddd, J = 7.7, 1.8, 1.2 Hz, 1H), 7.82 (td, J = 1.8, 0.6 Hz, 1H), 7.59 (td, J = 7.7, 0.6 Hz, 1H), 7.54 (dt, J = 7.7, 1.4 Hz, 1H), 4.60 (q, J = 7.3 Hz, 1H), 4.21 (q, J = 7.1 Hz, 2H), 2.43 (s, 3H), 2.27 (s, 3H), 1.51 (d, J = 7.3 Hz, 3H), 1.28 (t, J = 7.1 Hz, 3H). ^{13}C NMR (126 MHz, MeOD) δ 174.38, 169.62, 167.37, 159.97, 135.86, 131.97, 127.79, 117.32, 62.39, 50.33, 17.16, 14.48, 11.40, 10.63. Calculated for $C_{17}H_{21}N_2O_4$ $[MH]^+$ 317.1, LCMS $[MH]^+$ 317.2.

F.XII. 3-(3,5-dimethylisoxazol-4-yl)-N-(1-(ethylamino)-1-oxopropan-2-yl)benzamide **24**



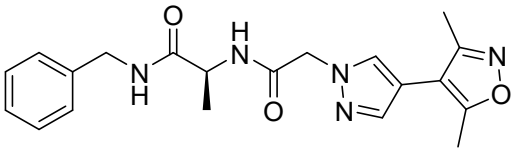
Synthesised from **23** (1.52 mmol, 0.48 g) general procedure C. Product was a white solid (0.43 mg, 90%). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.90 (dt, J = 7.8, 1.5 Hz, 1H), 7.84 (t, J = 1.8 Hz, 1H), 7.58 (t, J = 7.7 Hz, 1H), 7.52 (dt, J = 7.8, 1.5 Hz, 1H), 4.54 (q, J = 7.2 Hz, 1H), 3.24 (q, J = 7.2 Hz, 2H), 2.42 (s, 3H), 2.26 (s, 3H), 1.46 (d, J = 7.2 Hz, 3H), 1.13 (t, J = 7.3 Hz, 3H). ^{13}C NMR (126 MHz, Methanol-d4) δ 174.93, 169.35, 167.32, 159.95, 135.90, 133.55, 131.90, 130.16, 129.38, 127.83, 117.32, 51.21, 49.51, 49.34, 49.17, 49.00, 48.83, 48.66, 48.49, 35.33, 18.25, 14.79, 11.41, 10.64. Calculated for $C_{17}H_{21}N_3O_3$ $[MH]^+$ 316.1656, LCMS $[MH]^+$ found 316.3, HRMS $[MH]^+$ found 316.1653 (HPLC purity >99%).

F.XIII. (S)-N-benzyl-2-(2-(4-bromo-1H-pyrazol-1-yl)acetamido)propanamide **21-SI**



Synthesised from **20a** (0.98 mmol, 300 mg) and benzylamine (1.48 mmol, 158.62 mg) via general procedure C. The product was a white solid (210 mg, 59%). ^{1}H NMR (500 MHz, MeOD) δ 7.77 (d, *J* = 0.7 Hz, 1H), 7.47 (d, *J* = 0.8 Hz, 1H), 7.33 – 7.19 (m, 5H), 4.89 (s, 2H), 4.43 – 4.36 (m, 3H), 1.39 (d, *J* = 7.1 Hz, 3H). ^{13}C NMR (126 MHz, MeOD) δ 174.58, 168.67, 141.50, 139.75, 133.24, 129.53, 128.41, 128.19, 94.37, 55.38, 50.73, 49.51, 49.46, 49.34, 49.28, 49.17, 49.00, 48.83, 48.66, 48.49, 44.05, 18.22. Calculated for $C_{15}H_{18}{}^{79}Br\,N_4O_2$ $[MH]^+$ 365.1, LCMS $[MH]^+$ 365.3.

F.XIV. **(**S)-N-benzyl-2-(2-(4-(3,5-dimethylisoxazol-4-yl)-1H-pyrazol-1-yl)acetamido)propenamide **25**



Synthesised from **21-SI** (0.24 mmol, 89 mg) via general procedure B. Product was a white solid (31 mg, 34%). ^{1}H NMR (500 MHz, MeOD) δ 7.86 (s, 1H), 7.62 (s, 1H), 7.32 – 7.18 (m, 5H), 4.95 (d, *J* = 1.4 Hz, 2H), 4.43 (q, *J* = 7.1 Hz, 1H), 4.38 (d, *J* = 1.7 Hz, 2H), 2.43 (s, 3H), 2.28 (s, 3H), 1.41 (d, *J* = 7.1 Hz, 3H) ^{13}C NMR (126 MHz, MeOD) δ 174.58, 168.93, 166.50, 160.14, 139.81 (d, *J* = 12.1 Hz), 131.58, 129.51, 128.39, 128.17, 112.12, 109.13, 55.09, 50.73, 44.04, 18.29, 11.63, 10.95. Calculated for $C_{20}H_{24}N_5O_3$ $[MH]^+$ 382.1879, LCMS $[MH]^+$ 382.3, HRMS $[MH]^+$ found 382.1877 (HPLC purity >96%).

Structural Biology

Protein purification

All purification steps were performed using AKTA Pure at 4 ºC. For BRD4, harvested bacterial cells were resuspended in 50mM HEPES buffer (pH 7.4) containing 200 mM NaCl, 10 mM imidazole, 0.5 mg mL^{-1} lysosyme, and 0.2 mg mL^{-1} DNAse at 4ºC for 1h. After sonication and centrifugation (1h at 35,000 x *g*), the supernatant was purified by immobilized Ni^{2+} ion affinity chromatography. The peak fractions were pooled and incubated with TEV protease (50:1) at 4ºC overnight. The cleaved His-tag was separated by size exclusion chromatography using a Superdex 75 (26/60) column and eluted with 50 mM HEPES buffer (pH 7.4) containing 200 mM NaCl and 0.5 mM TCEP.

Surface Plasmon Resonance (SPR)

SPR-based ligand binding assays were performed using Biacore S200 (GE Healthcare) at 20 ºC using multi-cycle. Immobilisation of BRD4 was achieved using standard amine coupling on a CM5 chip surface. The surface was prepared through activation with EDC/NHS, followed by injection of 30 µg/ml BRD4 until target level 6000 RU was reached. The surface was then quenched using 1M Ethanolamine and washed with running buffer 10mM HEPES, 150mM NaCl, 0.01% TWEEN20, and 1% DMSO with a flow rate of 30 µg/ml. Compounds were injected in a dose-response manner (3-fold dilution over 11 points ranging from 0-33µM) in series across the control and BRD4 immobilised flow cells using solvent correction to account for bulk refractive index changes. The reference control channel was subtracted from BRD4 immobilised channel and dose-response data was fitted using a steady state 1:1 binding model to determine the K_d.

Compound	K_d Data (µM)	
	Mean (‡n=4, *n=2)	Standard Deviation
22	0.051‡	0.035
18a	35.049‡	1.900
18b	2.553*	1.240
24	14.619*	0.950
25	11.203*	1.760

Protein crystallography

Crystallization was performed at 20 ºC in the sitting drop vapour diffusion method dispensed using Mosquito (TTP labtech). Crystals of BRD4 were grown in the presence of 1 mM **22** from BRD4 7.5mg/ml, 90mM Nitrate Phosphate Sulfate, 100mM Tris Bicine pH8.5, 30%w/v Ethylene glycol PEG 8000. Crystals were harvested in cryoprotectant using additional reservoir and flash frozen in liquid nitrogen before data collection. X-ray diffraction data were recorded using Bruker D8-Venture MetalJet X-ray, Newcastle University (Newcastle, UK). Data processing was carried out using Proteum3 software (Bruker AXS 2015), POINTLESS/AIMLESS [PMID: 21460446] and other CCP4 programs [PMID: 15299374] run within the CCP4i2 GUI. PHASER [PMID: 19461840] was used formolecular replacement using pdb 5LRQ as a search model. Iterative rounds of model building and refinement was performed using COOT [PMID: 20383002] and REFMAC5 [PMID: 21460454], respectively. Figures were prepared using CCP4MG [PMID: 15572783].X-ray data collection and refinement statistics

	Compound **22** : BRD4 complex (pdb 8C11)
Data collection	
Space group	$P2_12_12_1$
Unit cell (Å)	a=38.1 b=43.2 c=79.3
Resolution (Å) (highest resolution shell)	18.2-1.8 (1.84-1.80)
Total observations	79924 (2514)

Unique	12689 (719)
R_{merge}	0.076(0.85)
Mean I/σ(I)	8.3 (1.4)
Multiplicity	6.3 (3.5)
Completeness %	99.8 (99.5)
CC(1/2)	0.99 (0.54)
Refinement	
Number of atoms (B-factor) protein other waters	 2,097 (14.2) 44 (43.3) 116 (22.1)
R (highest resolution shell)	0.195
R_{free} (highest resolution shell)	0.234
Rmsd bonds (Å)	0.0080
Rmsd angles (°)	1.550

The structures have been deposited in the PDB with accession codes 8C11.

Western blotting experiments

MM.1S cells were obtained from the American Type Culture Collection (ATCC) and maintained in RPMI-1640 medium containing 2 mM L-glutamine and supplemented with 10% (v/v) fetal bovine serum (FBS, Life Technologies). Cells were kept in culture for fewer than 25 passages. Cultures were tested for mycoplasma contamination every 3 months using the MycoAlert Mycoplasma Detection Kit (Lonza) and returned negative throughout. MM.1S cells 2 × 106 were seeded per well of six-well plates (Costar) and treated with compounds at various concentrations for 4 h. Cells were washed with PBS and soluble lysate was prepared in PhosphoSafe buffer (Merck Millipore) containing protease inhibitor

cocktail (Roche). Samples with 15 µg of total protein were loaded in Laemmli sample buffer containing final 2.5% (v/v) β-mercaptoethanol into 4-20% polyacrylamide Tris-glycine (TGX) gels (Bio-Rad) and transferred onto 0.45 µM Hybond nitrocellulose membrane (GE Healthcare). Primary c-Myc antibody (CST 5605) at 1:1000 dilution in 5% (w/v) BSA in Tris-buffered saline (TBS) containing 0.05% (v/v) Tween20 (TBST) (overnight incubation at 4oC), GAPDH antibody (Santa Cruz sc-47724) at 1:3000 dilution (1 h room temperature incubation), and secondary anti-rabbit or anti-mouse HRP-conjugated antibodies (Dako; P0448 or P0447, respectively) at 1:3000 dilution in 5% (w/v) non-fat milk in TBST (1 h room temperature incubation). The immunoreactive bands were detected with Clarity ECL (Bio-Rad) and visualized using a Fujifilm LAS3000.

Gel running buffer:

1X Tris-glycine running buffer (200 mM glycine, 3.5 mM sodium dodecyl sulphate, 25 mM Tris base

Transfer buffer from gel to membrane:

Tris-glycine transfer buffer (Life Technologies) containing 4% (v/v) methanol

ADME Screening

ADME assays were carried out as described previously: logD [PMID: 24168238], solubility [PMID: 26855285], metabolic stability [PMID: 29940120], MDCK permeability [PMID: 30222362].