Alcohol-derived DNA crosslinks are repaired by two distinct mechanisms

Authors: Michael R. Hodskinson[1†], Alice Bolner[2†], Koichi Sato[2‡], Ashley N. Kamimae-Lanning[1‡], Koos Rooijers[2‡], Merlijn Witte[2], Mohan Mahesh[1,5], Jan Silhan[1,6], Maya Petek[1,7], David M. Williams[3], Jop Kind[2], Jason Chin[1], Ketan J. Patel[1,4*], and Puck Knipscheer[2*]

Supplemental note 1

Methods

Description of all methods used in the study.

Methods References

Supplemental note 2

SI Figure 1

NMR spectra of compounds generated for the synthesis of the aminodiol (*5*) crosslinker.

a. Scheme for the synthesis of aminodiol. **b**. ^{1}H NMR spectrum and 13C{1H}, ^{13}C DEPT-135 and ^{13}C DEPT-90 NMR spectra of (*R*)-2-(Pent-4-en-2-yl)isoindoline-1,3-dione, 2. **c**. ^{1}H NMR spectrum and 13C{1H}, ^{13}C DEPT-135 and ^{13}C DEPT-90 NMR spectra of benzyl *N*-[(2*R*)-pent-4-en-2-yl]carbamate, 3. **d**. ^{1}H NMR spectrum and 13C{1H}, ^{13}C DEPT-135 and ^{13}C DEPT-90 NMR spectra of benzyl [(2*R*)-4,5-dihydroxypentan-2-yl]carbamate, 4. **e**. ^{1}H NMR spectrum and 13C{1H}, ^{13}C DEPT-135 and ^{13}C DEPT-90 NMR spectra of (4*R*)-4-aminopentane-1,2-diol, 5.

Supplemental note 3

SI Figure 2

This file contains uncropped source images with indication (when relevant) of the area where the image was cropped and size markers.

Supplemental note 1

Methods

Synthesis of site-specific crosslinks

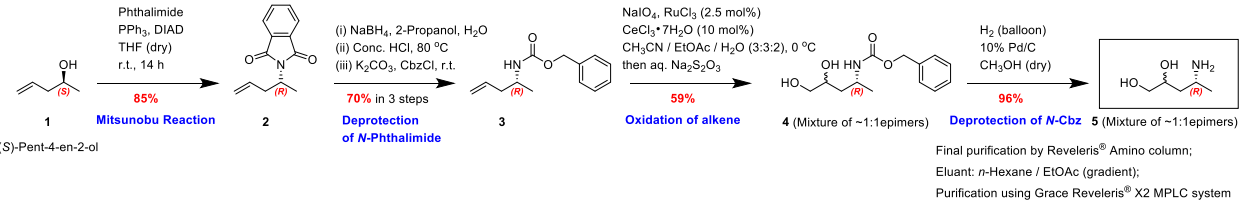
Reagents

Chemicals were supplied by Sigma (unless otherwise stated), except the reagents for HPLC and (4*R*)-4-aminopentane-1,2-diol synthesis, supplied by Link Technologies Ltd and Fluorochem Ltd. Culture medium was supplied by Gibco (Thermo Scientific)

Synthesis of the Cross Linker Aminodiol, **5**

We first approached the synthesis of the cross linker aminodiol **5**, following a protocol described by Cho *et al.*[9], who reported a 6-step synthesis of ^{13}C-labelled aminodiol **5** starting from (*R*)-2-amino-1-propanol and used K^{13}CN in one of the key step. However, in our hands, this 6-step protocol appeared to be cumbersome for scale up due to few non-reproducible, low-yielding steps and also involved handling highly toxic KCN and OsO_4 reagents in couple of key steps. Hence, we designed an alternative route for the synthesis of aminodiol **5** starting from commercially available (*S*)-pent-4-en-2-ol **1** (SCHEME 1). Each and every step in this synthesis of aminodiol **5** were optimised to afford good to excellent yields that were reproducible during scale-up. Additionally, in this modified synthesis of **5**, highly toxic reagents like KCN and OsO_4 were not used.

SCHEME 1: Synthesis of the Cross Linker Aminodiol **5**



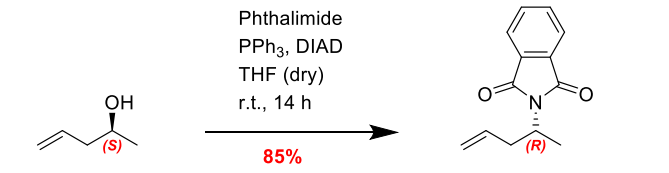
Abbreviations: DIAD = diisopropyl azodicarboxylate; Ph = phenyl; THF = tetrahydrofuran; r.t. = room temperature; h = hour(s); conc. = concentrated; °C = degree(s); Celcius; CBz = benzyloxycarbonyl; DMAP = 4-dimethylaminopyridine; aq.= aqueous

The (4*R*)-4-aminopentane-1,2-diol **5** was synthesised in 6 steps starting from commercially available (*S*)-pent-4-en-2-ol **1** with an overall yield of 34%. The synthesis commenced by stereospecific conversion of the enantiomerically pure alcohol **1** to the corresponding *N*-

phthalimide derivative **2** by reaction with phthalimide *via* a Mitsunobu reaction[26,27] using PPh_3 and DIAD reagents in anhydrous THF as a solvent in 85% yield. The *N*-phthalimide **2** was deprotected[26] to the corresponding (*R*)-amine under reductive conditions using $NaBH_4$ in 2-propanol and subsequent treatment with conc. HCl at 80 °C. The free amine was protected[26] with Cbz-Cl using K_2CO_3 as a base to afford the corresponding *N*-Cbz protected amine **3** in an overall yield of 70% in three steps. The alkene of **3** was subsequently oxidised[28] using $NaIO_4$, $RuCl_3$ (0.25 mol%) and $CeCl_3 \cdot 7H_2O$ (10 mol%) as a Lewis acid in a mixture of CH_3CN / EtOAc / H_2O (3:3:2) at 0 °C to obtain the corresponding diol **4** in 59% yield as a ~1:1 mixture of inseparable epimers. Deprotection of the *N*-Cbz group was achieved by hydrogenolysis of the diol **4** using H_2 gas and 10% Pd/C in CH_3OH to obtain the desired amino-1,2-diol **5** as a ~1:1 mixture of epimers in 96% yield after subsequent purification step. The product **5** was purified by Grace Reveleris® X2 system using *n*-Hexane/EtOH (gradient, using ELS detector) on Reveleris® amino cartridge and isolated as a clear colourless oil. This was stored in a freezer at -20 °C for several months without any decomposition.

(*R*)-2-(Pent-4-en-2-yl)isoindoline-1,3-dione,[26] **2** and benzyl *N*-[(2*R*)-pent-4-en-2-yl]carbamate,[26] **3** are known compounds from the literature,[26] while benzyl [(2*R*)-4,5-dihydroxypentan-2-yl]carbamate, **4** and (4*R*)-4-aminopentane-1,2-diol, **5** are novel compounds reported here.

(*R*)-2-(Pent-4-en-2-yl)isoindoline-1,3-dione, 2



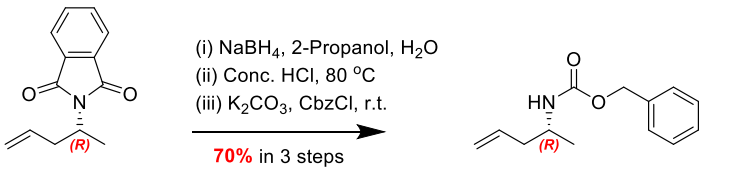
1 (*S*)-Pent-4-en-2-ol

2 Chemical Formula: $C_{13}H_{13}NO_2$
Exact Mass: 215.09463
Molecular Weight: 215.25200

(*R*)-2-(Pent-4-en-2-yl)isoindoline-1,3-dione, **2** was synthesised following a slightly modified protocol for Mitsunobu reaction described in the literature[26,27]. An oven-dried 500 mL 3-necked round-bottomed flask with a magnetic stirring bar was dried *in vacuo* using a heat gun and purged with argon. This procedure was repeated 3 times and cooled to r.t. under an argon atmosphere. The flask was then charged with (*S*)-pent-4-en-2-ol **1** (4.307 g, 50.0 mmol, 1.0

eq.), phthalimide (8.092 g, 55.0 mmol, 1.1 eq.) and PPh_3 (14.430g, 55.0 mmol, 1.1 eq.) and dissolved in dry THF (200 mL) under an argon atmosphere. The contents were cooled to 0 °C and stirred vigorously. DIAD (12.638 g, 62.5 mmol, 1.25 eq.) was added dropwise to the mixture for a period of 10 min at 0 °C under an argon atmosphere. The contents were warmed to r.t. for 30 min and then stirred for a further period of 2 h at r.t. (total time = 2.5 h). The contents were then evaporated to dryness and dissolved in EtOAc (400 mL) and sequentially washed with aqueous NaOH solution (0.5 M, 3 × 150 mL), aqueous HCl (1 M, 150 mL) and saturated brine solution (3 × 400 mL). The organic phase was then separated, dried over anhydrous Na_2SO_4, filtered and evaporated to dryness. The crude material was then subjected to flash chromatography on SiO_2 [gradient; eluant: EtOAc/*n*-hexane = 2:98 to 5:95] to obtain the desired (*R*)-2-(pent-4-en-2-yl)isoindoline-1,3-dione,[2] **2** as a colourless oil (9.190 g, 85%); $[\alpha]_D^{20}$ = -19.0° (*c* 1.0, $CHCl_3$); R_f = 0.48 (SiO_2 plate, EtOAc/*n*-hexane = 1:9); IR (neat) ν_{max} 3078, 2977, 2938, 1772, 1701, 1642, 1467, 1393, 1376, 1359, 1332, 1248, 1173, 1136, 1081, 1058, 1015, 994, 916, 895, 872, 793, 717 cm^{-1}; ^1H NMR (400.13 MHz, $CDCl_3$) δ 1.49 (d, *J* = 6.9 Hz, 3H), 2.51 (ddd, *J* = 14.1, 7.8, 6.8 Hz, 1H), 2.80 (dddd, *J* = 14.1, 8.4, 7.8, 1.3 Hz, 1H), 4.43 (ddq, *J* = 8.4, 7.8, 6.9 Hz, 1H), 4.95 (app. d, *J* = 10.1 Hz, 1H), 5.05 (dd, *J* = 16.8, 1.3 Hz, 1H), 5.71 (dddd, *J* = 16.8, 10.1, 7.8, 6.8 Hz, 1H), 7.68 (dd, J = 5.4, 3.0 Hz, 1H), 7.80 (dd, *J* = 5.4, 3.0 Hz, 1H); ^{13}C NMR (100.61 MHz, $CDCl_3$) δ 18.5 (CH_3), 38.3 (CH_2), 47.1 (CH), 117.8 (CH_2), 123.2 (CH), 132.1 (C), 133.9 (CH), 134.9 (CH), 168.6 (C); MS (ESI+) *m/z* (rel intensity) 238 [$(M+Na)^+$, 100%], 216 [$(M+H)^+$, 62], 205 (19), 201 (4); HRMS (ESI+) *m/z* calc'd for $C_{13}H_{14}NO_2$ $[M+H]^+$: 216.1019, found 216.1016 (Δ = -1.32 ppm). The structure of the product **2** was further characterised by 2D [1H, 1H] COSY, 2D [1H, 13C] HSQC, 2D [1H, 13C] HMBC NMR spectroscopic experiments.

<u>Benzyl *N*-[(2*R*)-pent-4-en-2-yl]carbamate, **3**</u>



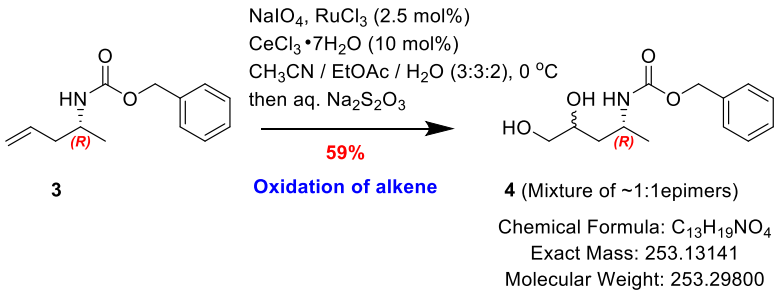
2 **3**

Chemical Formula: $C_{13}H_{17}NO_2$
Exact Mass: 219.12593
Molecular Weight: 219.28400

Benzyl *N*-[(2*R*)-pent-4-en-2-yl]carbamate, **3** was synthesised in 3 steps starting from the phthalimide **2** following a slightly modified protocol described in the literature by Gebauer *et*

al[26]. (*R*)-2-(Pent-4-en-2-yl)isoindoline-1,3-dione, **2** (4.951 g, 23.0 mmol, 1.0 eq.) was dissolved in isopropanol (170 mL) and H_2O (30 mL). $NaBH_4$ (8.701 g, 230.0 mmol, 10.0 eq.) was added portion wise (8 × 1.088 g) to the solution for every 5 min (5 × 8 = 40 min total) and the contents were stirred at r.t. overnight for 16 h. A white precipitate appeared in the reaction mixture and the reaction mixture was quenched by addition of conc. HCl (15.0 mL), stirred at r.t. for 10 min, then heated to 80 °C for 2 h and then cooled to r.t. A thick white precipitate was observed and the mixture was then neutralised with K_2CO_3 to pH 8.0, after which the mixture was cooled to 0 °C and benzyl chloroformate (5.885 g, 34.5 mmol, 1.5 eq.) was added dropwise for a period of 15 min. The mixture was stirred at 0 °C and DMAP (0.030 mg, 0.246 mmol, 0.0107 eq.) was added to it. The contents were then stirred at r.t. under an argon atmosphere and the progress of the reaction was monitored by TLC analysis (SiO_2 plate, EtOAc/*n*-hexane = 3:17). After 2.5 h, the reaction was judged to be complete and the reaction mixture was then quenched by addition of saturated brine solution (200 mL), extracted with Et_2O (3 × 200 mL). The aqueous phase was further extracted with EtOAc (50 mL). The combine organic extracts were washed with saturated brine solution (3 × 500 mL), dried over anhydrous Na_2SO_4, evaporated to dryness using a rotatory evaporator to obtain a colourless oil. This was purified by flash chromatography on SiO_2 [gradient; eluant: EtOAc/*n*-hexane = 1:19 → 1:9 → 3:17 → 1:4] to obtain the desired product benzyl *N*-[(2*R*)-pent-4-en-2-yl]carbamate,[26] **3** as a colourless oil (3.518 g, 70%) which crystallised upon freezing on long storage at -20 °C in freezer; $[\alpha]_D^{20}$ = +13.1° (*c* 1.0, $CHCl_3$); R_f = 0.48 (SiO_2 plate, EtOAc/*n*-hexane = 1:9); ^{1}H NMR (400.13 MHz, $CDCl_3$) δ 1.16 (d, *J* = 6.5 Hz, 3H), 2.22 (m, 2H), 3.81 (tq, *J* = 6.5, 6.5 Hz, 1H), 4.64 (broad s, 1H), 5.02-5.19 (2 × m, 4H), 5.77 (ddt, *J* = 16.9, 10.2, 7.2, 1H), 7.28-7.45 (broad m, 5H); ^{13}C NMR (100.61 MHz, $CDCl_3$) δ 20.6 (CH_3), 41.2 (CH_2), 46.6 (CH), 66.6 (CH_2), 118.1 (CH_2), 128.2 (2 × CH), 128.6 (CH), 134.3 (CH), 136.8 (C), 155.8 (C); IR (CH_2Cl_2) ν_{max} 3324, 3068, 3033, 2972, 1693, 1642, 1530, 1454, 1405, 1376, 1334, 1280, 1250, 1221, 1100, 1058, 1028, 994, 915, 776 cm^{-1} ; MS (ESI+) *m/z* (rel intensity) 242 [$(M+Na)^+$, 40%], 220 [$(M+H)^+$, 100], 176 (3), 159 (12), 117 (2), 88 (5); HRMS (ESI+) *m/z* calc'd for $C_{13}H_{18}NO_2$ $[M+H]^+$: 220.1332, found 220.1327 (Δ = -2.25 ppm). The structure of the product **3** was further characterised by 2D [1H, 1H] COSY, 2D [1H, 13C] HSQC, 2D [1H, 13C] HMBC NMR spectroscopic experiments.

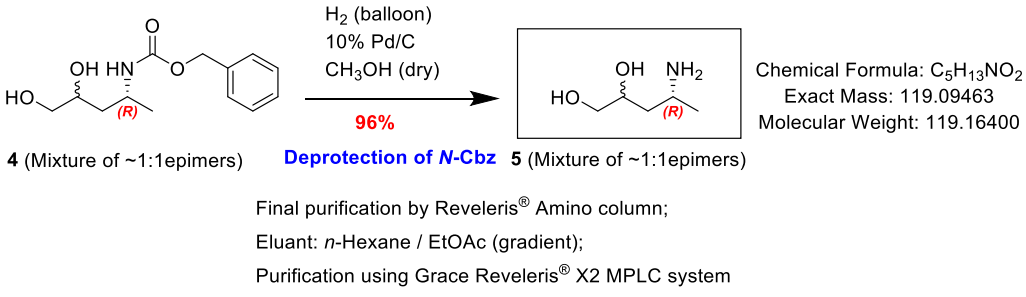
<u>Benzyl [(2*R*)-4,5-dihydroxypentan-2-yl]carbamate, **4**</u>



Benzyl [(2*R*)-4,5-dihydroxypentan-2-yl]carbamate, **4** was prepared by ruthenium tetraoxide (generated *in situ* using $RuCl_3/NaIO_4/CeCl_3 \cdot 7H_2O$) mediated oxidation of benzyl *N*-[(2*R*)-pent-4-en-2-yl]carbamate, **3** following a general procedure described in the literature by Tiwari[28] *et al*. A mixture of $NaIO_4$ (1.749 g, 8.175 mmol, 1.5 eq.) and $CeCl_3 \cdot 7H_2O$ (0.203 g, 0.545 mmol, 0.1 eq.) in H_2O (11 mL) was stirred at r.t. for 10 min in a 100 mL round-bottomed flask fitted with a magnetic stirring bar. The off-white mixture was cooled to 0 °C and stirred for 4-5 min, followed by sequential addition of EtOAc (16 mL), CH_3CN (33 mL) and $RuCl_3$ (0.0283 g, 0.1363 mmol, 0.025 eq.) at 0 °C. The hetereogeneous mixture was then stirred for 2 min and a solution of the benzyl *N*-[(2*R*)-pent-4-en-2-yl]carbamate, **3** (1.195 g, 5.45 mmol, 1.0 eq.) in EtOAc (16 mL) was added in one shot to this mixture. The contents were stirred at 0 °C and progress of the reaction was monitored by TLC analysis (SiO_2; Eluant: EtOAc/*n*-hexane = 3:17 and also using Eluant: EtOAc = 100%). After 22 min, the reaction mixture turned more purple brown from black and was stopped by diluting the mixture using saturated $Na_2S_2O_3$ solution (200 mL) on a separating funnel and addition of additional EtOAc (150 mL). The mixture was shaken vigorously in the separating funnel and the aqueous layer was further extracted with EtOAc (2 × 75 mL). The combined organic layers were washed sequentially with saturated aqueous $Na_2S_2O_3$ solution (200 mL), saturated aqueous $NaHCO_3$ solution (2 × 200 mL) and finally with saturated brine solution (200 mL). The organic layer was separated, dried over anhydrous Na_2SO_4 and evaporated to dryness to obtain a colourless clear oil. The crude material was subjected to flash chromatography on SiO_2 [gradient; Eluant: EtOAc/*n*-hexane = 3:1 to EtOAc = 100%) to obtain the desired product, *benzyl [(2R)-4,5-dihydroxypentan-2-yl]carbamate*, **4** as a colourless oil (0.817 g, 59%), which was found to be ~1:1 inseparable mixture of epimers and rotamers as observed in ^1H and ^{13}C NMR spectroscopic analysis; R_f = 0.35 and 0.41 (for 2 epimers); ^1H NMR (400.13 MHz, $CDCl_3$) δ (Mixture of epimers) 1.18 and 1.20 (2 × d, J = 7.0, 6.8 Hz, 3H), 1.24-1.38 and 1.40-1.76 (2 × m, 2H), 2.24 and 2.82 (2 × broad s, 1H), 3.08 and 3.40 (2 × broad s, 1H), 3.42-3.65 (2 × m, 2H), 3.70-3.80 (2 × m, 1H), 3.81-3.90 and 3.91-4.06 (2 × m, 1H), 4.23 and 4.95 (2 × app. d, J

= 2.6 and 8.3 Hz, 1H), 5.07 and 5.10 (2 × app. d, J = 11.8 and 12.1 Hz, 2H), 7.27-7.44 (broad m, 5H); ^{13}C NMR (100.61 MHz, $CDCl_3$) δ (Mixture of epimers) 21.6 (2 × CH_3), 40.3 (CH_2), 41.2 (CH_2), 44.1 (CH), 45.4 (CH), 66.6 (CH_2), 66.8 (CH_2), 67.2 (2 × CH_2), 68.8 (CH), 70.3 (CH), 128.2 (CH), 128.2 (2 × CH), 128.4 (CH), 128.6 (CH), 128.7 (CH), 136.3 (C), 136.6 (C), 156.3 (C), 157.3 (C); IR (CH_2Cl_2) ν_{max} 3401, 3325, 2970, 2933, 1690, 1516, 1455, 1411, 1377, 1339, 1275, 1248, 1088, 1055, 1020, 871, 843, 764, 757 cm^{-1} ; MS (ESI+) *m/z* (rel intensity) 276 [(M+Na)$^+$, 100], 254 [(M+H)$^+$, 12], 210 (14); HRMS (ESI+) *m/z* calc'd for $C_{13}H_{19}NO_4{}^{23}Na$ [M+Na]$^+$: 276.1206, found 276.1197 (Δ = -3.51 ppm) and *m/z* calc'd for $C_{13}H_{20}NO_4$ [M+H]$^+$: 254.1387, found 254.1384 (Δ = -1.28 ppm). The structure of the diol **4** was further characterised by 2D [1H, 1H] COSY, 2D [1H, 13C] HSQC, 2D [1H, 13C] HMBC NMR spectroscopic experiments.

(4*R*)-4-Aminopentane-1,2-diol, **5**



An oven-dried 100 ml round-bottomed flask with a magnetic stirring bar was fitted with a Quickfit® 3-way stopcock adapter and dried *in vacuo* using a heat gun and purged with nitrogen gas. This procedure was repeated thrice. The flask was cooled to r.t. and was charged with palladium on activated carbon (10% Pd, 0.500 g) under a nitrogen atmosphere and a solution of benzyl [(2*R*)-4,5-dihydroxypentan-2-yl]carbamate, **4** (0.778 g, 3.0715 mmol) dissolved in CH_3OH (25 mL) was added to it. The atmosphere inside the flask was evacuated using a diaphragm pump and purged with hydrogen gas (H_2 balloon) using the 3-way stopcock and this procedure was repeated thrice. The heterogenous mixture was then stirred under a hydrogen atmosphere (H_2 balloon) at r.t. and the progress of the reaction was monitored for the disappearance of the starting material **4** by TLC analysis (SiO_2 plate; Eluant:100% EtOAc) and LC-MS analysis (C18 reverse phase column, H_2O-CH_3CN as mobile phase, gradient). After 6 h, the reaction was judged to be complete and reaction mixture was filtered through Celite® 503 and washed with copious amount of CH_3OH (3 × 25 mL) and EtOAc (2 × 25 mL). The combined filtrate was evaporated to dryness to obtain *(4R)-4-aminopentane-1,2-diol*, **5** as a

clear colourless oil (0.366 g, quant. yield) which was observed to be a ~1:1 mixture of epimers as observed from its ^{1}H and ^{13}C NMR spectroscopic analysis. This was further purified and desalted on a Reveleris® Amino column using a Grace Reveleris® X2 MPLC system (gradient; using ELS detector; eluant: 100% *n*-hexane to *n*-hexane / EtOAc = 1:1; the product was isolated between *n*-hexane / EtOAc = 4:1 → 7:3) to afford *(4R)-4-aminopentane-1,2-diol*, **5** as a clear colour oil (0.353 g, 96% yield after purification). The aminodiol product **5** was stored in a freezer at -20 °C; ^{1}H NMR (400.13 MHz, $CDCl_3$) δ (Mixture of epimers) 1.12 and 1.14 (2 × d, J = 3.0 and 3.1 Hz, 3H), 1.23-1.40 and 1.44-1.65 (2 × m, 2H), 3.00-3.11 and 3.18-3.24 (2 × m, 1H), 3.19-3.35 (broad s, 4H), 3.36-3.48 and 3.49-3.58 (2 × m, 2H), 3.80-3.88 and 3.89-3.96 (2 × m, 1H); ^{13}C NMR (100.61 MHz, $CDCl_3$) δ (Mixture of epimers) 24.2 (CH_3), 27.0 (CH_3), 40.4 (CH_2), 40.7 (CH_2), 44.4 (CH), 47.3 (CH), 66.8 (CH_2), 66.9 (CH_2), 69.8 (CH), 72.5 (CH); IR (CH_2Cl_2) ν_{max} 3352, 3284, 2947, 2924, 2904, 2869, 1608, 1457, 1380, 1349, 1337, 1266, 1153, 1096, 1074, 1031, 988, 935, 853, 824, 732, 702 cm^{-1} ; MS (ESI+) *m/z* (rel intensity) 161 [{(M+CH_3CN)+H}$^+$, 7], 120 [(M+H)$^+$, 100], 102 (3); HRMS (ESI +) *m/z* calc'd for $C_5H_{14}NO_2$ $[M+H]^+$: 120.1019, found 120.1019 (Δ = -0.05 ppm). The structure of the aminodiol **5** was further characterised by 2D [1H, 1H] COSY, 2D [1H, 13C] HSQC, 2D [1H, 13C] HMBC NMR spectroscopic experiments.

Synthesis of *R*-methyl deoxypropanoguanine containing deoxyoligonucleotides

Although acetaldehyde forms both *R*- and *S*- stereoisomers of deoxypropanoguanine, the *R*-form is more stereochemically favourable for crosslink formation in a 5′-CpG-3′ on which we based this study[29]. A custom oligonucleotide, 5'-[phosphoryl]-GCACGAAAGAAGAGC-2FdI-GAAG was synthesized by Eurogentec at 1 µmol scale, using 5'-Dimethoxytrityl-2-fluoro-O6-p-nitrophenylethyl-2'-deoxyInosine,3'-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite and shipped on support. Support (20 mg, ca. 0.5 µmol oligo) containing 2FdIno, **6** was incubated with 12 mg (4*R*)-4-aminopentane-1,2-diol **5** in 280 µl DMSO with 140 µl TEA and incubated with agitation at rt overnight to give compound **7**. The support was washed three times with DMSO (200 µl) and three times with CH_3CN (400 µl), followed by deprotection of the O6-p-nitrophenylethyl group with 1 M DBU dissolved in CH_3CN (300 µl) at rt for 1 h, to give compound **8**. The support was washed three times with CH_3CN, dried with N_2 and treated with aq. 28% NH_4OH at 55 ˚C for 6 h to remove remaining protecting groups and elute the oligonucleotide from the support. The oligonucleotide was freeze-dried (Scanvac) and reactants separated by HPLC (Thermo Accela) on a AdvanceBio Oligonucleotide, 4.6 x

150 mm, 2.7 µm (Agilent) in 15 mM TEA/ 400 mM HFIP, (pH 7.0) at 0.5 ml min^{-1} at 60˚C, over 30 min, effecting a 15 – 27.5% MeOH gradient; **8** eluted at 19.2 min and **6** at 20.5 min. Fractions containing **8** were lyophilized and reacted with 50 mM $NaIO_4$ for 1 h at rt, and the product desalted over an Oligo R3 column (Thermo Scientific). The identity of **9** was confirmed by MALDI and/or LC-MS (calc'd mass: 6408.18, found 6409.0 ($\pm \leq 0.02\%$ error).

R-methyl deoxypropanoguanine crosslinking and purification

Unmodified oligonucleotides were supplied HPLC purified and verified by LC-MS (Integrated DNA Technologies, BVBA). Either deoxyguanine containing (5'-[phosphoryl]-CTTTCTTCTC-dG-CCTTCTCCC) or deoxyinosine (5'-[phosphoryl]-CTTTCTTCTC-dI-CCTTCTCCC) oligonucleotides were mixed at 1:1 molar ratio with the PdG containing oligo (5'-[phosphoryl]-GCACGAAAGAAGAGC-PdG-GAAG) in PBS, annealed (85˚C for 5 min, ramped to 25 ˚C, at - 0.1˚C s^{-1}) and incubated at 37˚C for 7-14 d to allow crosslink formation. The reaction progression was analysed by denaturing PAGE stained with Sybr Gold (Thermo Scientific). Final PdG-dG crosslink yields were approx. 50%. Crosslinks were purified by denaturing PAGE and/or HPLC.

For PAGE purification, bands were stained briefly with Sybr Gold, visualized by trans-illumination at 450 nm (Clare Chemical Research) and excised. DNA was electroeluted from the gel (Elutrap, Whatman), concentrated by ethanol precipitation and quantified by Nanodrop (260 nm).

HPLC purification of the crosslink followed the same conditions detailed above, crosslinked oligonucleotides eluted between 25.8 – 26.2 min, and were confirmed by denaturing PAGE analysis. Fractions containing crosslinked DNA were evaporated to dryness and reconstituted in 1 x PBS.

Crosslink Stability Assay

Crosslinked DNA was back-filled with exo-deficient Klenow (New England Biolabs) to incorporate a single ^{32}P-CTP radiolabel. Labelled oligos were purified by R3 Oligo resin and quantified by scintillation counting (Perkin Elmer). Labelled oligos were incubated in 1 x PBS and after the indicated temperature and time combinations samples were removed, mixed with formamide loading buffer and resolved by formamide-urea denaturing PAGE. Dried gels were

exposed on phosphorimager 24-48 h and visualized using an Amersham Typhoon Biomolecular Imager (GE Healthcare). Crosslink stability was calculated from band densitometry using Fiji software[30] and the results were plotted using Prism 7 (Graphpad Software) from three independent experiments.

Preparation of plasmid substrates

The plasmids containing site-specific interstrand crosslinks were prepared as described [31,17]. Briefly, duplexes containing ICLs derived from cisplatin, psoralen or acetaldehyde were ligated into a vector linearized with Bbs1. After ligation, the plasmid was purified using a cesium chloride gradient. To ligate the psoralen duplex, the pSVRLuc backbone was modified by site-directed mutagenesis to substitute an adenine to guanine present on one of the Bbs1 5' overhangs.

The plasmids containing the *lacO* sequence were prepared by ligating the crosslinked oligos duplexes into pSVR*lacO* vector (gift from Johannes Walter,[17]). To make pICL-AAreverse-*lacO*, the pSVR*lacO* was first digested with Not1 and a new insert containing Bbs1 overhangs in reversed position was inserted. This plasmid was used to insert the crosslinked duplexes.

To make pPdG, ligation of a duplex containing a PdG into a backbone linearized by Bbs1 was followed by cesium gradient purification, as described above.

***Xenopus* repair assay**

All animal procedures and experiments were performed in accordance with national animal welfare laws and were reviewed by the Animal Ethics Committee of the Royal Netherlands Academy of Arts and Sciences (KNAW). All animal experiments were conducted under a project license granted by the Central Committee Animal Experimentation (CCD) of the Dutch government and approved by the Hubrecht Institute Animal Welfare Body (IvD), with project license number AVD80100201711044. Sample sizes were chosen based on previous experience, randomization and blinding are not relevant to this study.

<u>*Xenopus* egg Extracts and DNA Replication</u>

Xenopus laevis female frogs (aged >2 years) were purchased from Nasco and used as a source of eggs. Preparation of *Xenopus* egg extracts and DNA replication were performed as previously described[32,33]. For DNA replication, plasmids were incubated in a high-speed supernatant extract (HSS) at a final concentration of 7.5 ng μl^{-1} for 20 min at room temperature

to license the DNA. Two volumes of nucleoplasmic extract (NPE) were added to start DNA replication. For nascent strand labeling, HSS was supplemented with ^{32}P-α-dCTPs. Where indicated, an unrelated non-damaged control plasmid (pQuant) was added at concentrations up to 0.8 ng ul^{-1}, to be used as internal control for quantifications. To block CMG unloading, p97 inhibitor (NMS-873, Sigma)[34] was added to NPE at a final concentration of up to 40 μM. To inhibit DNA replication recombinant Geminin was added to HSS at 300 nM. For the reactions with the pICL-*lacO* plasmids, the plasmids were incubated prior to the start of replication with purified biotinylated LacR protein at a final concentration of 17 μM for 45 minutes at room temperature[35]. For CMG uncoupling experiments, aphidicolin (Sigma) was added to the replication mix 5 minutes after the start of replication, at a final concentration of 12μM.

For analysis of undigested DNA replication products, replication reactions were stopped by adding five volumes of replication stop solution I (Stop I: 80 mM Tris pH 8, 5% SDS, 0.13% phosphoric acid, 10% Ficoll, 8 mM EDTA, 0.1% bromophenol blue). The samples were treated with proteinase K (1.5 μg μl^{-1}) for 40 min at 37°C and resolved by 0.8% native agarose gel electrophoresis. The gel was dried and visualized by autoradiography.

For extraction of DNA from replication reactions, aliquots of the reaction were stopped with 10 volumes of replication stop solution II (Stop II: 50 mM Tris pH 7.5, 0.5% SDS, 10 mM EDTA pH 8). Samples were then treated with RNase (0.13 μg μl^{-1} final concentration) for 30 min at 37°C, followed by Proteinase K (0.5 μg μl^{-1}) treatment for 1 h at 37 °C or overnight at room temperature. DNA was phenol/chloroform extracted, ethanol precipitated with glycogen (0.3 μg μl^{-1}), and resuspended in 10 mM Tris pH 7.5 in a volume equal to the reaction sample taken.

Antibodies and Immunodepletions

Antibodies against *xl*FANCD2, *xl*REV1, *xl*REV7 were previously described[12,14,19]. The *xl*NEIL3 antibody was raised against a C-terminal peptide of *Xenopus laevis* NEIL3 and affinity purified (New England peptide). FANCD2 and REV7 depletions were carried out as described[12,19]. To rescue the FANCD2 depletion 260 nM of recombinant FANCI-FANCD2 was added to NPE. For depletion of NEIL3, one volume of protein A Sepharose Fast Flow beads (GE Healthcare) was incubated with 2 volumes of *xl*NEIL3 affinity purified antibody (1 mg ml^{-1}) overnight at 4°C. For depletion from NPE, three rounds were performed in which 1

volume of antibody-bound PAS beads was incubated with 4.5 volumes of extract for 20 min at room temperature. HSS was depleted using the same volumes but for 2 rounds. Recombinant NEIL3 was added to NPE at 150 nM final concentration. For REV1 depletion, 1 volume of Dynabeads™ Protein A beads (Thermo Fisher Scientific) was pre-incubated with 0.5 volumes of antibody for 30 minutes at room temperature. To deplete NPE, 1 volume of antibody-bound beads was incubate with 1.5 volume of extract for 30 minutes at RT for two rounds. HSS was depleted with the same volume of beads in one round.

Protein purification

Recombinant *xl*FANCI-FANCD2 complex was prepared as previously described[12]. Biotinylated LacR protein was prepared as described[35].

NotI repair assay

Plasmids were replicated in *Xenopus* egg extract in absence of ^{32}P-α-dCTPs and intermediates were purified as described above. Replication reactions were supplemented with a modified pQuant plasmid (pQuant 28nt) that produces a 28 nt fragment upon NotI digestion, which was used as internal control for quantification. Repair intermediates were digested with NotI and 3' labeled by fill-in of the 5'-overhangs with Sequenase DNA polymerase (USB) in the presence of ^{32}P-α-dCTPs and non-labeled dGTP. One volume of denaturing PAGE Gel Loading Buffer II (Invitrogen™) was added, samples were denatured by incubation at 98 ˚C for 3 minutes, and the DNA fragments were separated by 20% Urea PAGE, products were visualized by autoradiography and quantified using ImageQuant software (GE Healthcare). NotI digested and labeled products that still contain an ICL are 88 nt in size while non-crosslinked molecules are 44 nt. The temporal reduction of the 88 nt product is caused by ongoing ICL unhooking but also by replication fork stalling because this results in a DNA structure that is partially single stranded and therefore not digested by NotI.

The temporal increase of the uncrosslinked 44 nt product was used as a readout for ongoing ICL repair because these are the products of TLS or HR. First, the intensity of pQuant 28 nt fragment was used to normalize the intensity of the 44 nt products, to make these products independent of amplification by replication/repair and to correct for extraction/loading variations. Then, the normalized linear 44 nt products were expressed as percentage of the total of 88 and 44 nt fragments at time point zero, which represents the percentage of the input DNA that is accumulating over time in the 44 nt band. Finally, the percentage of non-crosslinked

background was subtracted to yield the percentage of molecules that have undergone TLS or HR during repair. Of note, 2 or 3 bands can be seen at the height of the 44 nt fragment in some gels. First, the sequence difference between the bottom and top strand causes a slight difference in mobility. Second, upon ICL repair by the FA pathway the unhooked adduct will cause a shift in size (see also **Extended Data Figure 5k**).

Nascent strand analysis

Nascent strand analysis was performed as previously described[14]. In brief, DNA repair products were digested either with AflIII, with AflIII or EcoRI, or with AflIII and BamHI (New England Biolabs). After addition of one volume of denaturing PAGE Gel Loading Buffer II (Invitrogen™), the samples were separated on a 7% polyacrylamide sequencing gel, and visualized by autoradiography. The sequencing gel ladder was produced using the Thermo Sequenase Cycle Sequencing Kit (USB) and primer S (CATGTTTTACTAGCCAGATTTTTCCTCCTCTCCTG).

APE1 glycosylase assay

Plasmids were replicated in presence of pQuant. After extraction of the DNA repair intermediates, samples were digested with HincII (New England Biolabs) only and with HincII in combination with APE1 (New England Biolabs). The digested products were separated on a 0.8% native agarose gel and visualized by autoradiography. Quantification was done using Image Quant (GE Healthcare). The arm fragments were first normalized against pQuant and the highest value was set to 1.

HincII incision assay

ICL-containing plasmids were replicated in the presence of pQuant. Repair intermediates were extracted and digested with HincII. Fragments were separated on a 0.8% native agarose gel and visualized by autoradiography. Quantification was performed using ImageQuant software (GE Healthcare). The intensity of the HincII arm fragments was first normalized to pQuant, the highest value in the experiment was set to 1 and the data was plotted against time.

Strand-specific Southern blot assay

Strand-specific Southern blot was performed as described previously[14,17]. Briefly, purified replication intermediates were digested with AflIII and AseI, 0.5 volume of denaturing PAGE

Gel Loading Buffer II (Invitrogen™) was added, and the DNA fragments were separated on a 7% denaturing polyacrylamide sequencing gel. The DNA was transferred to a Hybond-N+ membrane (Amersham) overnight at 4°C. The membrane was rinsed once with 2X SSC buffer and UV-irradiated to crosslink the DNA to the membrane. The membrane was pre-hybridized with 12 ml of Ultrahyb buffer (Ambion) for 3 h at 42°C and probed overnight at 42°C. The membrane was then washed three times at 42°C in pre-warmed washing buffer (0.5X SCC, 0.25% SDS), dried and exposed to a phosphorimager screen. To generate the probes pCon was amplified by PCR using the primers XL-1 (CCTGCTGTCCATTCCTTATTCC) and XL-3 (GCATTGGTTCTGCACTTCCGC). The amplicon was purified and used as template for a primer extension reaction in the presence of ^{32}P-α-dCTPs to generate the probes[14] . Primer S was used for the probe that detects the bottom strand and primer XL-3 for the top strand.

Primer extension assay

Primer extension assay was performed as previously described[19]. In brief, late ICL repair products were isolated and purified as described above. Unreplicated pControl and pPdG were used as controls. Samples were digested with AflIII and BamHI, and the DNA fragments were used as template for primer extension. Primers (Forward: CTCGAGCGGAAGTGCAGAAC, Reverse: AATACGCAAACCGCCTCTCC) were first labelled radioactively with PNK at the 5' end with ^{32}P-γ-ATP and were subsequently annealed to DNA in a thermocycler. Each reaction contained only one primer (either forward or reverse). The primers were then extended using the Phusion High-Fidelity DNA polymerase (NEB) for one round. The resulting DNA fragments were first concentrated by ethanol precipitation (see above) and then separated on a 20% denaturing PAGE gel and visualized by autoradiography.

Sequencing analysis of DNA repair products

Sequencing of the DNA repair products was performed using a similar method as described previously[19]. Briefly, plasmids were replicated in *Xenopus* egg extract and late replication/repair intermediates (240 min for the crosslinked plasmids, 120 min for pCon and pPdG) were isolated. The DNA was amplified by PCR using the Phusion High-Fidelity DNA Polymerase (New England Biolabs) for 15 cycles using primers with unique barcodes. For experiment 1 we used forward primer: NNNNNN-barcode-AGAACCAATGCATGCGGC and reverse primer: NNNNNNATTCATTAATGCAGCGGATCG, where N represents random nucleotides and the barcode is composed of 3 nucleotides. For experiment 2 we used forward

primer: NNNNNN-barcode-CTCGAGCGGAAGTGCAGAAC and reverse primer: NNNNNN-barcode-AATACGCAAACCGCCTCTCC, the barcode was composed of 6 nucleotides. This generated a product of 96 nt for experiment 1 presented in the main figures, and 154 nt for experiment 2 in the supplement, with the lesion site roughly in the middle (Extended Data Item 8a). In addition to the replicated DNA samples, we also used 75 ng unreplicated pICL-Pt, pICL-AA and pPdG as template in independent PCR reactions. For each sample two independent PCRs were performed to create technical replicates. The products were separated by TBE PAGE and visualized by Sybr Gold (Thermo Fisher Scientific) staining. For experiment 1 we excised the 96 nt fragment including products up to ~10 nt smaller, for experiment 2 we excised the 154 nt band. Gel slices were broken into small pieces and incubated in elution buffer (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0, 300 mM NaCl) overnight at room temperature under agitation. The DNA was recovered using a Spin-X column (0.45 μM, Sigma) and further purified using paramagnetic beads (CleanPCR kit, CleanNA). The DNA was ethanol-precipitated in presence of 0.15 mg ml^{-1} glycogen and 0.3 M sodium acetate and resuspended in 10 mM Tris pH 8.0. Illumina TruSeq adapters were ligated to the DNA according to the manufacturer's instructions (Illumina), and the DNA was sequenced using paired end 75 bp NextSeq sequencing for dataset 1, and paired end 150 bp NextSeq sequencing for dataset 2 (Utrecht sequencing facility).

Demultiplexing

Data from experiment 1 was demultiplexed on 3 nt barcodes at the 5'-ends, downstream of a 6 nt random nucleotide segment of the first mate (R1) of read pairs, allowing no mismatches. The 3 nt barcodes were extended on the 3'-end with 9 nt of the forward PCR primer ("AGAACCAAT") to avoid reverse complement reads demultiplexing as forward reads. Reads not demultiplexed as forward reads were then demultiplexed on the second mate (R2) of read pairs, again allowing no mismatches, and again with the 3 nt barcodes which were extended with 9 nt of the forward PCR primer.

Data from experiment 2 was demultiplexed on 6 nt barcodes prefixed at the 5'-ends of the first mate (R1) of read pairs, allowing at most 1 mismatch (no indels). Barcodes were designed such that 1 mismatch could be tolerated without introducing ambiguity. For dataset 2, R2 of reads demultiplexed with forward primers and R1 of reads demultiplexed with reverse primers was taken the reverse complement of (using fastx_reverse_complement from FASTX_toolkit v0.0.14). For dataset 1, R2 of demultiplexed reads were trimmed by 6 nt. (using fastx_trimmer from FASTX_toolkit v0.0.14) and then taken the reverse complement of. To ensure R1 and R2

contained the same segment of the amplicon (which is larger than the read lengths in dataset 1) and did not dovetail, R1 was trimmed at the 5'-end using "ATGCGGCCGC" and allowing at most 1 mismatch and no indels, using cutadapt (v1.9.1[36]) and at the 3'-end using "GCGCGGCCGCG", again allowing at most 1 mismatch and no indels. Similarly, R2 reads were trimmed using "GCCGCGAAG" at the 5'-end and "GCATTAA" at the 3'-end. To make the data of dataset 1 comparable to dataset 2, R1 reads of dataset 1 were then prefixed with "CTCGAGCGGAAGTGCAGAACCAATGCATGCGGCCGC" and suffixed with "GCGCGGCCGCGATCCGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATT", and R2 reads were prefixed with "CTCGAGCGGAAGTGCAGAACCAATGCATGCGGCCGCGAAG" and suffixed with "GCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATT", reconstituting the amplicon of dataset 2.

Alignment to amplicon

Reads were mapped to the amplicon sequence of experiment 2 (Extended Data Item 8a) using blastn (from the BLAST suite, v2.4.0+, with parameters -dust no -parse_deflines -max_hsps 1 -num_alignments 1 -word_size 7 -evalue 100.0 -outfmt "6 qseqid sstart send qstart qend btop" -strand plus). The blast result generated a table with one line per read, for both the R1 and R2 read mates. From these tables, the Blast trace-back operations (BTOP) string was parsed to extract information about mismatches in the read relative to the amplicon sequence. These tables were further processed using python software. This software filters out read pairs where R1 and R2 do not agree, or do not fully contain the 130 nt amplicon. This software then summarized the BLAST tables into occurrences of each BLAST BTOP, for each sample. Read numbers for total reads, pair-matched reads, and reads with indels and insertions are listed in Extended Data Item 8b and c.

Post-processing of mutation count tables

Technical replicates (such as sequencing lanes and PCR replicates) were merged by summation. Because the reads of dataset 1 and dataset 2 are of different length, dataset 1 was only informative on mutations in the amplicon interval [40, 76) (0-based, half-open interval). Therefore, normalization was done using only mutations that were observed in both datasets (with at least 100 reads). Observed mutation (BTOP) counts were normalized by

$$N_{i,j} = \frac{C_{i,j} + \phi}{\sum_{i \in W} C_{i,j} + \phi |W|}$$

where $C_{i,j}$ is the count of mutation i in sample j, W is the set of mutations that meet the cutoff criterion in both datasets, ϕ is a pseudocount (set to 1 in further analyses) and $N_{i,j}$ is the normalized count of mutation i in sample j.

Cellular Survival Assays

Cell lines

Wild-type and *NEIL3*$^{-}$ HAP1 near-haploid human cells were purchased from Horizon Discovery and cultured at 37˚C and 5% CO2 in IMDM supplemented with 10% fetal calf serum and penicillin/streptomycin. *NEIL3*$^{-}$ cells were confirmed by immunoblotting against NEIL3. For targeting of *FANCL*, WT and *NEIL3*$^{-}$ cells were transfected using Turbofectin 8.0 (Origene) and the following plasmids: pX461 (Cas9 nickase) (Addgene), gRNA vectors (FANCL_left: CCTAATGCAATTCTGCGTGCTGT and FANCL_right: TTTTTCTGGCTCAAGTACCCAGG), and a *FANCL*-Puro targeting construct (Wellcome Trust Sanger Institute). Two days post-transfection, 3.5 µg ml^{-1} puromycin was added, and two days later, cells were plated in 96-well plates with puromycin at a density of 1 x 10^4 cells per well. After 14 days of incubation, individual clones were picked and analyzed for *FANCL* targeting using the SequalPrep Long PCR kit (Applied Biosystems) (using primers FANCL_ LR_FW2, TGTCTACCCCCTAAGTTCGTTGA; EF1a_R1, GCGATCTCTGGGTTCTACG TTAGTG). Targeted clones were then plated with 100 ng ml^{-1} mitomycin C overnight and analyzed by immunoblotting for FANCD2. *FANCL* knockouts were identified by failure to ubiquitinate FANCD2. All cell lines were confirmed mycoplasma negative using the MycoAlert Mycoplasma Detection Kit (Lonza).

Clonogenic Survival Assay

For colony survival assay, HAP1 cells were prepared at 2 x 10^5 cells ml^{-1}. Cells and acetaldehyde (diluted in culture media) were mixed in 96-well blocks (Greiner Bio-One Masterblock) and foil seals (Bio-Rad Microseal ‘F’) were applied before culturing cells at 37˚C for 2 h. Cells were then serially diluted in PBS using a multi-channel pipette to obtain 1:10 and 1:100 dilutions, and 100 µl of each were plated in duplicate in 24-well plates filled with 1.5 mL of culture media per well. Cells were cultured for 6 days before being stained with crystal

violet[37] and colonies were quantified by a GelCount colony counter (Oxford Optronix). Data presented reflects a minimum of three independent experiments, error bars denote S.E.M.

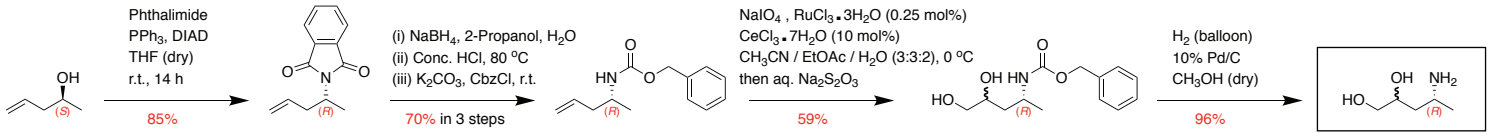
References

26. Gebauer, J., Rost, D. & Blechert, S. A highly efficient synthesis of (-)-pinidinol. *Heterocycles* **68**, 2129-2132 (2006).
27. Mahesh, M., Murphy, J.A., LeStrat, F. & Wessel, H.P. Reduction of arenediazonium salts by tetrakis(dimethylamino)ethylene (TDAE): Efficient formation of products derived from aryl radicals. *Beilstein Journal of Organic Chemistry* **5**(2009).
28. Tiwari, P. & Misra, A.K. An efficient stereoselective dihydroxylation of glycals using a bimetallic system, RuCl3/CeCl3/NaIO4. *Journal of Organic Chemistry* **71**, 2911-2913 (2006).
29. Lao, Y.B. & Hecht, S.S. Synthesis and properties of an acetaldehyde-derived oligonucleotide interstrand cross-link. *Chemical Research in Toxicology* **18**, 711-721 (2005).
30. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676-682 (2012).
31. Enoiu, M., Jiricny, J. & Scharer, O.D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Research* **40**, 8953-8964 (2012).
32. Lebofsky, R., Takahashi, T. & Walter, J.C. DNA replication in nucleus-free Xenopus egg extracts. *Methods Mol Biol* **521**, 229-52 (2009).
33. Walter, J., Sun, L. & Newport, J. Regulated chromosomal DNA replication in the absence of a nucleus. *Molecular Cell* **1**, 519-529 (1998).
34. Magnaghi, P. et al. Covalent and allosteric inhibitors of the ATPase VCP/p97 induce cancer cell death. *Nature Chemical Biology* **9**, 548-556 (2013).
35. Dewar, J.M., Budzowska, M. & Walter, J.C. The mechanism of DNA replication termination in vertebrates. *Nature* **525**, 345-350 (2015).
36. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2011).
37. Franken, N.A., Rodermond, H.M., Stap, J., Haveman, J. & van Bree, C. Clonogenic assay of cells in vitro. *Nat Protoc* **1**, 2315-9 (2006).

Supplemental note 2

SI Figure 1

a



1 (S)-Pent-4-en-2-ol

2

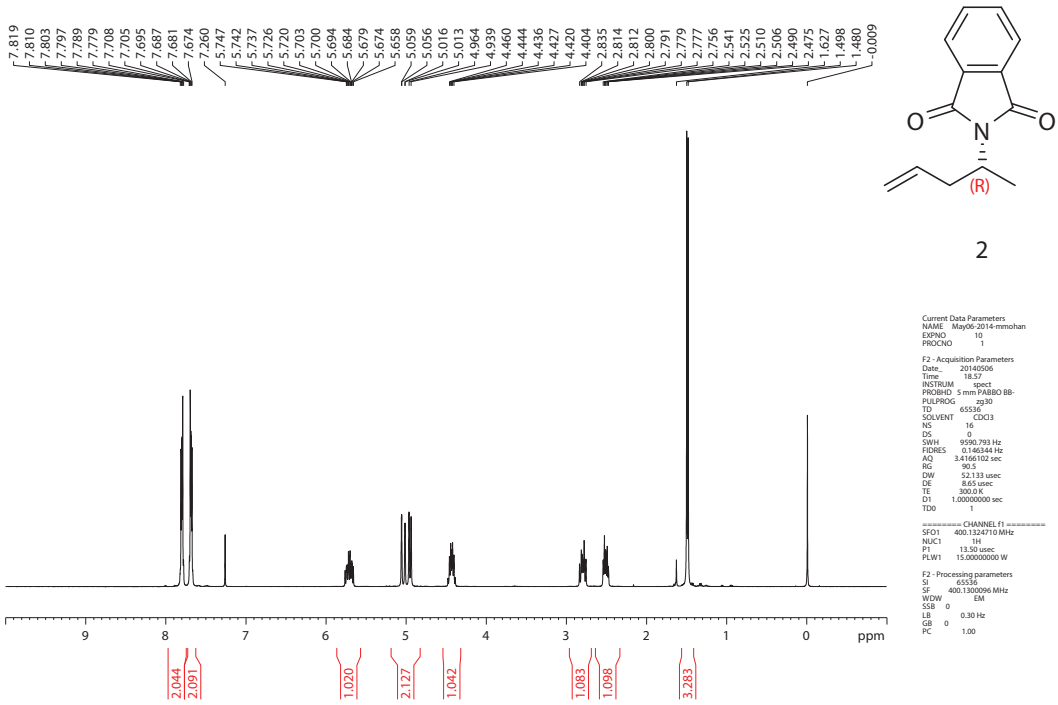
3

4 (Mixture of ~1:1 epimers)

5 (Mixture of ~1:1 epimers)

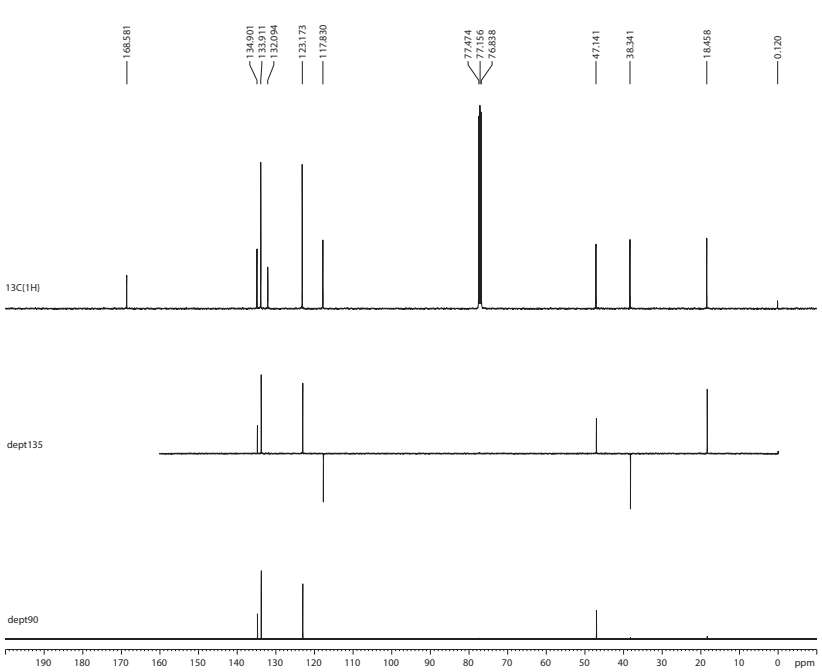
Final purification by Reveleris ® Amino column;
Eluant: n-Hexane / EtOAc (gradient);
Purification using Grace Reveleris ® X2 MPLC system

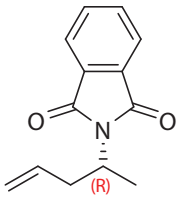
b ^{1}H NMR Spectrum



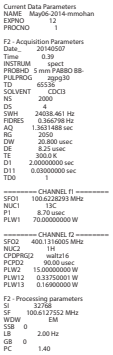
2

13C{1H}, ^{13}C DEPT -135 and ^{13}C DEPT -90 NMR Spectra



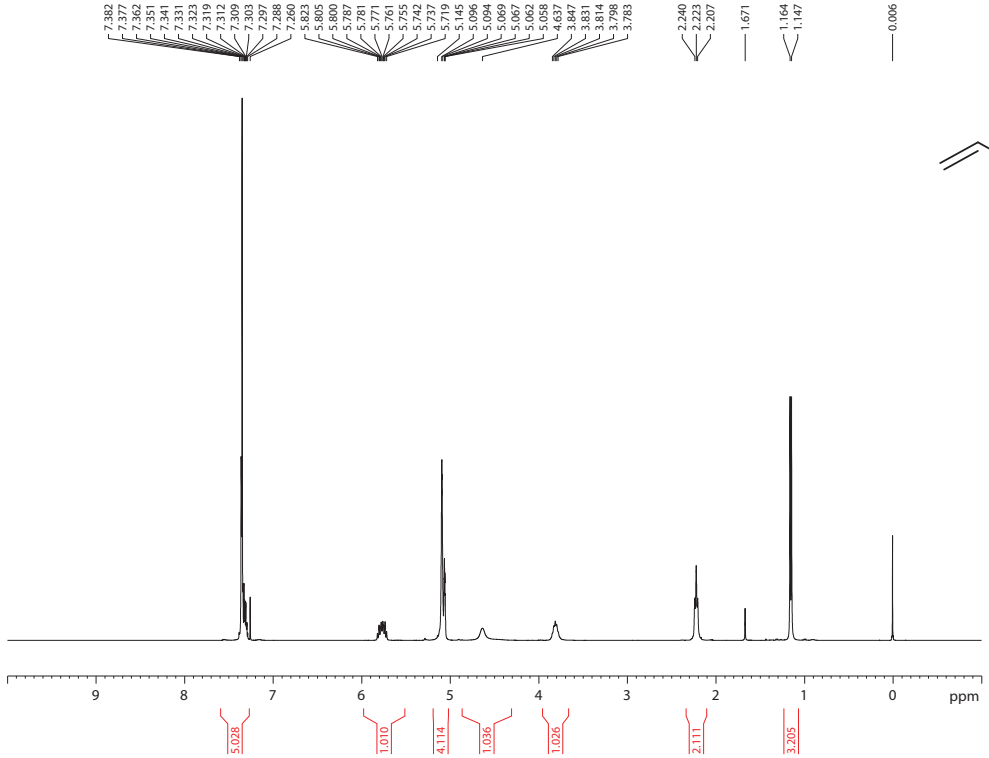


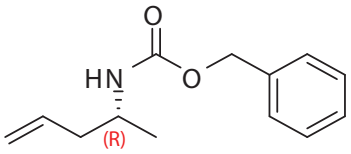
2



c

^{1}H NMR Spectrum





3

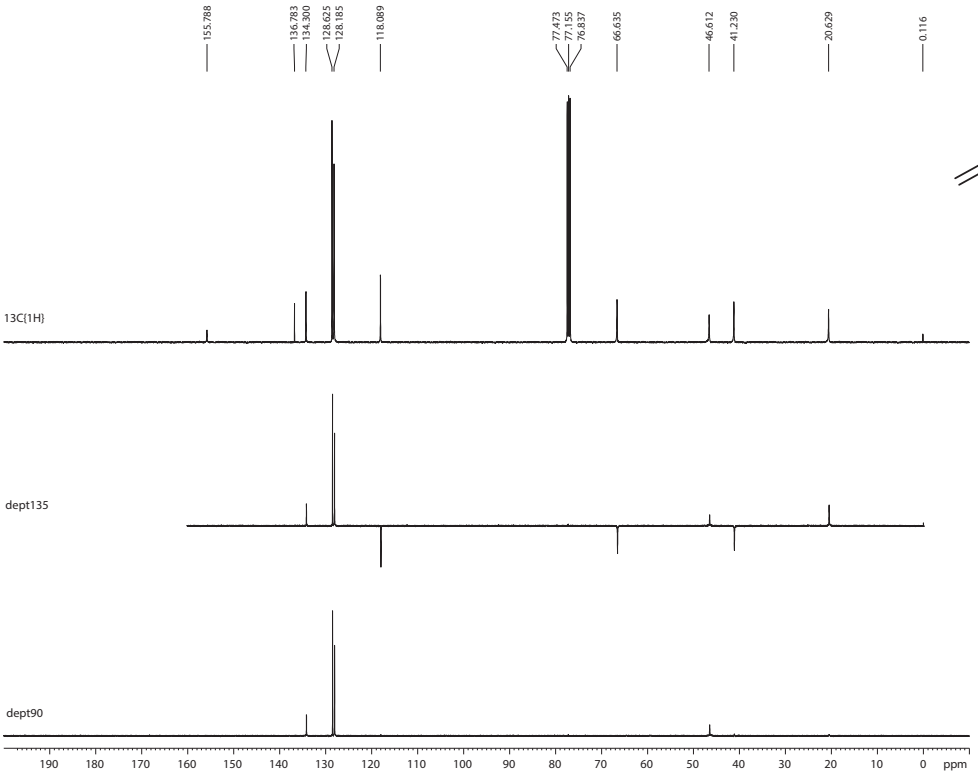
Current Data Parameters
NAME May09-2014-mmohan
EXPNO 10
PROCNO 1

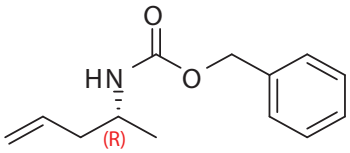
F2 - Acquisition Parameters
Date_ 20140509
Time 4.00
INSTRUM spect
PROBHD 5 mm PABBO BB-
PULPROG zg30
TD 65536
SOLVENT CDCl3
NS 16
DS 0
SWH 9590.793 Hz
FIDRES 0.146344 Hz
AQ 3.4166102 sec
RG 64
DW 52.133 usec
DE 8.65 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1324710 MHz
NUC1 1H
P1 13.50 usec
PLW1 15.00000000 W

F2 - Processing parameters
SI 65536
SF 400.1300096 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00

13C{1H}, ^{13}C DEPT -135 and ^{13}C DEPT -90 NMR Spectra





3

Current Data Parameters
NAME May09-2014-mmohan
EXPNO 11
PROCNO 1

F2 - Acquisition Parameters
Date_ 20140509
Time 5.55
INSTRUM spect
PROBHD 5 mm PABBO BB-
PULPROG zgpg30
TD 65536
SOLVENT CDCl3
NS 2000
DS 4
SWH 24038.461 Hz
FIDRES 0.366798 Hz
AQ 1.3631488 sec
RG 2050
DW 20.800 usec
DE 8.25 usec
TE 300.0 K
D1 2.00000000 sec
D11 0.03000000 sec
TD0 1

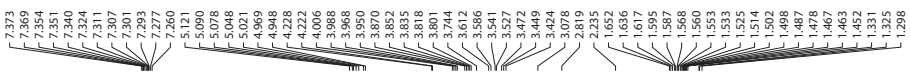
======== CHANNEL f1 ========
SFO1 100.6228293 MHz
NUC1 13C
P1 8.70 usec
PLW1 70.00000000 W

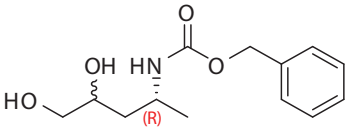
======== CHANNEL f2 ========
SFO2 400.1316005 MHz
NUC2 1H
CPDPRG[2 waltz16
PCPD2 90.00 usec
PLW2 15.00000000 W
PLW12 0.33750001 W
PLW13 0.16900000 W

F2 - Processing parameters
SI 32768
SF 100.6127574 MHz
WDW EM
SSB 0
LB 2.00 Hz
GB 0
PC 1.40

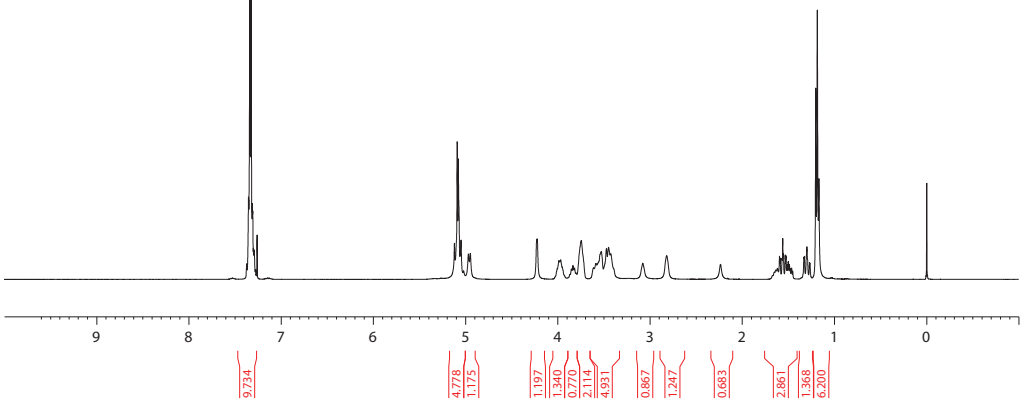
d

^{1}H NMR Spectrum



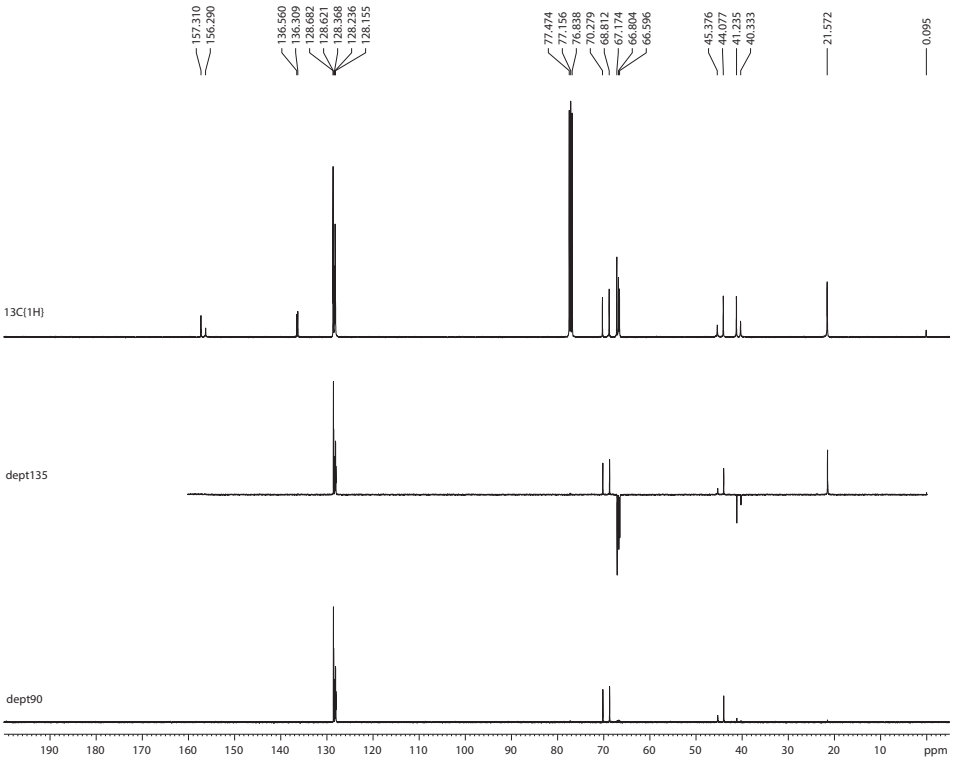


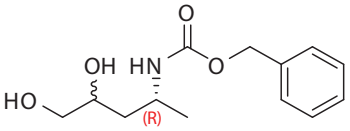
4 (Mixture of epimers)




```
Current Data Parameters
NAME      May21-2014-mmohan
EXPNO             10
PROCNO             1

F2 - Acquisition Parameters
Date_          20140522
Time              2.13
INSTRUM           spect
PROBHD  5 mm PABBO BB-
PULPROG           zg30
TD               65536
SOLVENT           CDCl3
NS                  16
DS                   0
SWH           9590.793 Hz
FIDRES        0.146344 Hz
AQ            3.4166102 sec
RG                50.8
DW               52.133 usec
DE                8.65 usec
TE               300.0 K
D1            1.00000000 sec
TD0                  1

======== CHANNEL f1 ========
SFO1       400.1324710 MHz
NUC1               1H
P1               13.50 usec
PLW1       15.00000000 W

F2 - Processing parameters
SI               65536
SF          400.1300095 MHz
WDW                EM
SSB   0
LB            0.30 Hz
GB    0
PC            1.00
```

13C{1H}, ^{13}C DEPT -135 and ^{13}C DEPT -90 NMR Spectra !



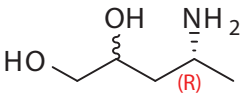


4 (Mixture of epimers)

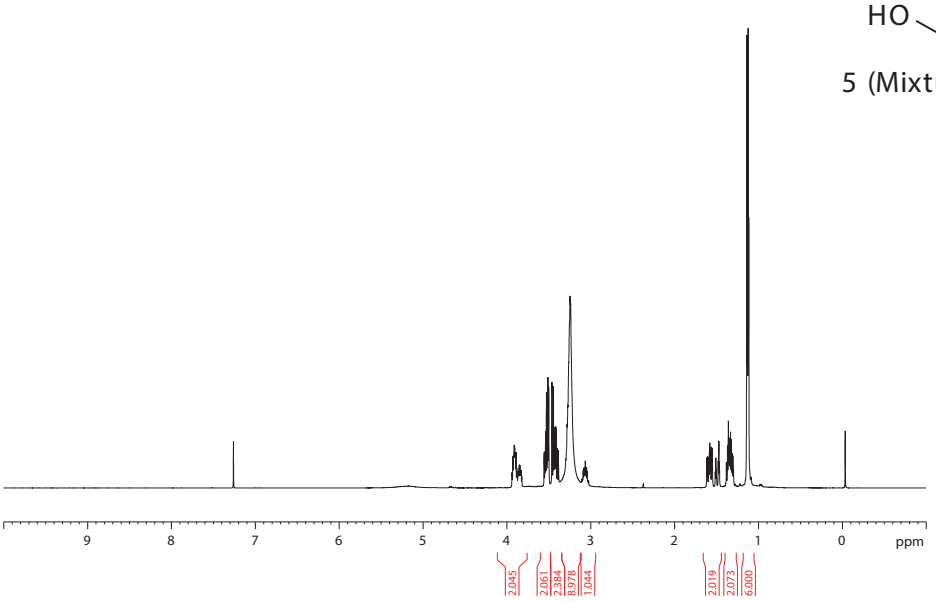
```
Current Data Parameters
NAME      May21-2014-mmohan
EXPNO             12
PROCNO             1

F2 - Acquisition Parameters
Date_          20140522
Time              23.34
INSTRUM           spect
PROBHD  5 mm PABBO BB-
PULPROG         zgpg30
TD               65536
SOLVENT           CDCl3
NS                2400
DS                   4
SWH          24038.461 Hz
FIDRES        0.366798 Hz
AQ            1.3631488 sec
RG                1820
DW               20.800 usec
DE                8.25 usec
TE               300.0 K
D1            2.00000000 sec
D11           0.03000000 sec
TD0                  1

======== CHANNEL f1 ========
SFO1       100.6228293 MHz
NUC1               13C
P1                8.70 usec
PLW1       70.00000000 W

======== CHANNEL f2 ========
SFO2       400.1316005 MHz
NUC2               1H
CPDPRG[2       waltz16
PCPD2          90.00 usec
PLW2       15.00000000 W
PLW12       0.33750001 W
PLW13       0.16900000 W

F2 - Processing parameters
SI               32768
SF          100.6127597 MHz
WDW                EM
SSB   0
LB            2.00 Hz
GB    0
PC            1.40
```

e

^{1}H NMR Spectrum



5 (Mixture of epimers)



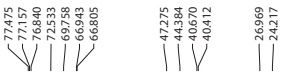
Current Data Parameters
NAME Oct30-2015-mmohan
EXPNO 10
PROCNO 1

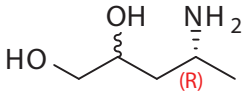
F2 - Acquisition Parameters
Date_ 20151030
Time 5.06
INSTRUM spect
PROBHD 5 mm PABBO BB-
PULPROG zg30
TD 65536
SOLVENT CDCl3
NS 16
DS 0
SWH 9590.793 Hz
FIDRES 0.146344 Hz
AQ 3.4166102 sec
RG 32
DW 52.133 usec
DE 8.65 usec
TE 300.0 K
D1 1.00000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 400.1324710 MHz
NUC1 1H
P1 13.50 usec
PLW1 15.00000000 W

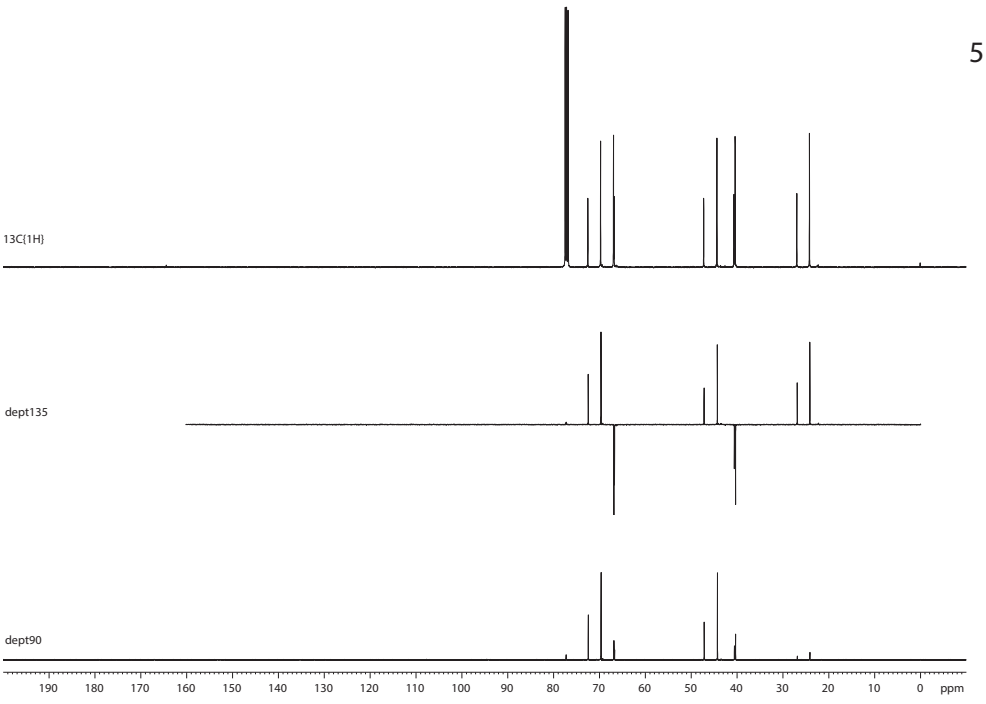
F2 - Processing parameters
SI 65536
SF 400.1300096 MHz
WDW EM
SSB 0
LB 0.30 Hz
GB 0
PC 1.00

13C{1H}, ^{13}C DEPT -135 and ^{13}C DEPT -90 !





5 (Mixture of epimers)



Current Data Parameters
NAME Oct30-2015-mmohan
EXPNO 12
PROCNO 1

F2 - Acquisition Parameters
Date_ 20151030
Time 8.11
INSTRUM spect
PROBHD 5 mm PABBO BB-
PULPROG zgpg30
TD 65536
SOLVENT CDCl3
NS 3000
DS 4
SWH 24038.461 Hz
FIDRES 0.366798 Hz
AQ 1.3631488 sec
RG 2050
DW 20.800 usec
DE 8.25 usec
TE 300.0 K
D1 2.00000000 sec
D11 0.03000000 sec
TD0 1

======== CHANNEL f1 ========
SFO1 100.6228293 MHz
NUC1 13C
P1 8.70 usec
PLW1 73.00000000 W

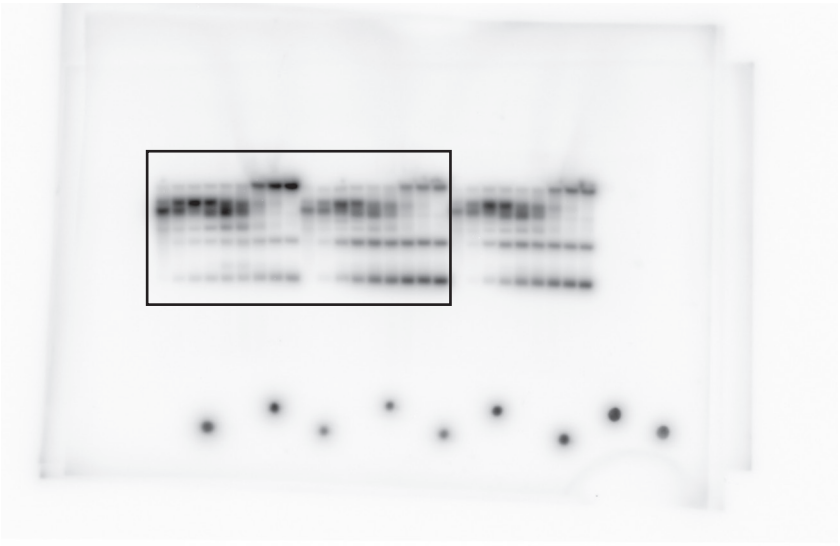
======== CHANNEL f2 ========
SFO2 400.1316005 MHz
NUC2 1H
CPDPRG[2 waltz16
PCPD2 90.00 usec
PLW2 15.00000000 W
PLW12 0.33750001 W
PLW13 0.16900000 W

F2 - Processing parameters
SI 32768
SF 100.6127590 MHz
WDW EM
SSB 0
LB 2.00 Hz
GB 0
PC 1.40

Supplemental note 3 - SI Figure 2

Fig 1c

left panel



right panel

Fig 1e

top panel

bottom panel

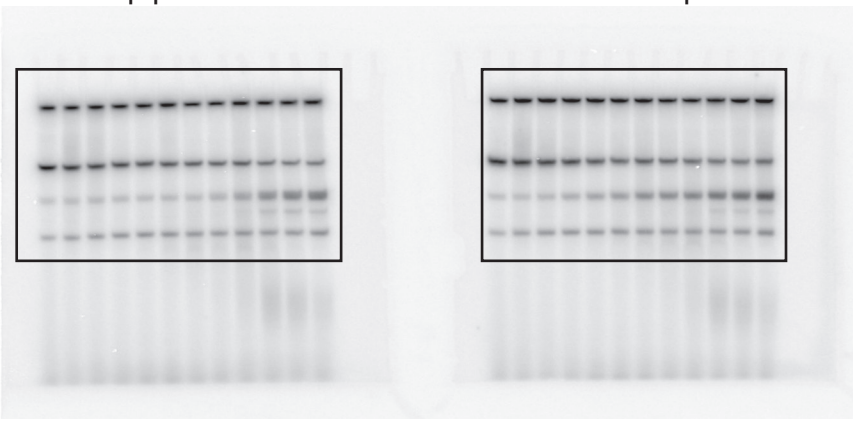


Fig 2c

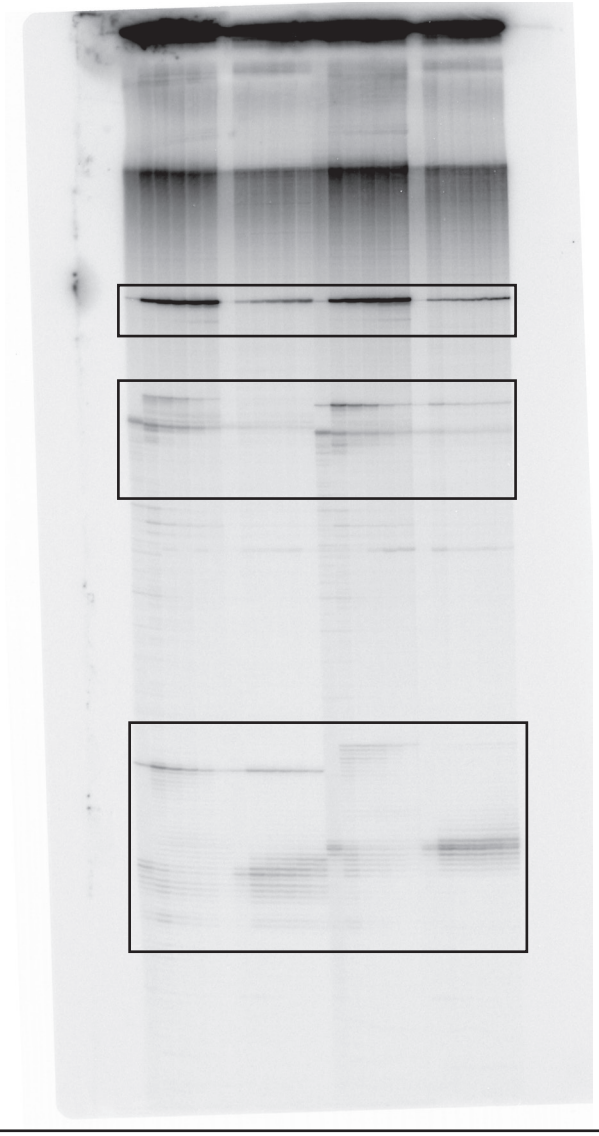
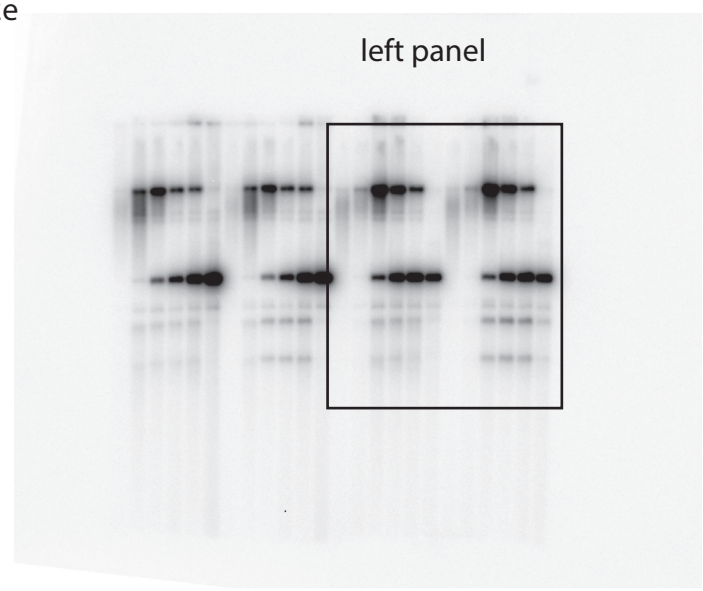


Fig 2e

left panel



right panel

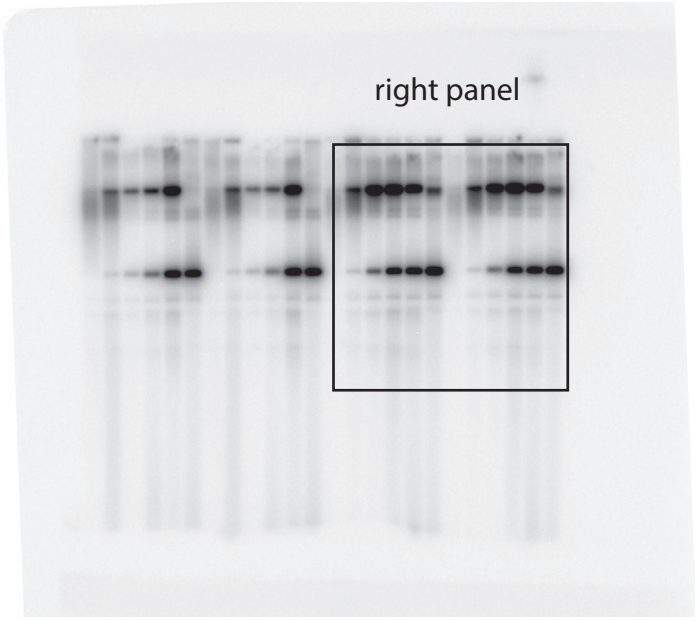


Fig 2f

* indicated samples are switched

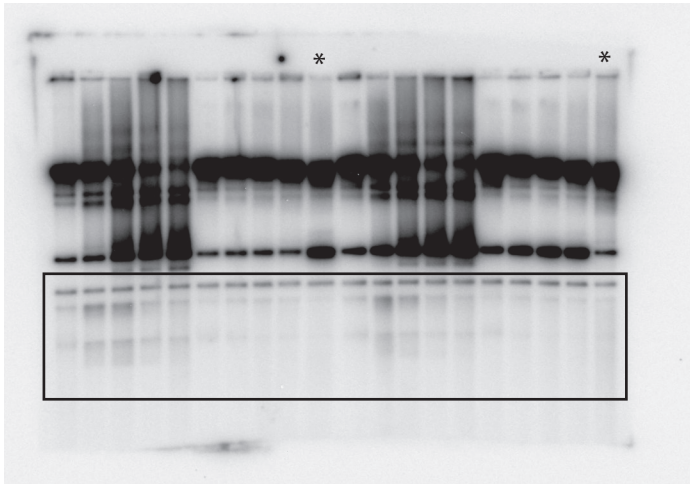


Fig 3b

REV1

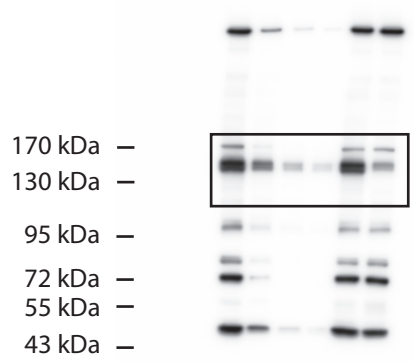


Fig 3e

left panel right panel

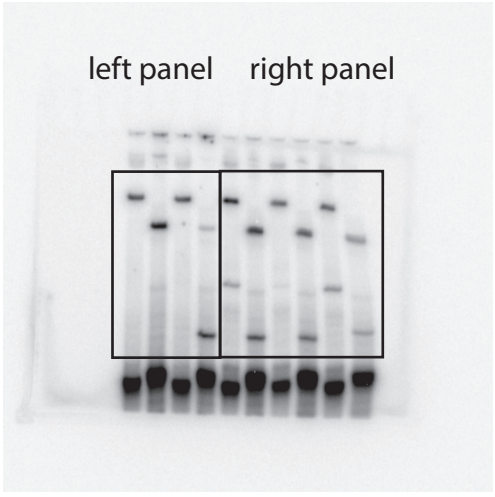
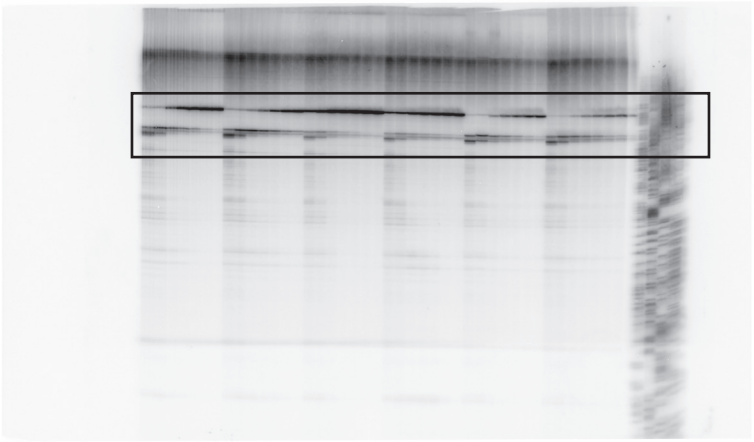
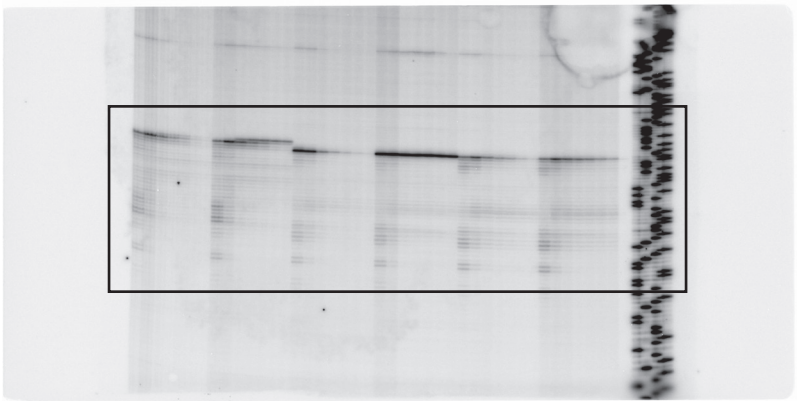


Fig 3c

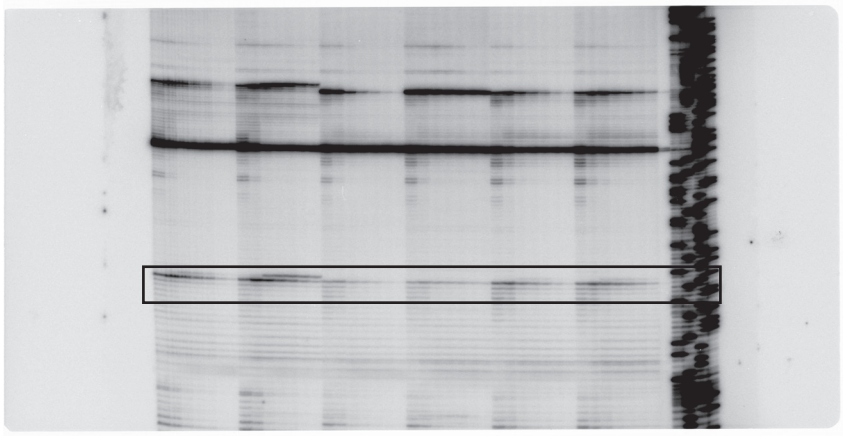
AflIII digestion



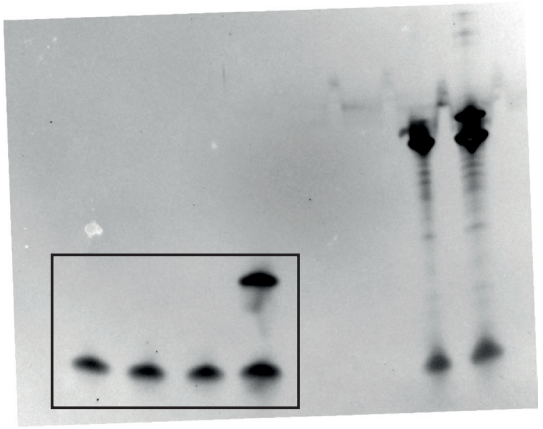
AflIII digestion



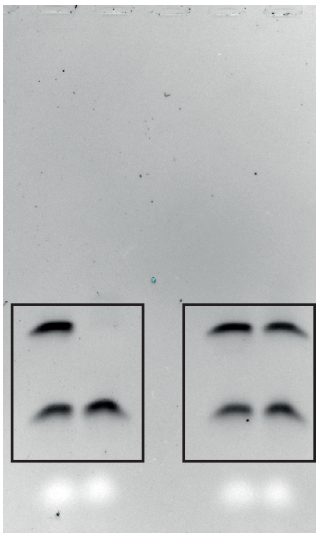
AflIII-BAMHI digestion



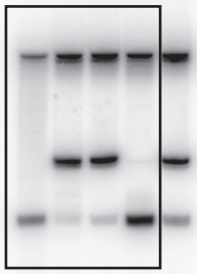
extended data 1c



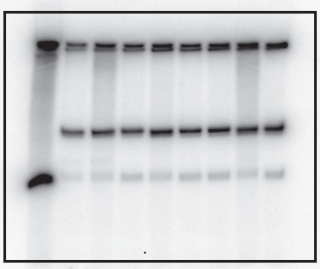
extended data 1f



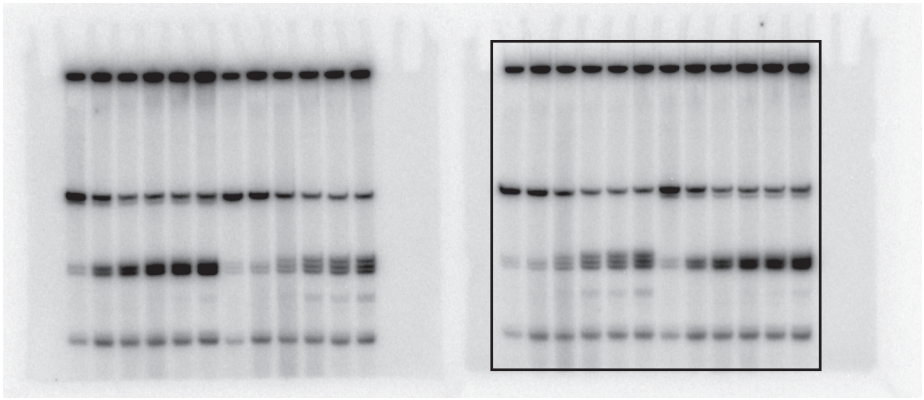
extended data 1h



extended data 1i

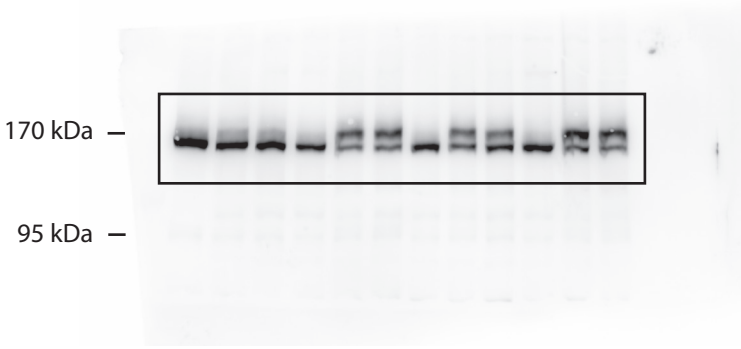


extended data 1k



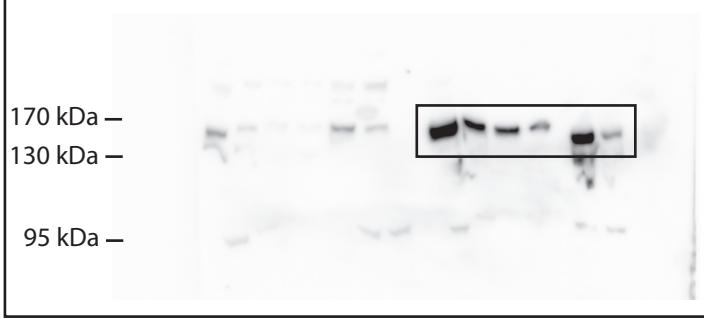
extended data 2b

FANCD2

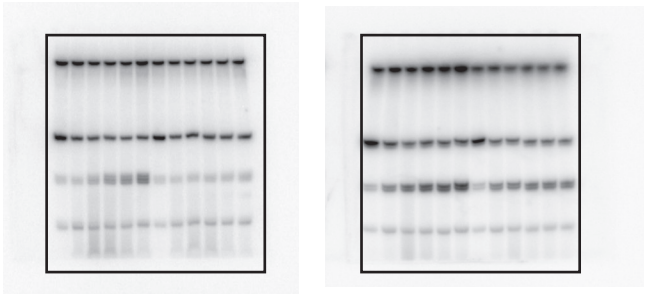


extended data 2c

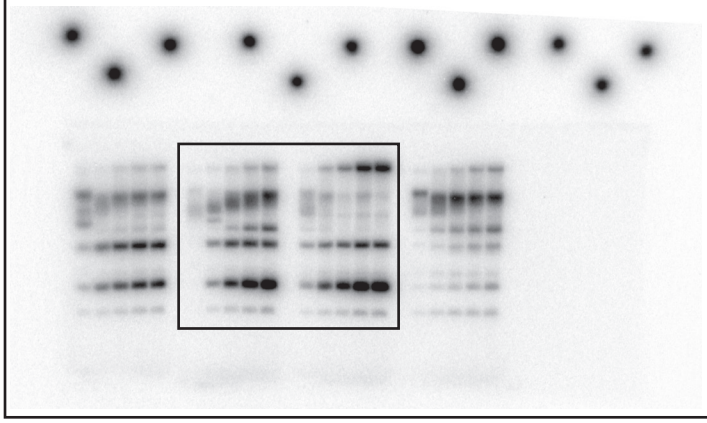
FANCD2



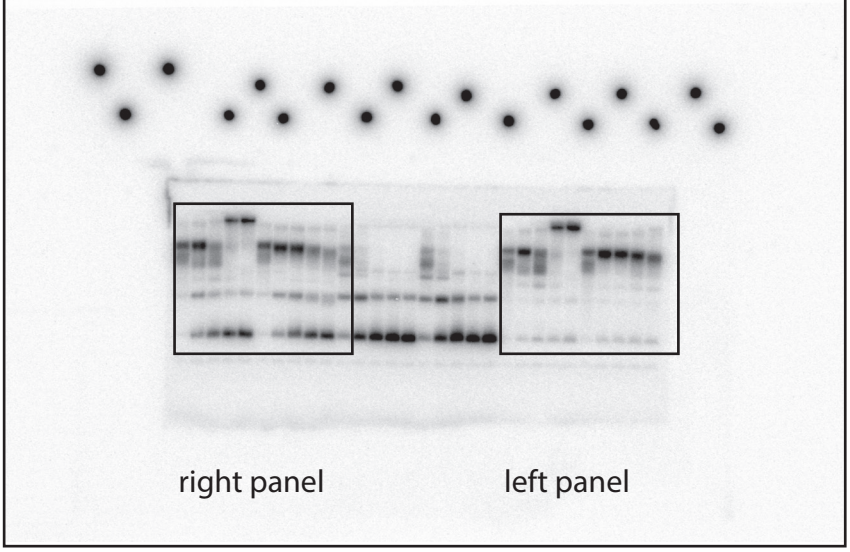
extended data 2e



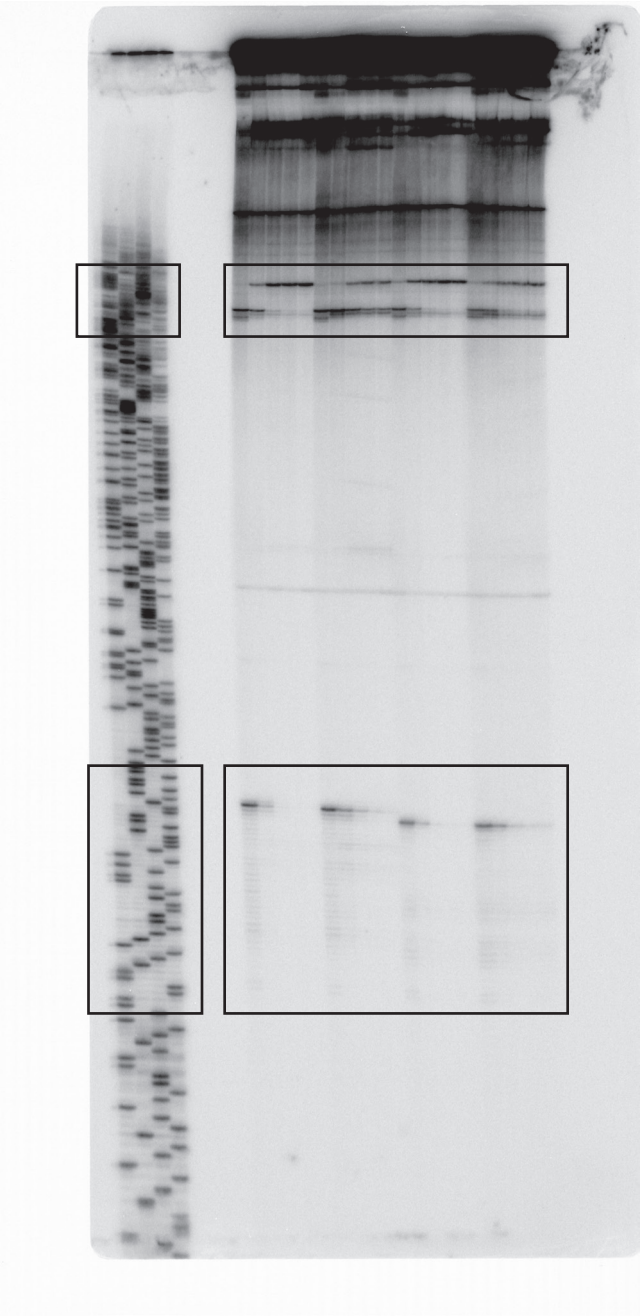
extended data 2h



extended data 2i



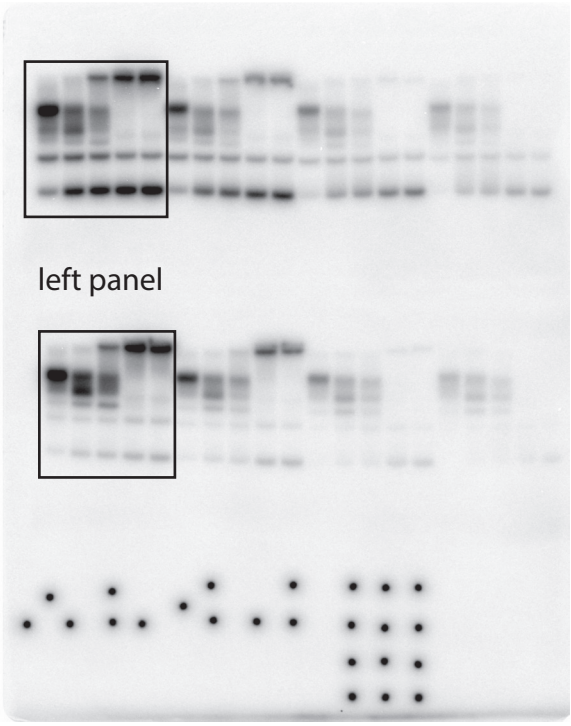
extended data 2d



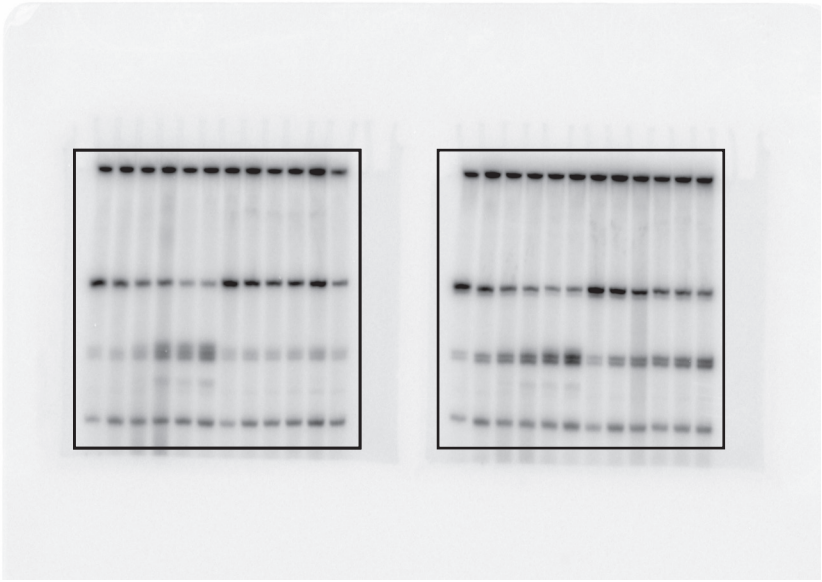
Extended data fig 3b

right panel

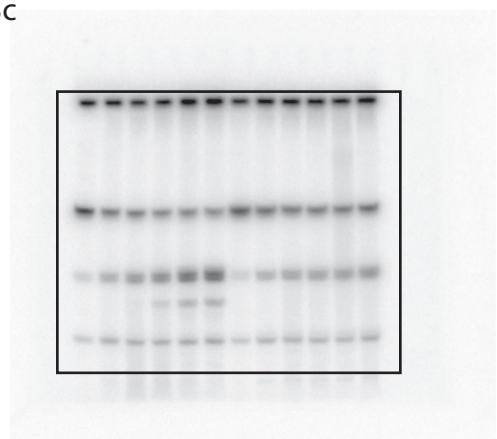
left panel

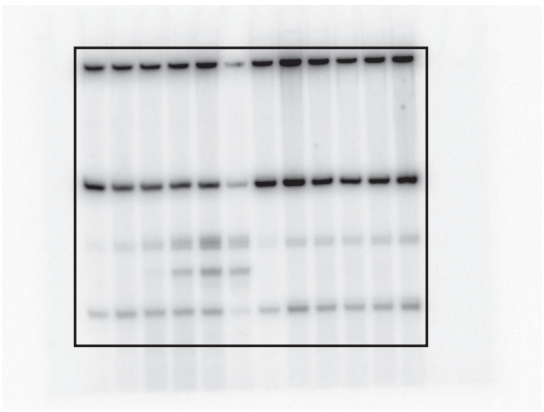


extended data 2j



extended data 3c

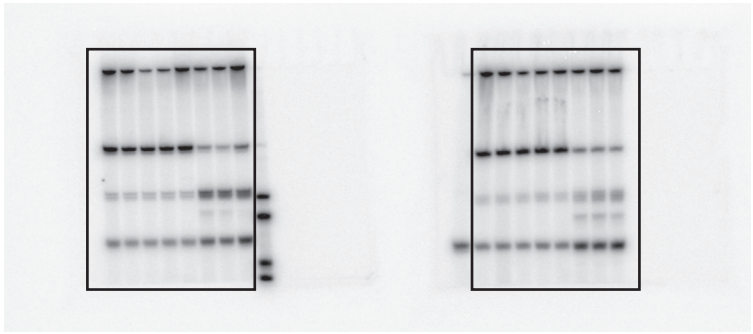




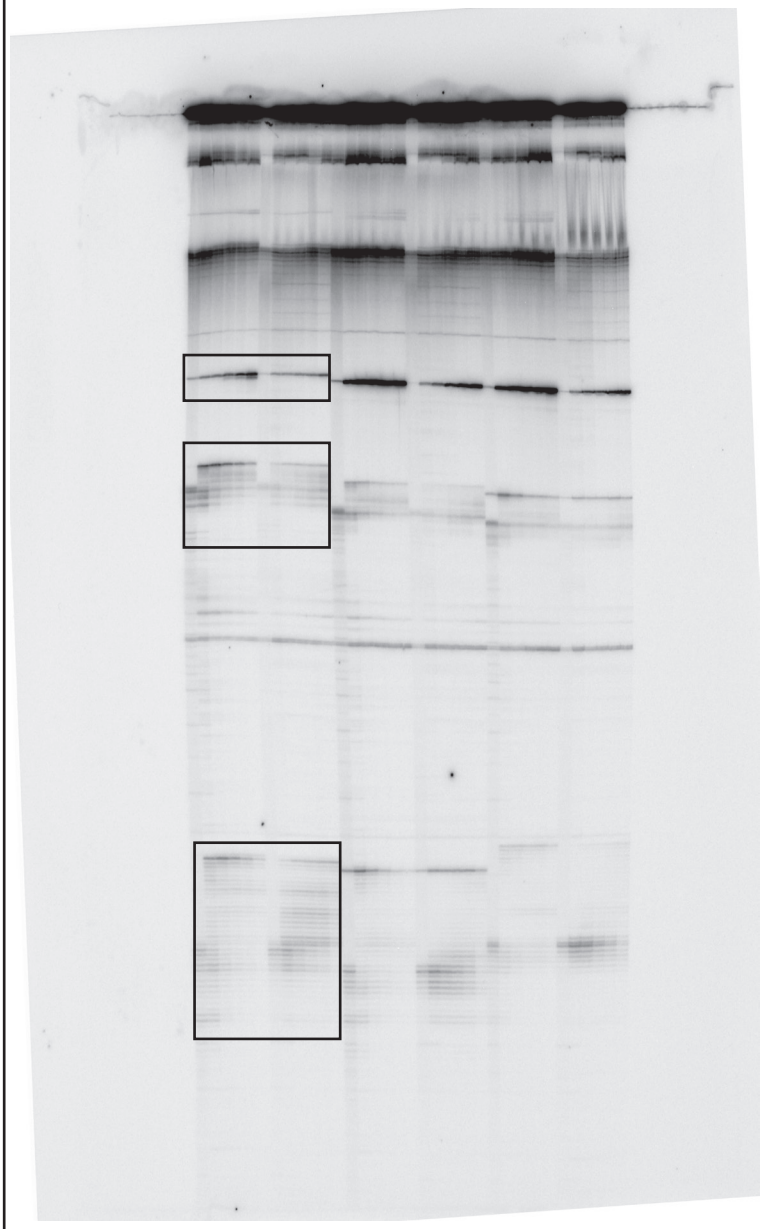
Extended DataFig 4a

right panel

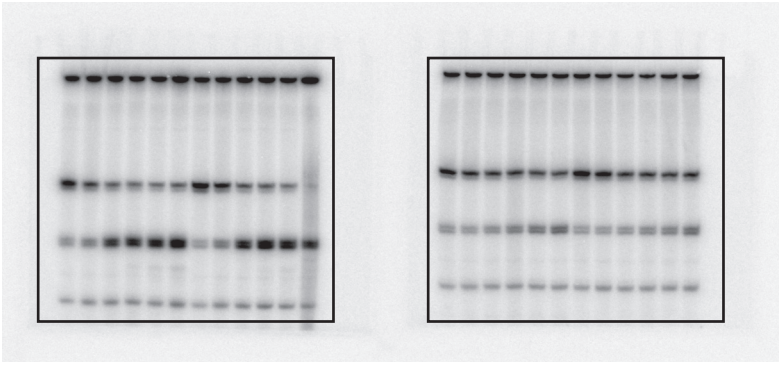
left panel



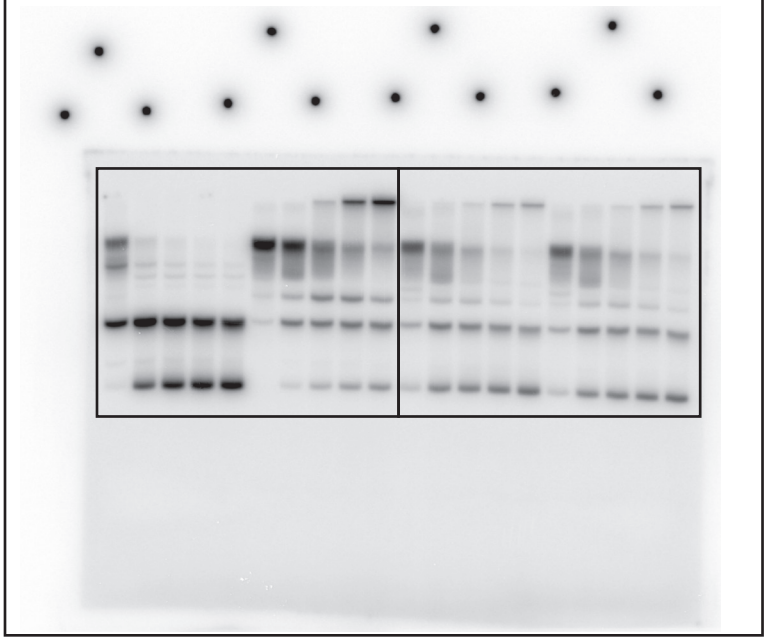
extended data 4d



extended data 4e



extended data 5b



extended data 5a

NEIL3

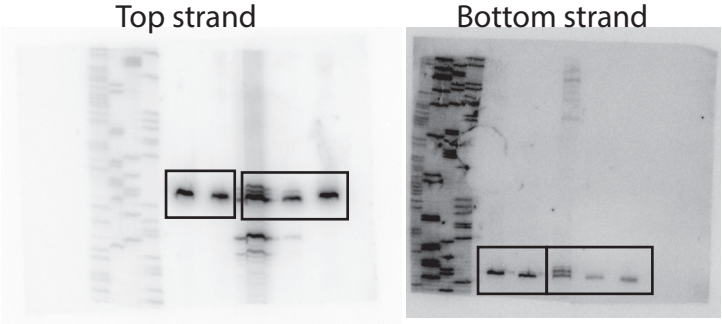
72 kDa

55 kDa

extended data 5k

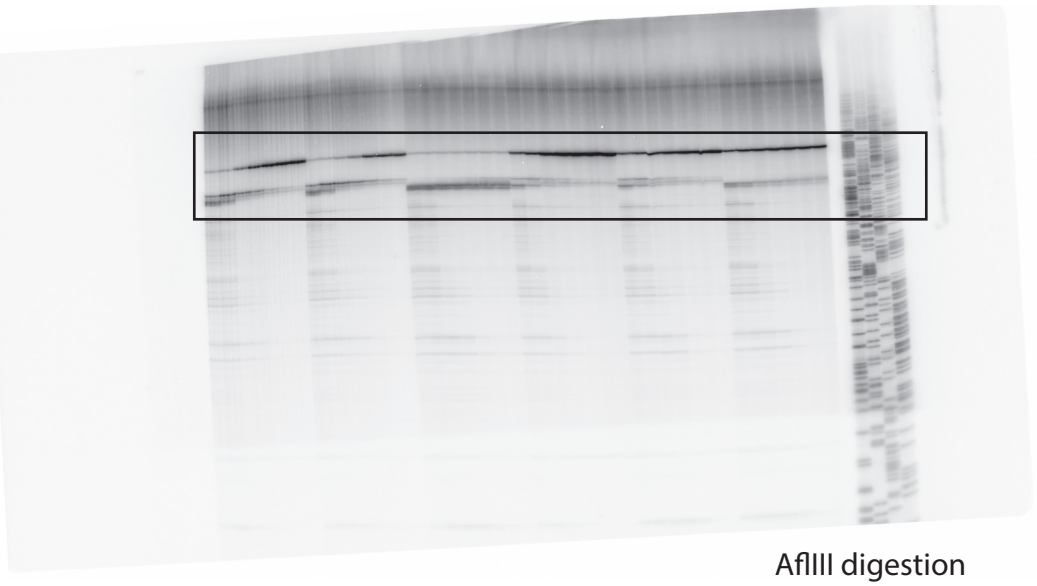
Top strand

Bottom strand

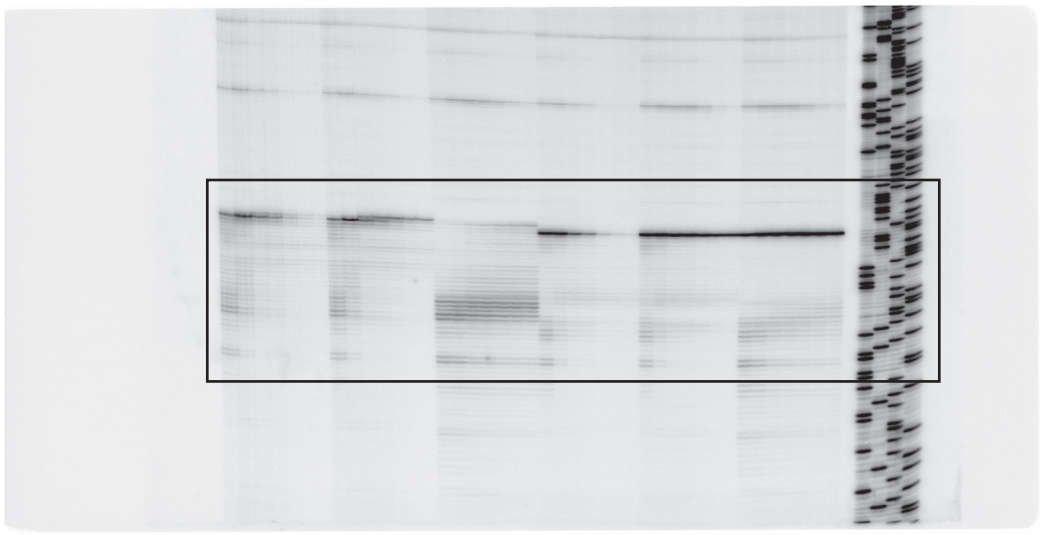


extended data 6A

AflIII digestion

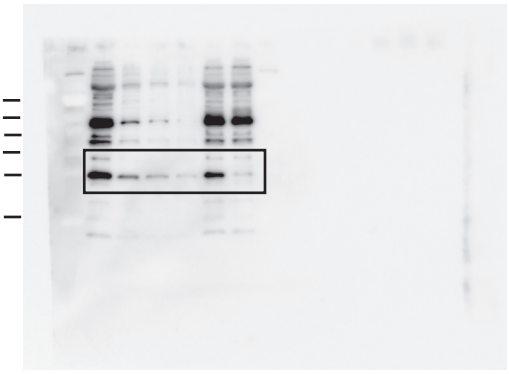


AflIII digestion

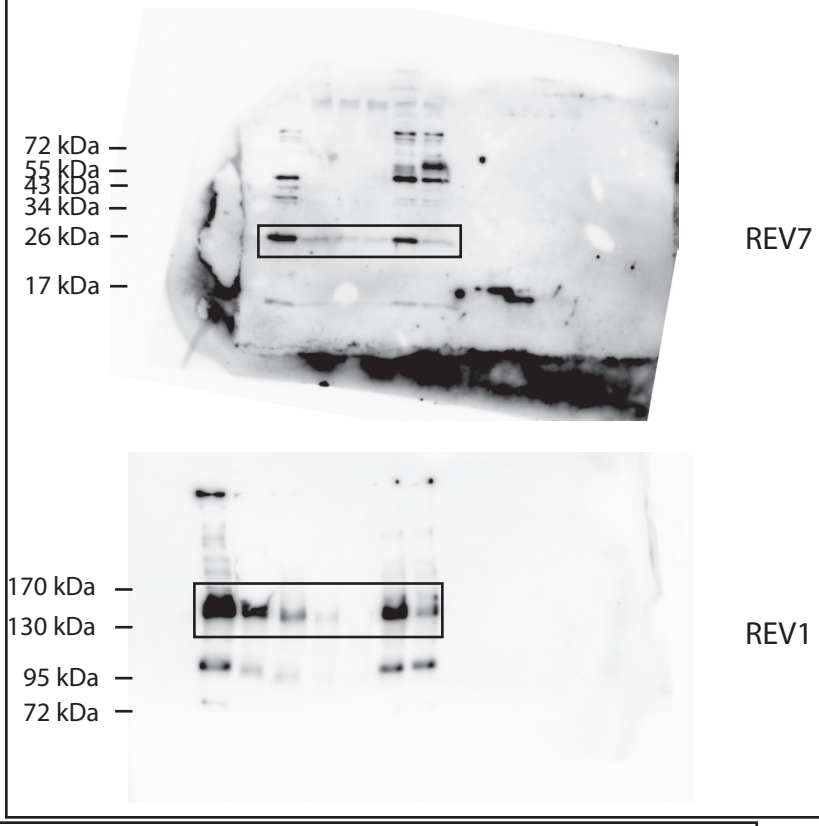


AflIII-BAMHI digestion

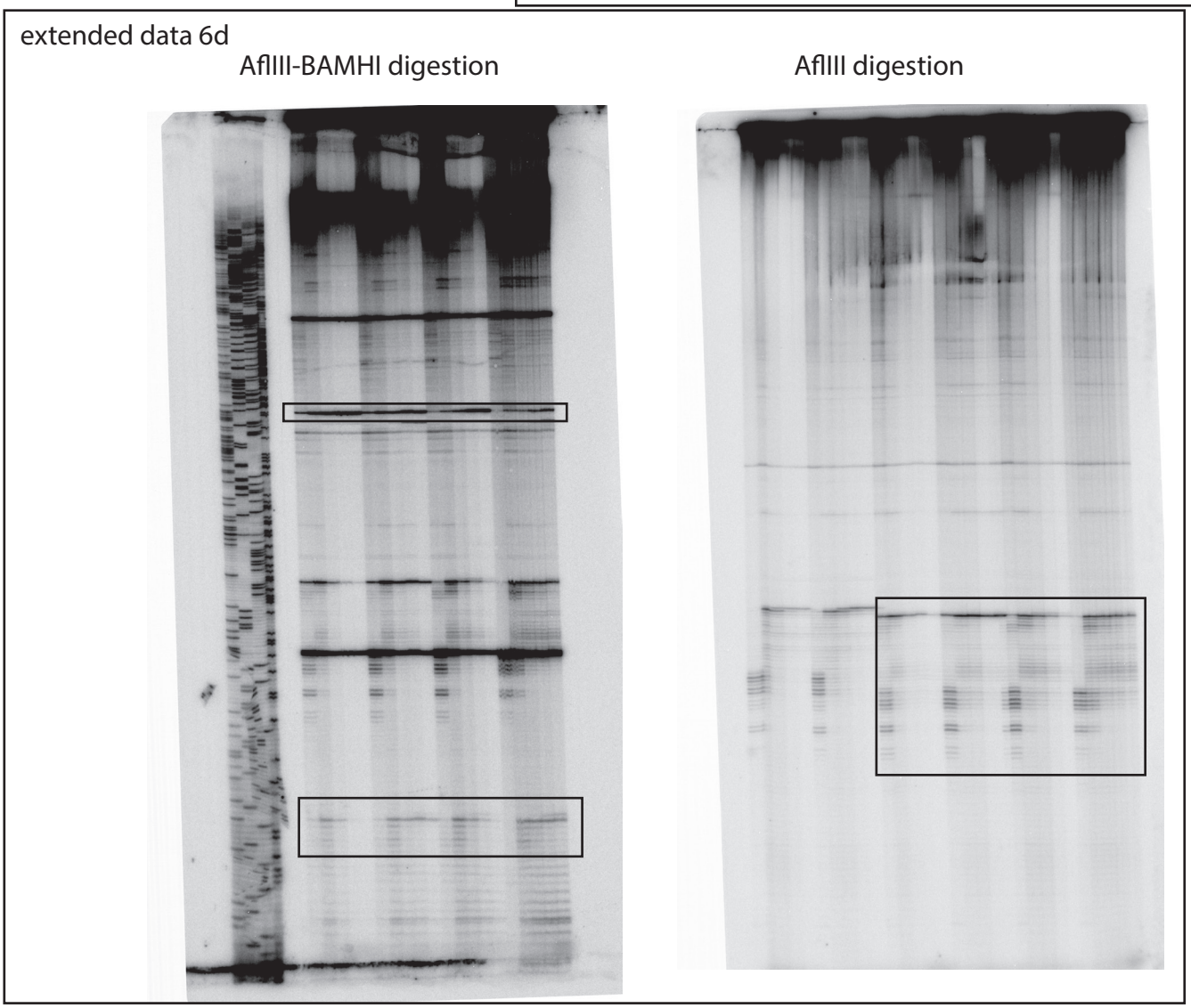
extended data 6B



extended data 6c



extended data 6d



extended data 6e