



Supporting Information

Discovery of New Carbonyl Reductases Using Functional Metagenomics and Applications in Biocatalysis

Sophie A. Newgas, Jack W. E. Jeffries, Thomas S. Moody, John M. Ward,* and Helen C. Hailes* © 2021 The Authors. *Advanced Synthesis & Catalysis* published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Supplementary Information

Discovery of new carbonyl reductases using functional metagenomics and applications in biocatalysis

Sophie A. Newgas,^a Jack W. E. Jeffries,^b Thomas S. Moody,^{c,d} John M. Ward^{b*} and Helen C. Hailes^{a*}

^a Department of Chemistry, University College London, 20 Gordon Street, London, WC1H 0AJ, U.K. E-mail: h.c.hailes@ucl.ac.uk

^b Department of Biochemical Engineering, Bernard Katz Building, University College London, London, WC1E 6BT, U.K. E-mail: j.ward@ucl.ac.uk

^c Almac Sciences, Department of Biocatalysis and Isotope Chemistry, Almac House, 20 Seagoe Industrial Estate, Craigavon, BT63 5QD, Northern Ireland, U.K.

^d Arran Chemical Company, Unit 1 Monksland Industrial Estate, Athlone, Co. Roscommon, Ireland.

Table of Contents

1. Tables

Table S1. The taxonomic and enzymatic assignments of the SDRs retrieved from the metagenome.

Table S2. DNA sequences of the 37 retrieved SDRs generated by sanger sequencing after Gibson assembly into pET29a.

Table S3. Amino acid sequences of the 37 retrieved SDRs.

Table S4. Table of primers used to retrieve SDRs from the oral metagenome.

2. Figures

Figure S1. **A.** Sequence alignments for SDRs 3,4,11,17,31,37. **B.** SDR percentage identity matrix for 3,4,11,17,31,37.

Figure S2. Graph displaying activity with (*S*)-**18** and (*R*)-**18** with SDR-17 and SDR-31 at a range of pHs.

Figure S3. SDS PAGE showing induced recombinant protein expression in *E.coli*.

Figure S4. HPLC calibration curves.

Figure S5. HPLC trace of the products from the reduction of *rac*-**WMK-18** using NaBH₄.

Figure S6. Characterisation data for (4*aR*,5*S*)-**22** from the scaled-up reaction using SDR-17.

Figure S7. ¹H and ¹³C NMR spectra of (4*aR*,5*R*)-**22**.

Figure S8. ¹H and ¹³C NMR spectra of (4*aS*,5*S*)-**22**.

Table S1. The taxonomic and enzymatic assignments of the SDRs retrieved from the metagenome. Assignments were generated by inputting the amino acid sequences of the SDRs into NCBI BLASTP. pQR numbers were assigned to recombinant plasmids for record keeping and identification within the UCL enzyme library.

1	pQR500	3-oxoacyl ACP reductases	<i>Porphyromonas somerae</i>
2	pQR501	3-oxoacyl ACP reductase	<i>Neisseria.sp</i>
3	pQR502	NAD(P)-dependent oxidoreductase	<i>Veillonella sp.</i>
4	pQR503	acetoin reductase	<i>Mycobacteroides abscessus</i>
5	pQR504	NAD(P)-dependent oxidoreductase	<i>Streptococcus parasanguinis</i>
6	pQR505	glucose-1-dehydrogenase	<i>Megasphaera micronuciformis</i>
7	pQR506	oxidoreductase	<i>Veillonella sp</i>
8	pQR507	3-oxoacyl ACP reductase	<i>Streptococcus infantis</i>
9	pQR508	oxidoreductase	<i>Rothia mucilaginosa</i>
10	pQR509	oxidoreductase	<i>Rothia mucilaginosa</i>
11	pQR510	oxidoreductase	<i>Rothia mucilaginosa</i>
12	pQR511	NAD(P)-dependent oxidoreductase	<i>Actinobacteria</i>
13	pQR512	NAD(P)-dependent oxidoreductase	<i>Actinomyces odontolyticus</i>
15	pQR513	oxidoreductase	<i>Atopobium parvulum</i>
16	pQR514	3-oxoacyl ACP reductase	<i>Rothia mucilaginosa</i>
17	pQR515	3-oxoacyl ACP reductase	<i>Streptococcus parasanguinis</i>
18	pQR516	NAD(P)-dependent oxidoreductase	<i>Neisseria mucosa</i>
19	pQR517	oxidoreductase	<i>Streptococcus sp</i>
20	pQR518	NAD(P)-dependent oxidoreductase	<i>Prevotella melaninogenica</i>
21	pQR519	3-oxoacyl ACP reductase	<i>Prevotella sp.</i>
22	pQR520	NAD(P)-dependent oxidoreductase	<i>Prevotella sp</i>
23	pQR521	NAD(P)-dependent oxidoreductase	<i>Prevotella sp</i>
24	pQR522	NAD(P)-dependent oxidoreductase	<i>Streptococcus parasanguinis</i>
25	pQR523	NAD(P)-dependent oxidoreductase	<i>Prevotella pallens</i>
26	pQR524	3-oxoacyl ACP reductase	<i>Neisseria sp</i>
27	pQR525	NAD(P)-dependent oxidoreductase	<i>Prevotella salivae</i>
28	pQR526	NAD(P)-dependent oxidoreductase	<i>Oribacterium sinus</i>
29	pQR527	NAD(P)-dependent oxidoreductase	<i>Veillonella atypica</i>
30	pQR528	enoyl-(acyl-carrier-protein)	<i>Porphyromonas sp</i>
31	pQR529	oxidoreductase	<i>Actinomyces graevenitzii</i>
32	pQR530	NAD(P)-dependent oxidoreductase	<i>Prevotella sp</i>
33	pQR531	3-oxoacyl ACP reductase	<i>Prevotella sp</i>
34	pQR532	NAD(P)-dependent oxidoreductase	<i>Prevotella histicola</i>
35	pQR533	oxidoreductase	<i>Haemophilus parainfluenzae</i>
36	pQR534	NAD(P)-dependent oxidoreductase	<i>Prevotella melaninogenica</i>
37	pQR535	oxidoreductase	<i>Haemophilus influenzae</i>
38	pQR536	3-oxoacyl ACP reductase	<i>Veillonella sp</i>

Table S2. DNA sequences of the 37 retrieved SDRs generated by sanger sequencing after Gibson assembly into pET29a. Sequences in italics originate from pET29a MCS, coding for 20 amino acids plus the hexa histidine tag.

SDR #	DNA sequence
1	<p>ATGAATCTACTTGCTAACAAGGTAGCCATCATCACCGGCGCAGGCCGTGGTATCGG CCGCGCTATTGCACTCAAGTATGCACAGGAAGGTGCTTCCGTAGTGATCACTGACCT CAAGATCGACGAGACTGTAGAAGCCTTCGTCAAGGAGCTCGAAGGGCTCGGTGTCA AGGCTAAGGCTTATGCCTCGAACGCAGCTAACTTCGAAGATGCTCACAAGCTCGTC GAAGCAGTCGTCGACACTTCGGCCGTATCGACGTCCTCGTCAACAACGCTGGTAT CACCCGTGACGGGCTGATGATGCGTATGACCGAAGAGCAGTGGGATCTCGTCATCA ACGTCAACCTCAAGAGTGCCTTCAACCTCATCCACGCTGTCACCCCGTTATGGTCA AGCAGCGTAGCGGTAGCATCATCAACATGGCCAGCGTCGTGGTGTCTCTGGCAAC GCAGGTCAGGCGAACTACTCCGCTTCTAAGGCTGGTATGATCGGTCTGGCAAAGAG CATCGCAAAGGAGCTCGGCGCTCGTGGTATCCGTGCCAACGCTATCGCTCCTGGCTT CATCATCACCGATATGACGGGTGCACTCTCTGAAGAAGTTCGCAAGCAGTGGGAAG TACAGATCCCCCTCCGTGCGTGGCGGTACGCTGAGGACGTAGCTAACGTAGCTACC TTCCTCGCAAAGCGACCTCTCCAGCTACGTCTCCGGTCAGACGATCCACGTCTGCGGT GGGATGAATATGCCATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGC GGCCGCACTCGAGCACCACCACCACCA</p>
2	<p>ATGAGTACACAAGATTTGAGCGGCAAAATTGCTTTGGTAACCGGCGCATCGCGCGG TATTGGTGCGGCGATTGCCGATACTTTGGCTGTGGCCGGTGCAAAAGTTATCGGCAC GGCCACCAGCGAAAGCGGCGCGGCTGCAATTAGCGAGCGTCTGGCGCAATGGGGC GCGAAGGCCGTGCATTGAATTCGCGCCGAACCCGAAACCATCGAAAACCTGATCGC CGACATCGAAAAAGAGTTCGGCAAGCTGGATATTCTGGTCAACAACGCCGGTATCA CCCGCGACAACCTCTTGATGCGTATGAAAGAAGAAGAGTGGGACGACATCATGCAG GTCAACCTCAAATCCGTGTTCCGCGCCTCCAAAGCCGTCTTGCGCGGCATGATGAA GCAACGCGCCGGCCGCATCATCAACATCACATCCGTTGTCGGTGTGATGGGCAATG CCGGTCAAACCAACTATGCCGCGGCAAAAGCGGGCTTAATCGGTTTCTCCAAATCC ATGGCGCGTGAAGTCGGCAGCCGCGCATTACCGTCAACTGCGTCGCCCCAGGCTT TATCGACACCGATATGACCCGCGCCCTGCCGGAAGAAACCCGCAAAACCTTTGAAG CGCAAACCTTCTTTGGGCAAATTCGGTGAAGCGCAAGATATTGCTGATGCAGTCTTGT TCTTGGCTTCCGATCNGCAAAATACATTACCGGTCAGACACTTCATGTCAACGGCGG CATGTTGATGCCTCCATGGCTGATATCGGATCCTGAATTCCAGCACACTGGCGGCCG TTAGTGGNTCCNANCTCGGTACCAAGCTTGATGCATAGCTTGAGTANTNCTAAC GCGTCNCCTAAATAGCTTGGNGTANCATGGNNATAGCTGTTTNC</p>
3	<p>ATGGCTCATAATATTTTTGTCAGTGGTGCAACGCTCGGTATCGGTCTTTGTATTGCTG AGGCTTATGCAAAGCATGGCGATAATGTGTTGATTTCTGGTCGTCGTGCTGAGTTAT TGGGCGAGGTACAGGCTCGTTTGTCTAAGGAATATGGCGTGCCTGTTGAGACTTTA GTTCTTGATGTGCGTAGTCGCGAGGATGTTGAAAGCAAGGTTCTGCAGCTATCGA GGCTTTTGGTGGCGTCGATGTGCTCGTTAATAATGCGGGTCTTGACAAGGGCTTGA TCCTTTCCAAGATAGTGCTGTTGATGATGCGGTGACTATGATTGATAACCAATGTGAA AGGCCTTTTATATGTAACAAAAGCAGTGCTACCTTTTATGATTGATAAAAATGAAG GTCATATTGTAATAATGGGTTCTACTGCAGGCATTTATGCATATCCTAATGGCGCAG TGTACTGTGCTACAAAGGCGGCGGTAAAGACATTAAGCGATGGCATTTCGCATGGAT ACCATAACTACAGATATTAAGGTTACTACCATTCACCAGGCATCGTAGAAACTCC ATTCAGTGAAGTACGCTTCCACGGTGATGCGGAACGAGCTAAATCTGTGTACGCTG GTATTGATGCAATCCAACCAGAGGATGTGGCCGATGTTGTAATAATGTAACAAAC CAACCTAAACGCTTGCAAATCTCTGATGTAACCATCATGGCGAACCAACAAGCGGC AGGCTTTATGGTGGCATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTG CGGCCGCACTCGAGCACCACCACCACCA</p>
4	<p>ATGTCTAAAGTAGCTATTGTTACAGGTGCAGGTCAAGGAATCGGTTTTGCAATCGC AAAACGCTTGGTCCAAGATGGTTTTAAGGTTGGAGTATTAGACTACAATGCTGAAA</p>

CAGCCGAAAAAGCAGTTGCTGAATTGTCAGCAGACAAGGCTTTTGCTGTTGTGGCT
GATGTGTCTAAACAAGCAGAGGTTGCTGCAGCTTTCCAAAAAGTTGTAGACCATTT
CGGAGACTTAAACGTTGTGGTAAACAACGCTGGTGTGCTCCGACTACACCTCTAG
ATACGATTACAGAAGAACAGTTCAATCTACCTTTGCTATTAACGTCGGTGGGGTAAT
CTGGGGAGCACAAGCTGCTCAAGCTCAATTCAAAGCACTTGGTCACGGTGGTAAAA
TCATCAATGCTACTTCTCAAGCAGGTGTTGTCGGAAATCCAAACTTGACTGTTTATG
GTGGTACAAAATTCGCTGTTTCGTGGTATTACTCAAACATTGGCGCGTGATTTAGCTG
ACTCAGGCATCACTGTTAATGCGTATGCACCTGGTATCGTTAAGACTCCAATGATGT
ATGATATCGCTCACGAAGTTGGGAAAAATGCAGGAAAAGACGACGAATGGGGTAT
GCAAACATTCGCAAAAGATATTACATTGAAACGCCTATCTGAGCCAGAAGATGTGCG
CTGCAGCGTTAGCTTCCCTGCTGGACCAGATTCAAACCTACATTACAGGTCAAACCA
TTATCGTCGATGGTGGTATGCAATTCCATCCATGGCTGATATCGGATCCGAATTCGAGC
TCCGTCGACAAGCTTGC GGCCGACTCGAGCACCACCACCACCAC

5 ATGTCAGAAACGATTTTAGTAACAGGAGCTTCAGCTGGTTTTGGTCAAGCGATTTGC
CGTCGCTTAGTAGCAGATGGATACCGTGTGATCGGATCAGCTAGACGCATCGATAA
ATTACAGGCACTTCAAGAAGAGCTGGGAGAAGCCTTTTATCCCCTGCAAAATGGATG
TGACGGATCTTTCTCAGGTAGATCATGCACCTGCCAGTTTGCTAAAAGCTTGGGAGA
AAGTGGATGTTTTGGTCAATAATGCTGGCTTGGCTCTAGGCCCTCGCCCCAGCTTATG
AAGCAGAGGTGCGCAGACTGGCTGACCATGATTCAGACCAATATTGTCGGCTTGACC
TATCTGACAAGGAAAATCTTGCTCAGATGGTGAACGAAATGATGGCTATATTAT
CAATTTGGTTCTACAGCAGGAACGTGCCTTATCCAGGGGCCAATGTTTACGGGG
CATCCAAGGCTTTGTCAAGCAATTCTCCCTCAATCTTCGGGCGGATCTAGCTGGCA
AGAAGATTCGTGTCAGCAATATTGAACCTGGTCTTTGCGAAGGGACAGAGTTCTCTT
CTGTTTCGCTTTAAAGGAGACGAAAAACGGGTAGAAGCCCTCTATCGAGATGCCCAT
GCCATTCAGTCTGAAGACATTGCCAACACTGTAGCTTGGTTGATCCAACAGCCCAA
GCATGTCAATGTCAACCGGATTGAAATCATGCCGTTTCCCAAACCTTTGGTCTCA
ACCCGTTTTTCGTCCATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGC
GGCCGACTCGAGCACCACCACCACCAC

6 ATGTACAGTGAATTAAGGAAAGGTAGCCGTCATTACAGGTGGATCTAAGGGAAT
CGAACGGCTATCGCCAAGCGTTTCGGGCAAGAAGGCATGAAGGTTGTCATTAAT
ATAATCCGATGCAGCCGGTGCAGAGTTGGCTGCAGAAGCGGTTTCGATGTGCCGGC
GGAGAGGCGTTACTGTCAAGGCTCATGTGCGTACGGAAGAAGGCGTGCAGTCTCT
GGTCGATGCGGCTGTAGAAAACACTACAGCGGTATCGACGTATGATTAACAATGCGG
GCATGGAAAACAAGGTGGCGACTCATGAAATGCCCTTAAGCGATTGGGAACGAGTT
ATTAACGTTAATTTAACAGGTGTATTTCTCGGTACACGTGCGGCTTTGACATATTTT
ATGGAGCATGATGTAAAGGGAAGTATTGTAAATATGTCATCAGTTCATGAGCAGAT
TCCTTGGCCGACCTTTGCGCATTACTGTGCATCCAAAGGCGGTATTAACCTTTTAC
GCAAACCTGTGGCCATGGAGTATGCCAAGTACGGAATTCGTGTAATGCCGTCGGTC
CCGGAGCTATTAATACGCCTATTAATGCGAAAAAATTCTCCGACCCCGTTCAGTATG
AAGAACTATGAGTATGGTTCCCATGAAGCGTATCGGTAAGCCTGAAGAAGTGGCA
GCTTGTGTGGCTTGGCTTCCGATGAAGCGTCATACGTAACAGGAATCACCCCTT
TTGTAGACGGCGGCATGACCCTGTATCCGGCTTTTCAAGACGGTAAGGGTCCATGG
CTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGC GGCCGACTCGAGCACC
CCACCACCACCAC

7 ATGCCTCATAAAAATGATGTTCAAGTGGCATTGATCAGCGGTGGTACATCGGGAAT
TGGATTTGCCACAGCTAAGCTATTATTGCAAGAAGGTTGGTGCCTGTTATCAATGG
CCGCGATGAACAAGCGGGACAAAGGGCAAAGATGAAATTGCGCCGATATTCTTCG
AAGGTGCGCTATGTGAAAGGCGATGTATCCAGTGTAGCGATTGTCAGCGCATCGT
AAAAGAAACGGTAGATTTCTTTGGTAGTATTTTCAGCTCTTGTAAACGGCGGCTGGATA
TTATGAAGAAGAGCTACTAGCCGATGTATCTGAATCAGCCTTTGATGAAATGTTTGG
AACAAATGTAAAAGGAACTGTATTCTTGTGCCAAGCAGCATTGCCTATTTACGTCA
AGTAAAGGGCAGTATTGTTACCGTATCTAGTGATGCAGGTCTACAGGGCAATGTAG
CCTGCTCAGTATATGGTGCCTCAAAGGTGCTATCGTAAGCTTTACAAAATCACTGT
CCCTTGAGATGGCGCCCCATGAGGTACGCGTTAACTGCGTGTGCCCTGGAGATGTA

GATACATCGCTTGTAGATAAGCAAATTGCCAATCCCATCAAGATGCGGAACAAGC
AAAAGAGGAGATGGGCCAACATTATCCGTTAGGACGTATTGCAAAACCTCATGAAA
TTGGCGAAGTTATCGCTTTTTTAATTAGTTCAAAAGCATCCTTTGTAACAGGTGCTG
CATGGACGATTGATGGAGGTCTTACGAGTCCATGGCTGATATCGGATCCGAATTCGAG
CTCCGTCGACAAGCTTTCGGGCCACTCGAGCACCACCACCACCA

8 ATGACTAAACGCGTCTTAGTTACAGGTGTTAGCTCAGGGATTGGTCTTGCACAAGC
ACGTCTCTTTTTAGAAAATGGTTATCAAGTTTATGGGGTGGATCAGGGTGAAAAAC
CTGACTTGCAAGGTAACCTTCACTTTTTACAACGTGATTTAACACTAGATTTAGAGC
CTATTTTACTGGTGTCTCAAGTTGATATTCTGTGTAATACTGCAGGAATTTTGG
ATGACTACAAATCTTTACTGGAACAATCAGCCCAAGAAATCCAAGAGATTTTGGAA
ATTAATATGTGACTCCAGTAGAGTTGACTCGCTATTATTTGACTCAAATGTTGGAA
TACAAGCGAGGAACCATCATCAATATGTGTTCCATTGCTTCAAGTCTGGCAGGCGG
AGGTGGGCACGCCTATACCTCTTCTAAACATGCCTTAGCAGGCTTTACCAAGCAATT
AGCTCTAGACTATGCTGAAGCTGGGATTACAGGTCTTTGGCATTGCTCCTGGTGC
TAAGACTGGTATGACAGCAGCTGACTTTGAACCAGGTGGTCTGGCAGACTGGGTAG
CCAGTGAAACACCTATCAAACGCTGGATTGAGCCAGAAGAAGTGGCAGAAGTCAG
CCTCTTTTTAGCAAGTGGAAAAGCAAGCGCCATGCAAGGACAAATCTTGACTATTG
ATGGTGGCTGGTCTTTGCCATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAG
CTTTCGGGCCACTCGAGCACCACCACCACCAC

9 ATGAGCTGGGTCATTGTGACCGGCGCGAACGGCGGTATTGGTGAGGCAACCGTCCA
CCAGCTCATCAAGAACGGCTATTCGGTGTTCGGGCTGACCTTTCGGGAGCAGCCCA
TTGCCCTCTTCGGTACCTACGGGGACGCAATCTTCGCTACCGCGCCGTAGACGTTA
CGAGCGAAGAGTCCGTCACCGCCCTCGCAGAAGCCGTTGCGGCGCTTGACGAACCC
ATCACCGGCGCGGTGCTCGCCGCGGGTATCGCCACAGTCAGCCGCTGCTGGAAAC
CAGCTTCGCCACCTGGAAGCGCCTGCACGCGGTCAACTCCGACGGCGTGTTCCTGT
GCCTGCGCGAATTCGCCCGCATCATGATTGACCAGCAGGAGTCCGACCCGACCAAC
AGCCGCTCCCTGGTGACGGTTCGCTCCAACGCGGCGCGTGTGCCCCGCGCGGAGTT
CGGCGCCTACGGTGCCTCAAGGCGTTCGGCGGCACGCGTGAGTTCCAGCTTCGGTC
TGCAGCTTTCGGCGCACGGTATTCGCGTGAACCTCGGTCTGCCCCGGCACCACCCGA
ACCCCATGGTGACCGACGCCTGGGAGGGCGAGGACCGCTCCGCCCTGCCGTTGC
CGGTAACCCGCAGACGTTCCGCCTGGGTATTCCGCTGGGTTCGCATCGCCGACCCTGC
CGACATCGCGGCGGTCAACGCCTTCCTCATTTCTGAGGCGGCACGCCACATCACCAT
GCAGGAGATTGTCGCCCCATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGC
TTGCGGCCACTCGAGCACCACCACCACCA

10 ATGACTGAGAACACCACATTCTGCATCTCCTTTCTGCCCCATACCGGTGACGGC
AAAGTTGCCGTGGTGACCGGCGCATCCAGCGGTATTGGCCGAGCCACCGTACAGCA
GCTAGTAGCAAGTGGCTGGACCGTCTATGCTCTGGCTCGTCGTACCGACCGTCTTTA
CACCTTTATGCGGAGACCGGCGCTCATCCGGTGACCTGCGATGTGACCGATGAGC
AGTCCGTTTCAAGTTCGTAGCCGAGCAGATCCTGGAAGAGCAGGGCACCATTGATGCC
CTGGTCAATATTGCTGGCGGTGCTATCGGTGTGGACAAGGTTGCAGAGGGTAAGCC
TGATGACTATCTGAAGATGTATCAGATGAATGTTCTTGGCATTCTGCATATGGTGCG
TGCTTCGCGGAGGCGCTGCGCCAGAATGGTTACGGTACGATTCTGAACCTGACCT
CCACCGCTGCGGAGCACGGTATGAGGGAGGCGCCGGGTACAACGCGGCGAAGTT
CGGCGCTCGCGGCTGACCGAGGCGCTTCGCTGGAGGAGGCGGAGAATAATATTC
GCGTGATTGAGATTTGCCCGGCATGGTGCACACCGAAGAGTTCTCTTAAACCGC
CTGGGGTCCAAGGAAGCAGCCGAGCGCTTACGCGGGCGTTGAGAAGCCGCTGA
CCGCAGAGGACGTGGCGCAGACCGTGACCTTTGCGTTGAACGTTCCGCACCATGTG
AACCTTGACCGTATTACGATTTCGCCCCGGTGGCCAGCCGAGTCAGTTCAAGGTGATT
CGCAAGGAGGGCCCATGGCTGATATCGGATCCGAATTCG
AGCTCCGTCGACAAGCTTTCGGGCCACTCGAGCACCACCACCACCAC

11 ATGCCAAGTCACTGCAACGCCCCGTCTTCTTCTACCGGCGCCAGCTCCGGTATC
GGCTACGACGTGGCGCCGCTACTGGTGCCTACGGCTACATCGTCTACGGTGCCGC

CCGCCGCGTTGAGAAGATTGAAGAGCTCGCCTCCGAGGGCGTGAAAGCGCTCAGCC
TGGACGTTACCGATGAAGCATCCATGGAAGCCGCCGTACAGCAGATCATTGATGCA
GAAGGCCGATTGACGTGCTCATCAACAACGCTGGCTACGGATCCTACGGCGCTAT
TGAAGACGTCCCCATCGACGAGGCGCGCCAGTTCGAGGTGAACCTCTTCGGCC
TGGCGCGCCTACCCAGCTGGTACTGCCGCATATGCGCGCACGGGGCAGCGGGCGC
ATCCTGAATATCTCCTCAATGGCGGGGCGTATTACCTCGCCCCTGGGTGCCTGGTAC
CACGCCACCAAGTATGCCCTCGAGGCGTTCAGTGATGCACTGCGCATGGAGGTTGA
AGAGTTCGGTATTGACGTGGTCATTATCGAGCCCGGGCGGCATCAAAAACCCCTGGG
GACTCATCGCGGCAGACCACCTCGAGGAATCCTCCCGAAACGGTGTCTACGCGGCT
CAGGCGCAGCGGGTTGCGGCAAATATGCGTAGGCTCTACTCTCCTTCATCCAATCTC
TCGGAGCCCAAGGTTATCTCTAACGCTATTCTGCGAGCCCTGGAGGCGCGCCGCC
CAAGACCCGCTACCTGGTAGGTTTCGGTGCTAAGCCCTCCGTGTTCTGCATACCGT
GCTACCGGACCGTCTCTTTGACAAGGTGGCGCGCCGTATCTTCCATGGCTGATATCG
GATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGGCCGACTCGAGCACCACCACCA
CCAC

12 ATGAGTAACGCTCCCGAGGTACGTGGGCTTTTCTCAAGGCCCTCGGCCGCCCGGTC
ATCGTGGCCCCAGCTCGGCGGAGCCGACAGTCACTTCGACGGTCCGCTACCGA
GGTGTGCTCCTGCTCCCTGAAGGAGACGGAGCTTCCCGTCGTCGTGCGCGCAGGCG
AGGAGACTTTCGAGGTTGCGGCCACTTCCACCGGCGAGCGCACCATCAACGGCCGC
GTCGCCCTCGTGACGGGCGGCGCCAGGGTTTTGGCGCGGAGATCGCGCGCGGCC
GGTCGACGCCGTTGCTTCGTGTTATCGCCGACCTGAACGGTGAGGGTGCCGCCG
CTAAGGCGGCCGAGCTCGGCGGCGAAGGCGTCGCACACCCGATCACGGTTAACGTC
GCGGACGAGGAGTCCGTCGCGGCGATGGCCGCCGAGATCGAGCGCGTACCGGCG
GCCTCGACCTGGTCGTGTCCAATGCGGGTATTGTGCGTGCCGGCTCCGTCCTCGAAC
AGGACGCTTCCGCGTTCCGCCTGTCCACGGACATCAACTACGTGGCCTTCTTCCTGG
TCACCAAGCACCTGGGCCAGCTTCTGGCCCGCCAGCACTCGACCGCCCCGAGTGG
CTGACCGACATCATCCAGATCAACTCCAAGTCTGGCCTCGTCGGCTCGAACAAGAA
CGCCGCCTACGCGGGATCGAAGTTCGGCGGCATCGGCCTAGTGCAGTCTTCGCC
TCGAGATGGTTGCTCACGGCGTGAAGGTCA
ACGCGATCTGCCCCGTAACCTTCTACGACGGCCCCGCTCTGGTCCGACCCGGATCGC
GGCCTCTTTGTCCAGTACCTGAACTCCGGCAAGGTCCCCGGCGGAAGACGGTCCG
CGACGTCAAGGAGTCTACGAGGCGAAGGTCCCCATGCGCCGCGGCGTCAAGGCA
TCGACGTTCTGCGCGGATCTTCTACATCGTCGAGCAGGAATACGAGACCCGGCCAG
GCCGTGCCCGTACGGGCGGCCAGGTGATGCTCTCCCATGG

13 ATGAACACTTCTCGTCGCGTCGTTGTCACTGGAGCCTCCACCGGAATCGGACAGGC
CACCGCCCGCCTGTTGGCAAAGCGGGGGTGAAGGTGTCGTCGCCGTGGCCCGGCGCC
GCGAACGCCTCGAGGCCCTCGCCGAGCAGATTGGCTGTGAGTACTGGGCGGCCGAC
CTACCGATGAGGCCAGGTGAAGGAGATGGCCGCCACGTCCTGGAGGGGGTCC
CGTGGATGCGGTTGTTAACAACGCCGGGGAGCGATCGGCGTCGACCGTGTGCCG
AGGGTGATCCCGCCCGGTGGAGCGCGATGTTTCGAGCGAACGTGCTACCGCCCTG
CACTGCTCGCGCGGTTCTTGCCGGGGATGCGCGAGCGCGGGGGAGACCTCGTGT
CCTGACCTCGACCGCCGCCACGACACCTACCCGGGTGGCGGCGGCTACGTGGCCG
CTAAGCACGCCGAGCGCATCATCGCCAACACGCTGCGTCAGGAGCTGGTGGGCGAG
CCTGTGCGCGTCATCGAGATTGCGCCCGCATGGTGGTACCGAAGAGTCTCGCTC
AACCGGCTCGGCTCCAGGAGGCGGCCGACCGGCTCTACGAGGGGGTCCGCCGCC
GCTCGTCGCCGAGGACATTGCGGAGGCGATCGTGTGGACCCTGGAGCGCCCCTCC
ACGTCAACATTGACTCGATGATCGTGCGCCCGGTGGCCAGGCGACAAACACGCTC
GTGGCCAGGAAGACCGCCCATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACA
AGCTTGCGGCCGACTCGAGCACCACCACCACCA

15 ATGGACATTGCAGATGCCGAGGGCGTCAAGAAGACCGTTGCTCAGATTCTTGAGGA
GTGCGGCGACGTCAACGTTCTTATCAACTGCGCTGGTGCATCAGTCCGTTCTTT
CACCGAGGTTGACGACAAAGAGTGGAACAACACTATCAACACCAACCTACCGGTA
CCTACAACGTAACCTACGCTCTCTGGCAGCACTTCATTGATCGTGGTGGCGCTCGCA

TTGTCAAC
GTTTCTTCCGTTGCAGGCAAGATTGGCGGCGGCCTCCTTGGTACTGTTGCTTACGCA
TCTTCTAAGGCTGGTATGAACGGCTTTACCAAGGCTATCGCTAAAGAGGGGCGGCAA
GTACGGCATTCTTGTAAACGCTGTTTGGCCATCCTTACCATTACAGATATGACCAC
TGCACTCTCCAACGATGAGGAGAAGTACAAGAAGGTCGTAGGCATTATTCCTCTCG
GCCGTCCTGCACAGGCATCTGAGCCTGCACAGATGGTACTGTTCTTTGCTTCCGACG
CTGCTAGCTTTGTCAACGGTGAGGTTGGCGACTGCGACGGCGGCATCGTGATGGAA
CCATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGA
GCACCACCACCACCA

16 ATGAGCGAGAACACTCCCAAGACCGTGCTGGTCACCGGCGGCAACCGAGGCATCG
GCTACGAAATTGCCAAGGAATTCCAGGCAGCAGGACACAACGTCTGCATCACCTAC
CGTAGCGGTGAAGCCCCGAAGAATTTTTCGCCGTCAAGGCAGATGTACGCGATGC
CGACAGCATTAACGAAGCATTCAAGGAAATTGAAGCAGAGTTCGGCCCCGTGGAG
GTGCTGGTCGCCAATGCAGGTATCACCCGCGACATGCTGCTGATGCGCATGAAAGA
ATCCGACTTACCGACGTGGTTCGACACCAACCTCACCGGTTCTTTCCGCGTGGTTCA
GCGCGCCATCAAGGGCATGCTGAAGCTCAAGCGCGGTCTGATCATTCTGGTTTCCCTC
CGTGGTGGGTCTGTACGGTCTCCCGGCCAGGTCAACTACTCCGCGTCCAAGGCCGC
ACTGGTCGGTATGGCGCGTCCATCACCCGTGAGCTGGGCGGTCTGTAACATCACCG
CTAACGTGGTGGCACCGGGCTTCATTAACACCGCAATGACCGAGGTTCTGCCCGAA
GAGACCAAGAAGAACTACCTGGCGTCCATCCCGGCTGGCCGTTTCGCTGANGCGGA
AGAGGTCGCTCGCGTGATTTCGCTGGCTCGCATCCGATGAGGCTAGCTACATTTCTGG
CGTGTATCCCCGTTGATGCGCGCTGGGCATGGGTACCCATGGCTGATATCGGAT
CCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCA
C

17 ATGGAACTTAAAAATAAAAAATGTTTTTGTAAACAGGTTCAACACGCGGAATTGGATT
GGCTGTGGCTCATAAATTTGCGAGTCTCGGTGCCAATGTTGTCCTAAATGGACGTTT
TGAAATTTCTGAGGACTTGGTTGCACAGTTTGTGACTATGGTGTGACTGTTGTTGG
TATTTCTGGGGATATTTCTAATGGCGAAGATGCGCAACGTATGGTAGCTGAAGCAA
TTGAAAAGCTTGGAAGTGTGATGTTTTGGTCAATAACGCTGGCATTACAAACGAC
AAGTTGATGTTGAAAATGACTGAAGAAGATTTTGAACGGGTCTTGAAAATCAACTT
GACCGGTGCCCTCAATATGACTCAAGCTGTCTTAAAACCTATGTCTAAGGCTCGTCA
AGGTGCCATTATCAACATGTCTCTGTAGTAGGTCTTATGGGGAATATCGGTCAAGC
AACTATGCAGCTTCAAAGGCTGGTTTGATTGGGTTTACCAAATCCGTTGCGCGTGA
AGTTGCGGCCCGTGGTGTTCGTTGAATGCCATTGCACCTGGTTTCATTGAATCAGAT
ATGACAGACGCTATTCCAGAGAAAATGAAAGATGCCATGCTAGCTCAAGTGCCAAT
GAAACGAATTGGTCAAGCTGAAGAAGTGGCAGAAGTTGCGGCTTTCTTAGCAGGTC
AAGAATATTTAACTGGTCAAACAATTGCCATTGATGGCGGAATGACTATGCAACCA
TGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCG
GCCGCACTCGAGCACCACCACCACCA

18 ATGGCAATTTTGATTACCGGTGCATCGGCAGGATTCGGCGCGGCAATGTGCCGTAC
TTTTGTTGCGGCAGGCTATCATGTATTGGCGCGGCACGGCGCGAAGACAAATTGC
AGCAGTTGGCAGAGGAATTGGGCGAGCAGTTTTACCCTTTGGAAATGGATGTTTCG
CGCACGGAGTCGATTCAAACGCTTTGAACAGCCTGCCCGAGCATTGTCCGAAAT
CGATTGCCTGATCAACAATGCAGGTTTGGCTTTGGGTTTGGATACTGCCGATAAAGC
CGATTTTGGCGATTGGGAAACCATGATTCAAACCAATATCATCGGCCTGACTTTTCT
AACCCGCCAGATTTTGGCGCAGATGGTAGCGCGCAAACAGGGCTATATCATCAATT
TGGGTTTCGATTGCCGGTAGTTATGCTTATCCGGCAGTAATGTCTATGGCGGACTA
AGGCTTTTGTGCGTCAATTCAGTATGAACCTGCGCGCCGAATTGGCAGATAAAAAT
ATCCGCATTACCAATATCGAACCGGTTTGTGCGGCGATACCGAGTTTTCCAATGTC
CGTTTCAAAGGCGACGACCAGCGTGCAGCCGAAGTTTATGAGAATGTCGAGTTTAT
CCAGCCGCAAGATATTGCTGATACCGCTTTGTGGTCTATCAGCGTCCGGCGCGGAT
GAACGTCAATTCATCGAAATCATGCCGGTTGCCAGACCTTTGCAGGCATGAAAG
TTTACCGCGACGAGCCTGCCCGGCGAAAGAAGAAACGTTTTGAAAACAAAGTATG

TCTTTGTTTGGAAAAATTAATCTTGGTTTAAGCCATGGCTGATATCGGATCCNAATTC
GAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCANCACCACCACCACC

19 ATGAATGATTGGTTAAACATAAAAAGGCAAAACAGTTCTTGTGACTGGCGCGTCATC
TGGGATCGGTAAGGCAATCGTTGAAGAATTGCTGGAATTAGGTGTCAATGTAGCGA
ATTTTCGATCTTAGCGACAACGATTTGCGTCATCCAAATCTATTATTTGTA AAAAGTTG
ATGTAACCTTCTCGCTCTGAAGTAGAAGAGGGTGTGCCAAAATAGTTGAAAGATTT
GGCAATATTGATGCAGTGGTCAATAATGCAGGGATTAATATCCCTAGATTATTAATC
GATGCAGAAAATCCTAAAGGTCCTTACGAATTGGACGATGAGACGTTTCGAAAAAGT
AACGATGATTAATCAAAAAGGTTTGTATCTGGTTAGTCAGGCAGTGGGACGTATTTT
AGTAAAAAATGGAAAAGGTGTAATTGTGAACATGGCTTCAGAAGCAGGCTTGAA
GGTCTGAAGGACAAAGTGCCTATGCTGCCACAAAAGCAGCAGTCTATAGTTACAC
TCGTTTCATGGGCAAAAGAGCTGGGTAAGCATGGTGTACGTGTGGTTGGAATTGCTC
CAGGAATTATGGAAGCTACGGGACTTCGAACTCTTTCTTATGAGGAAGCTCTTGCTT
ATACCCGTGGAAAGACGGTAGAAGATATTTCGAGCTGGTTATGCTTCAACTTCAACA
ACTCCTTTGGGACGAAGTGGTAAACTACGGGAAGTAGCTGACCTTGTAGCATTCTA
TATTTTCAGACCGTTCTAGCTATATAACTGGAGTCACTACAAATATTGCCGGANGGA
AAACTCGCGGTCCATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCG
GCCGCACTCGAGCACCACCACCACCACCA

20 ATGTATCTTTGCAAGCAAATATATCACACTGATAAACATAGAAAAACAAATATGGC
AAAAGTATTGGTAACAGGTGCTAACAAAGGCATTGGCTATGGCATCTGTAAGTTTT
TAGGCAAGAGTGGCTGGCAGATAATCGTTGGTGC GCGTAATAGTGAGCGTGCCGAG
GAAGCGATGAAGTCATTGAAAGCTGAGGGAGTGGACGTGATTGGATGGCAATACG
TCAATCTATCAGACAATGCTTCTTTGGAGCAAACAGCAAAAGAAGTAAAAGAGAAA
TATCATGACTTGGA ACTATTGGTAAACAATGCTGGTATTCCTGGAGATATGGAAGT
AGCAAGTTATGAGTCAGAGCTAAAGGATGTAATAGATACCGTACAAGTAAACTATG
TCGGAACATTCTGCTTGACAAAGGCACTTACCCCACTCTCTGCAAAACAAAGGA
AGAATTGTGAACATAACCGTACCATCTGAAGTAAGTCCTTACTGGCATCCAATGGC
TTACGTAGCCAGCAAAGCAGCA
CAGAATGCAATGACCAGCATTATGGCTATGGAATTTGAGAAGAACAACATACCTGT
TGAAATCTTCAATATCCACCCTGGAGCTACGACAACCGATTTAAATAACCATTATAC
AGGACCAGGTTCTCACTCGATAGATGTTGTTAGTGAGAAGATTGCAGAGGTTATTA
ATGATGGGGGAAAGCATCAAGGAGAGTTTGTGAACTATATCCTATCCCATGGCTGA
TATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCAC
CACCACCAC

21 ATGAAATTATTGGAAGGTAAGACAGCTCTGATTACAGGTGCTGCACGTGGTATTGG
TAAAGCAATCGCATTAAAGTTTGCAGAAGAGGGTGCAAACATTGCATTTACCGACC
TTGTTATCGATGAGAATGGTAAAGCAACAGAGGCAGAGATTGCTGCTAAGGGCGTA
AAAGTAAAGGGTTACGCAAGTAATGCTGCAGATTTTCGCACAGTCAGAAGAGGTTGT
AAAGCAGGTTAAAGAGGAATTTGGTTCTATTGATATTCTTGTTAAACAATGCAGGTAT
AACAAAAGATGGTCTTATGCTTCGTATGACTGAGCAGCAGTGGGATGCTGTTATAG
CAGTGAACCTTAAGAGTGCTTTCAACTTCATTCATGCTTGTGTTCCAGTAATGATGC
GTCAGCGTAATGGTAGTATTATCAATATGGCCAGTGTAGTCGGCGTACATGGTAAT
GGTGGTCAGGCTAATTATGCTGCATCAAAGGCTGGTATGATAGCCTTGCTAAGAG
TGTTGCACAGGAGATGGGCCCTAAGGGTATCCGTGCAAATGCTATCGCTCCTGGTTT
CATTGATACTGCAATGACACAGGCTTTGAACGACGACATTTCGTAAGGAATGGACAT
CAAAGATCCCTCTCCGCCGTGGTGGTACTGTTGACGATATTGCTAACACAGCAGTAT
ATCTTGGCTCTGAACTATCAAGTTATGTGTCTGGTCAGGTAATCCAGGTTGATGGCG
GTATGAATATGCCATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCG
GCCGCACTCGAGCACCACCACCACCACCA

22 ATGATGACGTCTCTCAGAAAAAAGATAATAGCTATTGGAGGTGCATTGGCTTACCC
ATCCATAAGTCTAACACCTGATATTACAAGAGAACA CTCTCTGGGAAGTGGGTTCT
TGTGACGGGTGCATCGCATGGAATAGGGAGAGCTTTAACCGAAAAAATAATCAATG

CAGGTGCAAATGTCTTTCTCATTGCTCGAAGCGAAGCTGACTTACGCCTCTTATGTG
CCAAAGCTAAGCAAATGGGTAGCAGTGCAGACTACTGCGCAATAGACCTAAGAGA
CAGGGAAAAGTTGGAACAGCTTTGCCAAAACTAAGGGAAACTACCACGGTTA
GATTACTTCTTTTGCAATGCCGGTAAATCTATCCATCGCAAGATCAATGATGCGCAG
GACCGCCTACACGACTATGACCGTACGATGGACCTCAACTATCGTTCCTTGGTTGCG
CTATCATTAGCTATCTTACCTGCCTTGAAAAGCAAGTAAAGGGAGTATCATCTATTCT
TCATCAGTGAGTACACTCTATCCAATGGCACCGGGCTGGTCGGCTTATCATGCCTCA
AAAAGTGCAGCAAACACATGGTGCAGACGGCAAACAGCGAGTTTGCGCCGTTAG
GTGTTACGTCAGATAGCTTATTTACCCTTGGTTCACACTGCGATGTGCGGATGTCA
ACGAACAATAACAAGCACTTACCTGCATACACACCAGCTGATGCAGCCAATATTCTA
TTGAAACTTGCTATACGGAAGGTCAGGACTTATAAACCTTGGTGGGCAAAGTTATC
AGCCCCAATAGCCTATCTCTTTGCTCCGATTATTACCTATATTATAAGCGACCATG
GCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAG
CACCACCACCACCAC

23 ATGAAAAAAGCAATCGTCGTAGGCGCCAGCAGCGGTATCGGACATGAAGTAGCAC
GACTGCTCATCGCACAAGGATGGGCGGTTGGTGTGGCTGCTCGTCGTATAGACAAG
TTGACAGACTTGCAAGCTATGGCACCAGAGCGGGTCTACACCGCCCAAATTGATGT
AAATAACGAAGATGCAGAGACTTCCTTACTGCAACTTATAGAACGTATGAACGGCA
TCGACCTCTATTTTCATGCCGCAGGAATCGGTTGGCAAACCCAAGTCTTAATGCTG
ACATAGAACTCAAAAACAATGGAAACCAATGCTGTTGGATTACACGAATGATTGGC
TGTGCCTATCGCTATTTTGCCAATAAGGGAGGTGGACACATCGCTTGTATCACTTCT
ATCGCAGGAACAAAAGGGCTCGGACCGGCTCCTGCTTATAGTGCAACAAAGGCGAT
GCAGAACACCTATCTACAAGCGTTGGAACAACCTCGCAGCTTGTAACATCATAACA
TCCACTTCACAGATATTCGTCCCGCTTTGTTGACACTCCCCTACTCGCTGGTACATC
TCACCTCCGATGCTGATGACTACAGAAAAGGTGGCACGCAGTATTATAAAGGCTA
TTAACAGCCGACGACACATCTGCGTCATCGATAGCCGTTGGTGCCTACTTACCTACT
TATGGCGACATATCCCTAACTGGATATGGAGGCGAATGAAGCTATGTCAACCATGG
CTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACA
CCACCACCACCAC

24 ATGAAGATGGAAGATACAATTAATAAAGGGATCATTATAGGAGCTACAGGAGGGA
TTGGTCGTCAATTAGCGAAAGAATTGGCACAACGTTTGGAGCATCTGATTTTAGTGA
GTCGAGATGCTGACAACTCTCCCAAGTCCAAAAGGAACTAACTGGGAGTAAGGCG
CAACTTTCAATCCTAACATTAGATATGCTAGATCAAGTGGCTCTAGAAGCTTTTGTG
GAGAGTCTAGATGCTGATCTTCTTGTAAATTGTGCAGGTTAGCTTACTTTTCCCGTG
AAAGTGACCTTGATTCAGCCAGTGAGCAAGACCTCTGGCAAGTCAACTACCATAGT
TCAGTCCAGCTAATCAAGCAGGTGGTGCAAAAGAATCAGAAGATCCAGCTGGTCCA
GCTGTCTTCGCTGGCAGCTCTTTTCCCTCATCCCTATCTAGCAGCCTACAGTGCTAGC
AAGGCTGCCTTGACAGACCTTCACTCTTGCTCTCCAGGAGGAACTGAGGCAATCAGA
ATCCCAGGTCCAGCTAGGTCTCTATATTCTTGGACCAGTCCAAACAGCTATTTTCC
TCCAAACTAGTAGAAGCATTGGGTGGCAGTCGCTTGCAGATGAAGTCTGAGAAAG
TCGCTCAGCAGTTGATTCGATTCATAGAAAGGGATACTTCCCTATGAGATTATCGGGC
TTCGCTATCGCTTGTAGTTTGGCTAGGTGCGCTGCTTCCCCAAAGGTGGATCATCC
GTCTCCTTGCTATATATTTAAAAAAGGGACTCCATGGCTGATATCGGATCCGAATTC
GAGCTCCGTCGACAAGCTTGCGGCCGCACTC GAGCACCACCACCACCACCAC

25 ATGAGTTATAACTTATTAATAAAGGTAAAAGAGGCATTATTTTCGGTGCCTCAACGA
GCAGTCAATTGCATGGAAAGTTGCAGAACGAGCTGTAGAAGAAGGAGCATCAATA
ACACTTTCCAACACTCCTATAGCCGTTAGAATGGGTACTGTAAATTCATTAGCAGAG
AACTGAATTGCGAAGTAATTGCAGCTGATGCTACAAGTGTAGAAGATTTAGAGAA
TGTATTTAAGCGTTCAATGGAAATTCTAGGTGGAAAAATTGACTTTGACTCCACTC
TATTGGTATGTCTCCAAACGTTTCGCAACATCGCACATACGACGATCTGGATTATAA
TATGTTGAATAACAACGTTGACATATCGGCTGTATCATTTTATAAAATGATTCAGAG
TGCTAAGAACTTGATGCCATAAGCGACTATGGTTCTATTCTAGCTCTATCGTACAT
CGCTGCGCAACGTACATTCTTTGGTTACAACGATATGGCAGACGCAAAGCATTAC
TAGAAAGTATTGCACGTAGCTTCGGTTACATATACGGTCGCGAGAAGAATGTACGT

ATCAACACAATATCACAATCTCCTACAATGACTACTGCTGGTCAGGGCGTAAAGGG
AATGGATAAATTGTACGACTTTGCTAACCGTATGGCGCCACTTGGTAATGCTTCTGC
TGCCGAATGTGCAGACTACTGTATTGTTATGTTCTCAGACTTAACAAAGAAAGTTAC
AATGCAAAATCTCTATCACGATGGCGGTTTCTTAACATTGGTATGAGTCTTCGTGC
AATGACAACCTATGAAAAAGGTATTGGAGACGAATATAAAGATGAGAACGGAAAG
ATTATCTACGGACCATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAG
CTTGCGGCCGCACTCGAGCACCACCACCACCAC

26 ATGAGTGAAACTATTTTAATTACAGGGTCTAATCGCGGTATCGGCAAAGCCGTCGC
GCTTGGTTTAGCGCAGGACGGCTTTGATATTGTTGTCCACTGCCGTAGCCGCCGTGA
TGAAGCGGAAGCTGTCGCGGAAGAAATCCGTGCATTAGGTAGAAATGCACGCGTGT
TGCAGTTTGACGTGTCCGACCGCGAAGCCTGCCGAGAAATTTGACTGCCGACATT
GAGGCAAACGGCACATATTACGGCGTGGTGTGAATGCCGGCCTGACGCGCGACAA
TGCGTTTCCTGCGTTACAGATGATGATTGGGACTTGGTGTGCTGCGTACCAATTTGGA
CGGTTTTTATAATGTGTTGCACCCTTTGACCATGCCGATGATTGCCGCCGCAAAGC
CGGACGGATTGTGTGCATGGCGTCAGTATCCGGTTTGACGGGCAACCGCGGTCAGG
TCAATTACAGTGCCTCAAAGCGGGTCTGATCGGCGCGGCAAAGCCTTGGCGGTT
GAGTTGGCAAACGAAAAATTACCGTCAACTGTGTGGCACCGGGTCTAATCGATAC
CGATATTATCGATGAGAACGTACCTGTGCAAGAAATCTTAAAGGCCGTCCCTGCCG
CGCGTATGGGATTGCCGGAAGAAGTGGCGCATGCGGTGCGCTTTTTGATGGATGAG
AAAGCGCGTACATTACGCGCCAAGTATTGCTGTGAACGGAGGTTTGTGTCCATG
*GCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACC
ACCACCACCACCAC*

27 ATGGCTTATAACTTATTTAAAAGGAAAACGTGGTGTATTTTCGGTGCCTGAATGATA
TGTC AATTGCATGGAAAGTGGCTGAACGCTGCGCAGAAGAAGGTGCAACTCTCGTGT
TGAGTAACACTGAAATGGCACTTCGTATGGGTTCACTTGATGAACTTTCCAAGAAGA
TTAATGCGCCAGTTATTGCGGCAGATGCGACCAATTACGATGATCTTGAGAATGTTTT
TGTTAAGGCTCAAGAGCTATTGGGAGGCAAAATAGACTTCGTTCTCCATTCCATAGG
GATGAGTCTAACGTTTCGTAAGCACCGTACTTACGATGATCTTGACTACAATTGGCTA
AATAAACGCTTGATATTTTCAGCCATTTCAATCCACAAGATGCTGCAAGCTGCAAAG
AAAGTTGACGCAATTGCAGAATATGGTTCAGTTCGTTGACGATGTTGGCTTCAC
ATCGCACATTCTTTGGCTATAACGACATGGCCGATGCAAAGAGTTTATTGGAAAGTA
TAGCACGTAGTTTCGGCTATATCTATGGCAGAGAGAAGAATGTTTCGATTAATACCA
TCTCACAATCTCTACAGAAACAACCTGCAGGTAAAGGCATTAAGGACATTGATAACA
TGATGGATTTTGCTGATAAGATGTCACCTCTCGGCAACGCTACAGCCGATGAGTGTG
CTGATTACTGTGTAACCTTTGTTTCAGTATCTAACCCGTAAGGTGACAATGCAGACCCT
TTTCCACGATGGAGGCTTCTCAAACATGGGTATGAGTCTGCGTGTATGAACCAATAT
AGTAAGACTTTGGATCCATGGCTGATATCGGA
TCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCA

28 ATGAACATTGCTATTATTACCGGTGCATCCCGAGGAATCGGTAAAGCCGCGGCCAAG
CGTTTTGCAAGAGAAGGTTACAGCCTCTTATTAATTGCGAAAAAACTGGACGCTT
TTGGAAGAATTGAAAAAGGAAATTCAGTCAGACTTGCCGGAAAACCTGCCGAAATT
TTTCTTTGCAAAGATTTGGGCACAAAAAAGGCCTTTCCCGTATTCTTGAAGGAAAA
AGCCTATCTAAGCTAATTCTTATCGCCAACAGCGGTAAGACGCTATCAAATTATTGC
AAGACTGCAGGGAGGAAGAAACAAAGGCTCTTTTGGAAACAAATCTTTTACAGCCCT
TTATTCTTTGCCAAAAGCTTCTTCCCTATCTTTTACAAGCTGAGGAAGGGAGAATTCT
ATTCAGTTCCTCTGTCTGGGGAAATGTGGGGGCTTCCATGGAAAGTCTTTATTCCCTC
ACCAAGGGCGGTATCTCCACCTTTGCGAAAGCTCTGGGAAAAGAGCTCGCCCCCTTCT
CACATCTCCGTTAATGCCGTAGCTTTCCGGTGTGTTGACACGGATATGAATTCCTGGC
TTAGCACCGAAGAAAAGCAAAGCTTAGAAGAGGGCATCCCCTACGGCAGAATGGCT
ACCGTAGAAGAAGCCGCAGACTTTCTTTACCTTCTCTCCAAGCCCCCTTACCTTA
CAGCCCAAGTGATACCTTTTCGACGGAGGCTGGATCCCATGGCTGATATCGGATCCGAAT
TGAGCTCCGTCGACAAGCTTGCGGCCGCACTC GAGCACCACCACCACCACCA

- 29 ATGAACCGCAAATTTTCGTTACAGGTGCTACTTCTGGTATCGGCCTTGAGTGTGCC
 GTGCTTTCGCTCAAGATGGAGATAATGTATTGATTGCTGGTCGTCGTCAGACCGCTT
 GGCGGCGATTAAGGAAGATTTTGAGCAACAATATGGCATTCTGTGTGGATACGCTTGT
 ACTTGATGTGAGCAAACGTGAAGATGTAGATGCTAAGGTAAAACCAGCTATCGAAGC
 TTTTGGTGGCATTGATGTACTTGTAAATAATGCAGGTCTGGCACAGGGACTCGATCCA
 TTTCAAGATAGTACCGTAGAAGATGCAGTCACAATGATTAATACGAATGTATTAGGC
 CTTCTTTATGTGACTAAGGCTGTATTGCCGTTTCATGATGGCTCAAAAATAGTGGTCATA
 TTGTGAACATGGGCTCTACTGCTGGAATTTATGCGTACCCAGGCGGCGCTGTGTACTG
 TGCTACAAAGGCAGCTGTCAAAATGTTGGCCGATGGCATTCTGATGGATACCATAGC
 TACAGATATTAAGGTTACTACCATTCAACCAGGTATCGTAGAAACTCCATTTAGTGA
 AGTTCGCTTTCACGGAGATGCTGAAAGAGCAGCTTCCGTTTATGCTGGCATTGAGGCT
 GTTCAACCAGAAGACGTAGCCGATGTAGTACTATATGTAACAAACCAACCTAAACGC
 TTGCAAATCTCTGATGTAACCATCATGGCGAACCAACAAGCGGCAGGTTTTATGCCAT
 GGCTGATATCGGATCCGAATTCGAGCTCCTC
 GACAAGCTTGGCGCCGACTCGAGCACCACCACCACCA
- 30 ATGAGCTATAATCTCTTAGCAGGAAAGAAGGGGATCATCTTCGGTGTCTCAACGAA
 CAGTCCATCGCTTGGCGCGTCGCAGAACGTGCTGTAGAAGAAGGCGCAGAAATCATC
 CTCACCAACACGGCTGTAGCTGTTTCGTATGGGCCAGCTGAATGAGCTGGGGCAGAAG
 CTCAATGCTAAGGTCGTCCCGCTGACGCTACGAAGGAAGAAGAAGTCTGAAGTAGTC
 TTCCAGGAAGCTATGAAGAGCTTCGGCAAGGTTGACTTCGTCCTCCACTCCATCGGTA
 TGAGCCCCAACGTCCGTAAGGGTCGTGTTACGATGACCTCGACTACAAGATGCTCC
 AGACGACCTTCGATATCTCGGCTGTATCTTCCACAAGATGCTGCAGGTAGCTAAGA
 AGCTCGACGCTATCGCAGAAGGTGGCTCTGTCTGTCGCTTACTTATATCGCTGCACA
 GCGTACCTTCGTCGGCTATAACGATATGGCTGATGCCAAGAGCCTCTTAGAGTCTATC
 GCTCGTAGCTTCGGTTACATCTACGGTCGTGACAAGGGTGTCCGTATCAATACCGTCT
 CTCAGTCTCCTACGGTCACGACTGCTGGTAGCGGTGTCAAGGGTATGTCTGACCTGCT
 CGACTTCGCAGAAGACCTCTCTCCTCTGGGTAACGCCGACGCTAACGACTGCGCAGA
 CTACTGTATCACGCTCTTCAGCGACCTGACGCGCAAGATCACGATGCAGAATCTCTTC
 AACGACGGTGGCTTCAGCTCAATGGGTATGAGCGCTGCTGCTATCGAAGCCTTCGCT
 ACGGGTTCGCGCAAACCGCGAGAAGCCATGGCTGATATCGGATCCGAATTCGAGCTCCGT
 CGACAAGCTTGGCGCCGACTCGAGCACCACCACCACCA
- 31 ATGACTAAGCATTACGCCGCGCCGTCGTCACCGGTGCCCTTAGCGGTATTGGCTGG
 GCCATAACTGAGCGCCTAGTGACCGAGGGCTGGCAGGTGGTAGGTATTAGCCGTACC
 GGCCAGGTGCCCGAGGGAGCCTTGAGTGTAAGCGCTGACTTGGCAGAGGACGGGGT
 AGGGGAGGCTATCACCCAAGCCAAAACCTGCTCGGCGGCGTCGACGCTTATGTGGG
 GGCGGCTGGAAGCACCTATGAGCAGCTGGCTGCCGCGCCGACTTGGAGCAGGTAA
 ATACCCAGCTGCGCCTGCACTACCTGTCTAACTATGAGGCTATTAGCCTGCTACTGCC
 GGGCATGGTCAGGGCGCGCTGGGGGCGGATTGTGTGCTTTTCTCAGTAGTGGCTCA
 AAGCGGTATGGCAGGTTTAAAGCGCCTATGGGGCGGCAAAGGTGCTCTAGAGGCGCT
 GGTTAAATCGCTGGCGCTGGAGGTGGGCCGGCGGCTATCACGGTTAATGCGGTGGC
 CCCC GTTATATCCAAACCCCATGACCCAGTCGCTAAGCCC GCGCCAGCAAGAACG
 CTACCTGCAACGCACCGGCGCCGCCAGGCCGGGCACGCCCCAGGATGTGGCCGGAC
 CGGTAGCTTTTTTGTCTTAGCGACGACGCCGCTACGTTAACGGGCAGATCCTGCATGT
 AGATGGGGCTATGGGGGTAGGAAACCCATGGCTGATATCGGATCCGAATTCGAGCTCC
 GTCGACAAGCTTGGCGCCGACTCGAGCACCACCACCACCA
- 32 ATGAACATCTTTATTACGGGCGGAACATCGGGTATCGGCTTGGCCTTGGCGCGGTTTT
 ATGCTGCCAAAGGCCATCGGGTGGGGTGTGCGGACGAAATACGGCACGAATAGAC
 AAGAGTGATGAGGTTAACAACACTGTTGCTTGCATACCAGCTGGATGTGTGTGATAAG
 GACGCATTAACCGTTGCCGTAGAAGTGTGTTGTGCCGATAAGGGGTTGGACATGATG
 ATTGTTGCTGCTGGATATTACCGTAACGGCGTTACTGAAGAAGTGGATTTT GAGCAG
 ACTTCACAAATGTTGAAGGTTAATATCGCGGGTGCATTGAATGCGATGGAAGTGGCA
 CGAGAAGCAATGAATGCATCCGGCGGACATTTGGTAGTGATTGCTTCGGTGGCAGGA
 TTGTTGCATTATCCGTGCGCAAGCGTATATGCCAAGTGCAAAAAGAGCATTGATACAG
 ATAGCCGATGCTTATCGCCGAAGTCTTGCCGACTATCAAATAACGGTTACAACACTT

GTTCCGGGCTACATAGACACGCCACGATTGAGAGAGATATACCGTAATGACCTTTCC
AAATGTCCGTTCTGCATGCCCTTAAATCGGGCTGTTGAAACAATGACAAAGGCCATA
GCGCAGCGCAAAGAACAGGTGGTGTTCGCCAAAAATGCGCCTTTCCATCGCTATT
CTTTGCTTTTACCCACTTGTTTGTATCGGCTTTTATGCACCGAAAGACATTATGGAG
CATTCCATGGCTGATATCGGATCCGAATTCGAGCTCC
GTCGACAAGCTTGCGGCCGACTCGAGCACCACCACCACCA

33 ATGAAGTATGCTTTAATAACTGGTGCGTCACGTGGAATAGGACGCTCTGTAGCCTTG
CTTTTAGCAGAACGGTACTCTATTATTATCAATTATCAGTCTAATGCAGAAGCTGCTC
AAGCAGTAAAACAGGAAATAGAGACCAAAGGCCGACATGTGGAACCTTCTCCTTTCC
ATGTTTCTGATCCGAAGGCGATAGAGGCAGCCATTGACACATGGGAAGCATCTCATC
CTGACGAGTTTATTTCTGTATTGGTAAACAATGCTGGTATTCGTCGTGATAATGTTAT
GTTTATGATGTCAGACGAAGACTGGCATAGCGTTCTTGACACGAATATGAATGGATT
TTTCTATATCACCCGCAGGCTTCTCAAGCACATGATGCCACGTAAGCGTGGTGGACG
AATCATCAATATGGCCTCATTGTCAGGTTTGAAAGGTTTGCCAGGGCAAGTAAATTA
TAGTGCAGCAAAGCTGCTCTGATAGGTGCCACGAAAGCTTTGGCACAGGAAGTGGC
TGCTCGTAAGATTACGGTGAATGCTGTTGCACCTGGTTTTATTAGACCGATATGACT
AAGGAAGTGCCTGAGGACGAATTGAAGAACTTGTCCTCCGTCGGTCGTTTTGGAACA
CCAGAAGAAGTTGCAGATGTCGTGGCTTTCTTGCCTCTGATGCTGCAGCATAACATCA
CAGGTGAGGTTATCAATGTGAATGGAGGTTTCTATCCATGGCTGATATCGGATCCGAAT
TCGAGCTCCGTCGACAAGCTTGCGGCCGACTC GAGCACCACCACCACCA

34 ATGAGAAAAACAGCATTAACTGGTGCTACCAGTGAATGGCGAGGCTTGTGCA
CGTAAATTCGCACAAGGTGGCTACGATGTTATCATTACTGGGCGTAACAAACAGCGT
CTTGCAGCACTCAAAGTGGAGCTTGAGACAGGCGAAACAAAGGTCCTTGCCTTGTCT
TTTGACGTCCGAAACCGTGCCGCAGCCACAAAGGCTATTAAGAGTTTACCGGCAGAA
TGGGCAAAGATAGATGTCCTTATCAACAATGCAGGGCTTGCTCTTGGATTGGAACCA
GAGTACGAAGGCGACTTTGAAGACTGGGATACGATGATCGATACCAACATAAAAAGG
ACTCCTGACGATGACCCGTCTCATTGTTCTAAGATGGTTGAGCGTAACAGTGGACA
CATTATTAATATAGGATCAGTGGCTGGTGCAGCTGCATATGCTGGTGGAAACGTCTA
TTGTGCCACGAAGGCTGCAGTAAAACCATCACAGATGGGCTCCGAATTGATGTTGC
TCACACTGCCGTAAGGGTTACGAATGTGAAGCCAGGGCTTGTAGAAACACACTTCTC
TAACGTCCGTTTCCACGGAGATGACAAGCGTGCGAACAGCGTTTATCATGGTATTAA
GCCGTTAACGGGTACCGATATTGCCGATGTTGCCTACTATGCAGCATCCGCTCCAGCC
CATGTGCAGATTGCAGAAGTGCTTGTCTGGCTACACATCAGGGCAGTGGCAGCGTC
ATTCACCCATGGCTGATATCGGATCCGAATTCGAGCTCC
GTCGACAAGCTTGCGGCCGACTCGAGCACCACCACCACCA

35 ATGGGTTTCTTAACTGGTAAACGTATTTTTAGTGGCAGGTCTTGCAAGCAACCGTTCTA
TTGCTTACGGGATCGCAAAAGCAATGAAAGAACAAGGCGCTGAACTTGCTTTCACTT
ATTTAAACGATAAATTACAACCACGCGTAGAAGAATTTGCAAAAAGAATTTGGTTCTG
ACATCGTACTTCTTTAGACGTAGCGACCGATGAAAGCATCCAAAAGCTTTGCTG
AATTAAGCAAACGTTGGGAAAAATTTGATGGTTTCGTACACGCTATCGCATTTCGCAC
CAGGCGACCAATTAGATGGTGATTACGTAAACGCAGCAACTCGTGAAGGCTACCGTA
TCGCTCACGATATCAGTGCATTACGCTTCGTTGCTATGGCACAAGCAGCACGTCTTA
CTTAAATCCAAATGCAGCATTATTAACCCTTTCTTACTTAGGTGCAGAGCGCGGATT
CCTAACTACAACGTCATGTGTTTAGCGAAAGCGTCTCTTGAAGCGGCAACTCGCGTG
ATGGCGGCTGATTTAGGTAAAGAAGGTATTCGTGTGAATGCGATCTCTGCTGGTCTT
ATCCGTACCTTAGCAGCATCAGGTATTA AAAACTTCAAGAAAATGTTTTCTGCATTTG
AGAAAACCGCAGCGTTACGCCGCACAGTTACTATCGAAGATGTGGGTA ACTCAGCAG
CATTCTTATGCTCTGATTTAGCATCTGGTATTACCGGTGAAATCGTTCACGTAGATGC
AGGTTTCAGCATCACCGCAATGGGCGAATtagGCGAACCATGGCTGATATCGGATCCGA
ATTCGAGCTCCGTCGACAAGCTTGCGGCCGACTCGAGCACCACCACCACCA

36 ATGGCAAAGTAATGGTAAACAGGTGCTAACAAAGGCATTGGCTATGGTATCTGTAAG
TTTTTAGGCAAAGAGTGGCTGGCAGGTAATCGTTGGTGC GCGTAATAGTGAGCGTGCC
GAGGAAGCGATGAAGTCATTGAAAGCTGAGGGAGTGGACGTGATTGGCTGGCAATA

CGTCAATCTATCAGACAATGCTTCTTTGGAGCAAACAGCAAAAAGAAGTAAAAGAGA
AATATCATGACTTGGAAC TATTGGTAAACAATGCTGGTATTCCTGGCGATATGAAAG
TAGCAAGTTATGAGTCAGAGCTAAAGGATGTAATAGATACCGTACAAGTAAACTATG
TCGGAACATTCTGCTTGACAAAAGGCACCTTACCCCACTCTCTGCAAACAAAGGAA
GAATTGTGAACATAACCGTACCATCTGAAGTAAAGTCCCTTATTGGCATCCAATGGCTTA
CGTAGCCAGCAAAGCAGCACAGAATGCAATGACCAGCATTATGGCTATGGAATTTGA
GAAGAACAACATAACCTGTTGAAATCTTCAATATCCATCCTGGTGCTACGACAACCGA
TTTAAATAACCATTATACAGGACCAGGTTCTCACTCGATAGATGTTGTTAGTGAGAA
GATTGCAGAAGTTATTAATGATGGGAAAAAGCATCAAGGAGAGTTTGTGAACTATA
TCCTATCGTAGACGAAGGACGACCATGGCTGATATCGGATCCGAATTCGAGCTCCGT
CGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCAC

37 ATGCAAAGAACAGCGTTAGTAACTGGCGCAACTGCCGGATTTGGTGCGGCAATTTGT
CGCACACTTATTGAAAATGGTTATCGTGTAATTGGCACGGGTCGCCGTGTGGCTCGTT
TAGAACAATTACAGCAAGAATTAGGTGAAAACCTCCACTTTCTTGCCTTTGATATTTT
AGATCGCCAAGCAACAGAAGATGCTTTCCATTCCCTTCCACTAATTGGCAATCCATT
GATTTATTGGTGAATAATGCAGGTTTAGGATTAGGCTTAGAAAAGTGCCGATAAAGCG
AGTTTAGACGATTGGATGCAAATGATTGATACCAATATTAAGGACTCGTCACCATC
ACTCGTCTTGTATTACCACAAAATGGTTGAACGCAATTCAGGTCATATTATTAATTTAG
GCTCAATTGCAGGTAATTATCCTTATCCAGGTGGCAATGTATACGGTGGCACTAAAG
CTTTTATTAACAATTTAGTTTAAATCTTCGAGCCGATCTTGCTGGAACCTCAAATTCG
CGTGACCAATGTAGAGCCTGGTCTTTGTGGTGGAACTGAATTTTCTAATATCCGCTTT
AAAGGTGATGATGCCCGAGCAAAAAAACTCTATGAAAATGTGGAATATGTCAGTCCA
CAAGATATTGCTAATATTGTGTTATGGCTCAATCAACAACCTGAACATGTCAATATTA
ATCGCATTGAAGTGATGCCAACTGCCAAACCTTTGCACCACTTAATGTCGCAAGAC
*CATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGC
ACCACCACCACCAC*

38 ATGCACTTAGAGGGTAAAGTAGCTCTCGTAACTGGCGCCTCTCGTGGTATCGGTCG
CGCCGTTGCTATTCAACTTGCACAATCTGGTGCTGATGTTGCTGTAACTATAGTG
GCAGTGAAGCTGCTGCTCAAGAAACAGTGGACGCTATCTTAGCATTGGGGCGCAA
AGCGATTAATAAATAAGCTAATGTTGCAAAATGCTGAAGAAGTTGCAGCAATGGTT
GAAGAAACGCATAAAACATTTGGTTCATCGATATCTTGGTCAATAATGCAGGTA
TCACACGTGATGGTTTACTTATGCGTATGAAAGATGAAGACTTTGATGCCGTTATC
GATATCAACCTTAAAGGTGTATATTTAGTAACAAAAGCAGTATCCAAAATCATGAT
GAAACAACGTGCTGGTTCATATTATTAACATGACTTCTGTTGTTGGTTTGGATGGGTA
ATGCTGGTCAAGCTAACTATGCAGCTTCTAAAGCTGGTGAATCGGTTTTACTAAA
TCTTGTGCAAAAGAGTTAGCTAGCCGTGGTATCACAGTTAATGCAATCGCACCTGG
TTTTATCAATACTGATATGACAGATGTATTGCCTGAAAAGGTAAGAAGCTATGG
TAACACAAATTCGTTGGGCCGTATGGCTAAAGCTGAAGAAGTGGCTGCCGTAAC
AACATTCCTTGCTAGTGATTTTGTAGCTATATTACAGGTCAAGTCATCAATGTAG
ATGGCGGCATGCCATGGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGC
GGCCGCACTCGAGCACCACCACCACCAC

Table S3. Amino acid sequences of the 37 retrieved SDRs (excluding the 20 amino acids and hexa histidine tag coded for by the pET29a vector).

SDR #	Amino acid sequence
1	MNLLANKVAIITGAGRGIGRAIALKYAQEGASVVITDLKIDETVEAFVKELEGLGVKA KAYASNAANFEDAHKLVEAVVADFGRIDVLVNNAGITRDGLMMRMTEEQWDLVINV NLKSAFNLIHAVTPVMVKQRSGSIINMASVVGVSNGAGQANYASASKAGMIGLAKSIA KELGARGIRANAIAPGFIITDMTGALSEEVRKQWEVQIPLRRGGTPEDVANVATFLASD LSSYVSGQTIHVCGGMNM
2	MSTQDLSGKIALVTGASRGIGAAIADTLAVAGAKVIGTATSESGAAAIISERLAQWGG GRALNSAEPETIENLIADIEKEFGKLDILVNNAGITRDNLLMRMKEEEWDDIMQVNLK SVFRASKAVLRGMMKQRAGRIINITSVVGVMGNAGQTNAAAAGLIGFSKSMARE VGSRGITVNCVAPGFIDTDMTRALPEETRKTFEAQTSLGKFGEAQDIADAVLFLASDQ AKYITGQTLHVNGGMLMPPWLISXPEFQHT
3	MAHNIFVTGATSGIGLCIAEAYAKHGDNLVLSGRRRAELLGEVQARLSKEYGVRVETLV LDVRSREDVESKVPAAIEAFGGVDVLVNNAGLAQGLDPFQDSAVDDAVTMIDTNVK GLLYVTKAVLPMIDKNEGHIVNMGSTAGIYAYPNGAVYCATKAAVKTLSDGIRMDT ITTDIKVTTIQGIVETPFSEVRFHGDAERAKSVYAGIDAIQPEDVADVVLVYVTNQP KRLQISDVTIMANQQAAGFMV
4	MSKVAIVTGAGQGIGFAIAKRLVQDGFKVGVLVLDYNAETAEKAVAELSADKAFVVA DVSKQAEVAAAQFQVVDHFGDLNVVNNAGVAPTTPLDTITEEQFNRTFAINVGGVI WGAQAAQAQFKALGHGGKIINATSQAGVVGPNLTVYGGTKFAVRGITQTLARDLA DSGITVNAAYAPGIVKTPMMYDIAHEVGKNAGKDDEWGMQTFAKDITLKRSEPEDVA AAVSF LAGPDSNYITGQTIIVDGGMQFH
5	MSETILVTGASAGFGQAICRRLVADGYRVIGSARRIDKLQALQEELGEAFYPLQMDVT DLSQVDHALASLLKAWEKVDVLVNNAGLALGLAPAYEAEVADWLTMIQTNIVGLTY LTRKILPQMVERNDGYIINLGSTAGTVPPGANVYGASKAFVKQFSLNLRADLAGKKI RVSNIIEPLCEGTEFSSVRFKGDDEKRV EALYRDAHAIQSEDIANTVAWLIQPP KHAVNVNRIEIMPVSQTFGPQPVYR
6	MYSELKGGKVAVITGGSKGIGTAIAKRFQEGMKVVINYNSDAAGAELAAEAVRCAG GEAVTVKAHVGTGEEGVQSLVDAAVENYSGIDVWINNAGMENKVA THEMPLSDWER VINVNLTVFLGTRAALTYFMEHDVKGSIVNMSSVHEQIPWPTFAHYCASKGGIKLFT QTVAMEYAKY GIRVNAVGPAINTPINAKKFSDPVQYEETMSMVPKMRIGKPEEVAA CVAWLASDEASYVTGITLFDVGGMTLYPAFQDGKG
7	MPHKNDVQVALISGGTSGIGFATAKLLLQEGWCVVINGRDEQAGQRAKMKLRRYSS KVRYVYKGDVSSVSDCQRIVKETVDFFGSISALVTAAGYEEELLADVSESAFDEMFGT NVKGTVFLCQAALPYLRQVKGSIVTVSSDAGLQGNVACSVYGASKGAIVSFTKLSLE MAPHEVRVNCVCPGDVDTSLVDKQIAQSHQDAEQAKEEMGQHYPLGRI AKPHEIGEVI AFLISSKASFVTGAAWTIDGG LTS
8	MTKRVLVTGVSSGIGLAQARLFLENGYQVYGVVQGEKPDQGNFHFLLQRDLTLDLEP IFDWCPQVDILCNTAGILDDYKSLLEQSAQEIQEIFEINYVTPVELTRYLQMLEYKR GTIINMCSIASSLAGGGGHAYTSSKHALAGFTKQLALDYAEAGIQVFGIAPGAVKTGM TAADFEPGGLADWVASETPIKRWIEPEEVAEVSFLASGKASAMQGQILTIDGGWSL
9	MSWVIVTGANGGIGEATVHQLIKNGYSVFAADLAEQPIASFGTYGDAIFRYRAVDVTS EESVTALAEVAALDEPITGAVLAAGIAHSQPLETSFATWKRLHAVNSDGVFLCLRE FARIMIDQQESDPTNSRSLVTVASNAARVPRAEFGAYGASKASAARVSSSFGQLAAH GIRVNSVCPGTTRTPMVTDAWEGEDRSALPVAGNPQTFRLGIPLGRIADPADIAAVNA FL ISEAARHITMQEIVA
10	MTENTHIPASPFLPHTGDGKVAVVVTGASSGIGRATVQQLVASGWTVYALARRTDRLY TLYAETGAHPVTCVDTDEQSVQFVAEQILEEQGTIDALVNIAGGAIGVDKVAEGKPPD

YLKMYQMNVLGILHMVRAFAEALRQNGYGTILNLTSTAAEHGYEGGAGYNAAKFGA
RGLTEALRLEEAEENNIRVIEICPGMVHTEEFSLNRLGSKEAAERVYAGVEKPLTAEDVA
QTVTFALNVP HHVNLDRITIRPVAQPSQFKVIRKEGP

- 11 MPKSLQRPVLLTGASSGIGYDVAPLLVRVYGYIVYGAARRVEKIEELASEGVKALSLD
VTDEASMEAAVQQIIDAEGRIDVLINNAGYGSYGAIEDVPIDEARRQFEVNLFLGLARLT
QLVLP HMRARGSGRILNISSMAGRITSPLGAWYHATKYALEAFSDALRMEVEEFGIDV
VIIEPGGIKTPWGLIAADHLEESSRNGVYAAQAQRVAANMRRLYSPSSNLSEPKVISNA
ILRALEARPKTRYLVGFGAKPSVFLHTVLPDRLFDKVARRIF
- 12 MSNAPEVRGLFLKALGRP VIVAPSSAEPTVTFDGPLTEVCSCSLKETELPVVVRAGEET
FEVRATSTGERTINGRVALVTGGAQGFAGAEIARGLVDAGCFVFIADLN GEGAAAKAA
ELGGEGVAHPITVNVADEESVAAMAAEIERVTGGLDLVVS NAGIVRAGSVLEQDASA
FRLSTDIN YVAFFLVTKHLGQLLARQHSTAPEWLT DIIQINSKSLVGSNKNAAYAGSK
FGGIGLVQSFAL EMVAHG VKVNAICPGNFYDGPLWSDPDRGLFVQYLSNGKVP GAKT
VADVKEFYEAKVPMRRGAQ GIDVLR AIFYIVEQEYETGQAVPVTGGQVMLS
- 13 MNTSRRVVVTGASTGIGQATARLLAKRGWKVVAVARRRERLEALAEQIGCEYWAAD
LTDEAQVKEMAAHVLEGGPVD AVVNNAGGAIGVDRVAEGDPARWSAMFERNV LTA
LHCSRAFLPGMRERGGDLVFLTSTAAHDTYPGGGGYVAAKHAERIIANTLRQELVGE
PVRVIEIAPGMVRTEEFSLNRLGSQEAADR VYEGVAAPLVAEDIAEAIWTLERPSHV
NIDSMIVRPVAQATNTLVARKTA
- 15 MDIADAEGVKKTV AQILEECGDVNV LINCAGRISSVPFTEVDDKEWNNTINTNL TGTY
NVTHALWQH FIDRGGARIVNVSSVAGKIGGGLLGT VAYASSKAGMNGFTKAI AKEGG
KYGISCNAVCPSFTITDMTTALSND EEKYKKVVGIIPLGRPAQASEPAQMVLFFASDAA
S FVNGEVGDCDGGIVME
- 16 MSENTPKTVLVTGGNRGIGYEIAKEFQAAGHNVCITYRSGEAPEEFFAVKADVRDADS
INEAFKEIEAEFGPVEVLVANAGITRDMLLMRMKESDFTDVVD TNLTG SFRVVQRAIK
GMLKLRGRILVSSVVG LYGSPGQVNY SASKAALVGMARSITRELGGRNITANV VAP
GFINTAMTEVLPEETKKNYLASIPAGRFAXAE EVARVIRWLASDEASYISG
AVIPVDGGLGMGH
- 17 MELKNKNVFTVGSTRGIGLAVAHKFASLGANVVLN GRSEISEDLLAQFADYGVTVVVG
ISGDISNGEDAQRMVAEAEIKLGSVDVLVNNAGITNDKLM LKMTEEDFERVLKINLTG
AFNMTQAVLKPMSKARQGAIINMSSVVG LMGNIGQANYAASKAGLIGFTKSVAREVA
ARGVRVNAIAPGFIESDMTDAIPEKMKDAMLAQVPMKRIGQAEEVAEVA AFLAGQEY
LTGQT IAIDGGMTMQ
- 18 MAILITGASAGFGAAMCRTFVAAGYHVIGAARREDKLQQLAEELGEQFYPLEMDVSR
TESIQNALNSLPEHLSEIDCLINNAGLALGLDTADKADFGDWETMIQTNIIGL TFLTRQI
LPQMVAR KQGYIINLGSIAGSYAYSGSNVYGATKAFVRQFSMNLRAELADKNIRITNIE
PGLCGDTEFSNVRFKGDDQRAAEVYENVEFIQPQDIADTALWLYQRPARMNVNSIEIM
PVAQTFAGMKVYRDEPAPAKEETFEKQSMSLFGKIKSWFK
- 19 MNDWLNKIGKTVLVTGASSGIGKAIVEELLELGVNV ANFDLSDNDRHPNLLFVKVD
VTSRSEVEEGVAKIVERFGNIDAVVNNAGINIPRL LIDAENPKGPYELDDETFEKVTMI
NQKGLYLV SQAVGRILVKNGKGVIVNMASEAGLEGSEGQSAYAATKAAVYSYTRSW
AKELGKHGVRVVG IAPGIMEATGLRTL SYEEALAYTRGKTVEDIRAGYASTSTPLGR
SGKLREVADLVAFYISDRSSYITGVTTNIAGXKTRG
- 20 MYLCKQIYHTDKHRKTNMAKVLVTGANKGIGYGICKFLGKSGWQIIVGARN SERAEE
AMKSLKAEGVDVIGWQYVNLSDNASLEQTAK EYHDELLVNNAGIPGDMEVA
SYESELKDV IDTVQVNYVGTFLTKALP LLSANKGRIVNITVPSEVSPYWHPMAYVA
SKAAQNAMTSIMAMEFEKNNIPVEIFNIHPGATTTDLNNHYTGPGSHSIDVVSE
KIAEVINDGGKHQGEFVELYPI
- 21 MKLLEGKTALITGAARGIGKAIALKFAEEGANIAFTDLVIDENGKATEAEIAAKGVKV
KGYASNAADFAQSEEVVKQVKEEFGSIDILVNNAGITKDGLMLRMTEQQWDAVIAVN
-

LKSAFNFIHACVPVMMRQRNGSIINMASVVGVHGNGGQANYAASKAGMIALAKSVA
QEMGPKGIRANAIAPGFIDTAMTQALNDDIRKEWTSKIPLRRGGTVDDIANTAV
YLGSELSSYVSGQVIQVDGGMNM

22 MMTSLRKKIIAIGGALAYPSISLTPDITREHFSGKWVLVTGASHGIGRALTEKIINAGAN
VFLIARSEADLRLLCAKAKQMGSSADYCAIDLRDREKLEQLCQKLRRETLPRLDYFFCN
AGKSIHRKINDAQDRLHDYDRTMDLNYRSLVALSLAILPALKASKGSIYSSSVSTLYP
MAPGWSAYHASKSAANTWCETANSEFAPLGVHVQIAYLPLVHTAMSDVNEQYKHLPL
AYTPADAANILLKLAIRKVRITYKPWWAKLSAPIAYLFAPIIHLYYKR

23 MKKAIVVGASSGIGHEVARLLIAQGWAVGVAARRIDKLTDLQAMAPERVYTAQIDVN
NEDAETSLQLIERMNGIDLYFHAAGIGWQNPNSLNADIELKTMETNAVGFTRMIGCAY
RYFANKGGGHIACITSIAGTKGLPAPAYSATKAMQNTYLQALEQLAACKHHNIHFT
DIRPGFVDTPLLAGTSHLPMLMTTEKVARSIKAINSRRHICVIDSRWCVLTY
LWRHIPNWIWRRMKLCQ

24 MKMEDTIKKGIIIGATGGIGRQLAKELAQRLEHLILVSRDADKLSQVQKELTGSKAQLS
ILTLDMLDQVALEAFVESLDADLLVNCAGLAYFSRESLDSASEQDLWQVNYHSSVQ
LIKQVVQKNQKIQLVQLSSLAALFPHYLAAYSASKAALQFTTLALQEELRQSESQVQ
LGLYILGPVQTAIFPPKLVEALGGSRLQMKSEKVAQQLIRFIERDTSYEII
GLRYRLVWLGRLLPQRWIIRLLAIYLKKGTT

25 MSYNLLKKGKRGIIFGALNEQSIWKAERAVEEGASITLSNTPIAVRMGTVNSLAEKL
NCEVIAADATSVEDLENVFKRSMEILGGKIDFVLHSIGMSPNVRKHRTYDDLDYNML
NTTLDISAVSFHKMIQSACKLDAISDYGSILALSIIAAQRTFFGYNDMADAKALLESIA
RSFGYIYGREKNVRINTISQSPTMTTAGQGVKGMKLYDFANRMAPLGNASAAECAD
YCIVMFSDLTKKVTMQNLYHDGGFSNIGMSLRAMTTYEKGIGDEYKDENGKIIYG

26 MSETILITGSNRGIGKAVALGLAQDGFIDIVVHCRSRRDEAEVAEEIRALGRNARVLQ
FDVSDREACREILTADIEANGTYYGVVNLNAGLTRDNAPFAFTDDDWDLVLRTNLDGF
YNVHLPLTMPMIRRRKAGRIVCMASVSGLTGNRGQVNYNSASKAGLIGAAKALAVELA
KRKITVNCVAPGLIDTDIIDENVPVEEILKAVPAARMGLPEEVAHA VRFLMDE
KAA YITRQVIAVNGGLC

27 MAYNLLKKGKRGVIFGALNDMSIAWKVAERCAEEGATLVLSNTEMALRMGSLDELK
KINAPVIAADATNYDDLENVFKAQELLGGKIDFVLHSIGMSPNVRKHRTYDDLDYN
WLNKTLDISAISFHKMLQAACKVDAIAEYGSVVALTYVASHRTFFGYNDMADAKSLL
ESIARSFGYIYGREKNVRINTISQSPTETTAGKGIKIDIDNMDFADKMSPLGNATADEC
ADYCVTLFSDLTRKVTMQTLFHDGGFSNMGMMSLRAMNQYSKTLT

28 MNIAITGASRGIGKAAAKRFAREGYSLLLNCEKNWTLLEELKKEIQSDLPENCPEIFLC
KDLGTTKGLSRILEGKSLSKLILIANSGKDAIKLLQDCREEETKALLETNLLQPFIQCQK
LLPYLLQAEGRILFSSSVWGNV GASMESLYSLTKGGISTFAKALGKELAPSHISVNAV
AFGAVDTDMNSWLSTEEKQSLEEGIPYGRMATVEEAADFLYLLSQAPLYLTAQVI
PFDGGWI

29 MNRKIFVTGATSGIGLECARAFAQDGDVNLVLIAGRRADRLAAIKEDFEQQYGIRVDTLV
LDVSKREDVDAKVKPAIEAFGGIDVLVNNAGLAQGLDPFQDSTVEDAVTMINTNVLG
LLYVTKAVLPMMAQNSGHIVNMGSTAGIYAYPGGAVYCATKAAVKMLADGIRMD
TIATDIKVTTIQPGIVETPFSEVRFHGDAERAASVYAGIEAVQPEDVADVVLVYTNQPK
RLQIS DVTIMANQQAAGFM

30 MSYNLLAGKKGIIFGALNEQSIWVAERAVEEGAEIILTNTAVAVRMGQLNELGQKL
NAKVVPADATKEEELEVVFQEAAMKSFQVDFVLHSIGMSPNVRKGRAYDDLDYKML
QTTFDISAVSFHKMLQVAKKLDIAEAGGSVVALTYIAAQRFTVGYNDMADAKSLESII
ARSFGYIYGRDKGVRINTVSQSPTVTTAGSGVKGMSDLLDFAEDLSPLGNADANDCA
DYCITLFSDLTRKITMQNLFNDGGFSSMGMSSAAIEAFATGRANREK

31 MTKHLRRAVVVTGASSGIGWAITERLVTEGWQVVGISRTGQVPEGALSVSADLAEDGV
GEAITQAQNLLGGVDAYVGAAGSTYEQLAARADLEQVNTQLRLHYLSNYEAISSLLL

GMVRARWGRIVLLSSVVAQSGMAGLSAYGAAKGALEALVKSLALEVGRRAITVNAV
APGYIQTTPMTQSLSPRQQERYLQRTGAARPGTPQDVAGPVAFLLSDDAAYVNGQILH
VDGAMGVGNP

32 MNIFITGGTSGIGLALARFYAAKGHRVGVCGRNTARIDKSDEVNKKLLAYQLDVCDK
DALTVAVEVFCADKGLDMMIVAAGYYRNGVTEEVDFEQTSQMLKVNIAGALNAME
VAREAMNASGGHLVVIVASVAGLLHYPCASVYAKCKRALIQIADAYRRSLADYQITVT
TLVPGYIDTPRLREIYRNDLSKCPFCMPLNRAVETMTKAIAQRKEQVVFPPKMRLS
IAILLSLLPTCLLSAFMHRKTLWSI

33 MKYALITGASRGIGRSVALLLAERYSIINIYQSNAEAAQAVKQEIETKGGHVPELLPFDV
SDPKAIEAAIDTWEASHPDEFISVLVNNAGIRRDVNMFMMSDEDWHSVLDTNMNGFF
YITRLLKHMMPKRGRGRIINMASLSGLKGLPGQVNYSAAKAALIGATKALAQEVAA
RKITVNAVAPGFIQTDMTKELPEDELKLVVGRFGTPEEVADVVAFLASDAAAAYITG
EVI NVNGGFY

34 MRKTALITGATSGIGEACARKFAQGGYDVIITGRNKQRLAALKVELETGETKVLALAF
DVRNRAAATKAIKSLPAEWAKIDVLINNAGLALGLEPEYEGDFEDWDTMIDTNIKGLL
TMTRLIVPKMVERNNSGHIINIGSVAGDAAYAGGNVYCATKAAVKTTIDGLRIDVAHTA
VRVTNVKPLVETHFSNVRFHGDDKRANSVYHGKPLTGTDIADVAYYAASAPAHVQ
IAE VLVLATHQGSVVIH

35 MGFLTGKRILVAGLASNRISYGIKAMKEQGAELAFYTLNDKLPQPRVEEFAKEFGSD
IVLPLDVATDESIQNCFAELSKRWEKFDGDFVHAIAPAGDQLDGDYVNAATREGYRIA
HDISAFSVMAMAQAARPYLNPNAALLTSLYGAERAIPNYNVMCLAKASLEAATRVM
AADLGKEGIRVNAISAGPIRTLAASGIKFNKMFSAFEKTAALRRVTIEDVGNAAFL
CSDLASGITGEIVHVDAGFSITAMGELGE

36 MAKVMVTGANKGIGYGICKFLGKSGWQVIVGARNSEAEAMKSLKAEGVDVIGW
QYVNLSDNASLEQTAKEVKEYHDLELLVNNAGIPGDMKVASYESELKDVIDTVQVN
YVGTFLTKALTPLLSANKGRIVNITVPSEVSPYWHPMAVYVASKAAQNAMTSIMAME
FEKNNIPVEIFNIHPGATTTDLNNHYTGPGSHSIDVVSEKIAEVINDGKKKHQGEFVELYP
IVDEGR

37 MQRTALVTGATAGFGAAICRTLIENGYRVIGTGRRVARLEQLQELGENFHFLAFDIS
DRQATEDAFHSLPTNWQSIDLLVNNAGLGLGLESADKASLDDWMQMIDTNIKGLVTI
TRLVLPQMVERNNSGHIINLGSIAGTYPYPGGNVYGGTKAFIKQFSLNLRADLAGTQIRV
TNVEPGLCGGTEFSNIRFKGDDARAKKLYENVEYVSPQDIANIVLWLNQQP
EHVNINRIEVMPTAQTFAPLNVAR

38 MHLEGKVALVTGASRGIGRAVAIQLAQSGADVAVNYSGSEAAAQETVDAILALGR
KAIKIVSKIMMKQRAGHIINMTSVVGLMGNAGQANYAASKAGVIGFTKSCAKELAS
RGITVNAIAPGFINTDMTDVLPEKVKEAMVTQIPLGRMAKAEVAAVTTFLASDFA
SYITGQVINVDGGMKANVANAEVAAMVEETHKTFGHIDILVNNAGITRDGLLMR
MKDEDFDAVIDINLKGYYLVTKA

Table S4. Table of primers used to retrieve SDRs from the oral metagenome. Italicised sequences match MCS from pet29a to facilitate Gibson assembly.

SDR #	Primers
1 FWD	<i>TAAGAAGGAGATATACATATGAATCTACTTGCTAACAAGGTAGCC</i>
1 RVR	<i>ATCCGATATCAGCCATGGCATATTCATCCCACCGCAGACG</i>
2 FWD	<i>TAAGAAGGAGATATACATATGAGTACACAAGATTTGAGCGGCAA</i>
2 RVR	<i>ATCCGATATCAGCCATGGAGGCATCAACATGCCGCCGT</i>
3 FWD	<i>TAAGAAGGAGATATACATATGGCTCATAATATTTTTGTCACTGGTGC</i>
3 RVR	<i>ATCCGATATCAGCCATGGCACCATAAAGCCTGCCGCTTGT</i>
4 FWD	<i>TAAGAAGGAGATATACATATGTCTAAAGTAGCTATTGTTACAGGTGC</i>
4 RVR	<i>ATCCGATATCAGCCATGGATGGAATTGCATACCACCATCGAC</i>
5 FWD	<i>TAAGAAGGAGATATACATATGTCAGAAACGATTTTAGTAACAGGAGC</i>
5 RVR	<i>ATCCGATATCAGCCATGGACGATAAACGGGTTGAGGACCAAAG</i>
6 FWD	<i>TAAGAAGGAGATATACATATGTACAGTGAATTAAGGAAAGGTAGCC</i>
6 RVR	<i>ATCCGATATCAGCCATGGACCCTTACCGTCTTGAAAAGCCG</i>
7 FWD	<i>TAAGAAGGAGATATACATATGCCTCATAAAAATGATGTTCAAGTGGCA</i>
7 RVR	<i>ATCCGATATCAGCCATGGACTCGTAAGACCTCCATCAATCG</i>
8 FWD	<i>TAAGAAGGAGATATACATATGACTAAACGCGTCTTAGTTACAGGTG</i>
8 RVR	<i>ATCCGATATCAGCCATGGCAAAGACCAGCCACCATCAATAGTC</i>
9 FWD	<i>TAAGAAGGAGATATACATATGAGCTGGGTCATTGTGACCGGC</i>
9 RVR	<i>ATCCGATATCAGCCATGGGGCGACAATCTCCTGCATGGTG</i>
10 FWD	<i>TAAGAAGGAGATATACATATGACTGAGAACACCCACATTCCTGC</i>
10 RVR	<i>ATCCGATATCAGCCATGGGCCCTCCTTGCGAATCACCTTG</i>
11 FWD	<i>TAAGAAGGAGATATACATATGCCCAAGTCACTGCAACGCC</i>
11 RVR	<i>ATCCGATATCAGCCATGGGAAGATACGGCGCGCCACCTTG</i>
12 FWD	<i>TAAGAAGGAGATATACATATGAGTAACGCTCCCGAGGTACGTG</i>
12 RVR	<i>ATCCGATATCAGCCATGGGAGAGCATCACCTGGCCGCC</i>
13 FWD	<i>TAAGAAGGAGATATACATATGAACACTTCTCGTCGCGTCGTTGTTAC</i>
13 RVR	<i>ATCCGATATCAGCCATGGGGCGGTCTTCTGGCCACGAG</i>
14 FWD	<i>TAAGAAGGAGATATACATATGGGAACAGCGCTGGTCACCGG</i>
14 RVR	<i>ATCCGATATCAGCCATGGCATGTGGGGGACGGAGCGCATG</i>
15 FWD	<i>TAAGAAGGAGATATACATATGGACATTGCAGATGCCGAGGGCGT</i>

15 RVR ATCCGATATCAGCCATGGTTCCATCACGATGCCGCCGTCGCA

16 FWD TAAGAAGGAGATATACATATGAGCGAGAACACTCCCAAGACCGTGCT
16 RVR ATCCGATATCAGCCATGGGTGACCCATGCCAGGCCGCCAT

17 FWD TAAGAAGGAGATATACATATGGAACCTAAAAATAAAAATGTTTTG
17 RVR ATCCGATATCAGCCATGGTTGCATAGTCATTCCGCCA

18 FWD TAAGAAGGAGATATACATATGGCAATTTTGATTACCGGG
18 RVR ATCCGATATCAGCCATGGCTTAAACCAAGATTTAATTTTTCCAAC

19 FWD TAAGAAGGAGATATACATATGAATGATTGGTTAAACATAAAAGGC
19 RVR ATCCGATATCAGCCATGGACCGCGAGTTTTCCCTCC

20 FWD TAAGAAGGAGATATACATATGTATCTTTGCAAGCAAATATATCACA
20 RVR ATCCGATATCAGCCATGGGATAGGATATAGTTCAACAAACTCTC

21 FWD TAAGAAGGAGATATACATATGAAATTATTGGAAGGTAAGACAGC
21 RVR ATCCGATATCAGCCATGGCATATTCATACCGCCATCAACC

22 FWD TAAGAAGGAGATATACATATGATGACGTCTCTCAGAAAAAAGAT
22 RVR ATCCGATATCAGCCATGGTCGCTTATAATATAGGTGAATAATCGG

23 FWD TAAGAAGGAGATATACATATGAAAAAAGCAATCGTCGTAGGC
23 RVR ATCCGATATCAGCCATGGTTGACATAGCTTCATTGCTCTCC

24 FWD TAAGAAGGAGATATACATATGAAGATGGAAGATACAATTA AAAAGGG
24 RVR ATCCGATATCAGCCATGGAGTCCCTTTTTTAAATATATAGCAAGGAG

25 FWD TAAGAAGGAGATATACATATGAGTTATAACTTATTA AAAAGGTAAAAGAGG
25 RVR ATCCGATATCAGCCATGGTCCGTAGATAATCTTTCCGTTCTC

26 FWD TAAGAAGGAGATATACATATGAGTGAAACTATTTTAATTACAGGGTC
26 RVR ATCCGATATCAGCCATGGACACAAACCTCCGTTACAGC

27 FWD TAAGAAGGAGATATACATATGGCTTATAACTTATTA AAAAGGAAAACGTG
27 RVR ATCCGATATCAGCCATGGATCCAAAGTCTTACTATATTGGTTCATAG

28 FWD TAAGAAGGAGATATACATATGAACATTGCTATTATTACCGGTGC
28 RVR ATCCGATATCAGCCATGGGATCCAGCCTCCGTCGAAAGG

29 FWD TAAGAAGGAGATATACATATGAACCGCAAATTTTCGTTACAGG
29 RVR ATCCGATATCAGCCATGGCATAAAACCTGCCGCTTGTTGG

30 FWD TAAGAAGGAGATATACATATGAGCTATAATCTCTTAGCAGGAAAG
30 RVR ATCCGATATCAGCCATGGCTTCTCGCGGTTTGCGCGA

31 FWD TAAGAAGGAGATATACATATGACTAAGCATTTACGCCGCG
31 RVR ATCCGATATCAGCCATGGGTTTCCTACCCCATAGCCC

32 FWD TAAGAAGGAGATATACATATGAACATCTTTATTACGGGCGGAAC
32 RVR ATCCGATATCAGCCATGGAATGCTCCATAATGTCTTTCGGTGC

33 FWD TAAGAAGGAGATATACATATGAAGTATGCTTTAATAACTGGTGCTTC
33 RVR ATCCGATATCAGCCATGGATAGAAACCTCCATTACATTGATAACCT

34 FWD TAAGAAGGAGATATACATATGAGAAAAACAGCATTAACTGGTGTC
34 RVR ATCCGATATCAGCCATGGGTGAATGACGCTGCCACTGC

35 FWD TAAGAAGGAGATATACATATGGGTTTCTTAACGGTAAACGTATTTAG
35 RVR ATCCGATATCAGCCATGGTTCGCCTAATTCGCCATTGCG

36 FWD TAAGAAGGAGATATACATATGGCAAAGTAATGGTAACAGGTGC
36 RVR ATCCGATATCAGCCATGGTCGTCCTTCGTCTACGATAGGATA

37 FWD TAAGAAGGAGATATACATATGCAAAGAACAGCGTTAGTAACTGG
37 RVR ATCCGATATCAGCCATGGTCTTGCGACATTAAGTGGTGCAAAGG

38 FWD TAAGAAGGAGATATACATATGCACTTAGAGGGTAAAGTAGCTCT
38 RVR ATCCGATATCAGCCATGGCATGCCCATCTACATTGATGAC

2. Figures

Figure S1.

A. Sequence alignments for SDRs 3,4,11,17,31,37. The catalytic triad Ser-Tyr-Lys and a fourth residue considered to be important are all marked red. The TGxxxGxG motif important in cofactor binding is marked in purple.

```

11      MPKSLQRPVVLLTGASSGIGYDVAPLLVRYGYIVYGAARRVEKIEELASE-----GVK
3      -----MAHNI FVTTGATSGIGLCIAEAYAKHGDNVLISGRRAELLGEVQARLSKEYGVRVE
37     -----MQRTALVTTGATAGFGAAICRTLIENGYRVI GTGRRVARLEQLQQL-----GENFH
4      -----MSKVAIVTTGAGQGIGFAIAKRLVQDGFVKVGLDYNAETAEKAVAELS-AD--KAF
17     --MELKNKNV FVTTGSTRGIGLAVAHKFASLGANVVLNGRS-EISEDLLAQFA-DYGVTVV
31     --MTKHLRRAVVTGASSGIGWAITERLVTEGWQVVGISRTGQVPEGALS-----
          .:***:  *: * :          *  *

11      ALSLDVTDEASMEAAVQQII DAEGRIDVLI NNAGYGS-YGAI EDVPI DEARRQFEVNLFG
3      TLVLDVRSREDVESKVPAAIEAFGGVDVLVNNAGLAQGLDPFQDSAVDDAVTMIDT NVKG
37     FLAFDISDRQATEDAFHSLPTNWQSIDLLVNNAGLGLGLESADKASLDDWMQ MIDTNIKG
4      AVVADVSKQAEVAAAFQKVVDHFGDLNVVNNAGVAP-TTPLDTITEEQFNRTFAIN VGG
17     GISGDISNGEDAQRMVAAEIEKLGSDVLVNNAGITN-DKMLMKMTEEDFERVLKINLTG
31     -VSADLAE-DGVGEAITQAQNLLGGVDAYVGAAGSTY-EQLAARADLEQVNTQLRLHYLS
          :  * : .          .          : :  .  **          : :  :  : .

11      LARLTQLVLPH-MRARGSGRI LNIS SMAGRITSP LGAWYHATKYALEAFSDALRMEVEEF
3      LLYVTKAVLPF-MIDKNEGHIVNMGSTAGIYAYPNGAVYCATKAAVKTLSDGIRMDTITT
37     LVTITRLVLPQ-MVERN SGHI INLGS IAGTY PYPGGNVYGGTKAFIKQFSLNLRADLAGT
4      VIWGAQAAQAQFKALGHGGKI INATSQAGVVG NPNLTVYGGTKFAVRGITQTLARDLADS
17     AFNMTQAVLKP-MSKARQGA I NMS SVVGLMGNIGQANYAASKAGLIGFTKSVAREVAAR
31     NYEAI SLLLPG-MVRARWGRIVLLS SVVAQSGMAGLSAYGAAKGALEALVKS LALEVGRR
          *  * :          *  . .          *  . : *  :  :  :  :

11      GIDVVIIEPGGIK-TPWGLIAADHLE-----E---SSRNGVYAAQAQRVAANMR
3      DIKVTTIQPGIVE-TPFSEVRFHGDA-----ERAKSVYAGIDAIQPEDVADVVL
37     QIRVTNVEPGLCGGTEFSNIRFKGDD-----ARAKKLYENVEYVSPQDIANIVL
4      GITVNAYAPGIVK-TPMMYDIAHEVGNAGKDDDEWGMQTFAKDITLKRLEPEDVAAAVS
17     GVRVNAIAPGFIE-SDMTDAIPEKMK-----DAMLAQVPMKRIGQAEVEVAEVA
31     AITVNAVAPGYIQ-TPMTQSLSPRQQ-----ERYLQRTGAARPGTPQDVAGPVA
          :  *          **          :          :  : *

11      RLYSPSS--NLSEPKVISNAILRALEARPKTRYLVGFGAKPSVFLHTVLPDRLFDKVAR
3      YVTNQPKRLQISDVTIMANQQAA-----G-----FMV-----
37     WLNQQPEHVNINRIEVMPTAQTF-----A-----PLN-----VAR
4      FLAGPDSNYITGQTIIVDGGMQF----H-----
17     FLAG--QEYLTGQTIAIDGGMTM----Q-----
31     FLLSDDAAYVNGQILHVDGAMGV----GNP-----
          :          .          :

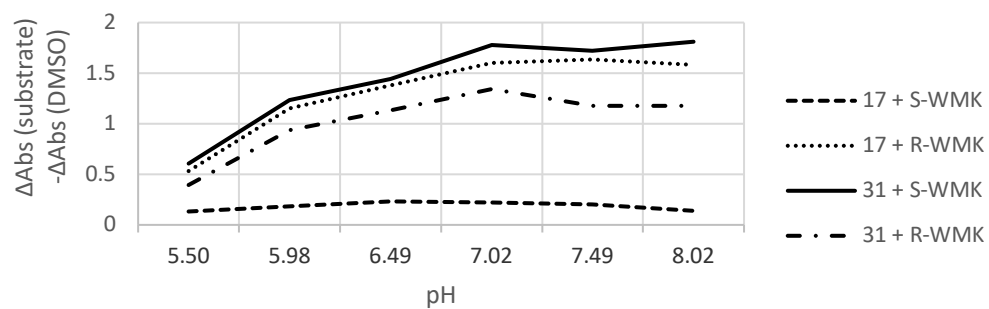
11      RIF
3      ---
37     ---
4      ---
17     ---
31     ---

```

B. SDR percentage identity matrix for 3,4,11,17,31,37

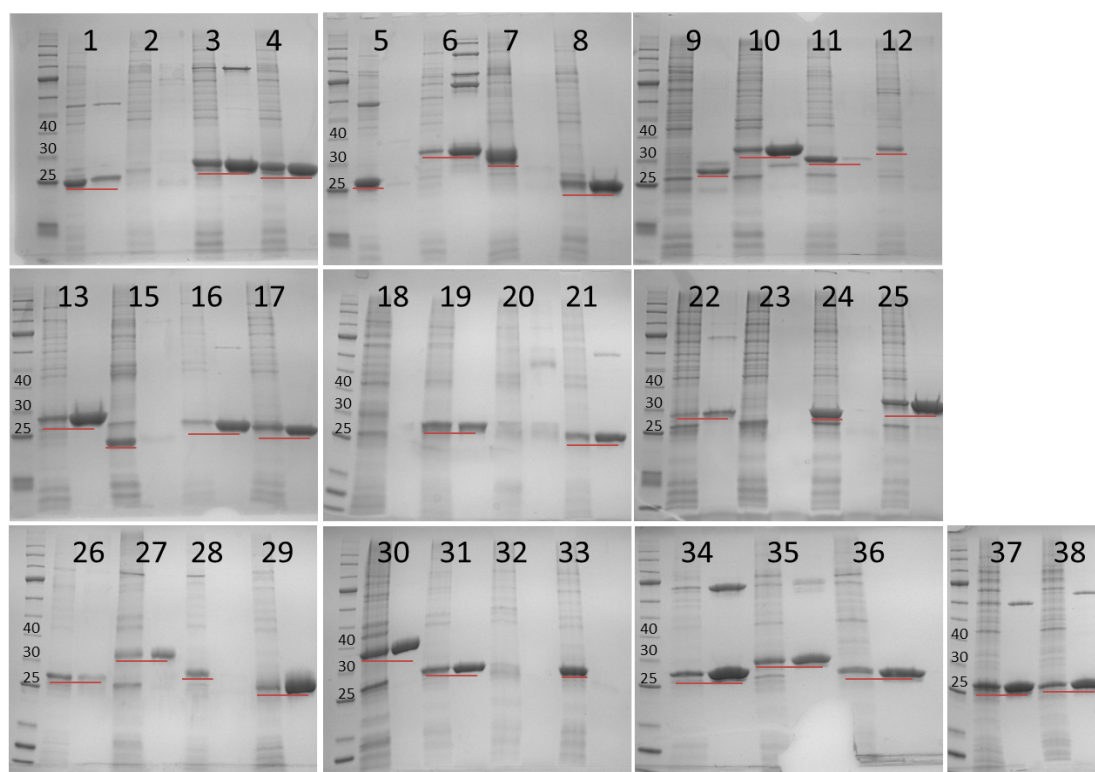
11	100.00	33.05	31.38	26.18	27.04	25.43
3	33.05	100.00	39.18	27.39	28.75	21.21
37	31.38	39.18	100.00	24.79	21.94	23.38
4	26.18	27.39	24.79	100.00	31.80	28.88
17	27.04	28.75	21.94	31.80	100.00	31.47
31	25.43	21.21	23.38	28.88	31.47	100.00

Figure S2. Graph displaying activity with (*S*)-**18** and (*R*)-**18** with SDR-17 and SDR-31 at a range of pHs



Reaction conditions (200 μL volume): **18** (5 mM), clarified cell lysate (0.4 mg/mL), NADPH (1 mM), KPi (100 mM), DMSO (10%, v/v). Reactions were shaken for 95 mins at 25 $^{\circ}\text{C}$, performed in triplicate and quantified by the spectrophotometric assay at 340 nm.

Figure S3. SDS PAGE showing induced recombinant protein expression in *E.coli*.



Numbers in the well line correspond to SDR number IDs. Each enzyme has two lanes, crude lysate and purified fraction from left to right. Bands corresponding to induced protein are underlined in red either in the crude lysate lane or both lanes in the case of successful purification. Markers are given in kDa (10-230 Broad range, NEB). Protein sizes in kDa calculated using the online tool expasy protparam as follows; 1) 29, 2) 27.5, 3) 29.5, 4) 29.4, 5) 30.2, 6) 30.7, 7) 29.8, 8) 28, 9) 28.9, 10) 31.5, 11) 33.2, 12) 35.6, 13) 29.6, 15) 22.9, 16) 28.4, 17) 28.5, 18) 33.3, 19) 31.2, 20) 30.1, 21) 29, 22) 34, 23) 29.7, 24) 31.5, 25) 34.3, 26) 29, 27) 33.1, 28) 29.1, 29) 29.2, 30) 32.6, 31) 27.4, 32) 30.1, 33) 28.96, 34) 29.2, 35) 30.7, 36) 28.5, 37) 30.2, 38) 28.5.

Figure S4. HPLC calibration curves for A. (*S*)-WMK, (*S*)-**18**. B. (*R*)-WMK (*R*)-**18**. C (4*aR*,5*R*)-**22**. D. (4*aR*,5*S*)-**22**. E. (4*aS*,5*S*)-**22**. Quantification by HPLC on a Chiralcel OJ column with 4% isopropanol/hexane mobile phase 0.5 mL/min flow rate and detection at 230 nm.

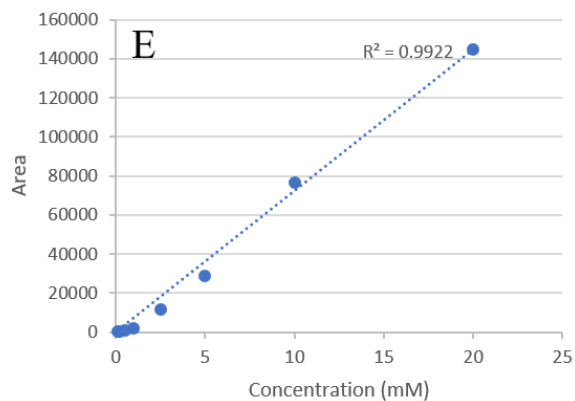
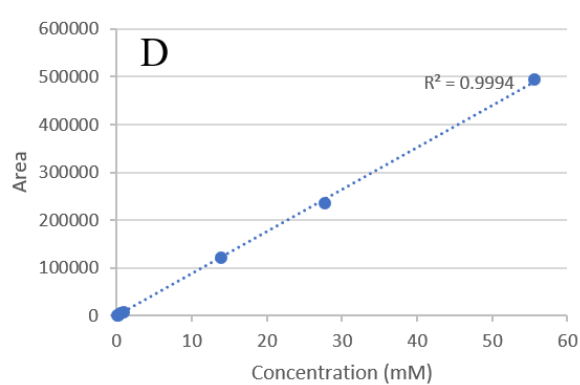
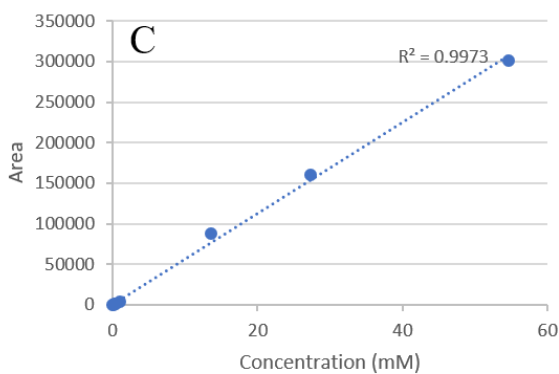
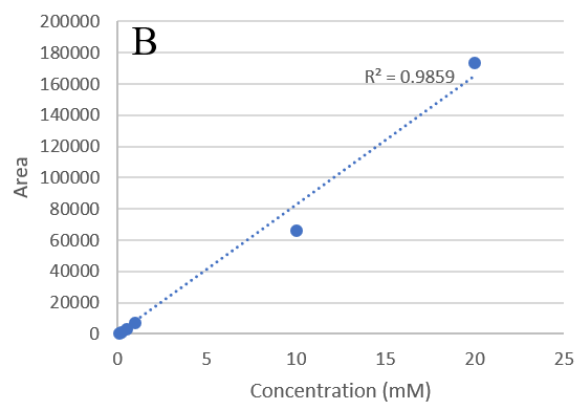
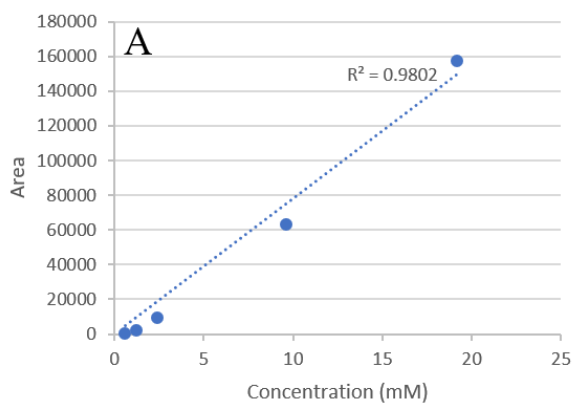


Figure S5. HPLC trace of the products of the reduction of *rac*-WMK-**18** using NaBH₄: (4*aR*,5*S*)-**22** Rt = 54.0 min, (4*aR*,5*R*)-**22** Rt = 61.2 min, (4*aS*,5*S*)-**22** Rt = 85.9 min. HPLC conditions were 4% iPrOH:hexane, 120 min, 0.5 mL/min at 230 nm, on a Chiralcel OJ column.

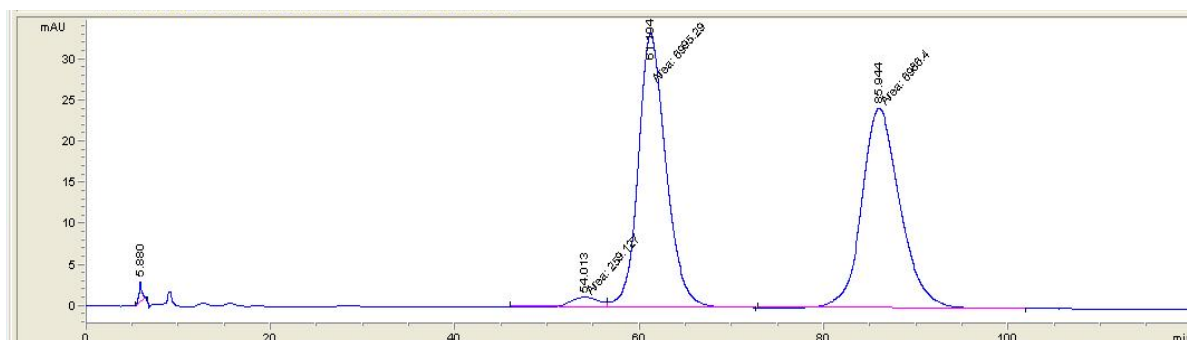
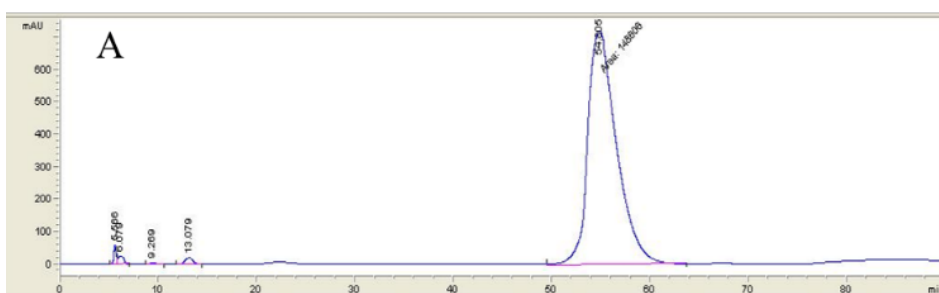


Figure S6. Characterisation data for (4*aR*,5*S*)-**22** from the scaled-up reaction using SDR-17. **A.** HPLC trace of (4*aR*,5*S*)-**22** starting from (*R*)-**18** using SDR-17. Rt = 54 min. HPLC using Chiralcel OJ column with 4% isopropanol/hexane mobile phase 0.5 mL/min flow rate and detection at 230 nm. **B.** ¹H and ¹³C NMR spectra of (4*aR*,5*S*)-**22**.



B

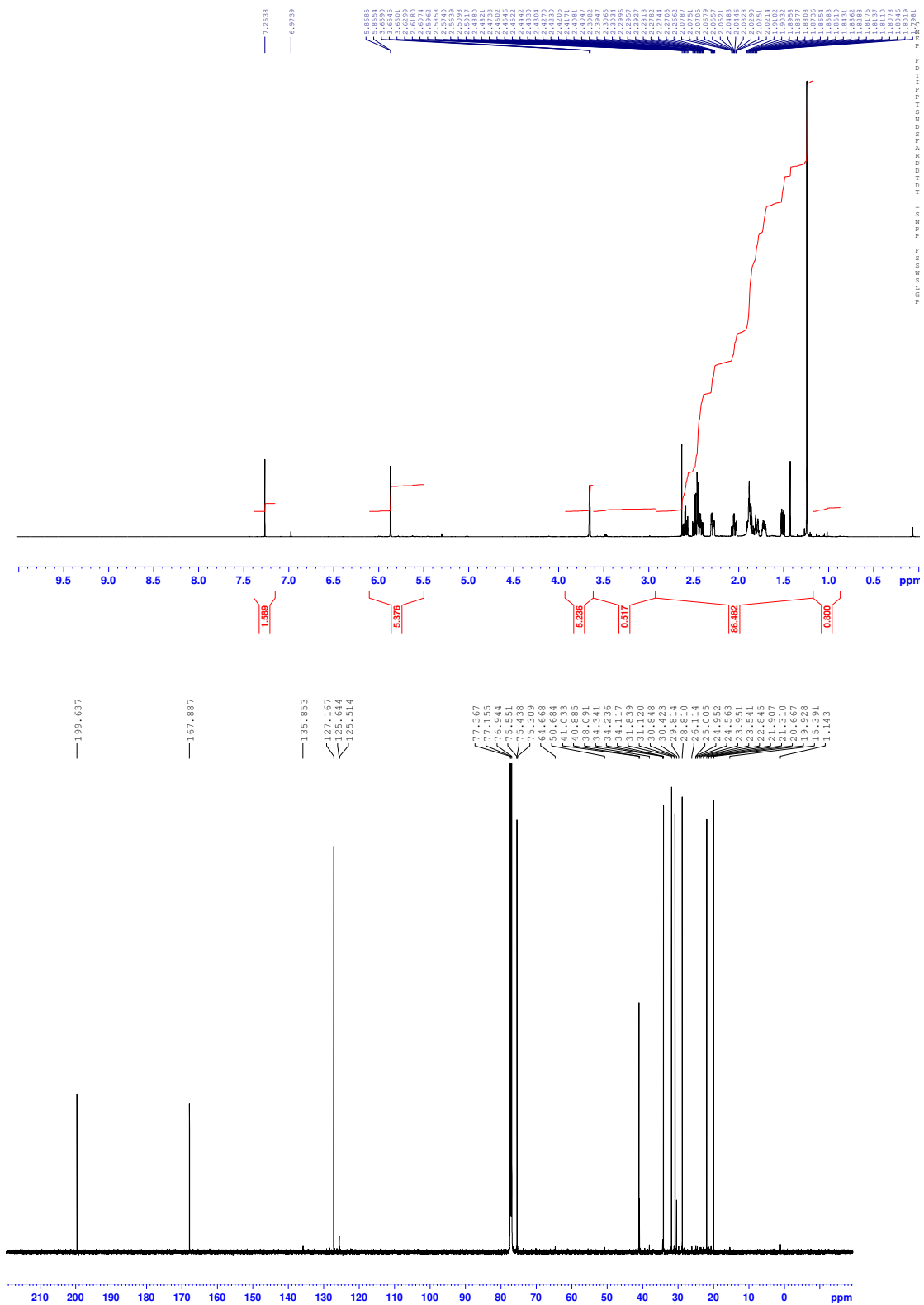


Figure S7. ¹H and ¹³C NMR spectra of (4aR,5R)-22.

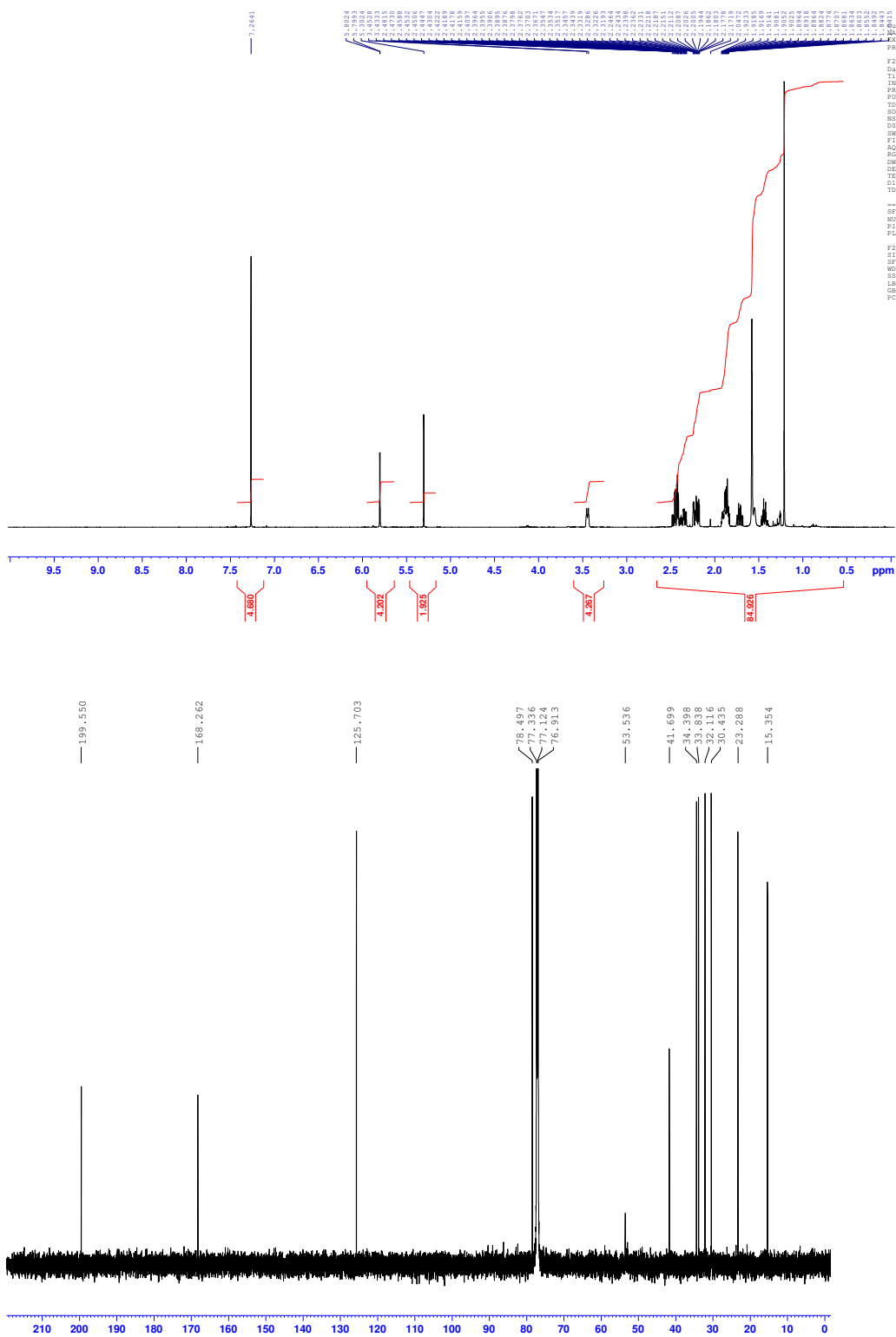


Figure S8. ^1H and ^{13}C NMR spectra of (4a*S*,5*S*)-22.

