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, Jack W. E. Jeffries, Thomas S. Moody, John M. Ward and Helen C. Hailes

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 (4aR,5S)-22 from the scaled-up reaction using SDR-17.

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

Figure S8. ¹H and ¹³C NMR spectra of (4aS,5S)-22.
|   | pQR   | Enzymatic Activity                          | Organism                  |
|---|-------|---------------------------------------------|---------------------------|
| 1 | pQR500| 3-oxoacyl ACP reductases                    | *Porphyromonas somerae*   |
| 2 | pQR501| 3-oxoacyl ACP reductase                     | *Neisseria* *sp.*         |
| 3 | pQR502| NAD(P)-dependent oxidoreductase             | *Veillonella* *sp.*       |
| 4 | pQR503| Acetoin reductase                           | *Mycobacteroides abscessus*|
| 5 | pQR504| NAD(P)-dependent oxidoreductase             | *Streptococcus parasanguinis*|
| 6 | pQR505| Glucose-1-dehydrogenase                     | *Megasphaera micronuciformis*|
| 7 | pQR506| Oxidoreductase                              | *Veillonella* *sp.*       |
| 8 | pQR507| 3-oxoacyl ACP reductase                     | *Streptococcus infantis*  |
| 9 | pQR508| Oxidoreductase                              | *Rothia mucilaginosa*     |
|10 | pQR509| Oxidoreductase                              | *Rothia mucilaginosa*     |
|11 | pQR510| Oxidoreductase                              | *Rothia mucilaginosa*     |
|12 | pQR511| NAD(P)-dependent oxidoreductase             | *Actinobacteria*          |
|13 | pQR512| NAD(P)-dependent oxidoreductase             | *Actinomyces odontolyticus*|
|15 | pQR513| Oxidoreductase                              | *Atopobium parvulum*      |
|16 | pQR514| 3-oxoacyl ACP reductase                     | *Rothia mucilaginosa*     |
|17 | pQR515| 3-oxoacyl ACP reductase                     | *Streptococcus parasanguinis*|
|18 | pQR516| NAD(P)-dependent oxidoreductase             | *Neisseria mucosa*        |
|19 | pQR517| Oxidoreductase                              | *Streptococcus* *sp.*     |
|20 | pQR518| NAD(P)-dependent oxidoreductase             | *Prevotella melaninogenica*|
|21 | pQR519| 3-oxoacyl ACP reductase                     | *Prevotella* *sp.*        |
|22 | pQR520| NAD(P)-dependent oxidoreductase             | *Prevotella* *sp.*        |
|23 | pQR521| NAD(P)-dependent oxidoreductase             | *Prevotella* *sp.*        |
|24 | pQR522| NAD(P)-dependent oxidoreductase             | *Streptococcus parasanguinis*|
|25 | pQR523| NAD(P)-dependent oxidoreductase             | *Prevotella pallens*      |
|26 | pQR524| 3-oxoacyl ACP reductase                     | *Neisseria* *sp.*         |
|27 | pQR525| NAD(P)-dependent oxidoreductase             | *Prevotella salivae*      |
|28 | pQR526| NAD(P)-dependent oxidoreductase             | *Oribacterium sinus*      |
|29 | pQR527| NAD(P)-dependent oxidoreductase             | *Veillonella atypica*     |
|30 | pQR528| Enoyl-(acyl-carrier-protein)                | *Porphyromonas* *sp.*     |
|31 | pQR529| Oxidoreductase                              | *Actinomyces graevenitzii*|
|32 | pQR530| NAD(P)-dependent oxidoreductase             | *Prevotella* *sp.*        |
|33 | pQR531| 3-oxoacyl ACP reductase                     | *Prevotella* *sp.*        |
|34 | pQR532| NAD(P)-dependent oxidoreductase             | *Prevotella histicola*    |
|35 | pQR533| Oxidoreductase                              | *Haemophilus parainfluenzae*|
|36 | pQR534| NAD(P)-dependent oxidoreductase             | *Prevotella melaninogenica*|
|37 | pQR535| Oxidoreductase                              | *Haemophilus influenzae*  |
|38 | pQR536| 3-oxoacyl ACP reductase                     | *Veillonella* *sp.*       |
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     | ATGAATCTACTTGTCAACAAGGGTAGGCTACATCACCGGCCAGGCGGTGATATCGGCGCTATTGCACTCAAGTATGCACAGGAAGGTGCTTCCGTAGTGATCACTGACCAGTGAATCTACTTGCTAACAAGGTAGCCATCATCACCGGCGCAGGCCGTGGTATCGGCCGCGCTATTGCACTCAAGTATGCACAGGAAGGTGCTTCCGTAGTGATCACTGACCTCAAGATCGACGAGACTGTAGAAGCCTTCGTCAAGGAGCTCGAAGGGCTCGGTGTCAAGGCTAAGGCTTATGCCTCGAACGCAGCTAACTTCGAAGATGCTCACAAGCTCGTGAAGCAGTCGTCGCAGACTTCGGCCGTATCGGACGTCCTCGTCAACAACGCTGGTATCACCCGTGACGGGCTGATGATGCGTATGACCGAAGAGCAGTGGGATCTCGTCATCAAGGTCAACCTCAAGAGTGCCTTCAACCTCATCCACGCTGTCACCCCCGTTATGGTCAAGCAGCGTAGCGGTAGCATCATCAACATGGCCAGCGTCGTCGGTGTCTCTGGCAACGCAGGTCAGGCGAACTACTCCGCTTCTAAGGCTGGTATGATCGGTCTGGCAAAGAGCATCGC
| 2     | ATGAGTACACAAAGATGGCAGGAAATTTGCTTTTGTAAACGGGCGATCGCCGCGG TATTGGTGGCAGGCAGATTGGCATTTTGGCTGTGAGGGTATACTTTGGCTGTGACACATGGCGAAATGGCAGACGAGGAGGTAGATCGGCAGCCGCTATTGGGCGAGGTACAGGCTGC TGGTTGTCTAAGGAATATGGCGTGCGTGTTGAGACTTTAGTTCTTGATGTGCGTAGTCGCGAGGATGTTGAAAGCAAGGTTCTGCAGCTATCGAGGCTTTTGGTGGCGTCGATGTGCTCGTTAATAATGCGGGTCTTGCACAAGGGCTTGA
| 3     | ATGGCTCATAATATTTTTGTCACTTGCTGGAAGTCTGGTGATATGCTGTATTTTGTAGTTATCAGGCTCAACTATATGCTGTTGATGATGCGGTGACTATGATTGATACCAATGTGAAAGGCCTATTATGTAACAAAAGCAGTGCTACCTTTTATGATTGATAAATGAAG
| 4     | ATGTCTAAAGTAGCTATTGTTACAGGTGCAGGTCAAGGAATCGGTTTTGCAATCGCAAAACGCTTGGTTCAAGATGGTTTTAAGGTTGGAGTATTAGACTACAATGCTGAAA |
5

6

7
TTGTCAAC
GTTTCTGTTGCAGATGGCAGCAGCTACGCTCTGGCCTGGTTGACTGTTGCTTACGCA
TTCTTCTAAGGCTGGTATGAACGGCTTTACCAAGGCTATCGCTAAAGAGGGCGGCAA
GTACGGCATTTCTTGGTAACGCTGTTTGCCCATCCTTCACCATTACAGATATGACCAC
TGCACTCTCCAACAGGCAATGGCTGCCAATAGTACTGGTCTCCACTGGGAA
CCATGGCTGATTCGGAATCCAGAATTCGAGCTGCTGGTGCAACAGCTGGCGCCACGCACG
GCCAACACACCACCCACCA

16
ATGAGCGGAAACAACCTTCCAAGACCGGTGCTGTCACCGGCGCAACCGGAGGACGACG
GCTACGAAAATTGGCAAGAATTCCGACGAGGACAACAGACGTCGACATACCCAC
CTGATCGGCTGAAAGCCCGGAAGATTTTTCGGCCTAGCAGGAGATGTCGAGCATGCG
CGAGCACATTAAAGGAAATTCGACATAAGGAAATGGAACGAGTTCGACGTCCCGTCAAG
GTGCTGTCGTCGCAATTGCTATACCCCGGCGACATGCTGTCGACTGGGAATTGCAAAGA
ATCCAGCTCACGCGAAGTACCGGTACATATTGCATACCTGGTATCCAGTCGAAAGCCGA
ACCAGGTGTCGGTCAATTCGCCAATTGGCGGCGTATCTCGTATTTCCCGGCCAG
CCGCGCATCAAAGGCGATGTCGAGCTCAAGGGCCTGCTGATATCGGATCCGAATTCGAGCTCC
TGCACTCGCACTCGGATCCAGCAGCTGCTGACATGCGGCGCCCTTGGGCCCAAG

17
ATGGAACTTAAAAAATAAAAAAGTTTGGTAACAGGTTCCAACACCGGGAATTGGAATT
GGCTGTGCTCAATAAATTGCGGCTCTGCAGTCTGGTCAATGGTCTTTAAATGACGCCGT
TGAATTGATTGGACACTTGTGGTCTGACAGTTTTGCTGAATATGGTCTTGTGTTGG
TATTCTGGGATTTATTCATTGGCGGAAGATCGCAGAACATGCTGACATGTCGAGCA
TTGGAAAAGCTTTGGAAGGTCTTGGTTTCAAGGTTGCTCTCAAAATCCTG
GACCGGTGTCCTCAATATGACCTCAAGCTGTCTTTAAACCTATCTGTCTAACAGCTGCTCA
AGTGCGCATTACATCAACATGTCCTCCTGTAGTATATAGGGAAATATGGCTCAAGC
AAACTATGGCATGCCTCAAAGGCGTCTTGGATTGGGTTACAAATCCGGTGCAGTGGTGA
AGTTGGCCTGGCCTGGTGTTTGGTATGGGAAATGCCATCTGGGCAAGATGACG
ATGGCAACTTACGCCAAGTACCGGTACATATTGCGGCGAGTCTGGTACTTTGGGA
CCGTTGGAATGGGCAGACGCTGCTGACATGCGGCGCCCTTGGGCCCAAG

18
ATGGCAATTGATTACCGGCTAGTACCGGCGGCAATGGTGCCTGATCTTGCGGCGGCAATGGTGCCTG
TTTGGCTAGGCGGAGCTATCGCTATCCGTGACCCTGGCCGAGCAACGAAATTGC
AGCATGGCGGAGGAATTTGGGCGGACGATTACTTTTGGGAATGGTCTTTCG
CGGCAAGGCTGATTATTACAAAGTCCTTGAAAACGCTGGCAGCAGCTTGGCGGCAAT
CATTGGCATGATCCAAACCACCTGAGGTTGGGTTTGATCCGGCTAGTAAAGCC
CGATTTTGGGCAGATTGGGAAACCGATTACAAACATCATCCGGCTGACTTTTCT
AACCAGGCAAGTTTGGGCGGAGATGTAACGCGGCAAAACAGGGCTATATACCATAAAT
TTGGGCTGATGGCTGCATTACGATGCTGCACTGGCAGTAAATTCGACTGCCGGAAGAA
AGGGCTTCTGGCGTCAATTCGAGCTAACCTGGCGGCAATGGGCAAGTAATAAT
ACCCGATTACAAATATGAGGCGGGTTTGCGGCGAACAGGCTTTTTTGGTCATGGGCTG
CGGTTGGAATGGGCAGACGCTGCTGACATGCGGCGCCCTTGGGCCCAAG
TTACCCGAGACGAGGCTGCGCCGCAAGAAGAAGAAGCTTGGGAAACGATGAT
ATCAACCAATAATCCCAATATCTCTCATAATGACATCTCTGGTCAGGGCGTAAAGGG
AATGGATAAATTGTACGACTTTGCTAACCGTATGGCGCCACTTGGTAAATAGGACTACTGATTGTTATGTTCTCAGACTTACAAAGAAAGTTAC
AATGCAAAATCTCTATCACGATGGCGGTTTCTCTAACATTGGTATGAGTCTTCGTGC
AATGACAACCTATGAAAAAGGTATTGGAGACGAATATAAAGATGAGAACGGAAAG
ATTATCTACCGGACATTGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAA
CTTGCGGCACACTGACACCACCAACCACCAACCACCAAC

26

ATGAGTGAAACTTTTTAATTACGGGTCTAATCGCGGTATCGGCAAAGCCGTCG
GCTTGGTTGTTGCGTGGACAGGCTGTGGTTGTAATTGGTCTCCACCTGGCTAGCCGCCTGA
TGAAGCGGAAGCTGCTGAGGAAAGTATGTGCTGACTTATAGGATGAAATAGGACTACTGCT
TGAATATTGCTGAGGCAAGCTGCTGAGGAAAGTATGTGCTGACTTATAGGATGAAATAGG
CTGCATTTCTCTCGTAATTACGGATGAATGGGTGCTGAGGAAAGTATGTGCTGACTTATAGG
GCTGATATCGGATCCGAATTCGAGCTCCGTCGACAACTTGCGGCCGCACTCGAGCACC
ACCACCAACCACCAAC

27

ATGGCTTATAACTTTAATTTTTAATTTTTGCTGCTAGTGAATGATA
TGCTAAATTGCAATGAAAATGGCGCATACGCGGACAGAGAAGTGTGCTGACTCTCGCT
TGAGTAAACATTGCAAATGATACGCACTTTCGTAATGGTCTGACTTATAGGATGAAATAGG
TGAATATTGCTGAGGCAAGCTGCTGAGGAAAGTATGTGCTGACTTATAGGATGAAATAGG
CTGCATTTCTCTCGTAATTACGGATGAATGGGTGCTGAGGAAAGTATGTGCTGACTTATAGG
GCTGATATCGGATCCGAATTCGAGCTCCGTCGACAACTTGCGGCCGCACTCGAGCACC
ACCACCAACCACCAAC

28

ATGAACATTGCTATTTTACGGGTCATCCCGGAGAATAGCTGCAAATGACTCTGGCTGACAC
ATCAGCACAATTCTCTTGCTATAAAGCAGACATGCGGCTAGGAAAGAGATTGATATTGGAAGATA
TACGACGTAGTTTCTCGGCTATAATCTATGGAAGAGAGATGCGTATTTTCTCGTAAATACCA
TCTCACAATCTCCTACAGAAAACATTGCAATTGAAGACAAATGCGTAAATACCA
TGATGAAATTCTCTGCTAAGAGTATGCTCACCCTCTCGGCAAACGCTACAGCCGATGAGTGTG
CTGATATTGCTGAAATCTGGAATGATCTGATCAGTAACTCAACCCGTAATGGTAAATGGAGACCCT
TTCTCAGATGAGGCTGCTCATTACAAATGGGTATGACTGTCCTGCTATGACACACATAT
ATGAAAGCTCTTTTGAGATGACGCTGATATCGGATCCGAATTCGAGCTCCGTCGACAACTTGCGGCCGCACTCGAGCACC
ACCACCAACCACCAAC
33 GTGCCGGGCTACATAGACACGCCACGATTGAGAGAGATATACCGTAATGACCTTTCC
AAATGTCCGTTCTGCATGCCCTTAAATCGGGCTGTTGAAACAATGACAAAGGCCATA
GCGCAGCGCAAAGAACAGGTGGTGTTTCCGCCAAAAATGCGCCTTTCCATCGCTATT
CTTTCGCTTTTACCCACTGTGTTTGTATCCTG6CTTTTATGACCGAAAGACATTAGGAG
CATTCCATGGCTGATATCGGA
CAGCTGATAACGTGCGACATCGGACTG
ATCGGACTGCTCGACGTCTCGACGTGATAGTGCAGCAGTCCGAGACTGGAAGACAT
34 ATGAGAAAAACAGCATTAATTACTGGTGCTACCAGTGGAATTGGCGAGGCTTGTGCA
GGTAAACTTCGACCAAAAGGTGCTAATGTTTATCAACTACCTGGGCTGAGGAT
CCGAAAAGATAGTGGACTTTGAGAGATGAGGATAGGATGAGGATAGGATGAGGATAGG
CATTATTATAATTAGCAGCGTGTGAGGCTGACTCGTACGTGCTGCGTACGTGCTGCGT
35 ATGGGAATATTGGATATCGGATCCGACTCCG
GTCGACAAGCTTGCGGCCGCACTCGAGCACCACCA
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     | MNLLANKVAIITGAGRGIGRAIALKYAQEGASVITDLKIDETVEAFVEKLEGLGVKAKAYSNAANFEDLKLYEAVADFGGRIDVLVNNAGITRGRMMLMRMTEEQWDLVINVNLKSFANLHAVAOTPVMQKRGQISNNVMGASNAGQYNSASKAGMIGLAKSIAKGELARGIRANAIAPGFIITDTMGALSEEVRKQWEVQIPLRGGTPEDVANVATFLASDLSYYVSQGTTIHCGGMNM |
| 2     | MSTQDLSGKIALVTGASRGIGGAAIADTLAVALAANVKEKAMMTSAAISGAAAIERSQVAWAGGRALNSAEPETIENLIAIDIEKEFKGDLVLYNNGAITRGRDNLMLRMKKEEEDDDIMQVNLKSVFRASKVLGMMKRQAGRNIITYSVVGVGNAGQTNNYAAKAGLIGFSKSMAREVGSRGTVNCVAPFIDTDMLPETRKFEAQTSGLTSEQADAVLFLASDQAKYTGQTILHNGMLMPWLIBXSEPQHT |
| 3     | MAHNIFVTGATGSIYGLCIAEAYAKHGNVLSLRARRAELGEFVQARSKEYGVRVETLVDVRSREDVESKVPAIEAGFGDVLYNAGLQAQLDFQDSAVDADVMTIDTNGKGLLYYTKAVLPMIDNENHVINMGTAGIYAPNYGVATKAATVIKTVSLGDVMDTITDIKIVTTIPQIVETPSFESVHFHDAERAKSVYAGIDAIQPEDVADVLYVTNQPKRLQISDVTIMANQAGFMV |
| 4     | MSKVAIVTGAGQGIIIFAKRQVLQDGFKVGLYDNYAETAEMKAVALADKAVICVAWAVYYFQKVDGFLVNNAGLAPARTPDDLTITIQFQRTFANVGVGWIWQAAQAQFKALGHGKIATTSQAGVGPLNTLYEGTFVQITGQTLARLDDSGITYNAYAPGIKTVPTIMDAHVEYKGNKSKIDGDEWMQTFADKIDLKRLSEPVEDVAAVSFSLAGSAITGQTIIIDVCGGMQFH |
| 5     | MSETILVTGASAGFQGICRRLVADGNYVRGSGARRIDKLAQLLOYEEGAFAYPLQMDVTDLSDQVHDHLSLAKWKEVDVLVNNAGLALPGAYEAEDWERTMQNIVGLTYTRKLPQFVLQHGGYNLHSTAGTGVYPPGANVYGASKAFQFSLNRADLGKIRVSNIEPLCGETFSSVFRKGDKEIKERAEYLYNQDHAIQFSEDIANTVAVLIIQPKHVNVRNEIPMSQTGQPPQPPVRV |
| 6     | MYSELKGGKAVATGGSKGIGTAIAKRFGQEMGKVVINYNSDAGAELAAEAVRCAGGEAIVKGVTEGGQSLVDAAYENYSGIDVWNAGMENKVCTHPEMPSLIVERVINNLTVGFLETRALTYMFREDVKGHSGNMNMSHEQPWPWTAYHCASKGGJIKLGFQRTQVAMEYAKYGRNVNGPAPIANTPAKFSFDPQYETETMSVMPMKRGKPEEEVAAVCAVLASDEAESVTGITLFVDDGMJTLYPAFDQK |
| 7     | MPHKNDVQVALISGSGTIGGTAKLLQLQEGWCVVINGRDEQAGQRAKMKRLRYSSKVRYYKGDVSSVSDCQIEVTDFFGGSISATVAAGYEEELADSEASDFEGMTNVKGGTLCVQALPRLYQVRKGSIVTVSSDAGLQNVACSVYGASKGAIIVTSLQLEMAPHEVRVNCVCPGDVDTLSVDKQIAQHSDQAEEKMAGQHYPLGRIAKPHIEGAFISSTGFCVTVATWDGGLTS |
| 8     | MTKRRVLTGVSQSGGLAQARLFLFENDGYQYVGYDVQDQGKPDQNGFHLFQRDLTLDLIEFDWCPQVSLTCNTAGILDDYKLSEQMESAQEIFEINEYTVPVELTVRLTYLQTMELYKRTTIINCMSSASLQGGHAYTSSKHALAGFTQKQLALDVAAGYIQVFGIAQSKGVTGMAADDPEFPGGLADVWNAVSETPIKREWIEEVEAEVSLFLASGKASAMQGQTLIDGGWSL |
| 9     | MSWVIVTGANGGIGEATVHLKINGYSVFAADLAEOAPIAFGTYGDARFRVARYDVTEESVTALAEVADLEPTAVGGLAAGIHASQPLLETSTFWAKRLHAVNSDVFLCLCREFARIMIDQESQDPSRDSLTVASNARVPAEFGAYGASKASAARRVSSSFLQLAAHGYRNSVCPTGTRTPMVTDNAMESPACEGDRSQLPAVANQPTFLGIPIGLGRIADPADIAAVNAFLISEAARHTIMEQIEVA |
| 10    | MTENTHIPSFLPHTGDGKVAVTVGASSGIGRATVQQLVLASGWTVYALARTRDRLTYLAETGAHPVTCVTFDEQSVQFVAEQILEEGTIDALVNIAGAIGVDKVAEGKPD |

S15
LKSAFNFIHACVPVMRQRNGSIINMASVVGHSVQGQANYAASKAMALAKSVAQ
EMGPKGIRANAIAPGFIDTAMTQALNDDIRKEWTSKPLRGLTYFCNFCA
AGSKISHRKCDQRHLDYRDTMDLNNYRSVLASLAILPALKASKGIIYSSSVSTLY
MAPGWSAYHASKSAANTWCEATANSEFAPLGVHVQIALPLVHTAMSDVNEQYKHL
AAYTPADAANILKCLAIRKVTRYPWWAKLSPIALYPHIPHHYKYR

MMTSLRKKIAGGALAYPSLTDTPREHFSGBKWLVTGASHGIRLACTEKINAGAN
VFLIARSEADLRLAACKAMQGMSADYCAIDLDRKREDLEQLCQKLRELTPRLDYFCN
AGSKISHRKCDQRHLDYRDTMDLNNYRSVLASLAILPALKASKGIIYSSSVSTLY
MAPGWSAYHASKSAANTWCEATANSEFAPLGVHVQIALPLVHTAMSDVNEQYKHL
AAYTPADAANILKCLAIRKVTRYPWWAKLSPIALYPHIPHHYKYR

MKKAIYVGASSGIGHEVARLIQAGWAVGVAARRIDKLDLQAMAPAVERVYTAQIDV
NEDAEHTSLLIQERMIANGHAVWGQPNLNADEILKTMENNAVGFIRMCAGY
RFYANKGCHHACTISIAATRGKLPAPAYSAKTAMQNTYLAQLEQAAKCWHKNHFT
DIRPGVFDTPPLAGTSHLPMLMTEKVARSIHAKINSSRHCIVIDSRWCVLTY
LWRHIPNWIIWRMRMCQ

MKMEDKKIIGGATGGIRGQLKELAQAQLREHLILVSRADDKLQVQKELTGSKAQLS
ILTLDLMDQVWALEVESLDALDVNCAGLAYSFRESLDSEASEQDLWQVNYHSQV
LIDQKNQKQKIQVQKLQVQSLALLFAPLPHLYAASAKQAALQTFTLQLQEEQVFQV
LQLYILVPVQAIFPPKLLELQGGRSLQMKSEKVAQQFLRIERDTSYEEI
GLRYLTVLWVLRLPRQWIIAILLAKKGT

MSYNNLKKGKRGIGFALGNEQSIAWKVAERAEEVAGASITLSNTPIAVRMGTNSLAELK
NCEVIAADATSVDELVENFVKSMEILGKIDFVLHSIGMSPIPEVRKHRYDDVNYMNL
NTTLDISAYFSFHKMQSAAKLDISDAYSILSYLIQAARFTTGFYMDADAKALESIA
RSFYGIGERYANKTVRSIPSTHTAQGGVKMKDLDFDNAMRAPLGNASAEACAD
YCVIMFSDLTKVTMQLNYHDGGFSNIGMLSRLAMTTYEKIGDEYDENGKIIYG

MSETILITGSNRIGKAVAVALGLAQDFGDFIVHCRRSRDEAEAVAAEIRALGRNRANVLQ
FVDSVSDREACREILTDIADNNTYVYGVVLNLAGTRDNAFAFTDDWDVLRLRNLDEG
YVNLHPLTMPMIRRRAKGRCMASVSGLTGNRQKVNYSAKIALAKALAVELA
KRKITYVNCAPGLDITDIDENPVPEEILKAVALMGLPEEVAHAVRFLMDE
KRAYITRQVIADVNGLC

MAYNLKLCGRGVIFGFALNMDMSIAWIKVAERAEEAGATLVLNSTEMALMRGSDELSEK
KINAPVIAADATSVNDFVKEQELGGKIDFVLHSIGMSPVRKHRYDDVNYMNL
WNKTLQDASAIFHKLQAANKDVAIAEYGVVALTYVASHRTTFGYMDADAKALESI
ESIARSEFGITYANKTVRSIPSTHTAQGGVKMKDLDFDNAMRAPLGNASAEACAD
ADVCYTFLFSDLTKVLMHTDDGFYGGFSNMGSLRANMQSTIKLD

MNIAYIITGASRGIGKAAAKRFAREGYSLLLNNECNKWTLLEELKKEIQLDSLPENCPEIFLC
KDLGTKKGSLRELEKSLKLILIANSGKDAIKLLQDCEEEKALLETNLLQPFLCQK
LLPYLLQQAEERGRILSSV gruntVNASMESLYSTLGKIGISF KALGKELASPHISNAV
AFGAVDTDMSNWLSTEEKQSSLGEGIPYGGRMTVEEADAFLYLLSQAFLYTAQV
PFDDGW

MNRKFIVTGATSGICLEARAFAAQDFGDNVLIAIGRARDLAAIKEDFQQYGIRVDTLY
LDVSVREDVAKVPAAEFGGIDVILVNNLAGAQQLDFPDQQSTDEDAVMTMINVLG
LLVTKAFLPMMAQSGHVNMGSTAGIAYYPGAYVCTAAVMKLAGDGRMD
TIAITDIKVTTPQGPIVETPEFSEVRFGDAAERASVAYGIEAQVFEDPVADVLYVLTQPK
RLQIS DVIITMANNQQAAGFM

MSYNLLAGKKGIGFALNMQSIAWRVARAEEVEAGAEILTTNATAVVRMGLQNLGKQL
NAKVPADATEKEELEVVQFEMAASFGKGDVFVLSHIGMSPNVLRKGRYDDLVDKML
QTFEDISAYFSFHKMQSAAKLDISDAYSILSYLQAARFTTGFYMDADAKALESI
NSFSGITYANKTVRSIPSTHTAQGGVKMELDFFDFAEDSLPLGNANDANDA
DYTICLTFSDLTKMUNQLNFDGGFSNMGSMFAQIEAFAFGSNAKGRNEK

MTKHLRAVVTGASSGIGWATERLVTENGWVQVPSRTGQVPEGSALVSAADAEDGV
GEAITHQANLGGDVAYVGAAGSTQEYQLARADLEQVNTQLRLHYLSNYEAISSL
GMVRARWGRIVLLSSVVAQSGMAGLSAYGAAKGALEALVKSLALEVGRRAITVNAV
APGYIQTPMTQSLSPRQQERYLQRTGAARPGTPQDVAGPVAFLLSDDAAYVNGQILHV
VGDMAGVGNP

32
MNIFITGGSIGGIARFYAAKGHVRVGCGGRANTIRIKSDSDEVNKLLLLAYQLDVICDK
DALTVAVEFCAKGLDMMIVAAGYRNNGVTEEVDFEQTSQMLKVNAAGALNAME
VAREAMNASHGGLVIALYHCASPSVYALKRALKALIQADAYRSLAYQITVT
TLVPYGYIDTPRLREIYRNDLSCFCMFPLNRAVEATMKIAAQKEQKVVFPPKMLS
IAILLLLPCLESAFMHRKTLWSI

33
MKYALITGASRGISVALLARERYSIIANYQSNAEAAQAVKQEIETKGGHVEVLLFLDV
SDPKAIEAADKHDPEFSMDFMSDSDWVSLDTNMGF
YITRRLKHMMPKRERGRIINMASLGLKGLPGQVNYSAAKAALIGATKALQAAAR
RKIVNNAVAPGFIQTDMTKELPEDELKLVVPVRGFPTPEEvADVAVIFLSDAAAYITG
EIVNNGGFFY

34
MRKTAIITGATSGIGEACARKFAQQGYDVIITGRNKQQLAAKVELETGETKVLALAF
DVRNRAAAATKIAKIDVLINNAGLALGLEPEYGFEGDFEWDITMTEYIKGL
TMTRLLPVKPMERNSSGHINIGSVAGDAAAYAGNNYCATKAAVTKITDGLRIDAHTA
VRVTNVPGLVEHFSNRFVGDKRANSYVHYIKPLTGDTIAADVAYAASAPAHVQ
IAEVLVLATHQSGSSVIIH

35
MGFLTGKRILVAGLSNRSAIYGIAMKEQGALFTYNLKDQLPVEEFAKFGSD
IVPLPDVATDESIOCNCFEALKSRWKEFDGVHAIAPQGDDQLDYVNAATREGYRI
HDIFASFVAMQAAPRYLNPNAAATLTSYLGAERAIPNYMVMLAKLAEAAVRM
AADDLGEKQIRVNAISAGPIRALTAAAGKNFKMFSAFEKTAAALRTTVTIEVDVNSAAL
CSDLASGITYEIPvHVGKSITAMGELGE

36
MAKVMVTGANKGIGYICKFLGKSGWQVIFVGARNSERAAEAMKSLKAEGVDVIGW
QYYVLNDNASLEQTAKEKEYHDLELLVNNAGIPDMKVASLKDVIDTVQVN
YVGFTFLKTAKLPLSSANKGRIVNITVPEVSPYWHPMAYVASKAAQANMTSIMAA
FEKNNIPVEINHPGATTTDLNNHYTGPShIDVVEKIAEVINGKKhQGEFVLEYP
IVDEGR

37
MRQTALVTGATAGFGAAICRRTLIENGYRIVGTRRVARLEQLQQLNELGENFHFHLADIS
DROQATADAIFHLPTNWSIDLVLNNAGIGGGLGLESADKASLDDQWMDQIMITNKGLVII
TRLVLPQVSNSSGRHIGLISAITGYPYPGNNYGTGTKIIFKQFSLNRADLAGTQRIV
NTVEPGLCGGTENSNIRFKGDARAKLKLYENVEYVSPQDIANIVLWNLQNP
EHNVNRIEVMPTQFTAPLNVAR

38
MHLEGKVVALTGASRGIAVIQQLAPQSGADVADAVNYGSEGAAAQETVDALALGR
KAIIYVSKMFKQRGHIINMTSVDVGLMNAGQQANYAASKAGVIGFTSKCAKELAS
RGITVNIAAPFINTDMTVDLPKEVKEAMVTIPLGMRAKEEAAVTTFLASDF
SYITGQVINVDGGMMKANVANAEAEEAAMVEETHKTFGHIDILVNNAGITRDGLLMR
MKDEXFDADOINLKGVYLTVKA
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 | TAAGAAGGAGATATACATATGAATCTACTTTGCTAACAAGGTAGCC |
| 1 RVR | ATCCGATATCACGCACTGGCCATATTCAATCCCACCAGGACAG |
| 2 FWD | TAAGAAGGAGATATACATATGAGTACACAAAGATTGAGCGGCAA |
| 2 RVR | ATCCGATATACGCACTGGAATCAGATACAACATAGCGGAGC |
| 3 FWD | TAAGAAGGAGATATACATATGGCTTCAATAATTTTTGTCACTGGTGC |
| 3 RVR | ATCCGATATACGCACTGGCCACCATCAAAGCCTGCGCCGTCG |
| 4 FWD | TAAGAAGGAGATATACATATGTCTAAAGCTAAAGCTTACAGGTGC |
| 4 RVR | ATCCGATATACGCACTGGCCACCATCAAAGCCTGCGCCGTCG |
| 5 FWD | TAAGAAGGAGATATACATATGCCTTATAAAAATGATGTTTCAATGGC |
| 5 RVR | ATCCGATATACGCACTGGCCACCATCAAAGCCTGCGCCGTCG |
| 6 FWD | TAAGAAGGAGATATACATATGAGCACTTGAGATAAAGGTGCAGGAC |
| 6 RVR | ATCCGATATACGCACTGGCCACCATCAAAGCCTGCGCCGTCG |
| 7 FWD | TAAGAAGGAGATATACATATGACTTACAAGCTTTTACTGCTGGCC |
| 7 RVR | ATCCGATATACGCACTGGCCACCATCAAAGCCTGCGCCGTCG |
| 8 FWD | TAAGAAGGAGATATACATATGACTTTACTGACTTGAGGAACAGG |
| 8 RVR | ATCCGATATACGCACTGGCCACCATCAAAGCCTGCGCCGTCG |
| 9 FWD | TAAGAAGGAGATATACATATGACTTTACTGACTTGAGGAACAGG |
| 9 RVR | ATCCGATATACGCACTGGCCACCATCAAAGCCTGCGCCGTCG |
| 10 FWD | TAAGAAGGAGATATACATATGACTTTACTGACTTGAGGAACAGG |
| 10 RVR | ATCCGATATACGCACTGGCCACCATCAAAGCCTGCGCCGTCG |
| 11 FWD | TAAGAAGGAGATATACATATGACTTTACTGACTTGAGGAACAGG |
| 11 RVR | ATCCGATATACGCACTGGCCACCATCAAAGCCTGCGCCGTCG |
| 12 FWD | TAAGAAGGAGATATACATATGACTTTACTGACTTGAGGAACAGG |
| 12 RVR | ATCCGATATACGCACTGGCCACCATCAAAGCCTGCGCCGTCG |
| 13 FWD | TAAGAAGGAGATATACATATGACTTTACTGACTTGAGGAACAGG |
| 13 RVR | ATCCGATATACGCACTGGCCACCATCAAAGCCTGCGCCGTCG |
| 14 FWD | TAAGAAGGAGATATACATATGACTTTACTGACTTGAGGAACAGG |
| 14 RVR | ATCCGATATACGCACTGGCCACCATCAAAGCCTGCGCCGTCG |
| 15 FWD | TAAGAAGGAGATATACATATGACTTTACTGACTTGAGGAACAGG |
31 FWD  TAAGAAGGAGATATACTATGACTAAGCATTTACGCGCG
31 RVR  ATCCGATATCGCCATGGAATGCTCCATAATGCTTTTCGGTGC

32 FWD  TAAGAAGGAGATATACTATGGAACATCTTTATTACGGGCGGAAC
32 RVR  ATCCGATATCGCCATGGAATGCTCCATAATGCTTTTCGGTGC

33 FWD  TAAGAAGGAGATATACTATGGAAGGTATGCTTTAATAACTGCTTTC
33 RVR  ATCCGATATCGCCATGGAATGCTCCATAATGCTTTTCGGTGC

34 FWD  TAAGAAGGAGATATACTATGGAAGATGCTTTAATAACTGCTTTC
34 RVR  ATCCGATATCGCCATGGAATGCTCCATAATGCTTTTCGGTGC

35 FWD  TAAGAAGGAGATATACTATGGAAGGTATGCTTTAATAACTGCTTTC
35 RVR  ATCCGATATCGCCATGGAATGCTCCATAATGCTTTTCGGTGC

36 FWD  TAAGAAGGAGATATACTATGGAAGGTATGCTTTAATAACTGCTTTC
36 RVR  ATCCGATATCGCCATGGAATGCTCCATAATGCTTTTCGGTGC

37 FWD  TAAGAAGGAGATATACTATGGAAGGTATGCTTTAATAACTGCTTTC
37 RVR  ATCCGATATCGCCATGGAATGCTCCATAATGCTTTTCGGTGC

38 FWD  TAAGAAGGAGATATACTATGGAAGGTATGCTTTAATAACTGCTTTC
38 RVR  ATCCGATATCGCCATGGAATGCTCCATAATGCTTTTCGGTGC
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.

| SDR | Alignment |
|-----|-----------|
| 11  | MPKSLQRPVVL|TGASSGIGYDVAPPLLVRGYGVYGVAAARVEKIEELASE--------GVK |
| 3   | --- | |MAHNIYFVGATSGILCIAEAYAKHDGNVLISGRRAEALLGEVQARLSKEYGVRV |
| 17  | --- | |MQTALVVTGATAGFAICRITLENGYRVIGTGRVVARLEQLQQEL----GENF |
| 31  | --- | |MSKVALVTGAGQQQFGAFAIARLVQDQFKEVGLVDYMANAESAEKAVSLSAD-CAF |
| --- | - | |MELINKNVVFTGVSTRIGLAVAHKFAALGANVNLQRS-ESIESDALLEPA-DYGVTV |

** Additional alignments for SDRs 3, 4, 11, 17, 31, 37 **

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

| SDR | 11 | 3 | 37 | 4 | 17 | 31 |
|-----|----|---|----|---|----|----|
| 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 °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 (4aR,5R)-22. D. (4aR,5S)-22. E. (4aS,5S)-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: (4aR,5S)-22 Rt = 54.0 min, (4aR,5R)-22 Rt = 61.2 min, (4aS,5S)-22 Rt = 85.9 min. HPLC conditions were 4% iPrOH:hexane, 120 min, 0.5 mL/min at 230 nm, on a Chiracel OJ column.

Figure S6. Characterisation data for (4aR,5S)-22 from the scaled-up reaction using SDR-17. A. HPLC trace of (4aR,5S)-22 starting from (R)-18 using SDR-17. Rt = 54 min. HPLC using Chiracel 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 (4aR,5S)-22.
Figure S7. $^1$H and $^{13}$C NMR spectra of (4aR,5R)-22.
Figure S8. $^1$H and $^{13}$C NMR spectra of (4a5,5S)-22.