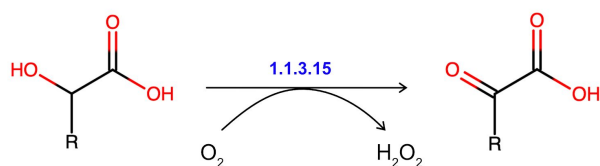
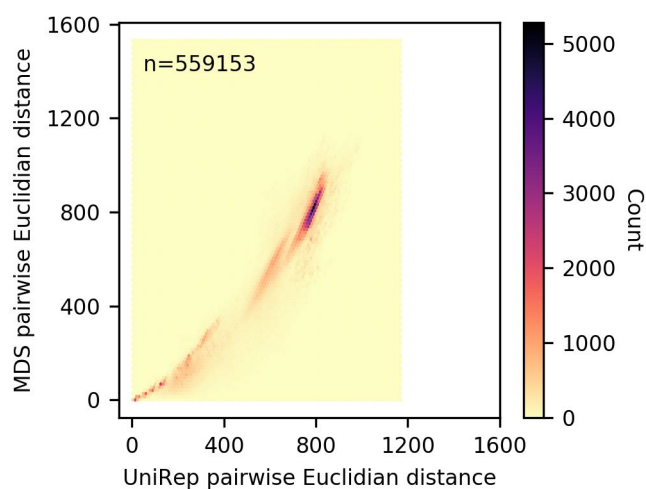


Supplement

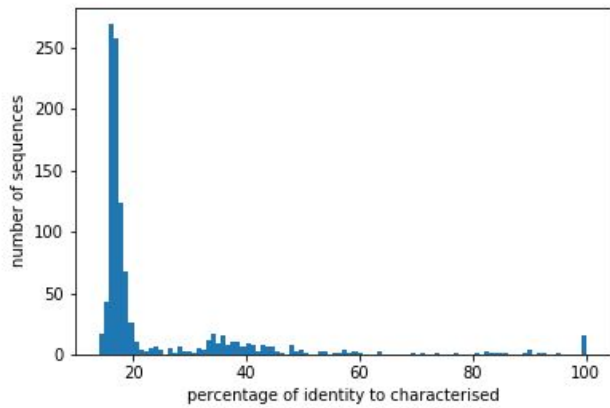
Supplementary figures



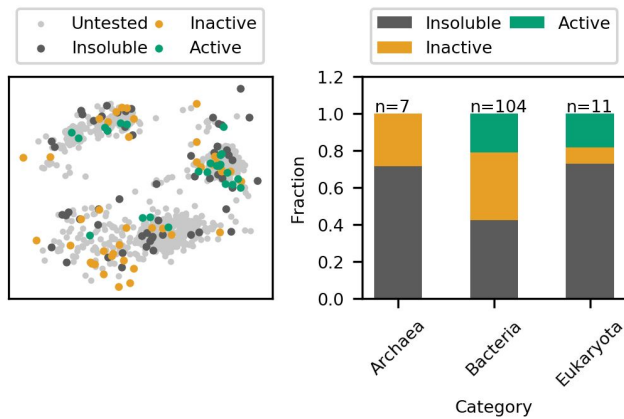
Supplementary figure 1 | Schematic representation of the reaction catalysed by S-2-hydroxyacid oxidases (EC 1.1.3.15).



Supplementary figure 2 | A hexbin plot indicating Euclidean pairwise distances between the 1058 EC 1.1.3.15 proteins. Clustering along the diagonal indicates that the multidimensional scaling (MDS) dimensionality reduction faithfully represents pairwise distances of the UniRep representations of these sequences. The total number of pairwise distances is indicated, corresponding to half of the distance matrix, without the diagonal.



Supplementary figure 3 | Identity of sequences annotated as EC 1.1.3.15 to the closest characterized S-2-hydroxyacid oxidase.



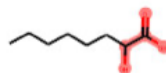
Supplementary figure 4 | (A) Distribution of the insoluble, active and inactive proteins throughout the sequence space. (B) Distribution of the insoluble, active and inactive proteins in the three superkingdoms.



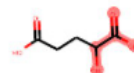
glycolate



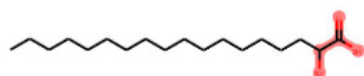
lactate



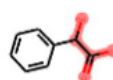
2-hydroxyoctanoate



2-hydroxyglutarate



2-hydroxystearate



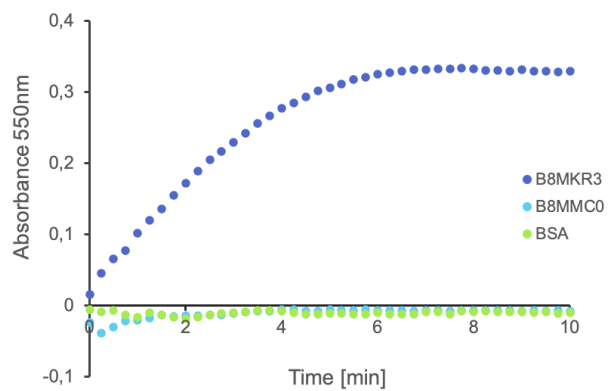
mandelate

Supplementary figure 5 | S-2-hydroxyacid substrates used for the screening of EC 1.1.3.15 sequence space. The donor group is marked in red.

[illegible]

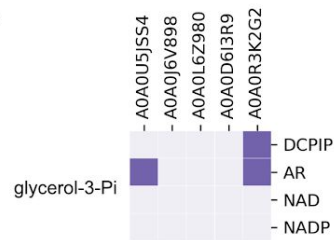
GOX_SPIO1 WQQTITS LPIIVKGVITADARLAVQ--HGAGIIVSNHGGRRLDYPVATIMALEEVKAAQ----GRIPVFLDGGVRRGDDVKALAMGAAGVFIGRP 310
 LOX_LACLA KKKKITTNPVIVKGVQSPIDADDANN--AGADGIIVSNHGGRRLDGGPASIDVPLAKSVN----HRVPIVFDGGVRRGHHVKALAGGADVVAVGRRP 337
 MDH_PSEFU WRDLWPHKLLVKGILLSADADRCIA--EGADGVIVSNHGGRRLDCAISPMEVLAQSVAKTG----KPVLLDGGFRRGSDIVKALAMGAAGVFIGRP 328
 LA2M_MYCSM WVRSTKMPVILKGIQHPDARRAVD--SGVDGIIVSNHGGRRLANGGLPADCLPEEVKASG----DTFVLFDSGIRTGADVVKALAMGASAVGIGRP 346
 CYB2_YEAST EKKKTKLPVIVKGVQRTEDVIKAAE--IGVSGVIVSNHGGRRLDPSRAPLEVAETMPILEQRNLKDKLEVFVDGGVRRGDDVKALAMGAAGVFIGRP 434
 X8B512 -----RRPG----- 237
 DIA2X8 WRREEWGQPFEMIKGVTRVDDAKRAVD--IGVTALIVSNHGGRRLDTPATIRLIPATAVG----DQVEVLLDGGVRRGDDVKALAMGAGAVLIGRP 336
 B8MCM0 YRKLTNPVIVKGVQSPIDAILAHK--HKVNGIIVSNHGGRRLDQAQAPLTLLEIKYAPQIITDKMQIFIDGGVRRGDDVKALAMGAGAVLIGRP 428
 B8MKR3 WTKVTNLPVILKGLQTHDDAYIASLYAPQVKGIIISNHGGRRLDTAPPAVHTLEIKRYCPEV--FDKIEVLVDGGIRRGDDVKALAMGAGAVLIGRP 429
 C4VMW0 KIKEMSGLPVIVKGVMMNADAYMAIG--AGADGIIVSNHGGRRLDTPATIDMPEFAAVN----GRVPIILDGGVRRGSHVKALAMGADLVGIGRP 321
 COXIJ3 RIANYTDLPIVVKGIESPDALYAIG--AGAGIIVSNHGGRRLHNGGPASFDVLEDAKAVN----GKVPVIFDSGIRRGDDVKALAMGADLVGIGRP 320
 C2K1F0 KQASYSLPVIKGIQHPADAVAAIT--AGAAGIIVSNHGGRRLDGPAPAEOLPATAAAVD----HRVPIIFDGGVRRGHHVKALAMGADLVGIGRP 319
 A9QH69 YLAQYSLPVIKGIQHPADAFRALE--AGAGIIVSNHGGRRLDGGPAAPFDSQEVAAVD----RRVPIVFDGGVRRGHHVKALAMGADLVGIGRP 327
 A4YVE0 LKDKFRPLIKGIATADDAIALD--HGVDNIIVSNHGGRRLDHGRGAMHVPPEVAAVK----GRKILVDGGFCRGGDIVKALAMGADLVGIGRP 309
 Q92NB8 LKLDSDPLVILKGIATVDDARIAVD--HGVDNIIVSNHGGRRLDHGRGTMDVPEPIDAVG--GQAKVMVDGGFCRGGDIVKALAMGADLVGIGRP 301
 E6SCX5 WRDQWPGRIIVKGIQHPDADAVAVD--LGDGIIVSNHGGRRLDRAATPEVLPDVAAVA----GRIEVLLDGGITDGGDIVKALAMGAGATGCLVGR 357
 C9Y9E7 WRREKGGKLLKGIQDVDDAVLAAQ--SGADAIIVSNHGGRRLDGPASSISALPATAAVG----DKLEVWMDGGIRSGDDVKALAMGAGKGTMIGRA 332
 W6W585 RDRDWTGPMIVKGVMMHADAIAALE--LGLNGVIVSNHGGRRLDPAAPATIDVPEIARTVG--TRGTVLFDGGVRRGDDVKALAMGAGAVLIGRP 350
 BIHZY7 EKKRRTNLPVILKGIHPDADAKLAV--NGVDGIIVSNHGGRRLDGVIGSDAPAPSVSAVK--GQPIILDGGVRRGDDVKALAMGAGAVLIGRP 333
 F4G5A4 WDLASTRLPVILKGIHPDADARELQA--AGAAGIIVSNHGGRRLDGPAPSRMPPARTAVG--AGYPLLLDGGVRRSGDDVKALAMGAGAVLIGRP 317
 B7RR92 RIADSPVPVIVKGIQHPADATRLID--LGAQGIIVSNHGGRRLDTPVAPITOLAAVVDVA--GAVPVVVDGGIRRGDDVKALAMGAGAVLIGRP 318
 Q15TJ7 WRRTNTHLPVILKGIHPDADALLAVE--SGCAGIIVSNHGGRRLDGLAPSIDLPPVRSAGV--EAFPIILDGGIRRGDDVKALAMGAGAVLIGRP 324
 F9Y7M7 EAIARAPLPVILKGVMSADAEKQCLD--AGAMGIIVSNHGGRRLDQPSAHSALPOVRAAG--PDVPIIFDSGIRRGSDIAAAVADGANAVLIGRP 316
 F6C110 EVVVSATRPVILKGIQHPDADAEALAD--AGAAIIVSNHGGRRLDTPGAAEVLPAAAVK--GRVTILADGGVRRSGDDVKALAMGAGAVLIGRP 292
 I7HM50 WRQQTNPVILKGIILSPDADALAID--HGVDNIIVSNHGGRRLDCTVPSMMVPEKAKIVSG--KNIKLFDVSGIYSDIVKALAMGAGAVLIGRP 297
 AOA087D1R1 WMLDLSGLPVIKGVMRPDAVDTAMN--AGAAGIIVSNHGGRRLDGPVAVAYAPPRKALG--RDABIFADGIRTGDDVFAALAMGAGAVLIGRP 295
 AOA082PSV8 WMLTITS LPIIVKGVITVDDTRVAIQ--AGAAIIVSNHGGRRLDYPVATIMALEEVKAAQ----GRIPVFLDGGVRRGDDVKALAMGAGAVLIGRP 185
 Q5FIR1 KYAASLEPPIVILKGVLSVADALKAKK--AGCKIIVSNHGGRRLDTPGAAEVLPAAAVK--GRVTILADGGVRRSGDDVKALAMGAGAVLIGRP 292
 I7I216 KYVASVDLPVILKGVLSVADALKAKK--AGVSIIVSNHGGRRLDGPVAVAYAPPRKALG--RDABIFADGIRTGDDVFAALAMGAGAVLIGRP 288
 GOX_SPIO1 VVFSLAA--EGEAGVKKVLMQMRDEFEELTMSLGGCRSKEKSRSHAADWDGPPSRAVARL----- 369
 LOX_LACLA VLYGLNL--GGAKGVQSVFELHNKELSIITMQLAGKNNDEETKHTSID----- 383
 MDH_PSEFU TLYGLAA--RGETGVDEVTLTKADIDRTLAQIGCPDITSLSPDYONEGVNTAPVDHLIGKGTHA----- 393
 LA2M_MYCSM YWAGLAL--GGSKGHEHVARSLAEADLLIMAVDGRNKKETIDARPTR----- 394
 CYB2_YEAST FLYANSCL--YGRNGEKAELIRDELEMSMRLLGVTSABEKKPDLDLSTLKARTVGVPNVDVLYNEVYEGPTLTFEDEA----- 511
 X8B512 -----RTSPGCARCGAGRSC-----SRA----- 255
 DIA2X8 YLWGLAA--NGQAGENVLDILRSGLDSAVLGLGRSSVHELPDDVIPPGRFRLAPGAS----- 393
 B8MCM0 FLYSMSSGYGEAGRRMIEIMREEIEQNMAIVGATKHSERRELENTSRRLRLDTTFIAKL----- 489
 B8MKR3 AHWGLGA--GGAGVERTLEILADETKTCMQLLGVKEISDGPPEYINSRIVEQQIYDGGPSGLEKARAKL----- 496
 C4VMW0 FLYGLAL--GGAKGVESVINQINNEFKILMQLTGCKTVEDVKHADIRQINYTANLPSNTDPSVRRAYPVTKENQMEGTQDAATGASKH 408
 COXIJ3 VLYGLAL--GGAQGVQSVFELDHLEIIMQLAGTKTSDVKNAKLNLIRY----- 369
 C2K1F0 FSYGLAL--GGWQGVKQVADHLKMEINIAMQLTGCTMADVQKMKVKTFFA----- 368
 A9QH69 VLYGLAM--GGSVGRQVFEKINDELKMMVMQLAGQTDDVKHFKRHNPNYDSSIPPSQNALKLD----- 390
 A4YVE0 QWALAA--AGENGIRMELLEDDEVIRALGLLGVTSFAELNASYHAATTNNMPSVFSAPFLADIEPYRY----- 378
 Q92NB8 QCYALAA--GGEAALIRMELLEDDEMLRSMALLGVPTGGEDRSYYPAPVPTIPPSALSAPFLIN----- 364
 E6SCX5 YLYGLMA--GGEAGVDRTLILRDOVTRTMRLLGVSSIDELTEHAVIAGSGPRARAQTP----- 415
 C9Y9E7 MYVGLGA--MGEAGVTKALQIHKELDVTMAFCGHTNQNVDNRIIVPGTF----- 381
 W6W585 FMLGLAA--LGDHGAHMAQAFIEBYRVTLGQSGLLNTDQARRAPVRHSSAWTQDQFQPHL----- 409
 BIHZY7 FLYGLAL--GQQQGVERVMTNIYDELKVSIALAGTSTEGRTITIVKNDGMEVK----- 386
 F4G5A4 QVYALAV--AGALGVAMHLMQMLVEELHACMAQAGCARSDITHDTITPSC----- 365
 B7RR92 VLNGLAV--DGPGRASQVLRRLRDELEVMTALCGCATVADITPDLITGFSGTGS----- 370
 Q15TJ7 VLNGLAV--AGALGVAMHSLTLTQQEELAMALTCGCTHSDITLDCY----- 369
 F9Y7M7 VEGGLAA--DGEAGVRRVLQILKDEYNTTLGHLGCNSTADLGPHVFI----- 361
 F6C110 LVIGAFG--GQTEGVKKVLEKLTGELKQAMILTGCASISDIDERVPRYPGMEC----- 344
 I7HM50 MLKALAA--GGTGESQYIARLYYDFQCCLASMGCSSEKMGQHCEYOVNSPELVKFSQDMPEQSVFRNKFIK----- 368
 AOA087D1R1 IQWALAA--GGEDGVDMIGKITDBLAHTMNSAGVSRHADDESFEVGGGAATIASYGVGF----- 355
 AOA082PSV8 VMFSLAA--DGKAGVRKVLRLMLVDELEVTIN----- 214
 Q5FIR1 ILPEVLK--NGTEAKNNKLQKMNEQLSEMMLYTCKKDMNSFDPTVHFKEK----- 340
 I7I216 ILSEILD--SGENAVIEKIEQMNQSELSELMMYTNVKDTKSFDRSVHIQ----- 335

Supplementary figure 6 | Multiple sequence alignment of previously characterised representatives of the FMN-dependant 2 hydroxyacid oxidase/dehydrogenase family and proteins characterised in the study (predicted to contain FMN_dh domain). Conserved residues around the active site are circled in red. Sequence of predicted heme binding domain is highlighted in green, the elongated loop 4 is highlighted in blue. MSA performed in PROMALS3D (Pei et al. 2008) and visualized with Multiple Align Show (<https://bioinformatics.org/sms/>).

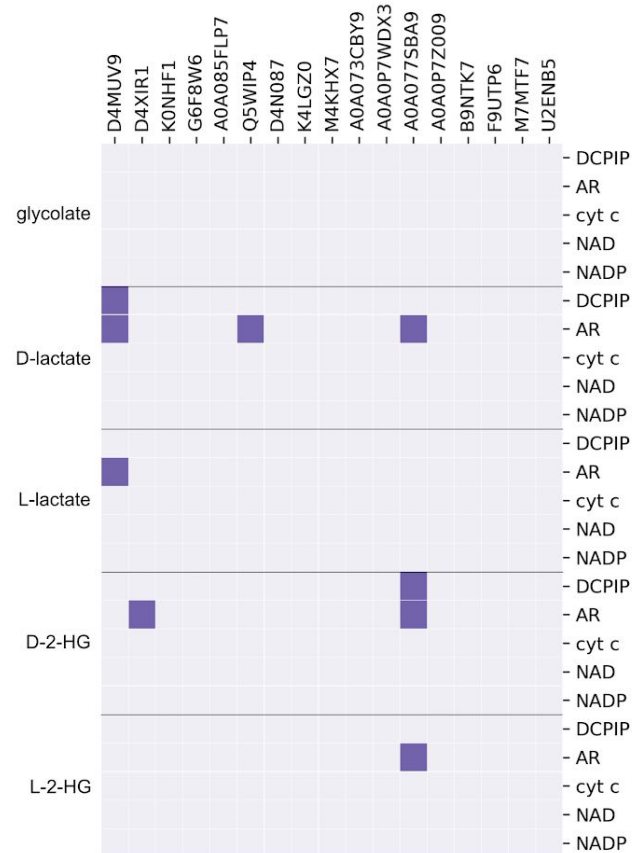


Supplementary figure 7 | Cytochrome c reduction assay of putative flavocytochrome b2 proteins. Increase of signal at the wavelength of 550 nm indicates reduction of cytochrome c and protein activity.

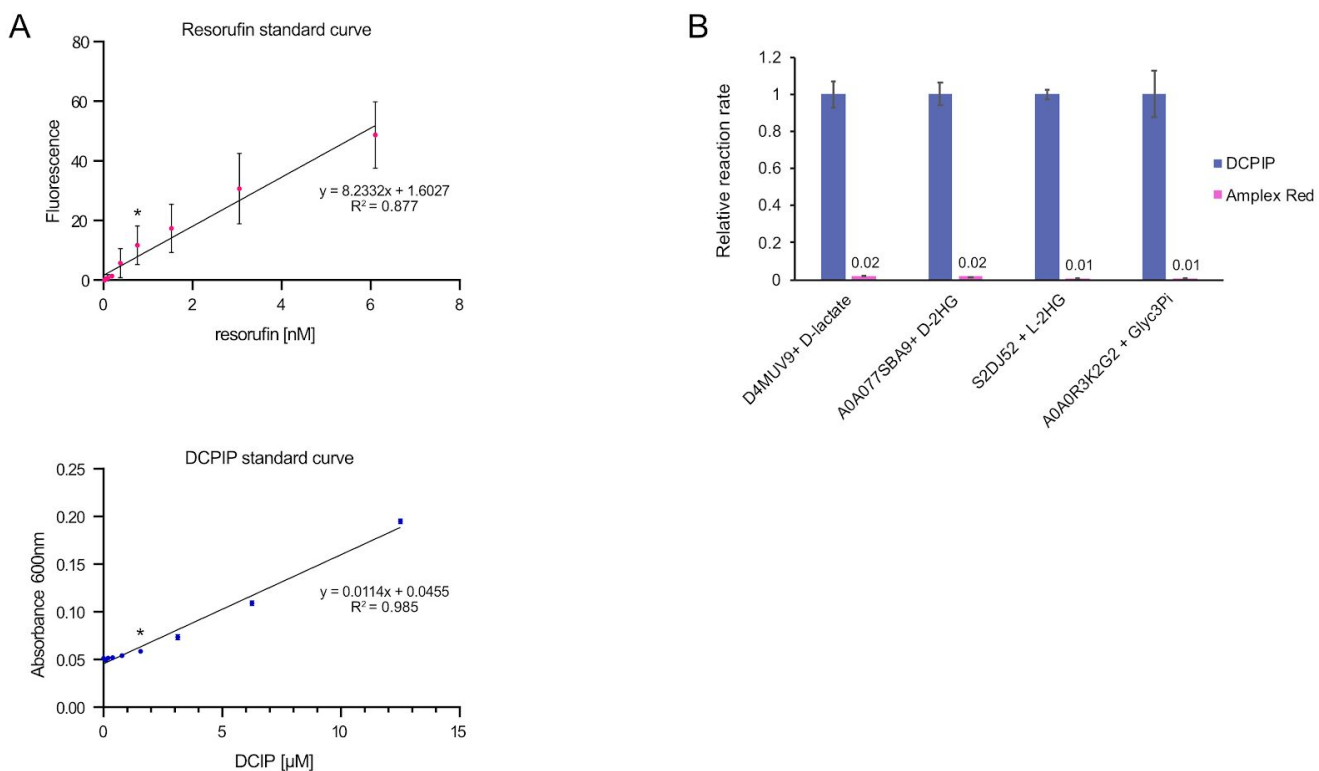
A



B

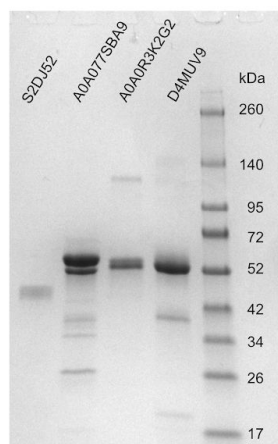


Supplementary figure 8 | Exploration of alternative activities of selected proteins. Presence of activity is marked with a dark purple square. (A) glycerol-3-phosphate dehydrogenase activity screen (B) 2-hydroxyglutarate dehydrogenase activity screen.

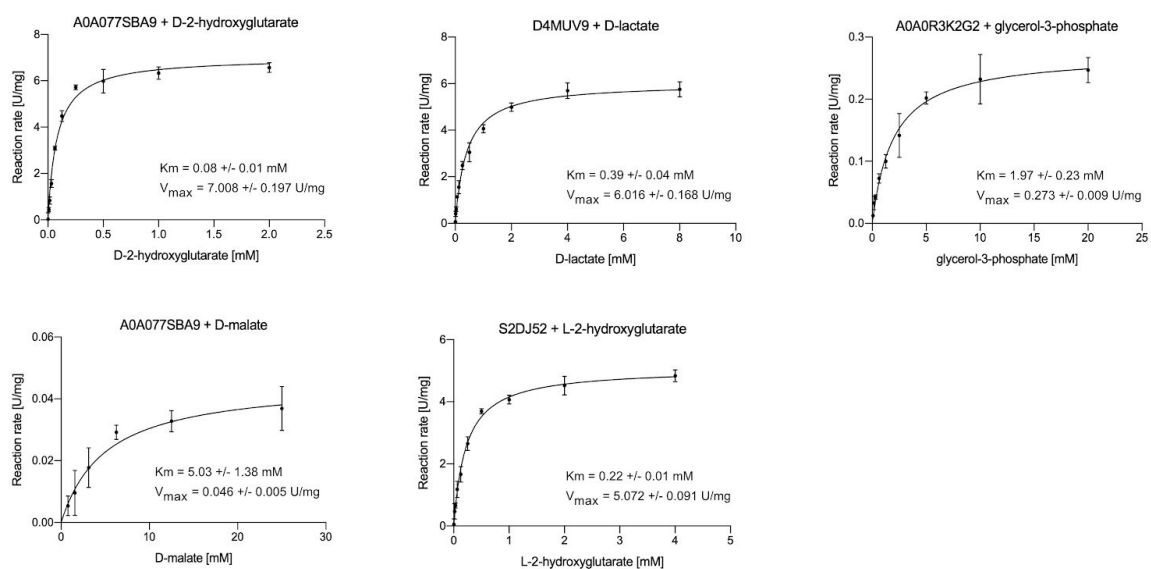


Supplementary figure 9 | Comparison of sensitivity of Amplex Red and 2,6-dichlorophenolindophenol (DCPIP)-based assays. (A) Standard curves of resorufin, a product of Amplex Red-based assay (upper panel) and DCPIP (lower panel). Indicated by asterisk are concentrations of detection limit, as calculated by Anova single factor test (0.76 nM resorufin, 1.56 μ M DCPIP). (B) Reaction rates of selected enzymes with the two electron acceptors, normalized to the reaction rate with DCPIP. Error bars in all figures represent standard deviation of the data obtained with three replicates.

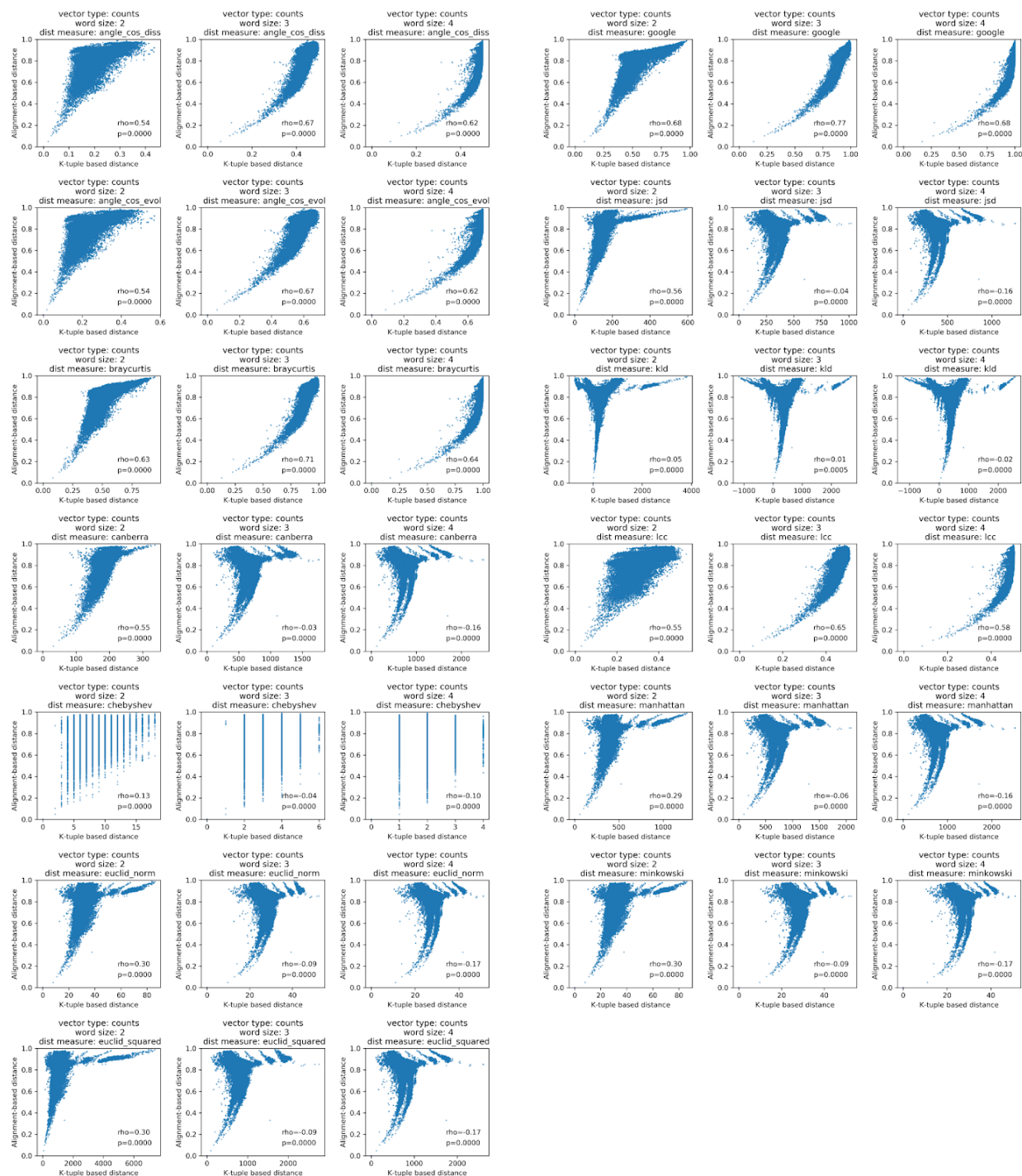
A



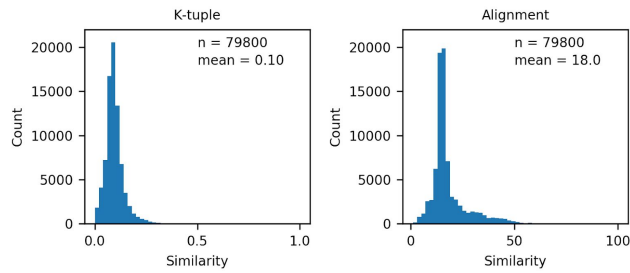
B



Supplementary figure 10 | Characterisation of proteins with activities alternative to 1.1.3.15. (A) SDS-PAGE gel of purified proteins chosen for kinetic characterisation. (B) Kinetic curves of the characterised enzymes.



Supplementary figure 11 | Test to find best k-tuple algorithm settings. Using 400 randomly selected protein sequences all pairwise distances were calculated using different word size and distance measures. These distances were compared to distances computed using pairwise alignments. Appropriate k-tuple settings will cause points to lie on a diagonal, thus showing a high degree of correlation with the alignment-based values. Spearman's rho and p-value is indicated for each plot.



Supplementary figure 12 | Average similarity between 400 randomly selected sequences from EC 1.1.3.15, using k-tuple scores (left panel) and pairwise alignments (right panel). The k-tuple score was computed using a word size of 3 and google as a distance measure. The mean alignment-based identity is 18 %. The total number of pairwise similarities is indicated, corresponding to half of the identity matrix, without the diagonal.