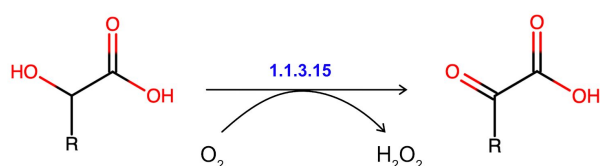
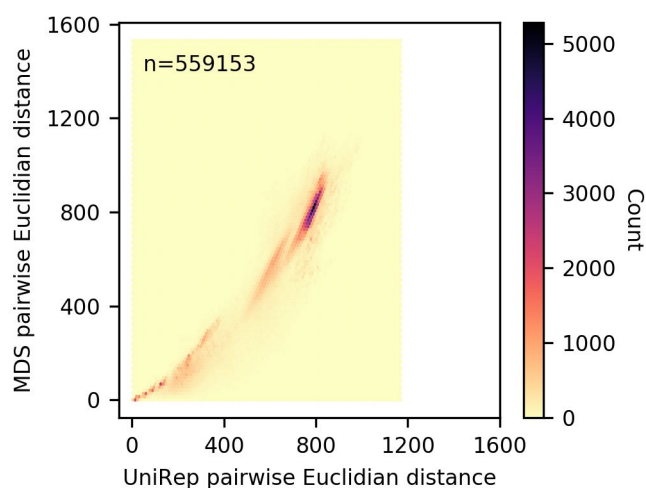


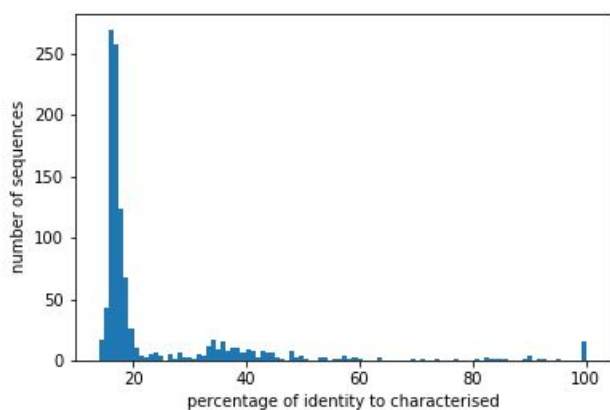
Supplementary figures



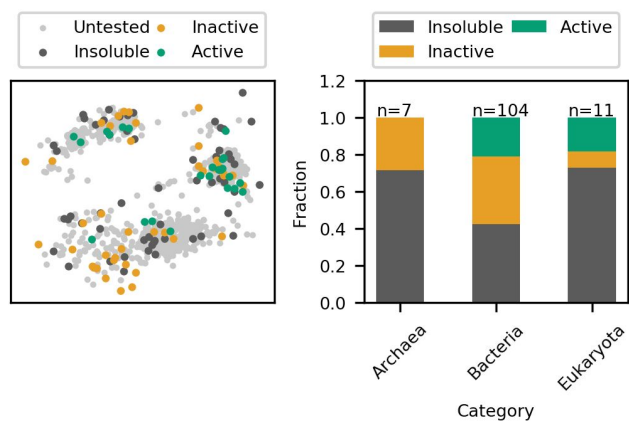
Supplementary figure 1 | Schematic representation of the reaction catalysed by S-2-hydroxyacid oxidases (EC 1.1.3.15).



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 characterised S-2-hydroxyacid oxidase.



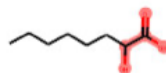
Supplementary figure 4 | (A) Distribution of the insoluble, active and inactive proteins throughout the sequence space. (B) Distribution of the insoluble, active and inactive proteins in the three superkingdoms.



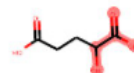
glycolate



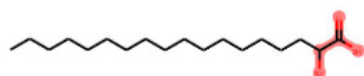
lactate



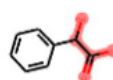
2-hydroxyoctanoate



2-hydroxyglutarate



2-hydroxystearate



mandelate

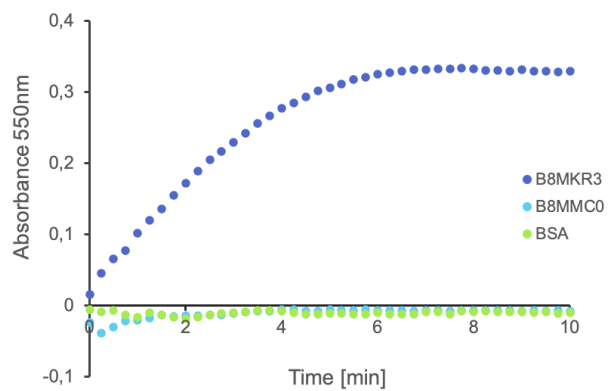
Supplementary figure 5 | S-2-hydroxyacid substrates used for the screening of EC 1.1.3.15 sequence space. The donor group is marked in red.

GOX_SPIOL ----- 0
LOX_LACIA -----MHLSTDVNFINKNRRTKM----- 18
MDH_PSEPU -----MSNWGD-----YENE----- 10
LAZM_MYCSM ----- 0
CYB2_YEAST EPKLDMNKQKISPAE-----VAKHNKPPDCWVINGVYDLTRFLPNHPPGGQDVIKFNAGKDVTAIFEPLHAPNVIDKRYIAPEKKLGLPQGGMPPE 91
X8B5I2 ----- 0
D1A2X8 ----- 0
B8MMCO -----MSQQNKGVQISGAEVAQHNNKDCSWIAVRGKQVYDVTDFLEENPGGARLLKLCAGRDATEDY--DAIHPAQILIEESLPDSAFKGTVDPSF 87
B8MKR3 -----MA--RVLD--AAEYAKHNTFESCHWILYGVYDVTETFISSHPGGVKVILRLAGTDAT-----EYDPIHPPTGLEENLKPEALL 75
C4VMW0 -----MTVYVK----- 6
COXIJ3 -----MVVVN----- 5
C2K1F0 -----MVDAV----- 5
A9QH69 -----MENKSEM-----INATTI----- 13
A4YV80 -----MNQO-----TPIE----- 8
Q92NB8 -----MSKGEE----- 6
E6SCX5 -----MSTR-----RRRPRWSELR----- 14
C9Y9E7 ----- 0
W6W585 -----MGGA-----GGDA----- 8
B1H2Y7 -----MTNT-----TDGD----- 8
F4G5A4 -----ME----- 2
B7RR92 -----MKKD----- 5
Q15TJ7 -----MNTD----- 4
F9Y7M7 ----- 0
F6CI10 ----- 0
I7HM50 ----- 0
A0A087D1R1 ----- 0
A0A082PSV8 ----- 0
Q5F1R1 -----MSKE----- 4
I7I2I6 -----MN----- 2

GOX_SPIOL -----MEITNVNEYEAIAKQKLP---KMVYVYASGAEDQ-----WTLAHRNNAFSLRFRPRIDVTN 57
LOX_LACIA -----TYKTSTENKALEIVNVKSLGKVKQSMEEAAGNKGGGYRGGSEDE-----WTLNENTSAFNKKQIMPRVGRGIDS 89
MDH_PSEPU -----MSQNLFNVEDYRKLQKRLP---KMVYVYEGGAED-----YGVKNRNDVQWQWFKPRRVDVSR 59
LAZM_MYCSM -----YGGQLGVGAVPTLPMSYADWEAHAQAALP---PGVLSYVAGGSGDE-----HTORANVEAFKHWGLMPMEMAATE 78
CYB2_YEAST LVCPPPYAPGETKEDIARKEQLKSLPLPLDNIINLYDFEYLASQTLT---KQAHYVYSSGAND-----VTHRENNHAYHRIFKPKWILVDVRK 176
X8B5I2 -----MADEWFETVAIAQQRARRRP---KSVYAAIAASEKG-----ITVSNVEAFAGELGFAPHVVGAPAK 60
D1A2X8 -----MPLKNPWFETVAEAQRAKKRLP---YMVYGAELAGSERG-----RTVQDNMTAFAGELGFAPRVVGHGA 62
B8MMCO LDRPKLADKSVRNAQKEGKKDPDSPPLSAMLNVDRFEKVAERHLA---ENSNAYYTAGADDE---YSKAEAEIAYRKVLFPRPRNVRG 172
B8MKR3 GTVDPTLTKPKDKQSSPEEEEGPPPMETLFLNLDEIEQVATKQVS---RKANGYVYSAADDL---ISKNNRREIYRSILLRPRVFDVVG 160
C4VMW0 -----GFPQSDRNEAIKMVNVDLEDRVKKVMP---EAAHYVYASGSENE---WTRNNATAFNHFQIVPRSTNMDN 73
COXIJ3 -----GYKQNEKKLNLVNLDLQLEKQAKEIIP---TGSGGYVYSGGSEDE---WTLRNNRAFTTHKQIVPRATNIEK 72
C2K1F0 -----KADPFGKVNAIDVLDLASLEAKRAKKVIL---RGEFGYVYSGGSDG---YTHHNTTAFQDVHMLPRVQGVEN 72
A9QH69 -----EFTSSAEGSVDFVNVFDLEKMAKQKVI---KGAPGYVYASGAGDT---FTLHNRIRSNHKLIVPRHKGVEN 80
A4YV80 -----AGWRNVELGASDEKQNLHEFIRKARQAQLN---QNAVYVYHGGETE---TTLRRNRMALDEIAFRPRVGRDVS 77
Q92NB8 -----ATCPPHLG---DFESLQEIQQAQALP---KEKQDSVYGAETE---TTLKRRNLIDSIAFKPRPRVGRVSV 70
E6SCX5 -----PLLRIEPPHLGSTRRRRLAASIDDLRRIARRTP---RSVYGYVYDGAEEQ---ISLRARARAFSRIEFRPRVGRDVS 88
C9Y9E7 -----MPVITNIEDRLVLAEKRPV---RMFYGYVYADSGSWTE---GTYRANEDDFQIKLPRVAVNMEN 78
W6W585 -----RTIETLAELRRRYPMIGDLEKRAKWRIP---RFAYEFLGGAGDE---TGLNRRNTLQAEVVPVRYGLDISA 75
B1H2Y7 -----LLKNITAQAPFFICFADLEKAVAEKIP---AGPFGYVYASGAGGE---QTLRNNRRAFAEKYSIVPRVNDVSN 75
F4G5A4 -----MSPSDRIPPGVWNAIDYERLAPQAMD---AGRHAYVYAGCGWMD---ATVANRRAAFAGWAVLPLPRLDVRA 65
B7RR92 -----SADLRDPDGMVPTLSDFEIDAAGRLS---ADLLAYVYEGGAAG---QSVTNRRAAFGRIGLLPRLSFCAG 67
Q15TJ7 -----YVLPKLSQIPADVIAISDYERLAKDYP---HGIYVYHAGGAGD---ITLNRNRRTAFDAIGNRPRVGRKFS 72
F9Y7M7 -----PIDNRYATGERFTTNLGIWRAAREACT---DEVYVYVYHGGDE---TTLRNRRTAFDQKQWKTFLFAGIGR 70
F6CI10 -----MNNKELREAAAREKLG---YCRVCFSCDGRACAGEVPGMGGTGTGASFKANIILARYRLNRTHHGAKN 67
I7HM50 -----MAQOES---SLARFYVYRAENG---ELANSRKYLSIKLLPRVMTSLSD 46
A0A087D1R1 -----MDWRTVHDKAVENMT---TAGLLYEDGADID---SGTNDVDTLAWRRYGLLPRTHFNVD 55
A0A082PSV8 -----MKTEMITNVTEYEAIAKEKLP---KMVYVYASGVT--- 33
Q5F1R1 -----NKSDVVTGASSETLDASMLENKKTTSPWP---GPGGMIPVYTSNADD---ANVNNRHYLDRLLVENRPRVDAVE 70
I7I2I6 -----EKLDAISDASSA---ASGHKSEATKSFPW---GPKGMIPVYTSQAEED---ANVNNRHYLDRLLVENRPRVDAVK 66

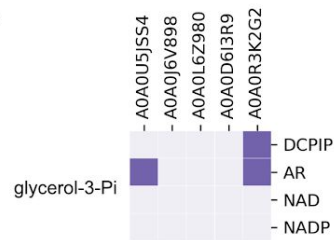
GOX_SPIOL IDMTTTLGFKKISPMIMHAPMAQ-KMAHP---EGEYVA---TARMAASAAGTIMTSS-WATSSVEEYASTGP---GIRFFQLVY---KDRNVVAQLVR 142
LOX_LACIA ADLSTSLFGLIKLKTPIIOAPVAAQ-GLAHA---EGEVA---TARMAAEVGSIFSTST-YGSTSVEDAAKAAP---DAPQFFQLVY---KDDKFNEFLK 175
MDH_PSEPU RSLQABVLEKRSQSMPLLLGGPGLN-GALWP---KGDLA---LARAATKAGIPFVIST-ASNMSHEDARQCD---GDLVQLVY---HREIAQGMVL 143
LAZM_MYCSM RDLSELVLEKKTAAAPMFFAPFGVI-ALCAQD-GHGDA---SAQASARTGVPYIST-LAVSSHEDARKHAG---DTPAYQLVY---BDRDIAESFR 166
CYB2_YEAST VDLSTMDLGSYVDVPEYYSATLCL-KLGNPLEGEXDVA---RSGGGGVTKVPMQIST-LASCSPEEIEAAPS-DKQIQWQLVY---SRKIDTDLVR 267
X8B5I2 RELATTVMGCEISMPLVLISSPGVQ-AVDP---DGEVA---VARAAARAGVIMGLISS-FASKPTTEVIAVNV---PKVYVQVYL---GGRDAILAEVR 147
D1A2X8 RDLSTVTVMGVPSTMVVISPGVQ-AVHP---DGEVA---VARAAARAGVIMGLISS-FASKPTTEVIAVNV---PNVYVQVYL---GGRDAILQRM 145
B8MMCO VDRTRTOLGQDVSLLPIYISAVGIA-KFAHP---QGECT---LAAAAGREGIAQLVAT-RSSMSHESHMARTGGPQPPFQDYVYH---KDAKISDATT 261
B8MKR3 CDLSTTLGHRVPLPIYISAPAMA-RLAHP---AGHAG---IAAACRGFGAMQMISN-NASMSPEOHVENAA---PDQVFGQLVY---MBRKKSEAMBA 247
C4VMW0 PSTETOFMGMIDLKTPIMICPACH-GIAHK---DAEVA---TAQGAAGAALFSSST-YANRSVEDHATATG---DSPKFFQLVY---KDWDFNKMVFD 159
COXIJ3 PELETNVFGFIPKLTPLFMVPAQAQ-GLAHV---KGEVD---TANGVAAVGGLMAQST-YSTSTHADTAASGT---GAPQFFQLVY---KDWDFNEALD 158
C2K1F0 PDOSTTFMGAKLASPLLIAPVASN-TLAHP---SGHLE---TANGVAAGGLMAQST-YSTSTHADTAASGT---GAPQFFQLVY---KDWDFNKMVFD 158
A9QH69 PSTEITFDGKLASPLLIAPVAAH-KLANE---QGEIA---SANGVKEFGTITYTSS-YSTSTHADTAASGT---GAPQFFQLVY---KDWDFNKMVFD 158
A4YV80 VDSASVEFGERRRLRPVVMAPVAGAL-EIFDP---AGHAG---VANGAGRFGAAMHMISS-VSEFGHEKTAEAAP---DALRIFQLVY---GDAFVEDVYV 163
Q92NB8 VDLSEIEFGERRRLRPVVMAPVAGAL-EIFDP---AGHAG---VANGAGRFGAAMHMISS-VSEFGHEKTAEAAP---DALRIFQLVY---GDAFVEDVYV 163
E6SCX5 VDSASRVVGSFSSPLVLIAPGGFT-RMMHH---EGHRA---VARAAARAGVIMGLISS-MGTVSVVEVAAAAP---GSELVQVYL---KRAASLELV 174
C9Y9E7 RTTATTFMGTAVKAMPVAPVAGLT-GMQHA---DGEIH---AARAAAEKFGIPFTIST-MSICSHEDHANTS---APFVQVYVYH---KDRNANAMHE 143
W6W585 VDSASVEFGHRYAAPIGISPGGFD-GMMWP---GATER---FAKAAQAHNIPYLVGT-LACATHEKVVEWAP---DVTVMQVYVYH---KDRNANAMHE 143
B1H2Y7 VHTSINLFGKTYPTPLIAPVGMN-GMVHE---EGHRA---VARAAARAGVIMGLISS-VSTYAHEDVAAAP---SATKWFQLVY---THEETAFSMAA 161
F4G5A4 GHTRLQAGHMDPLLIAPVAAH-RLAHP---DAEIA---TARMAAAQAGTGLVAST-LSCSTHEDHAAASG---PARVQVYVYH---KDRNANAMHE 150
B7RR92 GHTRTLTGKQAPHPIMVAPMAFO-NLFHP---QGEISA---TARMAAAQAGTGLVAST-LSCSTHEDHAAASG---PARVQVYVYH---KDRNANAMHE 150
Q15TJ7 GTEITAGGRDRFSMPPLIAPVAYO-SLLHP---QGEIA---TARMAAAQAGTGLVAST-LSCSTHEDHAAASG---PARVQVYVYH---KDRNANAMHE 150
F9Y7M7 PDTATQFLGHSLSFPAFIAPPGGGEYMLDA-EGHRA---TERRARDVGIRQIVYV-AAHSHEDHATASG---VAQMFFQLVY---GDAFVEDVYV 156
F6CI10 PDTSLTFLGKELSTPLIAPMT---GVTYN---MGGAL---TEREF---IGMINAS-ARQAGTGLMSGDG---GAPQFFQLVY---KDWDFNKMVFD 135
I7HM50 IDSTLTFEHLKSIPLVFIAPMAFH-CLFNQ---EGHVA---SAGACAEIGIFGCYVNYCYSRSRDSKHKVP---EGVYVYVYH---KDRNANAMHE 134
A0A087D1R1 IDTSVTVGLHLDTPFIAPPAH-RLASK---QGEIS---SONGHRAGSFIVYPT-NAGEPWEHFAQEAQ---GEWQVQVYVYH---KDRNANAMHE 134
A0A082PSV8 -----VFFPFSQVYVYH---KDRNANAMHE 134
Q5F1R1 PDLTTFEFGKYSAPLMPALSHL-NKILD---DKNRKPMQEKLAARENLNLMWGM-ETNEEYSEVYVYH---GDTIRIFVYVYH---KDRNANAMHE 134
I7I2I6 PDLTTFEFGKYSAPLMPALSHL-NKILD---DKNRKPMQEKLAARENLNLMWGM-ETNEEYSEVYVYH---GDTIRIFVYVYH---KDRNANAMHE 134

GOX_SPIOL RAERAG-FKAIAITDTPRLGRREADIKNRFLVLPPL-TLKNFEGIDLGM-----DKA-----NDSGLESSYVAGQIDRSLSWK-DVA 217
LOX_LACIA KAYSAG-VKAIITDSTLGGYREEDIVNHFFQPPFPM-PNLAASFESD-----GTGKGISEIYAAAKQGLVLE-DVA 244
MDH_PSEPU KAHTQO-YTTLVITDVAVNGYRGRLDHNRFKIPMSY-SARVVLDGCLLHPRNSLDVFRHGMPLQANFVSSQT--SLEMQAALMSRQMDASFNWE-AR 237
LAZM_MYCSM RAEAEAG-YDGLVITDTHIFGWRPRDLTISNFFFLRG-LCLTNVYVTDPPVQ--KKFKAHSGVAAEGLRD-----NPLAADFVHGLFGHSVITWE-DID 254
CYB2_YEAST NVEKLG-AVGLVITDAPSLGOREKDKMLKFSNTKAG-PKAMKKTINVEES-----QGASRALSKFIDPSLTWK-DHE 336
X8B5I2 RAERAG-AVGLVITDWSFSHGRDWSGSPKIPPEMNLRTILRLSPAVFKPRNLWKFGLTLRPPELRVPNQGRGEPGPPFFAYAGEMWGTPT-- 233
D1A2X8 RARNAQ-ARGLVITDWSFSHGRDWSGSPKIPPEMNLRTILRLSPAVFKPRNLWKFGLTLRPPELRVPNQGRGEPGPPFFAYAGEMWGTPT-- 233
B8MMCO KAYKAG-VKGLVITDVPVTKREERDERLKANVDVGE-ONDKIGKGPV-----QGVAKTSLSTSVSPVLTLDWD-TVE 240
B8MKR3 BEVEKLAIKCVITDAPVPPKREEDMDRTDNGKLLP-VSSAKVGAEREVELTDPGTP-----VPT---DGGGGVGVKQLFAGTAYDLTWKKTET 331
C4VMW0 AVKSAQ-YKGLVITDALVSGYREANLRTNFTFPVPL-DFDFFRYVAE-----EGGMSVAQMYANSQAQKIGA-DVA 228
COXIJ3 EAKRAG-VKGLVITDVAATVDGGRADINNFOFPIPM-ANLTKYSEDD-----GGGKGIAEYASAAQKIGSD-DVA 227
C2K1F0 EAKRAG-ALAIITDSTLGGYREKDVMMHHLKGR-ANLEGYNT-----QSGVGAGGLFKESMQKLDLG-LVE 226
A9QH69 BLKAEQ-VKGLVITDVAATVDGGRADINNFOFPIPM-PIVQVYLPN-----GAGKTMVYVYKATKQALSPPK-DIS 234

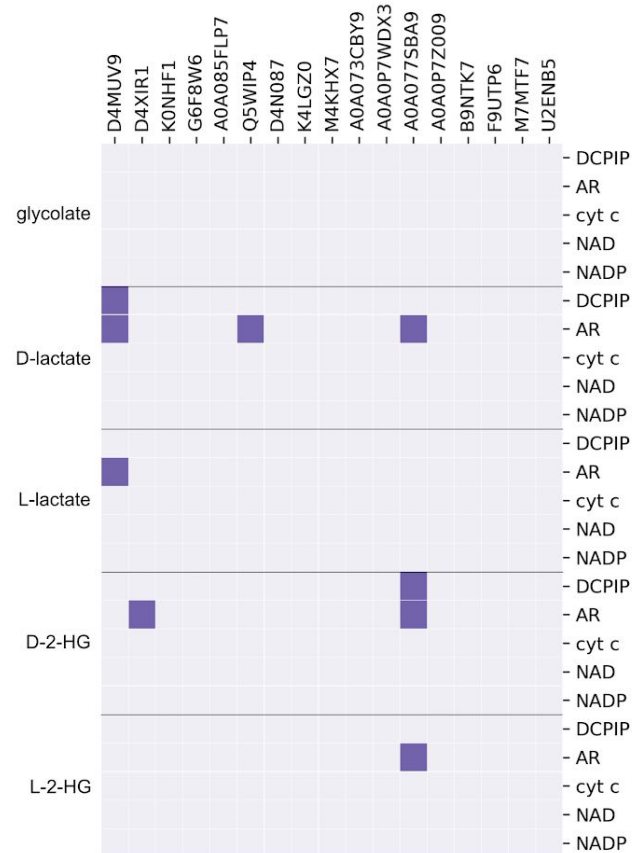


Supplementary figure 7 | Cytochrome c reduction assay of putative flavocytochrome b2 proteins. Increase of signal at the wavelength of 550 nm indicates reduction of cytochrome c and protein activity.

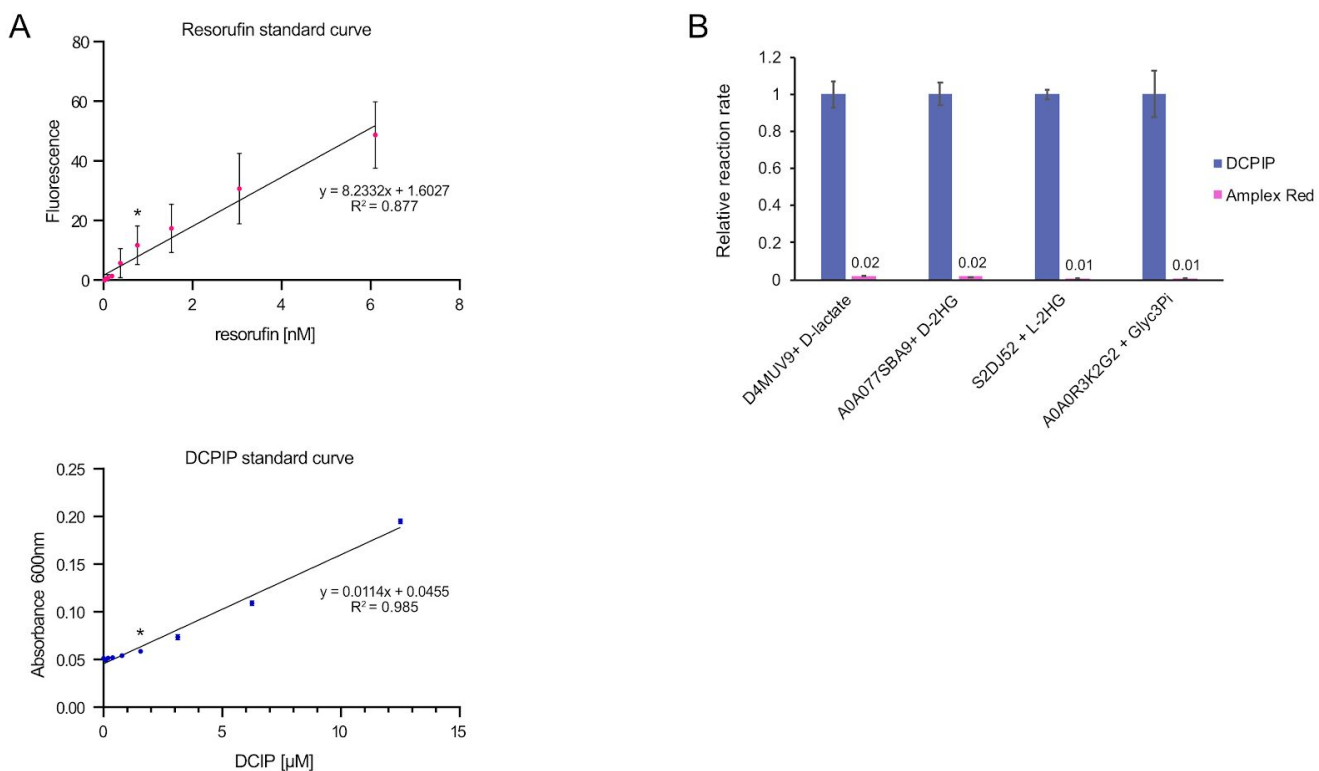
A



B

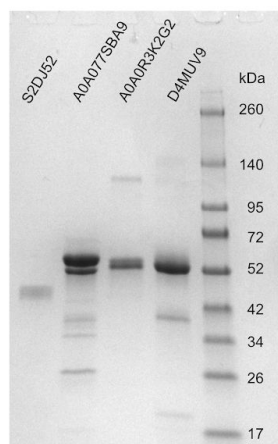


Supplementary figure 8 | Exploration of alternative activities of selected proteins. Presence of activity is marked with a dark purple square. (A) glycerol-3-phosphate dehydrogenase activity screen (B) 2-hydroxyglutarate dehydrogenase activity screen.

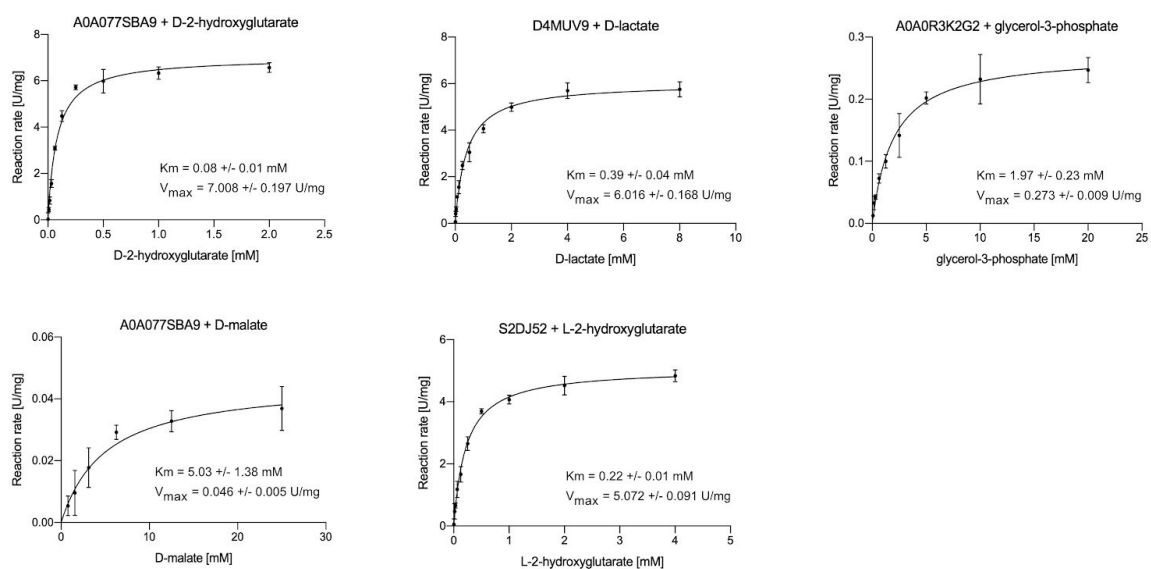


Supplementary figure 9 | Comparison of sensitivity of Amplex Red and 2,6-dichlorophenolindophenol (DCPIP)-based assays. (A) Standard curves of resorufin, a product of Amplex Red-based assay (upper panel) and DCPIP (lower panel). Indicated by asterisk are concentrations of detection limit, as calculated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

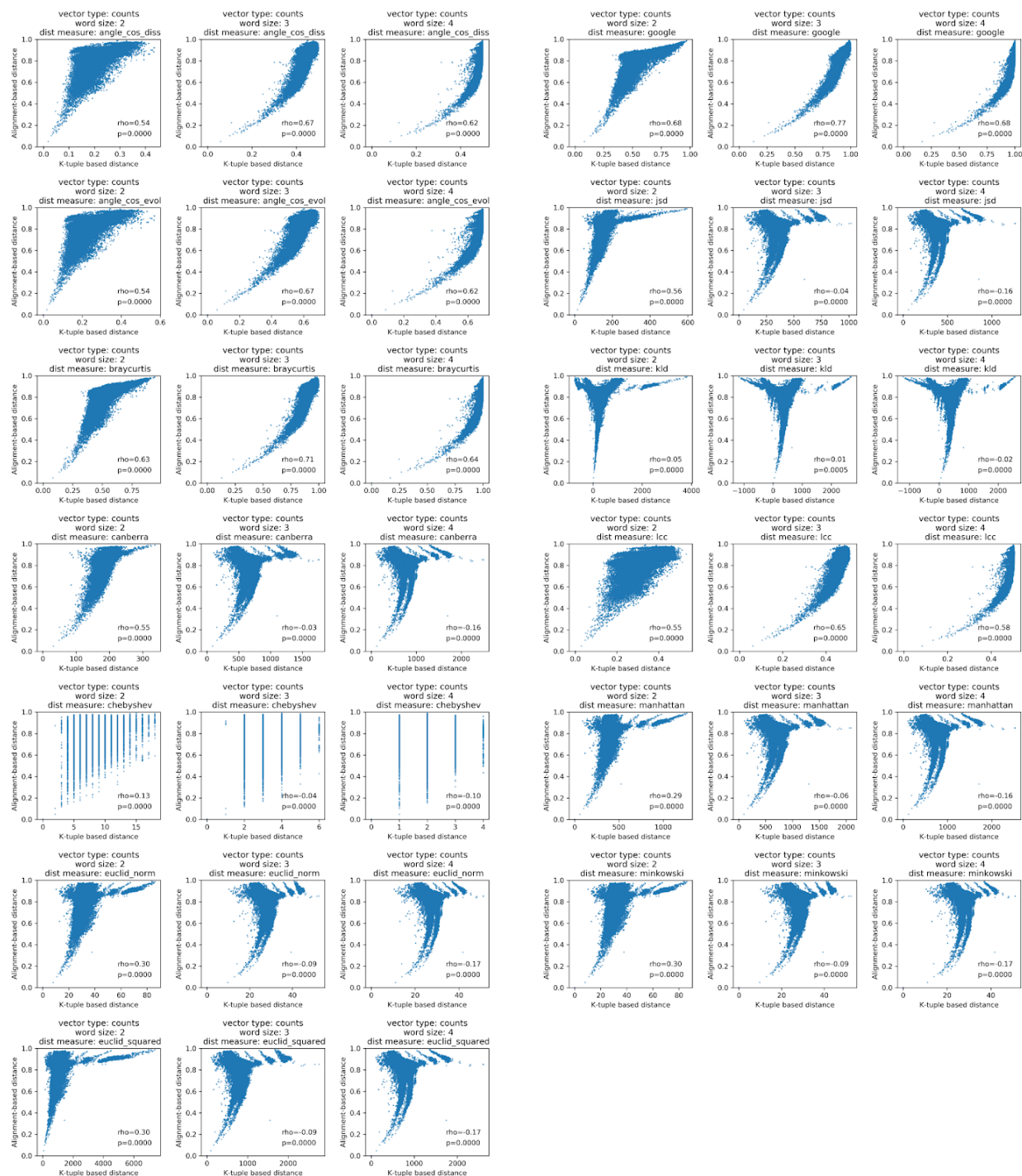
A



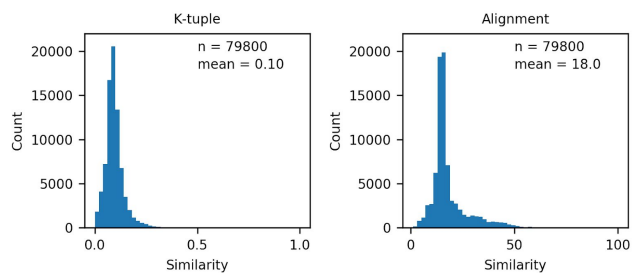
B



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. Error bars show standard error of three replicates.



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.