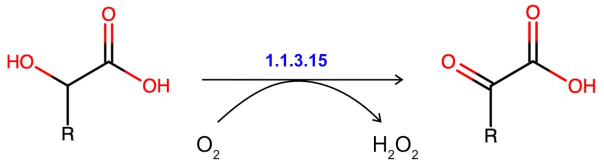
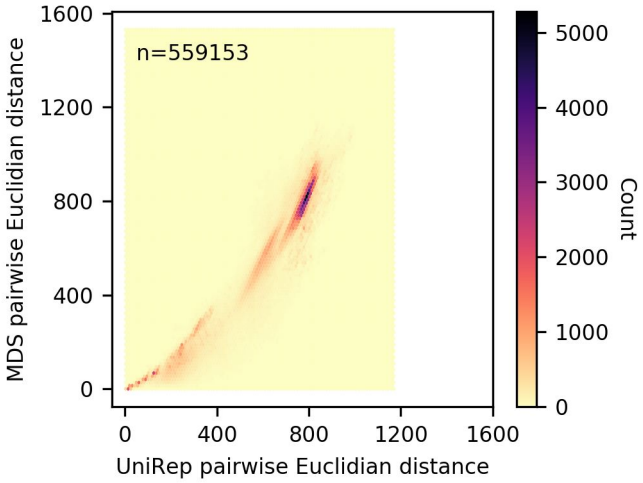
# Supplement

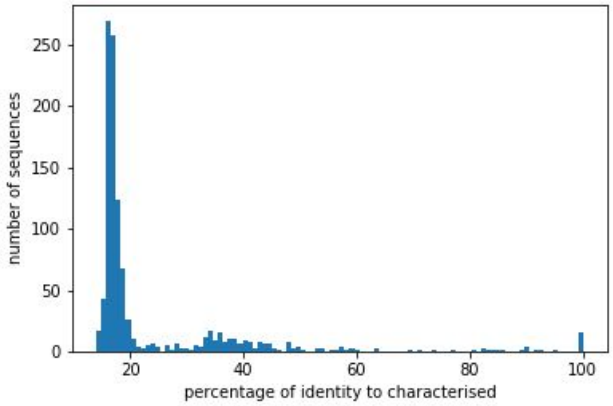
## Supplementary figures



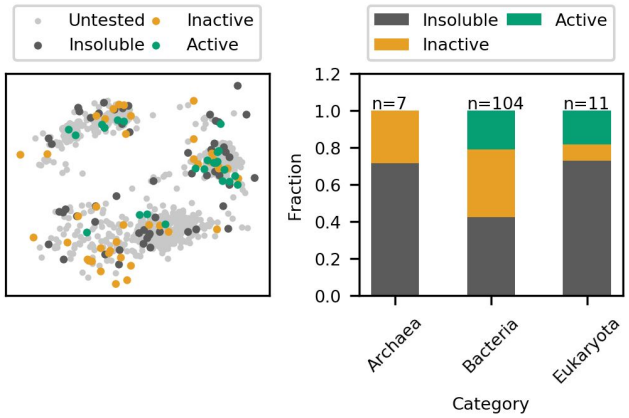

Supplementary figure 1 | Schematic representation of the reaction catalysed by S-2-hydroxyacid oxidases (EC 1.1.3.15).



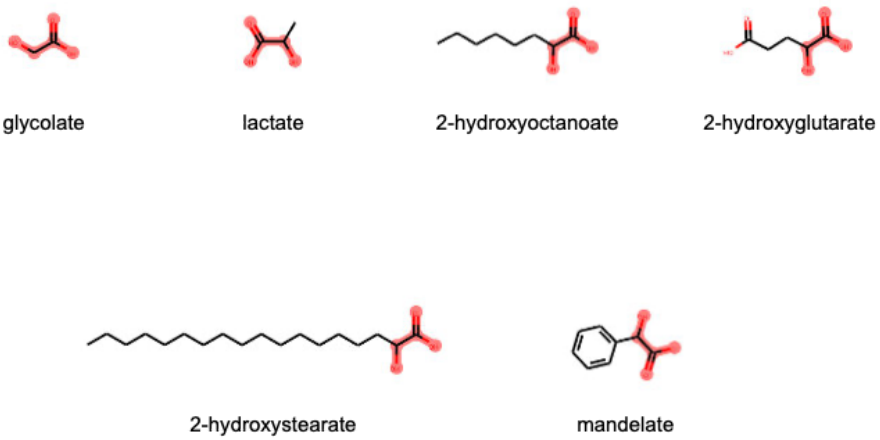

Supplementary figure 2 | A hexbin plot indicating Euclidean pairwise distances between the 1058 EC 1.1.3.15 proteins. Clustering along the diagonal indicates that the multidimensional scaling (MDS) dimensionality reduction faithfully represents pairwise distances of the UniRep representations of these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



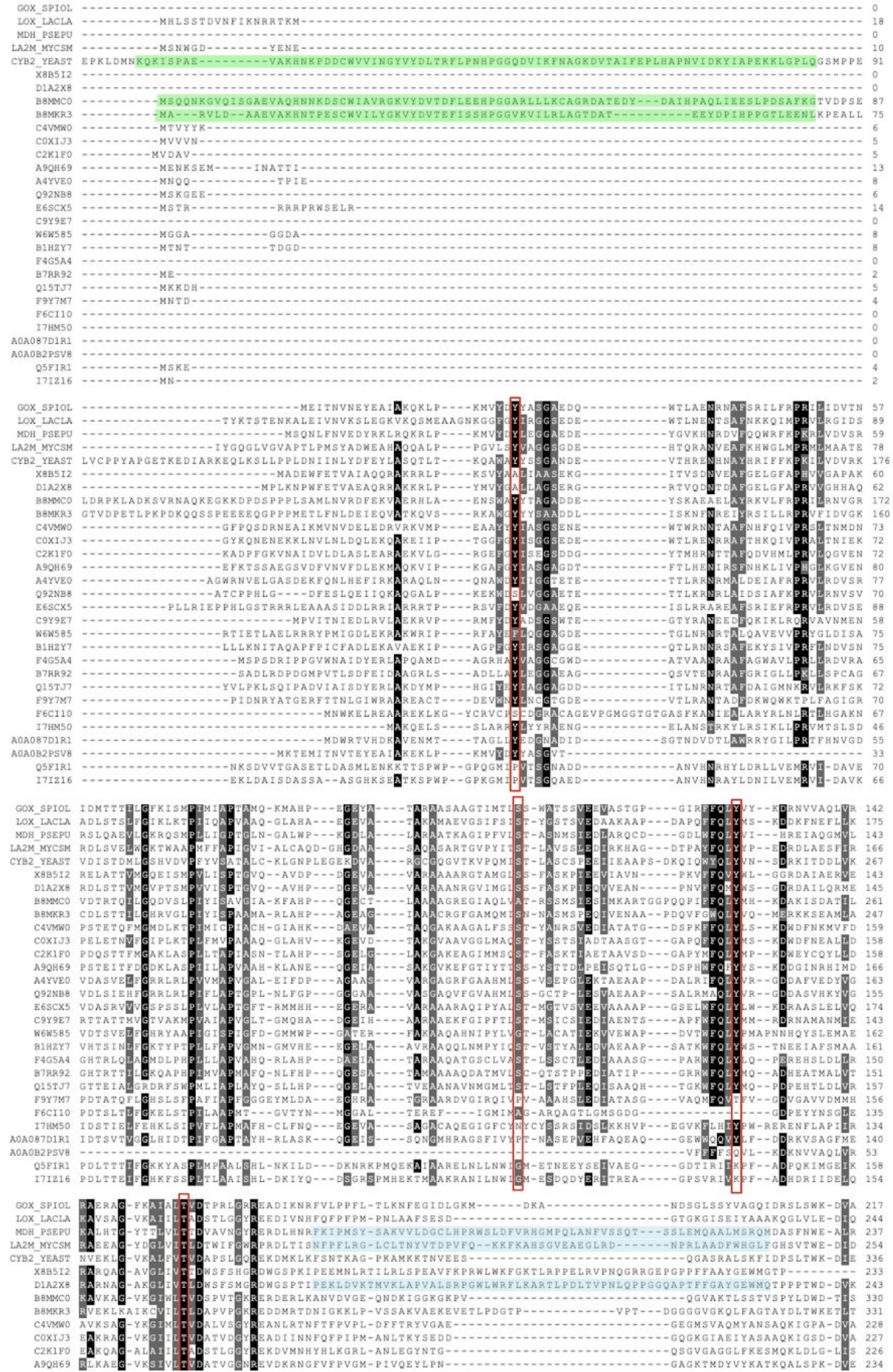

Supplementary figure 3 | Identity of sequences annotated as EC 1.1.3.15 to the closest characterized S-2-hydroxyacid oxidase.

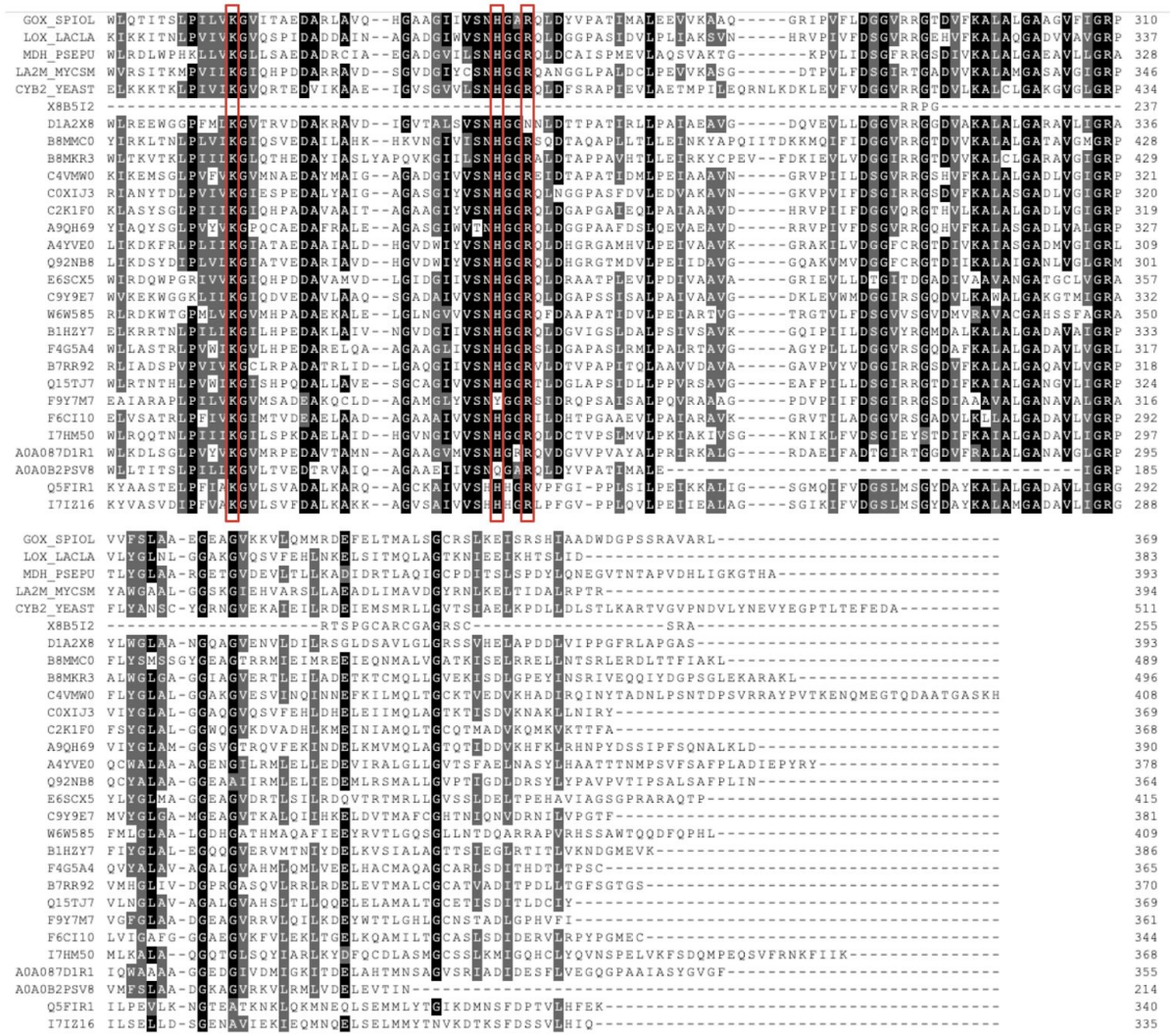



Supplementary figure 4 | (A) Distribution of the insoluble, active and inactive proteins throughout the sequence space. (B) Distribution of the insoluble, active and inactive proteins in the three superkingdoms.

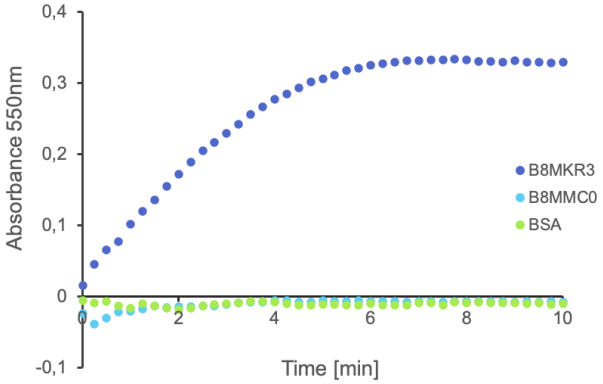



Supplementary figure 5 | S-2-hydroxyacid substrates used for the screening of EC 1.1.3.15 sequence space. The donor group is marked in red.

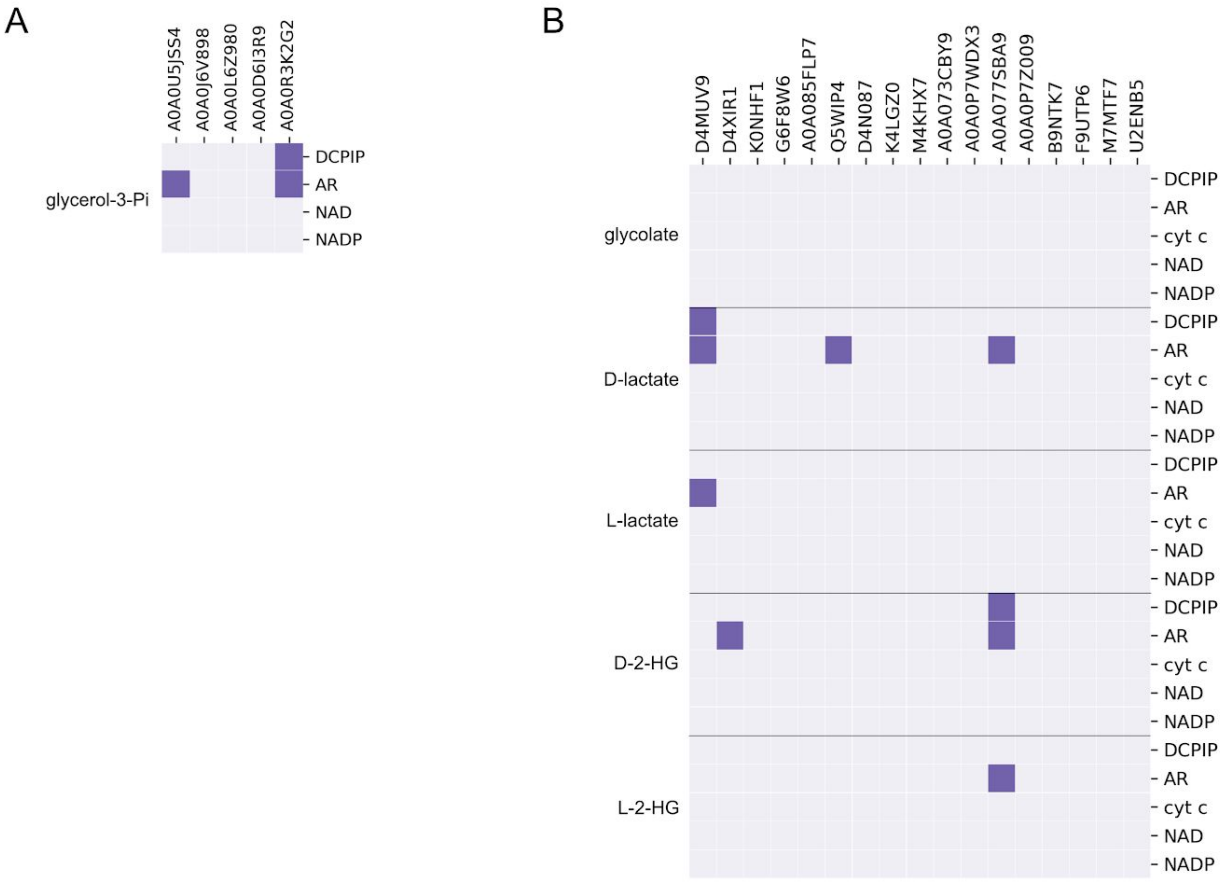

GOX_SPIOL
LOX_LACLA
MDH_PSEPU
LA2M_MYCSM
CYB2_YEAST
X8B5I2
D1A2X8
B8MMC0
B8MKR3
C4VMW0
C0XIJ3
C2K1F0
A9QH69
A4YVE0
Q92NB8
E6SCX5
C9Y9E7
W6W585
B1HZY7
F4G5A4
B7RR92
Q15TJ7
F9Y7M7
F6CI10
I7HM50
A0A087D1R1
A0A0B2PSV8
Q5FIR1
I7IZ16



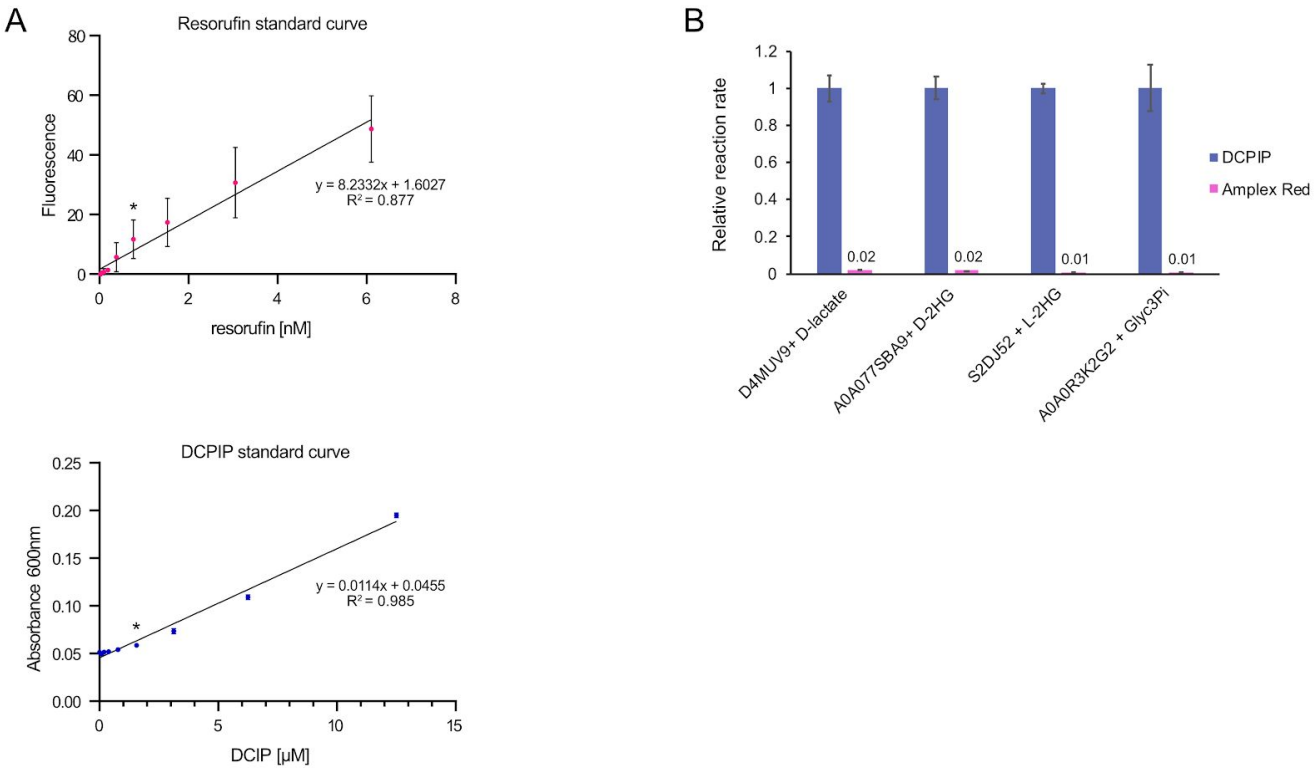
Supplementary figure 6 | Multiple sequence alignment of previously characterised representatives of the FMN-dependant 2 hydroxyacid oxidase/dehydrogenase family and proteins characterised in the study (predicted to contain FMN_dh domain). Conserved residues around the active site are circled in red. 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.

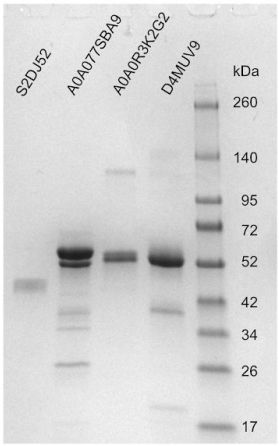



Supplementary figure 8 | Exploration of alternative activities of selected proteins. Presence of activity is marked with a dark purple square. (A) glycerol-3-phosphate dehydrogenase activity screen (B) 2-hydroxyglutarate dehydrogenase activity screen.

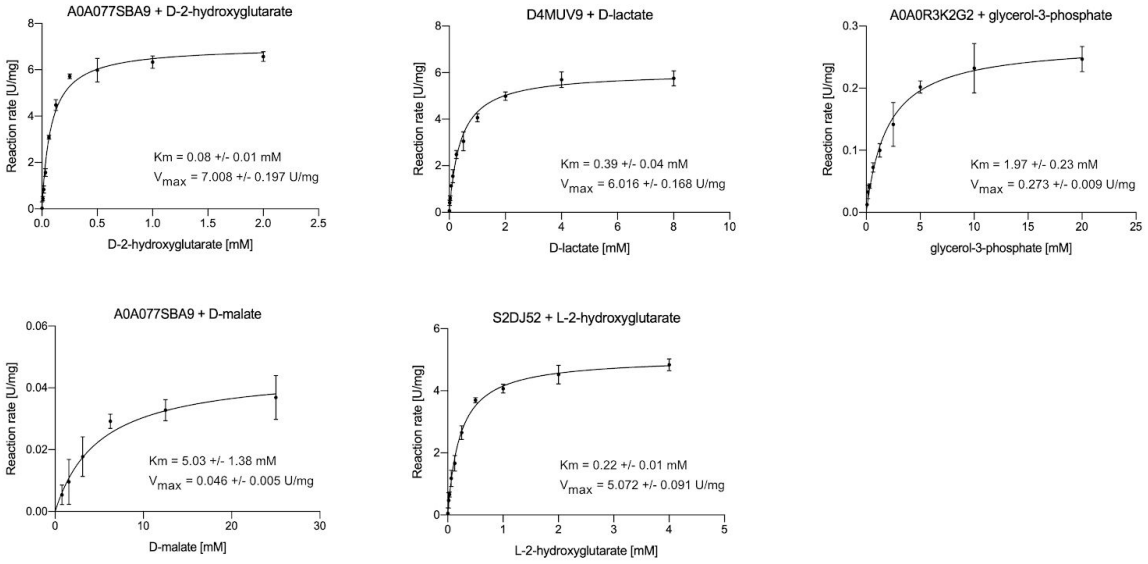



Supplementary figure 9 | Comparison of sensitivity of Amplex Red and 2,6-dichlorophenolindophenol (DCPIP)-based assays. (A) Standard curves of resorufin, a product of Amplex Red-based assay (upper panel) and DCPIP (lower panel). Indicated by asterisk are concentrations of detection limit, as calculated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

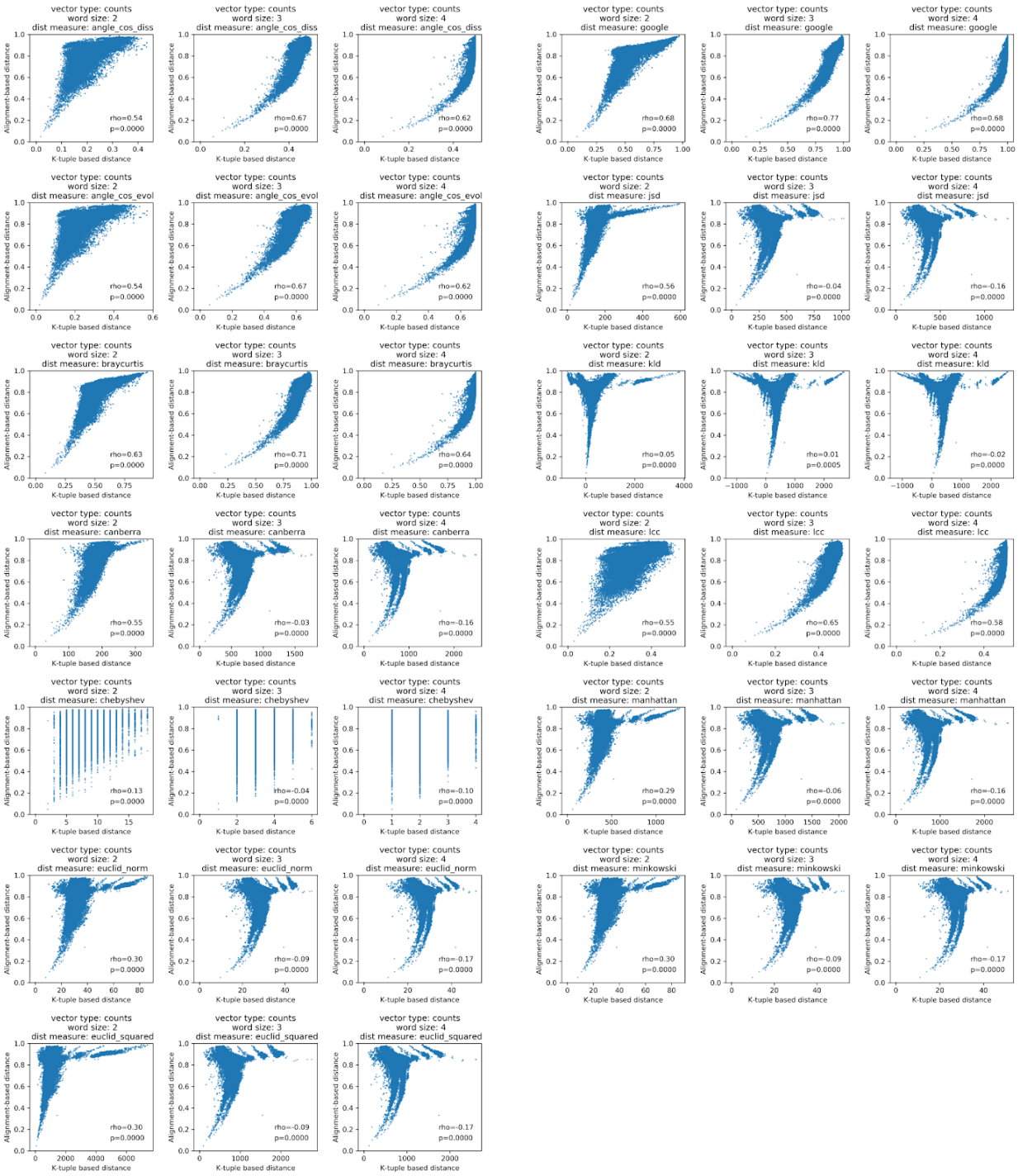
A



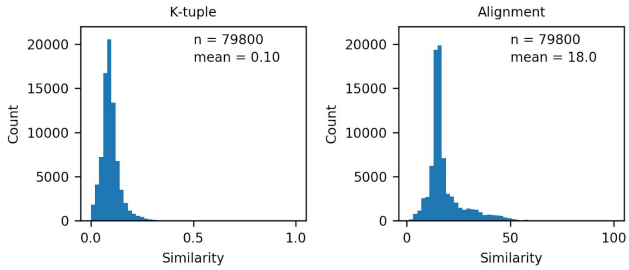

B




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.