

#### Additional file 4

The linkage of kinetic constants and amino acid composition for enzymes of the TCA pathway.

Relationship between the amino acid composition and  $K_M$  values for 10 enzymes (MDH1 (P17505), CIS1 (P00890), CIS2 (P08679), IDP1 (P21954), IDP2 (P41939), IDP3 (P53982), IDH1 (P28834), LPD1 (P09624), SDH1 (Q00711), KDG1 (P20967)) for yeast TCA pathway. The plot of actual  $K_M$  values (A) versus those predicted by the linear regression equation:  $\log(K_M) = 9.34992 - 0.291239*A - 0.506944*R - 0.663545*L - 0.618722*M + 0.258044*P$  ( $R^2_{adj.} = 91.82\%$ ,  $p = 0.0000$ ). Relationship between an increase in the percentage of explained variance and the number of independent variables (amino acid frequencies of occurrence) included in multiple regressions (B), where: 1 – L, 2 – L, R, 3 – L, R, M, 4 – L, R, M, A, and 5 – L, R, M, A, P.

