Supplementary Note

Evaluation of orthogonal enzyme-substrate pairs

Previously, we evaluated the orthogonality of a set of two enzyme-substrate pairs using the following equation (Jones *et al.*, *J. Am. Chem. Soc.* **139**, 2351-2358 (2017)):

$$O = \log_{10}\left(\frac{E_A S_1}{E_A S_2}\right) * \log_{10}\left(\frac{E_B S_2}{E_B S_1}\right) \tag{1}$$

Where $E_A S_1$ is the signal produced when enzyme A receives substrate 1, $E_A S_2$ is the signal produced with substrate 2 and enzyme A, *etc*. Though functional for pairs of enzymes and substrates, this equation does not allow searches for sets of greater than two. Thus, we sought a new way to quantify orthogonality that could utilize any number of enzyme-substrate pairs.

For rapid iteration and data sorting, it is important to represent each set of enzyme-substrate pairs as a scalar value. Such an operation can be completed using basic linear algebra. For the sake of this example, a set of three enzymes and three substrates will be used:

$$E_1, E_2, E_3 \text{ and } S_1, S_2, S_3$$

Consider that the reactivity of each substrate in each enzyme, r, can be represented in matrix form

$$M = \begin{array}{c} \\ E_1 \\ E_2 \\ E_3 \end{array} \begin{array}{c} \begin{array}{ccc} S_1 & S_2 & S_3 \end{array} \\ \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \end{array} \tag{2}$$

with enzymes arrayed into rows and substrates arrayed into columns. In this form, the ideal orthogonal case could be represented by the identity matrix. M_{ideal} shows three substrate-enzyme pairs where each substrate gives signal in only one enzyme.

$$M_{ideal} = \begin{array}{c} \\ E_1 \\ E_2 \\ E_3 \end{array} \begin{array}{c} \begin{array}{ccc} S_1 & S_2 & S_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \tag{3}$$

Comparison of this ideal case to a matrix of experimental data via evaluation of the RMSD (root-mean-square displacement) will yield a scalar value that can be used to rank enzyme-substrate pairings: the lower the RMSD, the better the pairing.

$$\sqrt{\langle (M_{exp} - M_{ideal})^2 \rangle} = O_{\mathrm{RMSD}} \tag{4}$$

In order to compare the ideal case to an experimental case, we must normalize and symmetrize the experimental matrix. Column-wise normalization is carried out to ensure that each column vector is the same length. This is done by dividing each vector component by the total vector length (determined via the Pythagorean Theorem), thus producing a set of unit vectors:

$$\vec{S_1} = \begin{array}{c} \\ E_1 \\ E_2 \\ E_3 \end{array} \begin{array}{c} \begin{array}{c} S_1 \end{array} \\ \begin{bmatrix} r_{11} \\ r_{21} \\ r_{31} \end{bmatrix} \end{array} \qquad\qquad \hat{S_1} = \frac{\vec{S_1}}{|S_1|}$$

Where $|S_1|$ is the length of $\vec{S_1}$, and $\hat{S_1}$ is the unit vector of $\vec{S_1}$. When this process is repeated for $\vec{S_2}$ and $\vec{S_3}$, the matrix can then be symmetrized via the equation:

$$M \cdot M^T = M_{sym} \tag{5}$$

Where '·' denotes the matrix dot product, $\hat{M}$ is the fully normalized matrix, and $\hat{M}^T$ is the matrix transform. Finally, as indicated in equation 4, the orthogonality of the set of vectors can be evaluated via comparison to the ideal case via RMSD.

$$\sqrt{\langle (M_{sym} - M_{ideal})^2 \rangle} = O_{\mathrm{RMSD}}$$

For simpler readability, RMSD can be converted to a positive orthogonality value representing the geometric mean of the resolution of each substrate by each compound. This is calculated from comparison of the experimental RMSD, O_{RMSD} and the worst possible RMSD (where M_{sym} is filled with ones), O_{worst} via a simple quotient.

$$M_{worst} = \begin{array}{c} \\ E_1 \\ E_2 \\ E_3 \end{array} \begin{array}{c} \begin{array}{ccc} S_1 & S_2 & S_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{array} \qquad O = 2 * \frac{O_{worst}}{O_{RMSD}} \tag{6}$$