Supplemental Information S3

We define M as the number of occupied bins, B as the number of bins, and N as the number of components. The length of a single bin is $0 < L \le 1$ noting that the total number of bins $B = L^{-2}$. Using normalized retention time pairs t_1, t_2, which are mapped onto a unit area, as described in the text as eq 1, each retention time pair is used to occupy bins one of the B bins. The starting point of this treatment is to define the probability of the number of occupied and non-occupied bins. As defined in the text, the surface coverage metric SC_G is defined as M/B.

Let ξ_i be a bivariate random variable with probability density function (pdf) of f(x,y) where x and y are the two normalized (between 0 and 1) retention time ranges and $1 \le i \le N$. Each bin has a random variable χ_j which is either 0 (empty) or 1 (occupied by one or more components) where the index j has limits of $1 \le j \le B$. Each bin is identified as β_j.

The probability that the *j*th bin is empty is:

$$\Pr(\chi_j = 0) = \left(1 - \int_{\beta_j} f(x,y)\, dx\, dy\right)^N \qquad \text{(s3.1)}$$

where the integration is over a small square patch defined by the spatial limits of β_j. The expected number of non-empty bins is thus

$$E[\chi_1 + \ldots + \chi_B] = E[\chi_1] + \ldots + E[\chi_B] = \sum_{j=1}^{B}\left(1 - \left(1 - \int_{\beta_l} f(x,y)\, dx\, dy\right)^N\right) \qquad \text{(s3.2)}$$

where E[] denotes the average or expectation operator. For small bins (small L) the integrals above can be approximated as

$$\int_{\beta_j} f(x,y)\, dx\, dy \approx f_j\, L^2 \qquad \text{(s3.3)}$$

where f_j is the value of the pdf evaluated at the center or any location within the bin. This simplifies the expression for the expected number of non-empty bins to

$$\sum_{j=1}^{B}\left(1 - \left(1 - f_j\, L^2\right)^N\right) \qquad \text{(s3.4)}$$

The ratio between the number of non-empty bins and the number of all bins is thus

$$\frac{\sum_{j=1}^{B}\left(1-\left(1-f_j\,L^2\right)^N\right)}{B}=\frac{\sum_{j=1}^{L^{-2}}\left(1-\left(1-f_j\,L^2\right)^N\right)}{L^{-2}} \qquad \text{(s3.5)}$$

In the special case when the number of components *N* is equal to the number of bins so that $N=B=L^{-2}$ and noting that $\lim_{N\to\infty}(1-1/N)^N = 1/e$

$$\begin{aligned} SC(N) &= \frac{1}{N}\sum_{j=1}^{N}\left(1-\left(1-{f_j}/{N}\right)^N\right)\approx\frac{1}{N}\sum_{j=1}^{N}\left(1-e^{-f_j}\right) \\ &\approx 1-\frac{1}{N}\sum_{l=1}^{N}\left(e^{-f_j}\right)\approx 1-\int_{\beta} e^{-f(x,y)}\,dx\,dy \end{aligned} \qquad \text{(s3.6)}$$

where the integral is over all bins. In the case where the pdf is uniformly distributed across all of the bin space, $f(x,y)=f_j=1$ and

$$SC(N,L)=\frac{L^{-2}\left(1-\left(1-L^2\right)^N\right)}{L^{-2}}=1-\left(1-L^2\right)^N \qquad \text{(s3.7)}$$

Equation s3.7 can be arranged into a form equivalent to eq 3 in the manuscript using the expressions $L^2=1/B$ and SC=M/B. It can also be converted into a logarithmic form s3.8 used to generate Figure 7 in the manuscript such that:

$$\log SC(N,L)=\log\left(1-\left(1-\frac{1}{B}\right)^B\right) \qquad \text{(s3.8)}$$

Mutual information calculation was carried out using following equations.

Total number of bins: $B = P \times Q$ (s3.9)

Indexing of bins: p=1…P, q=1…Q

Total number of components: $N = \sum_{p=1}^{P} \sum_{q=1}^{Q} n_{pq}$ (s3.10)

Column-wise (for q-th column) summation of components: $n_{+q} = \sum_{p=1}^{P} n_{pq}$ (s3.11a)

Row-wise (for p-th row) summation of components: $n_{p+} = \sum_{q=1}^{Q} n_{pq}$ (s3.11b)

Consequently: $\sum_{p=1}^{P} n_{p+} = \sum_{q=1}^{Q} n_{+q} = N$ (s3.12)

Mutual information is than calculated according equation s3.12:

$$MI = log(N) + \frac{1}{N} \times \sum_{p=1}^{P} \sum_{q=1}^{Q} n_{pq} log\left(\frac{n_{pq}}{n_{p+}n_{+q}}\right) \quad \text{(s3.13)}$$

A solved example of MI calculation is included in Excel sheet Supplemental SC(G) calculator S1 (worksheet "MI calc example").

Further simplification of relationship shown as equation 9 in the manuscript text:

Because L=B$^{-1/2}$ and for B=N the equation 9 can be rearranged as

$$\log SC = \log A + \left({}^{D}/_{2} - 1\right) \log N \quad \text{s3.14}$$

which for B=N=196 further simplifies as

$$\log SC_G = \log A + \left({}^{D}/_{2} - 1\right) \log 196 \quad \text{s3.15}$$

or

$$SC_G = A \times 196^{\left(\frac{D}{2} - 1\right)} \quad \text{s3.16}$$

The value of A needs to be obtained in the linear region of log *L* in equation 9. Beyond the linear region the equation 9 is not valid. See e.g. the plots in Figure 6D.