THE IMPACT OF 2D-LC/MS PEAK CAPACITY, ORTHOGONALITY AND DUTY CYCLE ON PROTEOMIC ANALYSIS

Martin Gilar, Petra Olivova, John C. Gebler
Life Sciences R&D, Waters Corporation

CHROMATOGRAPHIC PEAK CAPACITY

Peak capacity *P* is a maximum theoretical number of peaks that can be resolved within a gradient time t_g. For example, if the peaks (at 13.4 % peak height) are ~1 minute wide, then for t_g=100 minutes the peak capacity is 101 (equation 1).

$$P = 1 + t_g / w_{13.4\%} \qquad \textbf{(1)}$$

WHAT IS THE PEAK CAPACITY OF 1D LC?

A 150 x 4.6 mm, 5 µm C18 column used with a 100 min gradient a has peak capacity of ~350 (for separation of peptides). A mathematical model fitted with experimental parameters was used to predict chromatographic peak capacity [Gilar at al., J. Chromatogr. A 1061 (2004) 183]. Equation 2 shows that P is a non-linear function of column efficiency *N* and gradient time t_g. Both using a longer column and extending the t_g provides diminishing returns in terms of *P*.

$$P = 1 + \frac{\sqrt{N}}{4} \cdot \frac{B \cdot \Delta C}{B \cdot \Delta C \cdot (t_0 / t_g) + 1} \qquad \textbf{(2)}$$

Figure 1 shows that *P* tends to plateau with high values of *N* and t_g . We estimate that for practical purposes the maximum peak capacity of 1D LC is ~1400-1600.

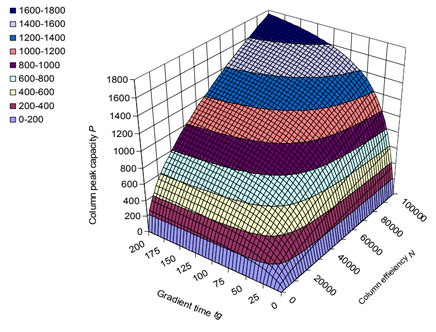


Figure 1: Peak capacity estimated for a well packed chromatographic column with mobile phase containing 0.1% TFA. The values for the nano LC proteomic experiment are likely to be lower (formic acid ion-pairing agent instead of TFA, and post column peak broadening).

WHAT IS THE PEAK CAPACITY OF 2D LC?

The simplest definition of peak capacity in 2D LC space is given by equation 3. The equation does take into account the impact factors such as limited orthogonality of 2D LC chromatographic modes, and the fraction collection frequency. The practical peak capacity is lower than predicted from equation 3.

$$P_{2D} = P_1 \times P_2 \qquad \textbf{(3)}$$

CAN WE ASSUME 100% ORTHOGONALITY IN 2D LC?

Orthogonality for 2D LC separation of peptides is not well known. We evaluated the orthogonality with a set of 196 tryptic peptides (5 protein digest) for selected LC modes [Gilar et al., J. Sep. Sci. 28 (2005) 1694]. The retention data were normalized, and plotted in 2D retention maps (Figure 2). Distribution of data points in 2D space indicates that no 2D LC scenario has an ideal orthogonality (some space is unused for separation). The geometric concept was developed for defining the degree of 2D LC orthogonality [Gilar et al, Anal. Chem. *77* (2005) 6426]. Briefly, the 2D retention maps in normalized space were overlaid with square bins of equal number to the number of peptides. The peak capacity of separation space (sum of all bins) was set to be equal to the number of data points. The area of a bin is equal to area occupied by peptide peak. The bins containing a data point were summed; the area used for separation is used as the measure of orthogonality (for details see paper listed above).

PRACTICAL PEAK CAPACITY IN 2D LC

Practical peak capacity N_p is defined by equation 4. *Pmax* is the total number of bins, ∑*bins* is the number of occupied bins.

$$N_p = P_1 \times P_2 \times \frac{\sum bins}{P_{\max}} \qquad \textbf{(4)}$$

Practical peak capacity N_p in 2D LC depends on the degree of orthogonality and peak capacities in both LC dimensions. When only a limited number of fractions is collected in first LC dimensions, the P_1 and practical peak capacity is reduced (see Table 1).

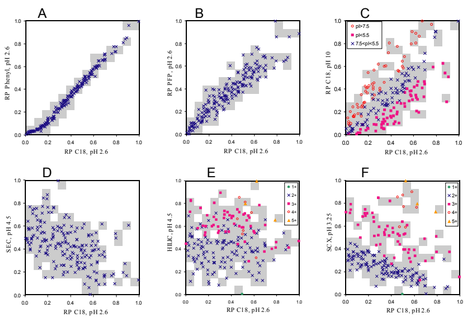


Figure 2: Retention maps of 196 tryptic peptides in normalized 2D LC space: Chromatographic modes: (A) Phenyl x C_{18}, (B) Pentafluorophenyl x RP C_{18}, (C) C_{18}, pH 10 x C_{18}, pH 2.6, (D) SEC x C_{18}, (E) HILIC x C_{18}, (F) SCX x C_{18}. The 2D-space was divided into 196 bins (14 x 14). Bins containing peptides are highlighted in grey. Greater surface coverage leads to a higher orthogonality (Table 1).

Table 1: Orthogonality and peak capacity of investigated 2D LC scenarios. For LC conditions and equations see [Gilar et al, Anal. Chem. *77* (2005) 6426].

Second LC dimension C18, pH 2.6 / First LC dimension	Phenyl	PFP	C18 pH 10	SEC	HILIC	SCX
1D-LC peak capacity	115	115	115	14	79	51
∑bins used	30	52	80	86	100	81
Used 2D area fraction	0.15	0.27	0.41	0.44	0.51	0.41
2D-LC orthogonality %	**13**	**31**	**53**	**58**	**69**	**54**
2D-LC theoretical peak capacity P_{2D}	13225	13225	13225	1610	9085	5865
2D-LC practical peak capacity N_p	1984	3571	5422	708	4633	2405
2D-LC practical peak capacity N_p (10 fractions collected, P_1=10)	172	311	472	506	587	472

WHAT IS THE LC-MS PEAK CAPACITY ?

MS is another separation dimension in LC-MS proteomic experiments. The peak capacity can be defined as a number of MS/MS scans, which, in the ideal case, is equal to the number of identified peptides. The speed of MS/MS data acquisition is important; however, only 10-25% of MS/MS scans provide identifiable MS/MS spectra. If we assume that 1 successful MS/MS scan can be accomplished per second, the peak capacity of 1 hr LC-MS is 3600. For 10 hr 2D LC-MS experiment (10 fractions analyzed in the second dimension), the peak capacity is 18000, assuming limited orthogonality (only 50% of chromatographic time is covered by eluting peaks). The MS defined peak capacity is, to a certain degree, independent of chromatographic peak capacity.

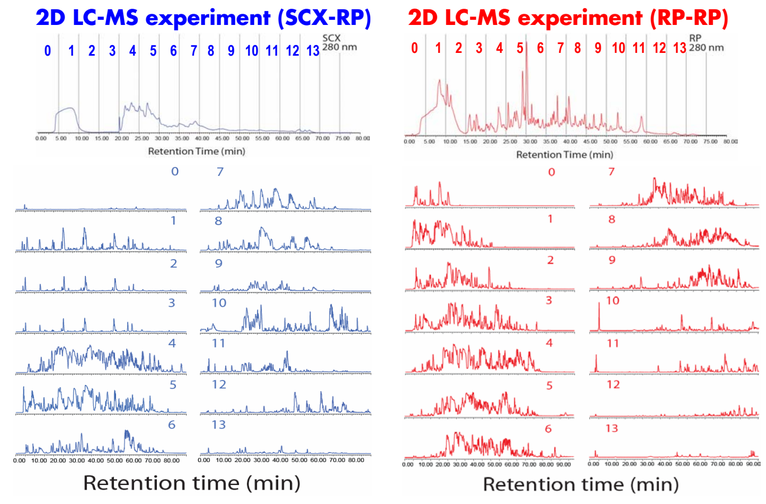
DATA DEPENDENT ACQUISITION (DDA) VERSUS MS^E

Though a powerful technique, DDA has some limitations. The number of MS/MS scans is typically lower than the complexity of the sample (undersampling). Also, the algorithm preferentially selects the most abundant ions for MS/MS analysis. Consequently, the low abundant species may not be identified, although present in sample in detectable quantities. Novel MS^E approach [Silva et al., Mol. & Cell. Proteomics 5 (2006) 589] uses a parallel collection of MS and MS/MS data. The dedicated software assigns the fragments to the appropriate parent ions (precursors) according to the accurate mass and retention time values. MS^E peak capacity is comparable to DDA. It provides greater sequence coverage, and better detection limits of peptides/proteins. Data acquisition is highly reproducible.

ANALYSIS OF UNDEPLETED HUMAN SERUM DIGEST USING MS^E - COMPARISON OF RP-RP AND SCX-RP 2D LC METHODS.

A recently developed RP-RP 2D LC method was compared to conventional SCX-RP 2D LC. The equivalent of 50 µl of undepleted human serum digest was fractionated in the first LC dimension (2.1x150 mm). The equivalent of 1 µl of serum was injected onto a 0.3x150 mm C18 column. Only pepties identified in 2 out of 3 repetitive analyses with MASCOT score >20 and comparable retention (within 2.5 minutes window) were accepted as correct hits. The results indicate that RP-RP provides similar or better results compared to SCX-RP 2D LC; **616**/**180** and **514**/**149** **peptides**/**proteins** were identified, respectively. The overlap between the experiments was **114** proteins and **317** peptides; these represent the more abundant species. Lower abundant ions (near LOD) show greater variability.

Figure 3: Comparison of 2D LC-MS experiments; analysis of human serum.



WHY DON'T WE IDENTIFY 18 000 PEPTIDES?

The majority of peptides cannot be detected because they are present at concentrations below the LOD of a given MS instrument. Greater mass load would result in column overloading (with abundant peptides). Depletion of 6-20 abundant serum proteins improves the LOD by 1-2 orders of magnitude; however, it remains to be seen how many additional proteins can be reliably identified in such experiments. The ultimate LOD depends on the amount of serum available for analysis.