The Other Sanger Sequencing

George W. Preston

E-mail: george.preston@kcl.ac.uk

Earlier this year, a plaque honouring the chromatographers Archer J. P. Martin (1910-2002) and Richard L. M. Synge (1914-1994) appeared in Headingley, Leeds, United Kingdom (Fig. 1A). The plaque, which was unveiled by Leeds Philosophical and Literary Society, can be found on Headingley Lane, near to where the laboratories of the Wool Industries Research Association once stood. It was in these laboratories in the 1940s that Martin and Synge developed partition chromatography – a method for which the pair would receive the 1952 Nobel Prize in Chemistry (Martin, 1952; Synge, 1952). Taking Martin and Synge's work as a starting point, this article will examine the role of chromatography in efforts to elucidate primary structures of proteins and, latterly, to explore proteomes.

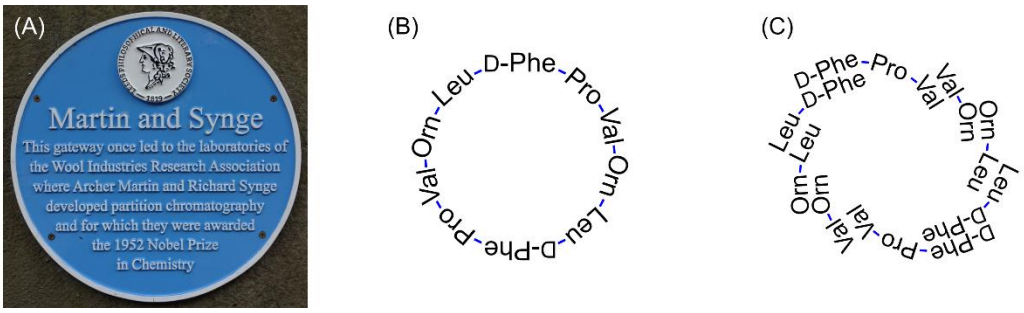

Fig. 1. The work of Martin and Synge. (A) The plaque on Headingley Lane; (B) the structure of gramicidin S; (C) the structure of gramicidin S as a series overlapping fragments.

PARTITIONING PEPTIDES

Partition chromatography is the separation of substances based on how they distribute among different phases. Martin and Synge's innovation was to support an aqueous phase on a material such as silica or paper, and to pass over this *stationary phase* an organic *mobile phase*. They imagined the resulting system as a stack of miniature liquid-liquid extractions, or 'theoretical plates', between which the mobile phase (and any analytes dissolved in it) could migrate (Levine, 1963).

By enabling the breakdown products of proteins – i.e., amino acids and peptides – to be analysed independently of each other, chromatographic methods greatly accelerated research into protein structure. Methods including partition chromatography allowed the sequences of individual peptides to determined, and this in turn allowed complete protein sequences to be pieced together. Martin, Synge and colleagues' work on the cyclic decapeptide gramicidin S (Fig. 1B) is an example of a sequence can be built up in this way, and how chromatography can assist in the endeavour. The team partially hydrolysed gramicidin S using a mixture of acetic and hydrochloric acids, and separated the resulting di- and tripeptides using two-dimensional partition chromatography on paper (Consden et al., 1947). Individual peptides were identified using an assay involving deamination, hydrolysis and further chromatography, and the sequence of intact gramicidin S was deduced from overlaps of the peptides' sequences (Fig. 1C).

THE OTHER SANGER SEQUENCING

Martin and Synge's work was to impress one Frederick Sanger (1918-2013), who at the time was working to determine the sequence of bovine insulin (51 amino acid residues). Nowadays, Sanger is perhaps better known for his use of dideoxynucleoside triphosphates to sequence nucleic acids, and it is with this later method that 'Sanger sequencing' is synonymous. His sequencing of insulin was, however, a major technical accomplishment in its own right, and represented a significant advance in the understanding of protein structure. As Sanger put it in reference to his 1949 report in *Biochemical Journal*, the work "showed proteins were real chemicals with a defined sequence" (Brownlee, 2014).

Sanger's report describes the partial sequencing of insulin's 'phenylalanyl chain' – one of two chains in the molecule (Sanger, 1949). Sanger separated this chain from its counterpart, the glycyl chain, by oxidising the disulfide bonds that held the two chains together. He derivatised the chain's α-amino group with 1-fluoro-2,4-dinitrobenzene, then partially hydrolysed the derivative with hydrochloric acid. The resulting *N*-2,4-dinitrophenyl (DNP) peptides were then fractionated by partition chromatography and subjected to further analyses. The simplest DNP-peptide was found to consist only of valine and DNP-phenylalanine, meaning that the chain's N-terminal dipeptide had to be Phe-Val. By

extending this approach, Sanger and his colleagues were able to deduce the full sequences of both of insulin's chains, and to determine the positions of the molecule's three disulfide bonds (Sanger, 1988).

The American biochemists William H. Stein (1911-1980) and Stanford Moore (1913-1982) picked up where Sanger left off, setting their sights on the sequencing of ribonuclease (124 amino acid residues) (Kresge et al., 2005). Together with associate Werner Hirs, they oxidised bovine ribonuclease and digested it with trypsin, chymotrypsin or pepsin. The resulting peptides were separated using ion exchange chromatography, and individual peptides were subjected to amino acid analysis and Edman degradation (see Hirs et al., 1960 and references therein). Central to the team's approach was the use of different enzyme specificities to produce series of overlapping peptides. Starting with the N-terminal heptapeptide of the trypsin series, for example, the sequence could be extended via the N-terminal octapeptide of the chymotrypsin series, then via the second peptide of the trypsin series, and so on until the C-terminus was reached.

FROM PROTEINS TO PROTEOMES

Fast-forward 60 years, and ion exchange chromatography is still being used to separate peptides. It is sometimes used as the first of two dimensions of separation, the second being some sort of high-pressure liquid chromatography. Peptides' sequences are now often analysed using mass spectrometry, a method that can be used 'online' (i.e., to analyse eluate directly, on the timescale of the chromatography). With the sequences of thousands of proteins now known, this kind of technology can be used to identify proteins from enzymatically-digested proteomes – i.e., very complex mixtures of peptides.

A recent report by Kwiatkowski et al. illustrates the sophistication of modern methods. Writing in *Analytical Chemistry*, the authors describe the application of a technique known as displacement chromatography to the separation of peptides (Kwiatkowski et al., 2018). Displacement chromatography is based on the principle that, if forced to compete for binding sites on a stationary phase, analytes of higher affinity will displace ones of lower affinity. If competition is effected under flow conditions, analytes will spread out along the length of the stationary phase in order of descending affinity. The analytes can then be eluted as bands, in order of ascending affinity, by injecting portions of a 'displacer' molecule. The displacer outcompetes the analytes nearest the top of the stationary phase, which are then free to displace analytes bound further down. These are then free to compete with the next band of analytes, and so on until the lowest-affinity analytes elute from the end of the stationary phase (Fig. 2).

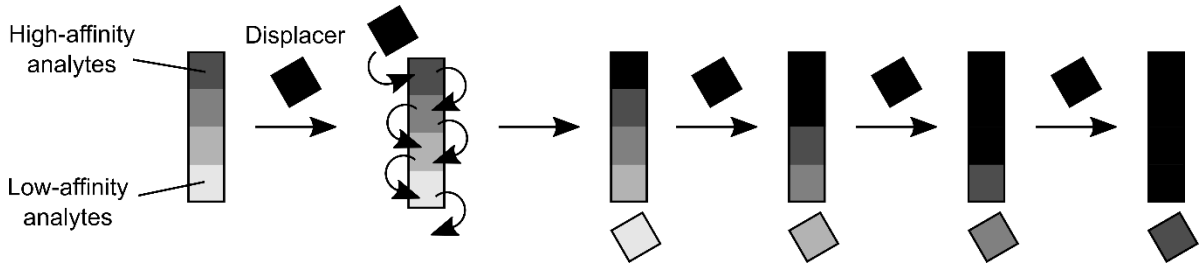

Fig. 2. Displacement chromatography. Analytes spread out along the length of a stationary phase in order of decreasing affinity, and are eluted using a displacer molecule.

Kwiatkowski et al. digested proteins from cultured human cells with trypsin, and loaded the resulting peptides onto a cation exchange resin in a mobile phase of pH 2.3. Under these conditions, most tryptic peptides (i.e., peptides formed by the action of trypsin) would be expected to carry at least two positive charges – one at each terminus, plus one for each internal basic group. Using the polyamine spermine as their displacer, and analysing their eluted peptides with reversed-phase liquid chromatography and mass spectrometry, the authors observed separation of peptides with between one and five charges (singly-charged peptides eluting first, peptides with five charges eluting last). Compared to a more conventional method involving gradient elution with ammonium acetate, the displacement method effected better separation and enabled more peptides to be identified. It was particularly successful in analysing doubly-charged peptides.

REFERENCES

Brownlee GG (2014) *Fred Sanger – Double Nobel Laureate: A Biography*. Cambridge: Cambridge University Press. doi: 10.1261/rna.055590.115

Consden R, Gordon AH, Martin AJP, Synge RLM (1947) Gramicidin S: The sequence of the amino-acid residues. *Biochem. J.* 41, 596-602. PMID: 16748221

Hirs CHW, Moore S, Stein WH (1960) The sequence of the amino acid residues in performic acid-oxidized ribonuclease. *J. Biol. Chem.* 235, 633-647. PMID: 14401957

Kresge N, Simoni RD, Hill RL (2005) The elucidation of the structure of ribonuclease by Stanford Moore and William H. Stein. *J. Biol. Chem.* 50, e47. http://www.jbc.org/content/280/50/e47.full.pdf

Kwiatkowski M, Krösser D, Wurlitzer M, Steffen P, Barcaru A, Krisp C, Horvatovich P, Bischoff R, Schlüter H (2018) Application of displacement chromatography to online two-dimensional liquid chromatography coupled to tandem mass spectrometry improves peptide separation efficiency and detectability for the analysis of complex proteomes. *Anal. Chem.* 90, 9951-9958. PMID: 30014690

Levine J (1963) Column partition chromatography in pharmaceutical analysis. *J. Pharm. Sci.* 52, 1015-1031. PMID: 14079626

Martin AJP (1952) The development of partition chromatography. Nobel Lecture. http://www.nobelprize.org/uploads/2018/06/martin-lecture.pdf

Sanger F (1949) The terminal peptides of insulin. *Biochem. J.* 45, 563-574. PMID: 15396627

Sanger F (1988) Sequences, sequences and sequences. *Annu. Rev. Biochem.* 57, 1-28. PMID: 2460023

Synge RLM (1952) Applications of partition chromatography. Nobel Lecture. http://www.nobelprize.org/uploads/2018/06/synge-lecture.pdf