Nanopores and nucleic acids: prospects for ultrarapid sequencing

David W. Deamer and Mark Akeson

DNA and RNA molecules can be detected as they are driven through a nanopore by an applied electric field at rates ranging from several hundred microseconds to a few milliseconds per molecule. The nanopore can rapidly discriminate between pyrimidine and purine segments along a single-stranded nucleic acid molecule. Nanopore detection and characterization of single molecules represents a new method for directly reading information encoded in linear polymers. If single-nucleotide resolution can be achieved, it is possible that nucleic acid sequences can be determined at rates exceeding a thousand bases per second.

Sometime in the next decade, DNA sequencing will reach a pinnacle when the three billion base sequence of the human genome becomes available to anyone with a link to the Internet. Achieving this goal has required 25 years of effort, following on from the pioneering publications of Sanger *et al.*[1], and Maxam and Gilbert[2], who first described methods for sequencing DNA. The most efficient current methods can sequence ~30 000 bases per day per instrument, at an approximate cost of US$0.50 per nucleotide for a finished sequence.

The Sanger method, which is in common use for DNA sequencing, typically requires two working days and approximately 10^{10} nucleic acid fragments to produce a detectable band by gel electrophoresis. By contrast, in the nanopore method, an electric field can cause individual nucleic acid molecules to move through a single nanopore, 2 nm in diameter, on a microsecond to millisecond timescale. Because the pore is so narrow, it is restrictive, and the molecules are translocated necessarily as single strands, rather than duplexes, and move through the pore in strict linear sequence. Furthermore, a nucleic acid molecule generates a distinctive electrical signal as it enters and passes through the pore.

The concept of nanopore sequencing is dependent on the possibility that each base in the nucleic acid will modulate the signal in a specific and measurable way as it passes through the pore. If this were the case, the base sequence in a nucleic acid molecule could be determined at rates between 1000 and 10 000 per second, analogous to the decoding of magnetic signals on a tape as it passes through the head in a tape player. Granted, this scenario is hard to imagine, but in fact all of the steps except the last step (single-base resolution) have been achieved. Research on the nanopore sequencing of single-stranded nucleic acids was initiated in 1991 by David Deamer (University of California, Santa Cruz, CA, USA) in collaboration with Daniel Branton (Harvard University, Cambridge, MA, USA), and independently for double-stranded DNA by George Church and Richard Balderelli (Harvard Medical School, Boston, MA, USA). Because the nanopore concept requires a very small pore of molecular dimensions to force a long strand of DNA to move through lengthwise, the first question to be addressed was what kind of pore could be used, and how could a molecule of DNA be threaded through such a narrow constriction?

In the late 1980s, John Kasianowicz (National Institute for Standards and Technology, Gaithersburg, MD, USA) was working on α-hemolysin, a 33 kD protein isolated from *Staphylococcus aureus*, which self-assembles in lipid bilayers to form a channel with a relatively large pore. Kasianowicz found that the α-hemolysin channel remains open at neutral pH and high ionic strength. Furthermore, unlike most membrane channels, α-hemolysin passes a steady ionic current in the range of 100 picoamps (pA) with an applied voltage of 100 mV; by contrast, other biological channels carry currents of just a few pA, under similar conditions. The presence of a current of this magnitude suggested that the limiting pore of α-hemolysin had a diameter of ~2 nm. This estimate was confirmed when Song *et al.*[3] determined the structure of the channel at 1.9 Å resolution (Fig. 1). The important structural features are the mouth of the channel (2.6 nm), which leads into a larger vestibule, and the stem of the channel containing a pore with an internal diameter of ~2.2 nm. The opening between the vestibule and the stem forms the limiting aperture of 1.5 nm diameter, and is composed of a ring of 14 alternating lysine and glutamate side chains.

The basic question of whether or not the α-hemolysin pore could detect nucleic acids was first tested using homopolymers of RNA (Ref. 4). These were chosen because they have a minimal tendency to form secondary structures that might block the channel. The expectation was that individual strands of the homopolymers would be drawn into the pore by the standing electrical field, and would thereby block the ionic current that otherwise flows unimpeded through the pore. The satisfying result was that ionic current blockades appeared immediately after polymers such as polyadenylic acid (polyA) or polyuridylic acid (polyU) were added to the *cis* side of a lipid bilayer containing a single α-hemolysin channel (Fig. 2). The blockade events required 100 mV applied voltage to be observed, positive on the *trans*-side of the membrane, and were hundreds of microseconds to several milliseconds in

D.W. Deamer (deamer@hydrogen.ucsc.edu) and M. Akeson (makeson@hydrogen.ucsc.edu) are at the Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064, USA.

 0167-7799/00/$ – see front matter © 2000 Elsevier Science Ltd. All rights reserved. PII: S0167-7799(00)01426-8

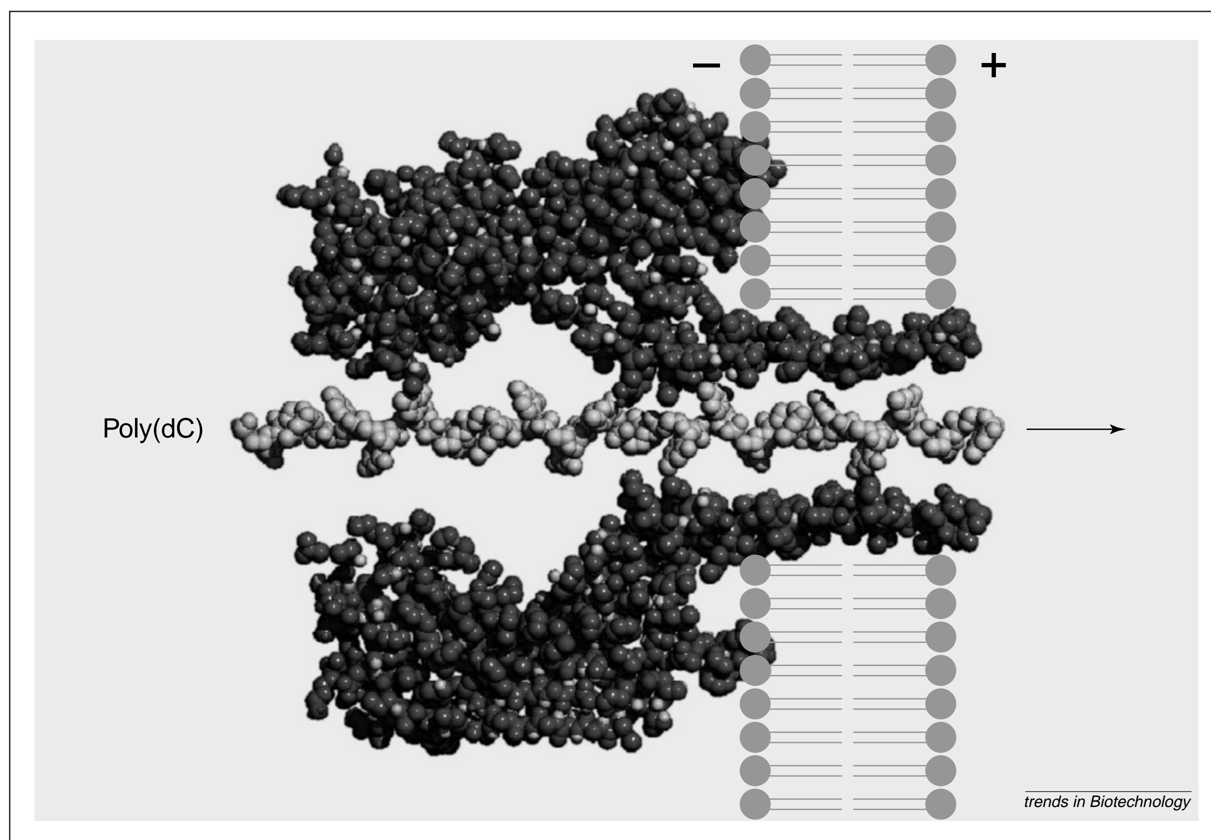

Figure 1
A single α-hemolysin channel is shown in cross-section embedded in a lipid bilayer. A voltage has been imposed, *trans*-side positive, and a single strand of poly(dC) DNA is being driven through the pore by the electrical field. (Image prepared from coordinates provided by Song *et al.*[3])

duration. They entirely disappeared when the polarity was reversed, as would be expected if electrophoresis were pulling anionic nucleic acids through the pore. The addition of ribonuclease A to a polyA solution had no effect, but the same addition to a polyU solution was followed by an enormous burst of blockades for one to two minutes, which then rapidly decreased to approximately zero after 10–15 minutes. This is exactly what would be expected if the enzyme cut the polyU first into multiple fragments that produced increased numbers of blockades, and then into ever-smaller fragments that finally could not be resolved; the substrates for ribonuclease A are pyrimidine-containing ribopolynucleotides such as polyU.

Although the preliminary evidence was exciting, it was still difficult to believe that a seemingly small applied potential of 100 mV could capture the ends of individual nucleic acid molecules and draw them through a 1.5 nm pore. It was necessary to ensure that the nucleic acid molecules actually traversed the channel and produced an ionic current blockade by occupying the pore. This experimental approach was developed by Dan Branton (Harvard University) and involved a comparison of the ability of DNA as single-stranded (ssDNA) and double-stranded (dsDNA) molecules to pass through the pore. Quantitative PCR was then used to make a direct measurement of nucleic acids appearing on the *trans* side of the membrane, and the numbers were compared with the number of blockade events that occurred during the same time interval. It was found that only ssDNA molecules appeared on the *trans* side; dsDNA molecules remained on the *cis* side of the pore. The ratio of single-stranded DNA molecules appearing on the *trans* side to the number of blockade events was close to 1. Furthermore, the duration of the blockade was linearly proportional to the length of the nucleic acid molecules between 75 and 430 nucleotides in length. The rate of translocation was constant, ~1 microsecond per nucleotide for polyU, for example, or ~20 microseconds per nucleotide for polyA. The most reasonable explanation for these observations was that single-stranded nucleic acid molecules were captured by the electrical field and driven through the pore[4].

To take the next step towards sequencing, it was necessary to demonstrate that the nanopore can distinguish between the bases present in nucleic acids. Thus, homopolymers were again used to determine whether or not they produce specific and distinguishable blockade signatures. There are two primary components of a blockade event caused by a nucleic acid molecule passing through the pore: amplitude and duration. Therefore, these two parameters were measured for short homopolymers of RNA, and the remarkable result was that the pore easily detected measurable differences in both amplitude and duration[5]. For instance, polyA produced blockades of ~85%, and polyC ~95%. Furthermore, there were dramatic differences in the velocity with which the polymers traversed the pore. PolyC, for instance, passed through at three microseconds per nucleotide; polyA rates were 20 microseconds per nucleotide. This information can be summarized for homopolymers of similar lengths as an event diagram in which blockade amplitude is plotted against

duration. One such diagram for polyA and polyC is shown in Fig. 3. Each dot represents the amplitude and duration of a single RNA molecule passing through the nanopore. The plot clearly demonstrates the extraordinary ability of the nanopore to distinguish between two RNA species that differ only in that one has a purine base and the other a pyrimidine base.

The next step was to determine whether or not the nanopore would detect a difference between purine and pyrimidine bases when a single molecule of a nucleic acid passed through the pore in several hundred microseconds[5]. The approach was to synthesize an RNA molecule from a DNA template to produce a sequence of 70 cytosine ribonucleotides (Cs) followed by a sequence of 30 adenine ribonucleotides (As). Representative blockades produced by this molecule are shown in Fig. 4, and are compared with blockades from the homopolymers of polyC and polyA. Most of the blockades showed a clear transition from a 95% amplitude, characteristic of polyC, to an 85% amplitude, characteristic of polyA. This represented a signature characteristic of a nucleotide sequence, not of single nucleotides in a polymer, but of a group of pyrimidines followed by a group of purines.

However, adenine is a purine, with a chemical structure composed of fused heterocyclic five- and six-membered rings; by contrast, cytosine, a pyrimidine, has only a six-membered ring. Therefore, all other things being equal, why would polyC, with its smaller base, produce a larger blockade than polyA, yet move through the pore nearly six times faster? The answer probably lies in the fact that homopolymers of polyC and polyA RNA (but not polyU) are able to form stable single-stranded helices[6,7]. The helix of polyC is close to 1.3 nm in diameter, which barely fits through the 1.5 nm limiting aperture of the pore; the polyA helix is 2.2 nm in diameter. Our working model is that the polyC helix nearly fills the pore as it passes through, blocking most of the ionic current and producing a characteristic 95% blockade. PolyA, by contrast, can apparently get into the vestibule and produce the partial blockade that is commonly observed. Occasionally, one end of the polyA enters the limiting aperture, but only if it is extended from the helix structure. The energy required to extend the helix would slow its

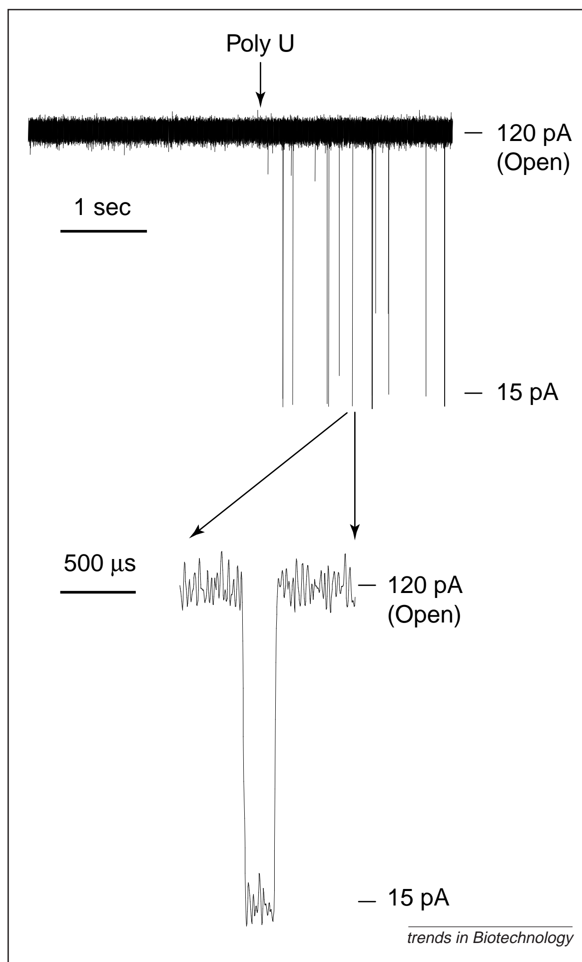

Figure 2
(a) Ionic current blockades produced by polyU RNA homopolymers, each ~150 nucleotides in length. **(b)** A magnified portion of a single blockade event.

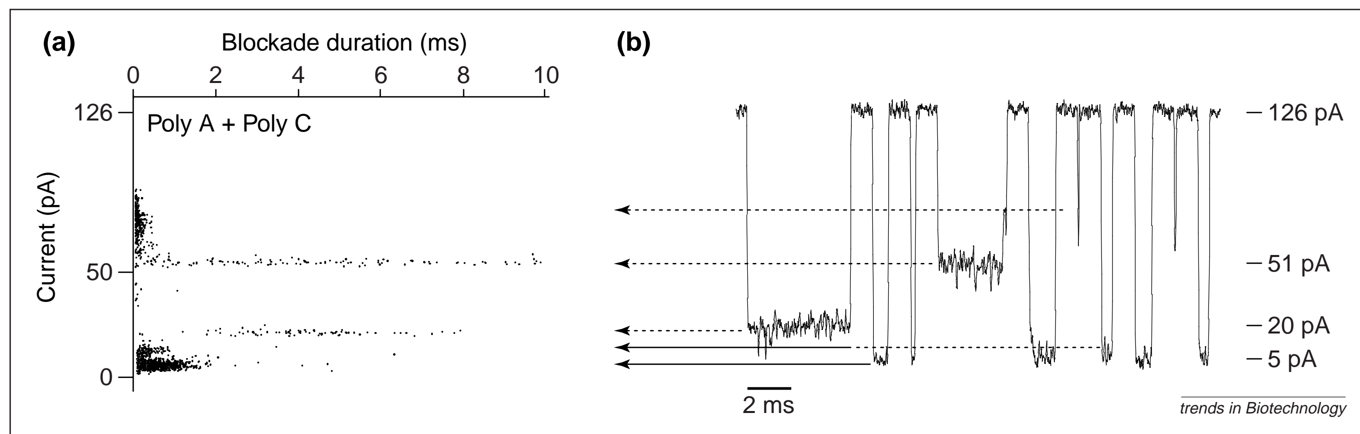

Figure 3
(a) Event diagram showing the amplitude and duration of polyC and polyA blockades in a 1:1 mixture of the two homopolymers. The upper scale (*x* axis) is in milliseconds (ms), and the *y* axis is in picoamps (pA), where 0 represents complete blockage. **(b)** Examples of individual blockades that are summarized in **(a)**. PolyC (solid arrows) produces relatively short blockades with amplitudes of 91% and 95%; polyA (dashed arrows) produces blockades of 85% that are longer in duration. The long 55% partial blockades associated with polyA are not length dependent and are probably caused by the polymer entering the pore vestibule, then falling out before translocation occurs. The very short duration 35–50% polyA blockades most likely represent transient collisions with the pore opening.

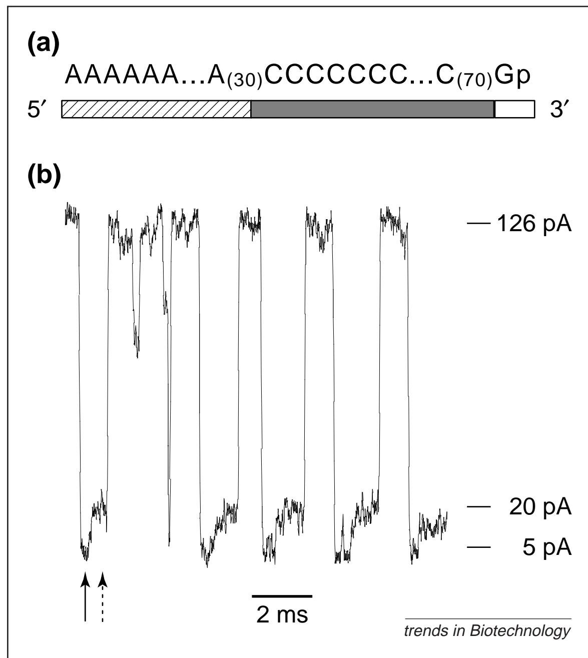

Figure 4

Blockades produced by a block copolymer of RNA containing two segments composed of purine and pyrimidine nucleotides (A_{30} and C_{70}). The single G at the 3′ end is a remnant of the enzymatic process that produced the synthetic 100mer of RNA. Note that virtually every full blockade has two residual current levels, one at 5 pA, characteristic of polyC (95% blockade; solid arrow), and one at 20 pA, characteristic of polyA (85% blockade; dashed arrow).

passage, although at the same time the extended form would occupy less volume in the pore than polyC, thus producing a smaller blockade of 85%. DNA has less tendency to form single-stranded helices, and both poly(dC) and poly(dA) pass through at faster rates than polyC and polyA RNA homopolymers.

Future prospects

A successful method for nanopore sequencing would enable one to determine the sequence of entire genomes at rates of thousands of bases per second and would have obvious applications in research and clinical diagnostics. However, there are still major hurdles to be overcome, the most important being single-nucleotide resolution. The limit of resolution at present is that the nanopore can certainly detect pentanucleotides in free solution, and can distinguish the transition between a 30-nucleotide segment of A and a 70-nucleotide segment of C in an RNA 100mer. In addition, polynucleotides pass through the pore at rates ranging upwards from one nucleotide per microsecond. However, the actual number of ions involved in the transition between a single purine and a single neighboring pyrimidine is only ~100 ions per microsecond. Because the time interval available for a measurement is a few microseconds per base, this difference is lost in the noise. A straightforward solution would be to slow down the transport rate by 100–1000-fold, so that each base occupies the pore for up to a millisecond; this is presently being investigated.

Another problem arises from the pore structure itself. First, it is composed of a complex protein embedded in a lipid bilayer and is intrinsically unstable. A more robust nanopore will be required for research and clinical applications. Furthermore, even though there is a limiting aperture of 1.5 nm in α-hemolysin, this aperture leads to a 5 nm neck with a diameter of 2.2 nm. Depending on conformation (stacked vs extended chains), about 10–15 nucleotides are present in the neck at any given time and these will contribute to the amplitude of the blockade. The ideal solution, of course, is to make a 1.5 nm pore in an insulating material that is approximately the same thickness as a single nucleotide, that is, 0.3–0.4 nm. This has never been done before, but novel approaches are being explored.

Although it might be several years before actual nanopore sequencing becomes a possibility, the research effort to develop the present instrument has already opened a wealth of possible applications. For example, because the nanopore detects single molecules in solution, the instrument has the potential to measure the concentration of pure DNA in microscopically small volumes. The nanopore can also provide rough estimates of chain length, and would have no difficulty in analysing populations of synthetic DNA fragments that vary by 10–20 nucleotides.

Perhaps most intriguing is the ability of the nanopore to distinguish between segments of different chemical composition. This leads to the concept of targeted molecular bar codes (TMBCs). Suppose, for example, that we wanted to detect and measure a known set of cell-surface antigens, such as the human-leukocyte-antigen group used in tissue typing. It is possible to build a series of molecular bar codes, each capable of producing a specific signal as it passes through the nanopore. The bar codes could be block copolymers of any linear polymer bearing an ionic charge. For example, bar codes composed of DNA can be produced on a DNA synthesizer, using both normal and artificial monomers to produce readable segments. Each bar code would then be covalently linked to a targeting agent, such as the Fab fragment of an antibody specific for one of the cell-surface antigens. To perform an assay, the entire mixture of TMBCs would be added to the cells to be typed, followed by centrifugation. Those ten or so TMBCs that were specific for the cell-surface antigens would bind to their targets, while the rest of the TMBCs would be washed away. In the last step, the bond linking the bar code to the targeting group would be broken and the mixture analysed by the nanopore. The antigenic sites present on the cell surface would be identified by the TMBC present, and their relative numbers could be directly calculated.

To summarize, single-base resolution in DNA or RNA molecules remains the goal of nanopore sequencing, and several significant intermediate applications are now within reach. These include the simple detection of DNA in microscopic volumes, as well as TMBCs. Although nanopore analysis is just emerging as a new research tool, the rapidly developing technology has immense promise as an analytical method for measuring and characterizing linear polyions such as nucleic acids.

Acknowledgments

The authors thank H. Olsen for preparing the computer graphic representations of α-hemolysin.

References

1 Sanger, F. *et al.* (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–5467
2 Maxam, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U. S. A.* 74, 560–564
3 Song, L. *et al.* (1996) Structure of staphylococcal α-hemolysin, a heptameric transmembrane pore. *Science* 274, 1859–1865
4 Kasianowicz, J.J. *et al.* (1996) Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13770–13773
5 Akeson, M. *et al.* (1999) Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys. J.* 77, 3227–3233
6 Fasman, G.D. *et al.* (1964). The helical conformations of polycytidylic acid: studies on the forces involved. *Biochemistry* 3, 1015–1021
7 Saenger, W. *et al.* (1975). A structural model for the polyadenylic acid single helix. *J. Mol. Biol.* 93, 529–534

Single-cell MALDI: a new tool for direct peptide profiling

Lingjun Li, Rebecca W. Garden and Jonathan V. Sweedler

Matrix-assisted laser desorption-ionization (MALDI) mass spectrometry (MS) is a rapid and sensitive analytical approach that is well suited for obtaining molecular weights of peptides and proteins from complex samples. MALDI-MS can profile the peptides and proteins from single-cell and small tissue samples without the need for extensive sample preparation, except for the cell isolation and matrix application. Strategies for peptide identification and characterization of post-translational modifications are presented. Furthermore, several recent enhancements in MALDI-MS technology, including *in situ* peptide sequencing as well as the direct spatial mapping of peptides in cells and tissues are discussed.

There has been a long-standing interest in elucidating the complex chemical interactions that occur within individual cells. Because of the intricacy and minute volumes of most cells, microseparation techniques are often used. Ideally, capillary separation methods such as capillary liquid chromatography (LC) or capillary electrophoresis separate nanoliter-volume samples into their individual components for assay by a chemically information-rich detector. The inherent sensitivity and high information content of mass spectrometry (MS) makes it a desirable detector for single-cell assays. In combination with microseparation strategies, soft ionization methods, such as electrospray ionization and matrix-assisted laser desorption-ionization (MALDI; see Glossary), offer unparalleled chemical information. The use of these methods has revolutionized the field of protein and peptide chemistry[1,2] and is having an increasing impact on single-cell measurements[3]. One notable study by Hsieh *et al.* combined the resolving power of capillary LC with the structural information of MS to characterize neuropeptides from a single identified neuron of the freshwater snail *Lymnaea stagnalis*[4].

During the past decade, MALDI-MS (Box 1) has also developed into an important tool for the direct analysis of complex, unfractionated samples[2]. Significantly, this technique provides the exciting possibility of studying the pattern of neuropeptide gene expression at and below the level of a single cell. One important application of MALDI-MS is direct peptide profiling, which involves measuring the molecular weights of all the proteins and peptides present at significant concentrations directly from a cell or tissue sample without a

L. Li, R.W. Garden and J.V. Sweedler (jsweedle@uiuc.edu) are at the Department of Chemistry and Beckman Institute, University of Illinois, Urbana, IL 61801, USA.

Glossary

Collision-induced dissociation (CID) A fragmentation technique used to obtain primary structural information. This approach uses collision with a target gas to induce fragmentation of a mass-selected ion.
In-source decay (ISD) A fragmentation technique used with matrix-assisted laser desorption-ionization–time-of-flight mass spectrometry (MALDI–TOF MS) to obtain structural information, typically from peptides greater than 1500 Da. The decay of the precursor ions occurs in the MALDI ion source, yielding a series of product ions indicative of amino acid sequence.
Matrix-assisted laser desorption-ionization (MALDI) A sample introduction and/or ionization method that generates ions by desorbing them from a solid matrix material with a pulsed laser beam. This ionization method is tolerant of high levels of impurities.
Post-source decay (PSD) A fragmentation technique used with MALDI–TOF MS to obtain structural information, typically from peptides less than 2000 Da. The decay of a mass-selected precursor ion occurs after the ion leaves the ion source in the flight tube, with the fragment ions being separated in the reflectron.
Reflectron A device used in a time-of-flight mass spectrometer to increase the mass resolution. This device retards and then reverses ion velocities in order to correct for the flight times of ions having different kinetic energies and to increase the effective flight path.
Time-of-flight (TOF) One of the simplest mass analysers that is used to separate ions based on their mass-to-charge ratio (*m/z*). The time of an ion traveling through the flight tube is correlated to its *m/z*, with lighter ions arriving earlier and heavier ions arriving later.

0167-7799/00/$ – see front matter © 2000 Elsevier Science Ltd. All rights reserved. PII: S0167-7799(00)01427-X