STOCHASTIC SENSING OF BIO-POLYMERS USING PATTERNED NANOPORES

J. A. Cohen, A. Chaudhuri and R. Golestanian
The Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK.

Bio-polymers are the building materials of life. Polymers are types of molecules that are formed by linking different chemical units together in a chain, much like a series of carriages along a train. One familiar bio-polymer is DNA, comprised of four units named *bases* commonly abbreviated to A, C, G and T. One can think of these bases as letters which are written along DNA to form words, sentences and even volumes of building 'instructions' for the individual components in a living organism. The total set of instructions for the human body is known as a genome and consists of over 3 billion bases. If stored digitally, this information would occupy under 1 gigabyte (GB) of memory, allowing several to be stored on a single iPod. The words formed by individual bases, known as genes, reveal the important differences between the makeup of one person to another. If one could know his or her individual set of genes then it is possible to predict: if any genetic diseases are carried, the likelihood of developing certain types of cancers and even how you may respond to certain forms of treatment. Medicine could be tailored to fit one's gene type by recommending a treatment that responded well to others with a similar genetic makeup, analogous to how online vendors such as Amazon recommend books you might like by comparing your previous purchases to those made by others. Such a 'recommendation engine' could be implemented in medicine as long as each patient's genome is easily and cheaply obtainable. The problem now becomes the technical challenge of reading DNA accurately and at low cost. Bio-polymers exist at the nanoscale in a fluid mixture of water and salts. To put this scale into perspective imagine the following: a human hair has a width of around one tenth of a millimetre, if you split this hair into a hundred equally sized pieces they would be a micrometer wide. Splitting a single micrometer slice equally into a further thousand pieces would take you to the nanoscale. Physics of the nano world is very different to that experienced in everyday life. Movement alone is very challenging, it would be somewhat like trying to cross a crowded high street if other fellow pedestrians only changed their direction on collision. Your path is no longer a direct line but consists of many erratic turns, perhaps in the opposite direction to that intended. The time it would take you to cross the street is no longer a definite single value but could take a range of times with assigned probabilities, call this the **distribution of translocation times**. This kind of motion is random, or *stochastic*. We must consider the distribution of times and the probabilities with which they occur when working at the nano-scale. A current active area of research involves measuring the crossing time for a bio-polymer through a hole or channel of similar diameter called a *nanopore*. These nanopores can be made by firing a stream of electrons into a thin *silicon wafer* to 'mill' a hole. The time taken for a bio-polymer to cross the nanopore is found by conducting an electric current across this hole. Once the bio-polymer enters it will block the pore and stop the flow of current. This drop in current can then be used to reconstruct the crossing time. Experiments on the nanoscale are extremely challenging and, in some sense, are performed blind because light, as used in microscopy, is much too large to resolve nano sized features. In fact the only way to 'see' what is happening is by computer simulation providing a 'virtual nano-scope'. In collaboration with A.C. and R.G., I have created a unique simulation environment that can be used to investigate the movement of bio-polymers through pores of different structure and composition. By experimentation and analysis of the results of over 100, 000 simulations, we have shown that by modifying the inside of the nanopore to include patterns of different attractive and repulsive patches, such that the DNA base A would be more attracted to the patch than the base C, we can control the **distribution of translocation times**. As a result of these studies, I have proposed and theoretically demonstrated a system that uses the crossing time measurements of an array of different nanopore types to identify an unknown bio-polymer. This is achieved by utilising the different responses of the specifically engineered pores. A specially developed algorithm then considers how likely the measured times are for a particular structure by using the **distribution of translocation times**. This systematically eliminates unlikely structures to determine the most likely structure. We have shown that this method can be used to identify simple bio-polymers to any required accuracy. Our system is novel in its approach as it allows for fast measurements by exploiting the stochastic environment, unlike other nanopore based methods that try to eliminate the 'noisy' environment by slowing down the crossing process to give a 'clean' measurement. Further development of this work could lead to accurate, affordable and rapid DNA sequencing, heralding a new era of globally unified, statistically based treatment, tailored to the individual by learning from our genetic partners.