GENOMICS

Sequencing in a flash

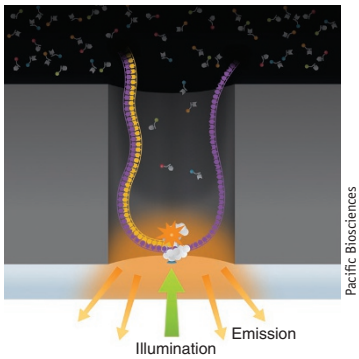
A new system capable of simultaneously monitoring thousands of individual polymerase enzymes enables accurate, multiplex DNA sequencing in real time.

Scientists would be hard-pressed to engineer a reagent for DNA synthesis that rivals the speed and efficiency of DNA polymerase, and a platform that fully exploits this enzyme's potential as a driver for fast and accurate real-time sequencing would be a powerful asset. However, developing such a system has posed a daunting technical challenge, requiring the capability to directly monitor the activity of individual enzymes without impeding their function.

Steve Turner and Jonas Korlach first started collaborating to tackle this task over a decade ago, as graduate students at Cornell University. Korlach had initially approached Turner, hoping to benefit from his expertise in nanostructure design, to develop a system for imaging individual polymerases. "I was originally interested in trying to find a way to see these machines in real time and understand their dynamics," recalls Korlach. "And then very quickly, we realized that if you could do that and identify which base is incorporated at any time, then you could have a potentially very powerful DNA sequencing technology."

Now, Korlach and Turner—who subsequently founded a company, Pacific Biosciences—have transformed their thought exercise into a working instrument. Their system makes use of nanostructures called zero-mode waveguides (ZMWs), tiny wells in which individual polymerase molecules have been immobilized in fixed orientation. These ZMWs establish small volumes in which each enzyme can be readily imaged with minimal background noise in the presence of appropriately high concentrations of fluorescently tagged nucleotides.

These latter also represent an important advance: conventionally labeled nucleotides generally impede polymerase processing or terminate chain synthesis, so Turner and Korlach's team developed nucleotides labeled at the terminal phosphate, which is released as bases are appended to the nascent DNA chain.



Individual polymerase molecules immobilized within ZMWs in an aluminum substrate generate detectable pulses as fluorescently labeled bases associate with the enzyme before their addition to the nascent DNA chain.

In the final system, each ZMW contains a polymerase molecule, synthesizing a complementary chain to the template molecule. As

PROTEIN BIOCHEMISTRY

A MEASURE OF FLOPPINESS

By analyzing the data generated by the Northeast Structural Genomics Consortium (NESG), researchers quantified the physical properties that control protein crystallization.

Since 2000 researchers at the NESG and three other large-scale structural genomics consortia sponsored by the Protein Structure Initiative of the National Institutes of Health have been developing rigorous methods for protein preparation and crystallization. These high-throughput projects systematically collected data and performed quality control analyses on huge numbers of proteins in the crystallization pipeline. Overall, only about 12% of the purified proteins yielded crystals good enough for structure determination, so crystallization is clearly the roadblock to crystal structure solution.

"Within a couple of years we had concrete data showing that crystallization is rare among even monodisperse, biochemically well-behaved proteins," says John Hunt of Columbia University, a participant in the NESG. "The overall results of the initiative sharpened the focus on this as an important technological problem, and the organization of the projects gave us the resources to look into it."

To understand the physical properties that control crystallization, Hunt and his colleagues analyzed the data generated by the NESG. Their major finding was that the mobility or entropy of surface residues is a major determinant of crystallization propensity: well-ordered epitopes on the protein surface presumably mediate interprotein packing interactions and thereby promote formation of good crystals for structure determination. Although this had been widely suspected by crystallographers, Hunt points out that this work "provides experimental measures of disorder at the surface and a quantitative estimate of how much this affects the probability of getting a crystal structure."

The group found that some glycines, alanines and phenylalanines also independently correlate with increased crystallization propensity, and hypothesized that these residues effectively mediate surface packing interactions. The presence of glycine in this list was particularly surprising because the backbone dihedral angles of this small residue are the least constrained, and it is often found in flexible, disordered loops. This naturally leads to the expectation that the presence of glycines should increase surface entropy and inhibit crystallization. And although this was the case in some locations, the favorable effect on crystallization was dominant in surface loops. "We are not sure exactly what they are doing, but they seem likely to be promoting packing interactions, perhaps by directly contacting the neighboring molecule or by creating some local conformational flexibility enabling adjacent residues to optimize their contacts," suggests Hunt.

Also surprising was the lack of correlation with thermodynamic stability or, in other words, the tendency of the protein to stay in

each labeled nucleotide enters the polymerase, it generates a transient fluorescent pulse that becomes visible in the ZMW, then vanishes once the new base is attached. The entire process is simultaneously monitored in every ZMW in real time with a specially designed multiplexed confocal imaging instrument.

Despite the many technical challenges that needed to be overcome, the instrument performs strongly, delivering accurate real-time sequences for both circular and linear templates at a rate of 2–4 bases per second. Although errors are a routine problem for individual reads, the multiplexed nature of the system makes it simple to achieve an accurate consensus, and the team is continuing to optimize the platform. "Whereas in the paper we required 15-fold coverage for a consensus with 99.3% accuracy, we've now gotten a consensus of 99.97% with 11-fold coverage," says Turner, "and this is still improving." Notably, the system can also achieve longer individual reads than existing sequencing-by-synthesis methods—up to 4 kilobases, in this demonstration.

The authors anticipate that their first-generation commercial instrument—slated for release in 2010—will dramatically streamline the sequencing process but believe that future versions will be able to tackle a host of other applications as well. "If in half an hour you can exhaustively sequence all the nucleic acids in a sample, you can emulate a microarray platform or a quantitative PCR or other things," says Turner. "We plan on reducing a whole series of problems in biology to essentially a software problem."

Michael Eisenstein

RESEARCH PAPERS
Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).

the folded state. "Many crystallographers would have assumed that there is some dependence on protein stability; one reflection of that is that people trying to crystallize something difficult will try a hyperthermophilic ortholog or homolog believing that's more likely to get them a crystal structure," explains Hunt. But the investigators noted only a small advantage for hyperthermophilic proteins, which might be attributable to the fact that potentially floppy surface residues participate in cooperative interactions more frequently than their counterparts in mesophilic proteins, making them less floppy.

Based on ongoing research inspired by their published findings, Hunt's group hopes to develop specific recommendations for engineering protein surfaces. For now, he suggests first evaluating sequences of all functionally related proteins to determine which have the highest crystallization propensity using the NESG's Pxs protein crystal structure propensity prediction server, which is based on data presented in their current analysis, as well as the Joint Center for Structural Genomics web tool, Protein Sequence Comparative Analysis, which provides complementary data. Next he suggests using the DISOPRED2 server to predict the ordered protein core for crystallization trials. For the still-elusive protein crystals, the next step is to vary the protein termini or to make mutations as suggested by the surface entropy reduction server.

Irene Kaganman

RESEARCH PAPERS
Price, W.N. II *et al.* Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nat. Biotechnol.* **27**, 51–57 (2009).

NEWS IN BRIEF

PROTEOMICS

MaxQuant software for proteomics

Cox and Mann introduce MaxQuant, a software package for managing large quantitative mass spectrometry datasets. MaxQuant efficiently detects peaks, isotope clusters and stable amino acid isotope (SILAC)-labeled peptide pairs. It yields very high mass accuracy information by integrating multiple mass measurements, facilitating more effective database searching and thus greater peptide identifications in complex proteomes. The software is available at http://www.maxquant.org/.

Cox, J. & Mann, M. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

CELL BIOLOGY

SUPER bilayers

Membrane fission is essential for both organelle and lipid vesicle formation but remains technically challenging to study. Pucadyil and Schmid now report supported bilayers with excess membrane reservoir (SUPER) templates, which are compatible with both real-time microscopy and biochemical assays. They use the SUPER templates to reevaluate the relationship between dynamin and GTP during membrane fission.

Pucadyil T.J & Schmid S.L. *Cell* **135**, 1263–1275 (2008).

GENE TRANSFER

Bypassing the blood-brain barrier

The ability to deliver therapeutic genes to the brain and spinal cord would likely have a great impact on medicine, but it has been very challenging to develop strategies to bypass the blood-brain barrier. Foust *et al.* discovered that the adeno-associated virus 9 vector effectively bypasses the blood-brain barrier, which they demonstrated by delivering and expressing *GFP* in the central nervous system in both neonatal and adult mice.

Foust, K.D. *et al. Nat. Biotechnol.* **27**, 59–65 (2009).

BIOINFORMATICS

Finding protein functional regions

Structural genomics efforts have yielded a large number of protein structures for which their functions are unknown. To begin to characterize these proteins, Nimrod *et al.* present PatchFinder, an algorithm that detects clusters of highly conserved residues on protein surfaces, which likely correspond to functional regions. They also created N-Func, a database containing 757 structures of previously unannotated proteins. PatchFinder and N-Func are available at http://patchfinder.tau.ac.il/.

Nimrod, G. *et al. Structure* **16**, 1755–1763 (2008).

MICROFLUIDICS

Paper and tape microfluidics

Cheap, simple and rugged devices are needed for diagnostic applications in the field and in developing countries. Martinez *et al.* describe the fabrication of three-dimensional microfluidic devices made from layers of paper and water-impermeable double-sided tape, patterned such that fluid is wicked within and between the layers of paper. Such devices cost only 3 cents to make and may find applications in environmental monitoring as well as healthcare.

Martinez, A.W. *et al. Proc. Natl. Acad. Sci. USA* **105**, 19606–19611 (2008).