

Albert Crewe's Dream Realized: Sequencing DNA with STEM

David C. Bell¹, W. Kelley Thomas², Katelyn Murtagh³, and William R. Glover³

1. School of Engineering and Applied Sciences, and Center for Nanoscale Systems,
Harvard University, Cambridge, Massachusetts 02138

2. Hubbard Center for Genome Studies, University of New Hampshire, Durham NH 03824

3. ZS Genetics, North Reading MA, 01864

A non-aberration-corrected Scanning Transmission Electron Microscope (STEM) was the method of choice for Crewe and co-workers to originally image single heavy atoms in 1970 [1]. In a related *Science News* article he was quoted as saying that “in one or two years he and his colleagues will determine the sequence of component bases” [2], it took a little longer than he thought but now after 40 years we can finally achieve this goal. The issues have been both the instrument, and the sample. The greatest difficulty has been in attaching heavy metal atoms to the bases, allowing a stable sample for STEM observation, and preparing the sample in a new way that allows the bases to be “read” by stretching out the double helix that then offers the ability to take direct images of genetic material identifying its labeling with one or a few key unique identifying labels (Fig. 1). Recently there have been evolutions and revolutions in DNA sequencing technologies. These changes have dramatically increased the speed and reduced the costs of sequencing, at a pace even faster than Moore's law operates in semiconductors. The drawbacks to these technologies are that they can only read a few hundred bases on a molecule. The result is massive parallelism, and many analytic challenges simply cannot be resolved by the new short-molecule technologies.

The advent of higher resolution, aberration-corrected, and more stable instruments have allowed single atom annular dark field (ADF) imaging to become straightforward. This imaging platform has the potential to be used in a number of different applications in genetic analysis. The idea is simple: to sequence DNA by taking digital images of the labeled DNA bases with a high-resolution electron microscope with the eventual aim being a fully automated imaging system. In this way a sequence thousands of base-pairs long can be read directly from the image with developed pattern recognition software. We have been able to consistently reproduce, prepare and image hundreds of linearized and labelled DNA molecules following our labelling and linearization process (Fig. 2). A critical distinguishing factor in identifying these molecules is their general morphology. Specifically, at low TEM magnification (12,000 to 80,000x) the labelled DNA molecules are 2 nm in width and several microns in length. Both the length and linearity correspond to the length of the DNA molecules used in each experiment. Atomic-level imaging was conducted by ADF STEM in aberration-corrected microscope, Carl Zeiss Libra 200-80kV also using the in-column energy-filter retaining only zero energy-loss electrons. Figure 2 shows the corresponding heavy atom-labelled and linearized DNA molecule imaged with ADF STEM. A production analysis system could be expected to read a complete haploid of human genome in days, perhaps hours, with 10X coverage, at a cost of a few thousand dollars. The path to the \$1,000 or less genome will then be in sight.

References

1. Crewe A.V., Wall, J. & Langmore, J. Visibility of single atoms. *Science* 168, 1338-1340 (1970).
2. “Individual Atoms Photographed”, *Science News*, Vol 97 p 524 (1970).

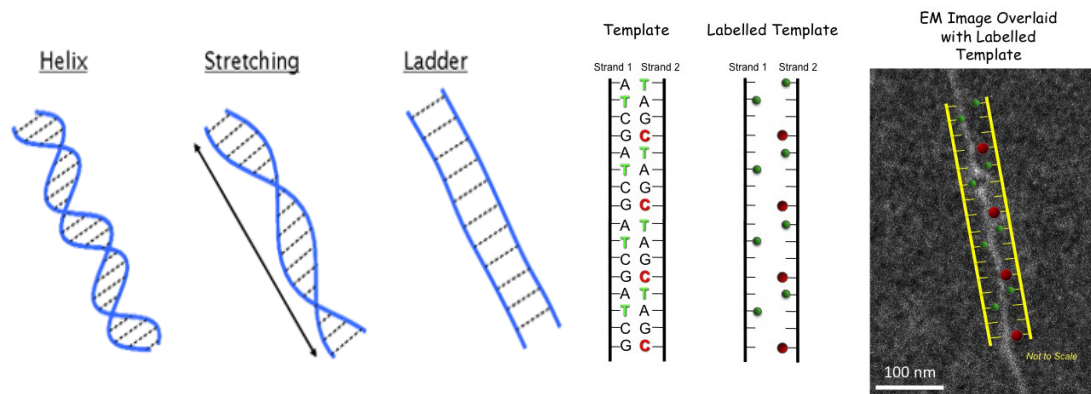


Figure 1. The fundamental sample preparation problem, stretching the DNA helix to lay flat, and allow comprehensive imaging; suitable labeling that is applied can then be used as a mechanism to “read” sequences. The labeled DNA template now able to be read directly using high-resolution electron microscopy, in this case an ADF STEM image is shown.

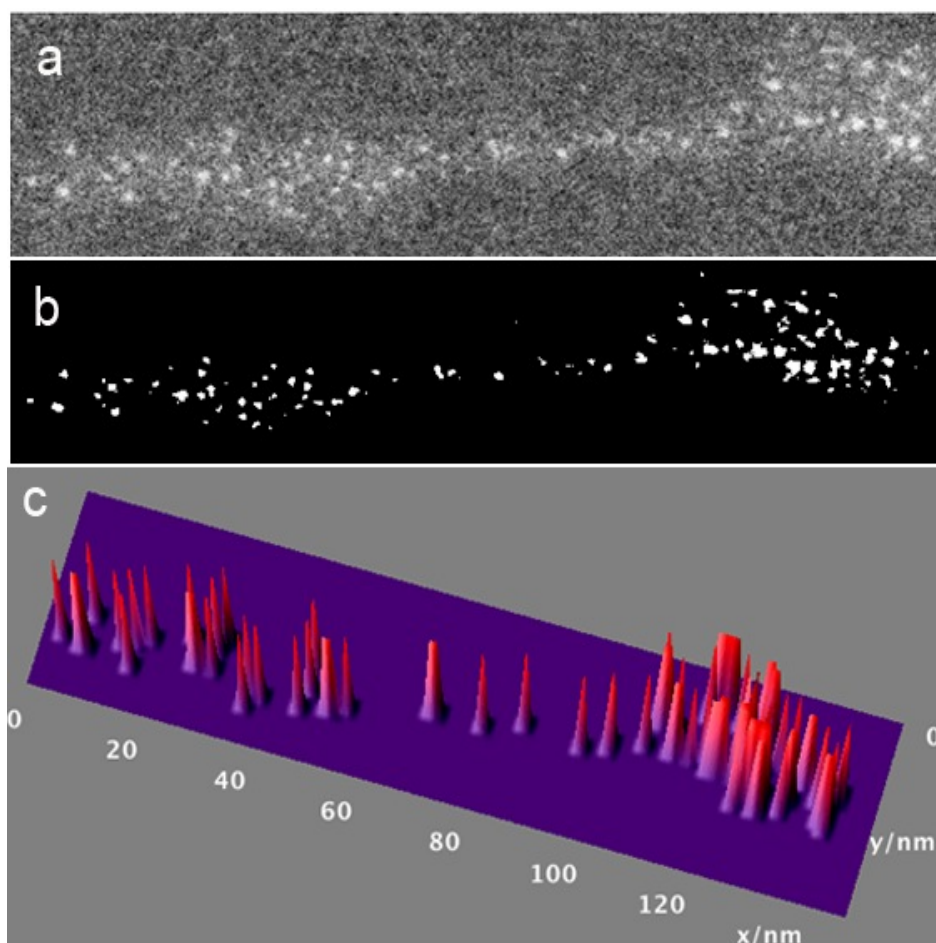


Figure 2. Identification of individual single-atom DNA labels. a-c, methodology to obtain individual label spacing from the substrate noise: a) Area selection of STEM ADF image of stretched DNA molecule; b) ROI following thresholding of intensities; c) Pseudo 3D contour plot of image intensities indicating peaks at heavy atom, single-atom and clusters in locations of DNA labels.