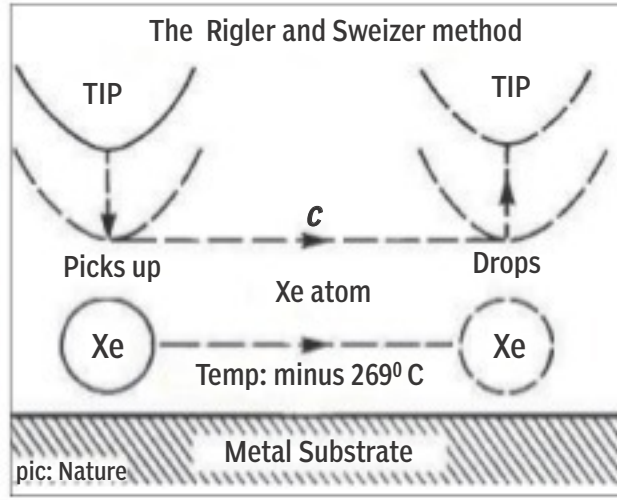
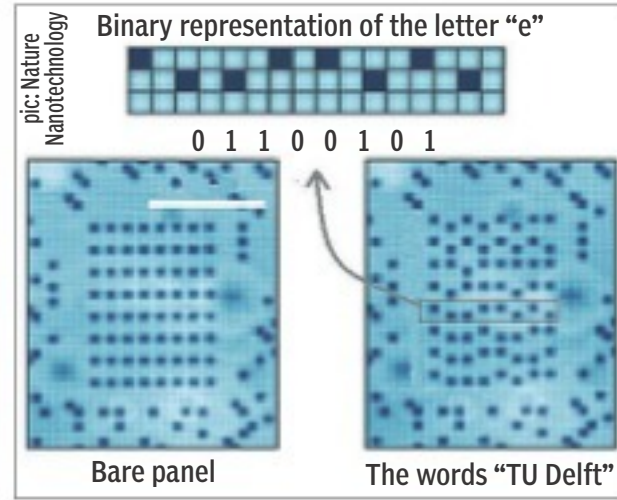


# Atom-level data storage

COMPUTER STORAGE MAY NOT GET MORE MINIATURISED THAN WHAT HAS, IN PRINCIPLE, BEEN REACHED, WRITES ANANTHANARAYANAN

The space taken for electronic data storage has shrunk as fast as computing capacity has increased. While computing speed is now at billions of a second, the size of the components in microcircuits has dropped from microns to as low as 45 nanometres, or thousandths of a micron. The first hard disks were marvels that stored 260 MB and now they store in TB, or approaching a million times more than what they used to. Keeping up the pace of miniaturisation, however, has become challenging as the dimensions of the storage components are fast approaching the physical limits to how small things can get and still function.



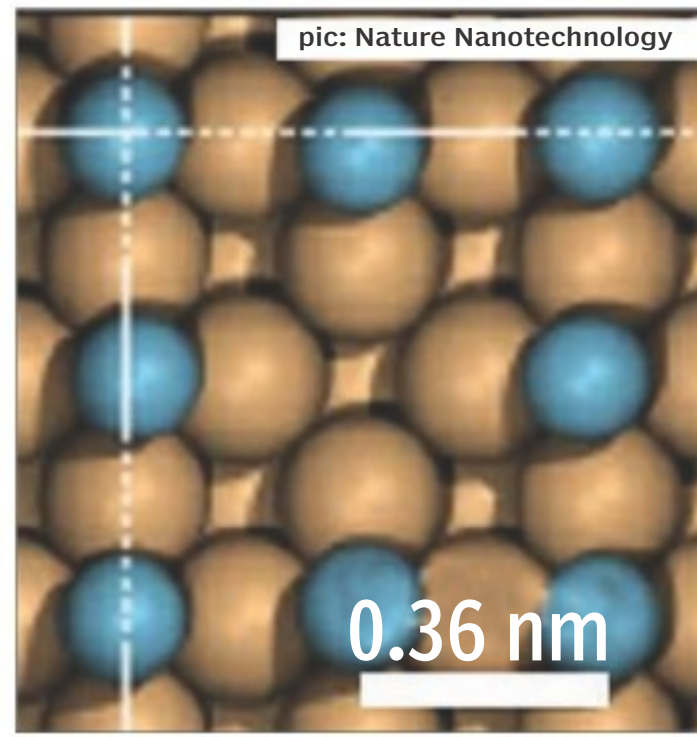
But an achievement in miniaturisation, which has just been reported, may well be the end of the line. FE Kalff, MP Rebergen, E Fahrnenfort, J Girovsky, R Toskovic, JL Lado, J Fernández-Rosier and AF Otte, from the Kavli Institute of Nanoscience, Delft, the Netherlands, the International Iberian Nanotechnology Laboratory in Portugal and the Universidad de Alicante in Spain report in the journal, *Nature Nanotechnology*, their success in storing a reasonable

quantity of data in a format reduced to the size of individual atoms! The best data storage used so far works with the help of magnetic regions that are just below a micrometre in size and along the length of a film of magnetic material. Each of these regions can be magnetised in one of two directions, and the sequence of regions can represent the digits "0" or "1", so that data can be stored in binary form. Each of these magnetised regions, in turn, consists of some hundreds of crystal grains and each grain, of course, consists of millions of atoms. A good part of the progress and improvements in storage technology has been to use smaller regions for each "bit", or the option of "0" or "1" of information in such a way that one region does not disturb the other and also that it is practical for each region to be "written" and "read".

A great advance, through the manipulation of individual atoms, either for fabrication of components or for recording data, became a possibility with the development of the *Scanning Tunneling Microscope*, which is used to detect matter at the level of atoms. The actual microscope consists of a metallic tip, a needle that has been sharpened so that its tip is just a single atom. Now the free electron cloud associated with atoms of a metal, which the microscope is to study, extends just beyond the metal

surface. What this implies is that if the tip of the STM is brought really close to the metal surface, electrical forces start acting between the tip of the needle and the surface. And while just in this position of extreme proximity, if a small electric potential is applied to the metallic tip, an electron could move from the tip to the metal surface, "tunneling" through the repulsive electron cloud, and an electric current would flow. As the current rises sharply if the tip comes even an atomic diameter closer, the current becomes a measure of the evenness of the surface as the tip scans the surface.

It is the same technology that the authors of the paper have used to manipulate atoms to record data. With only this difference, that in place of measuring a spurt in the current, they



Lattice of chlorine atoms on the surface of crystalline copper

supplied the tip of the STM with a small voltage when they wished to nudge a chlorine atom in a honeycomb of copper atoms from one slot to the next.

The first time the STM was used to move atoms was in 1990 when Don M Eigler and EK Schweizer stationed xenon atoms on a nickel surface in a pattern that spelt the letters "IBM". Ever since then, the *Nature Nanotechnology* authors say, atomic arrangements have been assembled in this way by lateral positioning of

atoms and molecules that have been deposited on a flat metal surface. This method, they say, is good with up to some hundreds of components, but because there is no framework on the metal surface, creating complex and close-set patterns is not feasible. And, further, the procedure needs to be carried out at exceedingly low temperatures. Instead, the new procedure they have developed allows accurate placing of chlorine atoms in a regular and extended array and is possible at a more practical temperature.

The basis of the method is a raft of chlorine atoms that assemble themselves into a regular lattice on the surface structure of crystalline copper. This arrangement of chlorine atoms, according to the spacing of the copper atoms, forms a large grid on which specific shapes can be drawn. There could also be vacant spaces, or vacancies between the chlorine atoms. And then the lattice remains stable at comparatively higher temperatures, that is, as high as minus 196° Celsius. This temperature can be maintained by using liquid nitrogen, which is easier than lower temperatures that need liquid helium. But the best feature is that the position of the chlorine atoms can be readily changed with a nudge using the Scanning Tunneling Microscope.

These features, the authors say, allow for a very large pattern of occupied and vacant spaces in the grid to be built up with great reliability and speed. Representing data according to the pattern of the grid would, hence, allow data to be written onto the copper surface and also to be deleted or edited as desired.

Comparably large atomic-scale data storage has been demonstrated using individual silicon atoms placed on a silicon-gold surface, the authors say. But this arrangement, unlike the present one, cannot be modified at will and needs to be completely regenerated to change even one bit, they say.

The method now developed is to fix a point on the copper base as a "marker" and distribute the chlorine atoms, or vacancies in different blocks. The positions were ranged in rows of eight pairs, each pair consisting of either a chlorine atom or a vacancy with separators all around each pair. As each pair could be either "atom-vacancy" or "vacancy-atom", each could represent either "1" or "0" and the row of eight pairs would represent the eight bits that computers use to store an item of data.

The researchers have succeeded in creating a panel of such rows to represent 1,016 characters, that is to say, the letters, punctuation and spaces of a whole text passage within a 96x126 nanometre chip. The passage, which can be readily read and rewritten with an automatic procedure, is a whole KB, or 8,000 bits, and the storage density works out to 502 terabits per square inch, which is hundreds of times more dense than the best hard disks.

The technology is still far from ready for the industry as it has been tested only in the laboratory and does need a temperature of nearly minus 200° Celsius. A new paradigm of data storage has, however, been proved in principle and shows the way to very high storage density, which would be useful for archiving massive data and also for high capacity miniature devices.

The technology is still far from ready for the industry as it has been tested only in the laboratory and does need a temperature of nearly minus 200° Celsius. A new paradigm of data storage has, however, been proved in principle and shows the way to very high storage density, which would be useful for archiving massive data and also for high capacity miniature devices.

The technology is still far from ready for the industry as it has been tested only in the laboratory and does need a temperature of nearly minus 200° Celsius. A new paradigm of data storage has, however, been proved in principle and shows the way to very high storage density, which would be useful for archiving massive data and also for high capacity miniature devices.

The technology is still far from ready for the industry as it has been tested only in the laboratory and does need a temperature of nearly minus 200° Celsius. A new paradigm of data storage has, however, been proved in principle and shows the way to very high storage density, which would be useful for archiving massive data and also for high capacity miniature devices.

The technology is still far from ready for the industry as it has been tested only in the laboratory and does need a temperature of nearly minus 200° Celsius. A new paradigm of data storage has, however, been proved in principle and shows the way to very high storage density, which would be useful for archiving massive data and also for high capacity miniature devices.

The technology is still far from ready for the industry as it has been tested only in the laboratory and does need a temperature of nearly minus 200° Celsius. A new paradigm of data storage has, however, been proved in principle and shows the way to very high storage density, which would be useful for archiving massive data and also for high capacity miniature devices.

## PLUS POINTS

### First photograph

A subject of cautionary tales rather than scientific inquiry, the giant squid was considered part of mythology in the late 19<sup>th</sup> century. In 1874, a group of fishermen in Logy Bay near St John's port in Newfoundland, USA, brought naturalist Moses Harvey a whole giant squid that had died thrashing in their nets.



The first photograph of most of a giant squid on display in the Smithsonian National Museum of Natural History.

"Those fishermen had obviously heard that Harvey had paid \$10 for a tentacle the

previous year and thought, 'Well, goodness, what would he pay for the entire thing?'" says Matthew Gavin Frank, author of a book on Harvey called *Preparing the Ghost*. "The answer was also \$10."

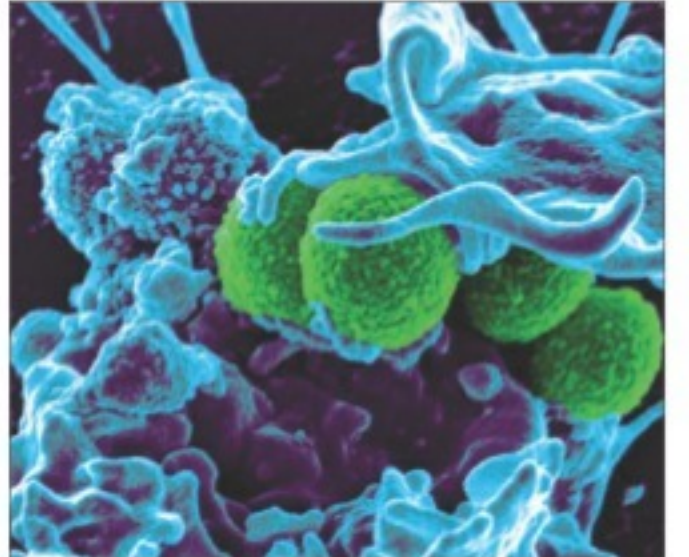
Harvey took the specimen back to his house and draped its 27-foot-long body over a curtain rod above the bathtub. He arranged its tentacles, set up a photograph of the cephalopod hanging from the rail and then relegated the animal to a vat of brine in the backyard.

In the end, Harvey sent his squid to Yale University zoologist Addison Emery Verrill, who was horrified by the specimen's condition. Frank says, "The carcass had shrunk, desiccated a bit, and some of the suckers had fallen off."

THE SCIENTIST

### Nasal antibiotic

A new type of antibiotic that is capable of fighting off some superbugs has been discovered in the human nose. The new drug, named lugdunin,



originates from bacteria present in human nostrils. Importantly, it gave no indications of allowing the bacteria it attacks to mutate and develop resistance. A human form of the drug has not yet been developed — it has only been tested on mice — but if successfully formulated, it would be a hugely significant scientific breakthrough.

New antibiotics have not been developed since the 1980s and antimicrobial resistance to drugs is a growing concern. Most existing antibiotics are sourced from the soil, where there is a constant war between different microorganisms. But lugdunin originated from a type of bacteria called *Staphylococcus lugdunensis*, which is found in about nine per cent of human noses.

Scientists from the University of Tübingen in Germany made the discovery while examining the *Staphylococcus aureus* bacteria, which can cause drug-resistant infections like Methicillin-resistant *Staphylococcus Aureus*.

WILL WORLEY/THE INDEPENDENT

### Simian voice control

An orangutan named Rocky impressed scientists at Durham University, UK, by mimicking more than 500 vowel-like sounds in human speech, the researchers reported in *Scientific Reports*. It's the first time a non-human primate has demonstrated this level of vocal fold control, hinting out how humans evolved speech after they split off from great apes.

"The long-held notion that great apes lack this ability can now be thrown



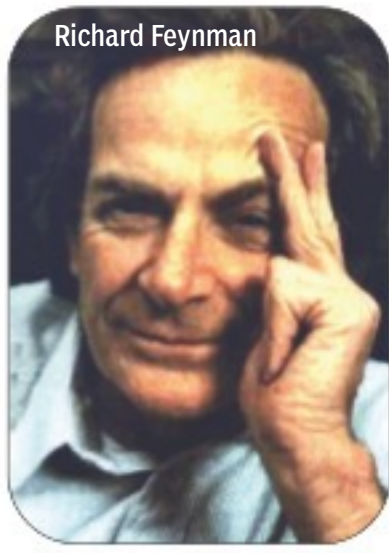
into the trash can," study co-author Adriano Lameira of Durham University said. Lameira and colleagues trained Rocky to produce vowel sounds that mimicked the pitch and tone of human speech. The researchers compared Rocky's vocalisations with a database of thousands of hours' worth of calls of wild and captive orangutans, finding that Rocky was able to control his voice in novel ways and learn new sounds.

The findings could help explain how humans evolved the sophisticated control of their voices that allowed them to develop speech, the team added.

## Plenty of room at the bottom

EIGLER and Schweizer chose to write the letters "IBM" with atoms because their work was in the IBM laboratory. The 1,016-character passage that was written into the atomic storage by Otte, et al, has been taken from the text of Richard Feynman's seminal 1959 lecture, "There's plenty of room at the bottom", where he first suggested the world should be discovered at small dimensions, or the science of nanotechnology.

"Why cannot we write the entire 24 volumes of the Encyclopaedia Britannica on the head of a pin?" Feynman asked in 1960. "Let's see what would be involved. The head of a pin is a 16th of an inch across. If you magnify it by 25,000 diameters, the area of the head of the pin is then equal to the area of all the pages of the Encyclopaedia Britannica. Therefore, all that is necessary is to reduce in size all the writing in the Encyclopaedia by 25,000 times. Is that possible? The resolving power



Richard Feynman

of the eye is about 1/120 of an inch — that is, roughly the diameter of one of the little dots on the fine half-tone reproductions in the Encyclopaedia. This, when you demagnify it by 25,000 times, is still 80 angstroms\* in diameter — 32 atoms across, in an ordinary metal. In other words, one of those dots still would contain in its area 1,000 atoms.

"So each dot can easily be adjusted in size as required by the photoengraving, and there is no question that there is enough room on the head of a pin to put all of the Encyclopaedia Britannica." ■ 80 Angstroms = eight nanometres

# GENETIC EVIDENCE OF THE DNA CODE

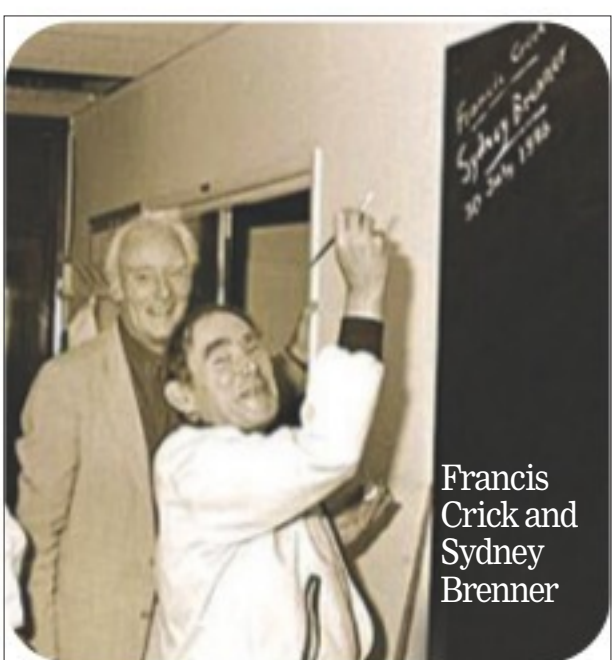
FRANCIS CRICK AND SYDNEY BRENNER'S ABILITY TO DEDUCE THE CORRECT EXPLANATION FROM THEIR ANALYSIS OF PROFLAVIN-INDUCED MUTATIONS IS AN ESPECIALLY INSPIRING EXAMPLE OF THE CAREFUL, OFTEN INGENIOUS REASONING THAT ALMOST ALWAYS ACCOMPANIES SIGNIFICANT ADVANCES IN SCIENCE, SAYS TAPAN KUMAR MITRA

Given a sequence relationship between DNA and proteins, you might want to know how many nucleotides in DNA are needed to specify each amino acid in a protein. We know that the information in DNA must reside in the sequence of the four nucleotides that constitute it: A, T, G and C. These are the only "letters" of the DNA alphabet. Because DNA language has to contain at least 20 "words", one for each of the 20 amino acids found in protein molecules, the word coding for each amino acid must consist of more than one nucleotide. A doublet code involving two adjacent nucleotides would not be adequate, as four kinds of nucleotides taken two at a time can generate only 4<sup>2</sup>=16 different combinations.

But with three nucleotides per word, the number of different words that can be produced with an alphabet of just four letters is 4<sup>3</sup>=64. This number is more than sufficient to code for 20 different amino acids. In the early 1950s, such mathematical arguments led biologists to suspect the existence of a *triplet code* — that is, a code in which three base pairs in double-stranded DNA are required to specify each amino acid in a polypeptide. But direct evidence for the triplet nature of the code was not provided until 10 years later. To understand the nature of that evidence, we first need to become acquainted with frameshift mutations.

In 1961, Francis Crick, Sydney Brenner and their colleagues provided genetic evidence for the triplet nature of the code by studying the mutagenic effects of the chemical *proflavin* on bacteriophage T4. Their work is well worth considering, not just because of the critical evidence it provided concerning the nature of the code but also because of the ingenuity of the deductive reasoning that was needed to understand the significance of their observations.

Proflavin is one of several *acridine dyes* commonly used as mutagens (mutation-inducing agents) in genetic research. Acridines are interesting mutagens because they act by causing the addition or deletion of single base pairs in DNA. Sometimes, mutants generated by acridine treatment of a wild-type ("normal") virus or organism appear to revert to the wild type when treated with more of the same type of mutagen. Closer examination often reveals, however, that the reversion is not a true reversal of the original mutation but the acquisition of a second mutation that maps very close to the first. In fact, the two mutations are in the same gene, and if they are separated by genetic recombination each still gives rise to a mutant phenotype.



Francis Crick and Sydney Brenner

These mutations display an interesting kind of arithmetic. If the first alteration is called a plus (+) mutation, then the second can be called a minus (-) mutation. By itself, each creates a mutant phenotype, but when they occur close together as a double mutation they cancel each other out and the virus or organism exhibits the normal, wild-type phenotype. (Properly speaking, the phenotype is said to be *pseudo wild-type* because, in spite of its wild-type appearance, two mutations are present.) Such behaviour can be explained using the analogy.

Suppose that line 1 (see illustration) represents a wild-type "gene" written in a language that uses three-letter words. When we "translate" the line by starting at the beginning and reading three letters at a time, the message of the gene is readily comprehensible. A plus mutation is the addition of a single letter within the message (line 2). That change may seem minor, but since the message is always read three letters at a time, the insertion of an extra letter early in the sequence means that all the remaining letters are read out of phase.

There is, in other words, a shift in the *reading frame*, and the result is a garbled message from the point of the insertion onward.

A minus mutation can be explained in a similar way because the deletion of a single letter also causes the reading frame to shift, resulting in another garbled message (line 3). Such *frameshift mutations* are typical effects of acridine dyes and other mutagens that cause the insertion or deletion of individual base pairs.

other's effect, particularly when they are located in close proximity. The insertion caused by the plus mutation compensates for the deletion caused by the minus mutation, and the message is intelligible from that point on (line 4). Notice, however, that double mutations with either two additions (+/+; line 5) or two deletions (-/-; line 6) do not cancel in this way. They remain out of phase for the remainder of the message.

When Crick and Brenner generated T4 phage mutants with proflavin, they obtained results similar to those in the hypothetical example involving a language that uses three-letter words. They found that minus mutants, which exhibited an abnormal phenotype, could acquire a second mutation that caused them to revert to the wild-type (or, more properly, pseudo wild-type) phenotype. The second mutation was always a plus mutation located at a site different from, but close to, the original minus mutation. In other words, phage reverting back to the wild-type phenotype exhibited a +/- pattern of mutations.

Crick and Brenner observed many examples of +/- (or +/+) mutants exhibiting the wild-type phenotype in their experiments. But when they generated +/+ or -/- double mutants by recombination, no wild-type phenotypes were ever seen.

The effect of frameshift mutations can be illustrated with an English sentence. The wild-type sentence (line 1) consists of three-letter words. When read in the correct frame, it is fully comprehensible. The insertion (line 2) or deletion (line 3) of a single letter shifts the reading frame and garbles the message from that point onward. (Garbled words due to shifts in the reading frame are underscored.) Double mutants containing a deletion that "cancels" a prior insertion have a restored reading frame from the point of the second mutation onward (line 4). However, double insertions (line 5) or double deletions (line 6) produce garbled messages. Triple insertions (line 7) or deletions (line 8) garble part of the message but restore the reading frame with the net addition or deletion of a single word.

Crick and Brenner also constructed triple mutants of the same types (+/+ or -/-) and found that many of these did revert to wild-type phenotypes. The reading frame (based on three-let-

Wild type	1	OURBIGREDDOGBITTHEOLDMAN	OUR BIG RED DOG BIT THE OLD MAN
Single mutants	+	OURBIBIGREDDOGBITTHEOLDMAN	OUR XBI GRE DDO GBI TTH EQL DMA N
	-	OURBIBREDDOGBITTHEOLDMAN	OUR BIR EDD OGB ITT HEO LDM AN
Double mutants	+/-	OURBIBIGREDDOGBITTHEOLDMAN	OUR XBI RED DOG BIT THE OLD MAN
	+/+	OURBIBIGBIREDDOGBITTHEOLDMAN	OUR XBI GYR EDD OGB ITT HEO LDM AN
Triple mutants	-/-	OURBIBIRDDOGBITTHEOLDMAN	OUR BIR DDO GBI TTH EQL DMA N
	+/-	OURBIBIGBIREDDOGBITTHEOLDMAN	OUR XBI GYR EDZ DOG BIT THE OLD MAN
	-/-	OURBIBIRDDOGBITTHEOLDMAN	OUR BIR DOG BIT THE OLD MAN

ter words) at the beginning and end of that hypothetical message remains the same when three letters are either added or removed. The portion of the message between the first and third mutations is garbled, but provided these are sufficiently close to each other, enough of the sentence may remain to convey an intelligible message. Applying this concept of a three-letter code to DNA, they concluded that adding or deleting a single base

pair would shift the reading frame of the gene from that point onward, and a second, similar change shifted the reading frame yet again. Therefore, from the site of the first mutation onward, the message is garbled. But after a third change of the same type, the original reading frame is restored and the only segment of the gene translated incorrectly is the segment between the first and third mutations. Such errors can often be tolerated when the genetic message is translated into the amino acid sequence of a protein, provided the "affected" region is short and the changes in amino acid sequence do not destroy protein function. This is why the individual mutations in a triple mutant with wild-type phenotype map so closely together. Subsequent sequencing of "wild-type" polypeptides from such triple mutants confirmed the slightly altered sequences of amino acids that one would predict.

Based on their finding that wild-type phenotypes are often maintained in the presence of three base-pair additions (or deletions) but not in the presence of one or two, Crick and Brenner concluded that the nucleotides making up a DNA strand were read in groups of three. In other words, the genetic code is a triplet code in which the reading of a message begins at a specific starting place (to ensure the proper reading frame) and then proceeds three nucleotides at a time, with each such triplet translated into the appropriate amino acid, until the end of the message is reached. Their ability to deduce the correct explanation from their analysis of proflavin-induced mutations is an especially inspiring example of the careful, often ingenious reasoning that almost always accompanies significant advances in science.

THE WRITER IS ASSOCIATE PROFESSOR, HEAD, DEPARTMENT OF BOTANY, ANANDA MOHAN COLLEGE, KOLKATA, AND ALSO FELLOW, BOTANICAL SOCIETY OF BENGAL, AND CAN BE CONTACTED AT tapanmitra59@yahoo.co.in

