The Challenges of Sciences. A Tribute to Carlos Chagas
Pontifical Academy of Sciences, *Scripta Varia* 103, Vatican City 2002
www.pas.va/content/dam/accademia/pdf/sv103/sv103-perutz.pdf

THE CHALLENGES OF SCIENCE

MAX PERUTZ

I thought I would just tell you about a great adventure which Carlos Chagas and myself shared. I got to know him a little better. I do not remember which year the Pope made an appeal against nuclear weapons to the Heads of the nuclear States, stimulated I think mainly by Ricky Vasco who had persuaded Chagas to propose this to the Pope, and the Pope enthusiastically endorsed it. So the Pope wrote letters to Reagan, Brezhnev and the Queen of England, and the idea was that delegations from the Academy would actually submit his appeal personally to the Heads of State.

To my surprise, Chagas asked me whether I would accompany him to deliver this to the Queen. Now, I felt a bit of a sort of Charlie Chaplin about this, a sort of immigrant to Britain of Jewish extraction acting as a messenger from the Pope to the Queen of England; it looked a bit funny. But anyway the next thing that happened was that the Foreign Office decided that it was a political message and that it should be delivered to the Prime Minister and not to the Queen. So Chagas and Hermann Bruck, another member of the Academy, and I, made an appointment with Margaret Thatcher at number 10 Downing Street. She received us in her drawing room dressed to the nines and with not a single hair out of place, looking very smart indeed, and Chagas delivered this message, a passionate message against nuclear weapons – it was really marvellous to see him in action – to which she replied with an impromptu well-reasoned statement that she would not accept any such move unless there already existed a reliable system to verify that nuclear disarmament really had taken place. However, she allowed us to deliver this message for the Queen and so Chagas gave the message to Margaret Thatcher, she read it and she said: "I shall decide, I shall tell the Queen what to reply". With that we were accompanied out and of course were really amused by this outcome, although a little saddened

because we realised that we could not get any further. But as I said I admired Chagas' determination and his persuasiveness on this occasion. I think he would have melted anybody else's stony heart but he did not melt Margaret Thatcher's.

Now, to come to my story, in 1949 Linus Pauling and others published a sensational paper in *Science* entitled 'Sickle-cell Anaemia, a Molecular Disease'. Sickle-cell anaemia is a blood disease which affects black people and others living in malarial regions. It manifests itself by a reversible change. As we can see from the diagram, on the left is shown the normal shape of a red blood cell, a sort of saucer shape. On loss of oxygen the shape of the red cell changes to this sickle shape, which is why it is called sickle-cell anaemia. The venous circulation contains the sickle cells, the arterial contains the normal ones, and what Pauling showed in his paper was that the normal and sickle-cell haemoglobin differ by the electric charge which the haemoglobin molecule carries. The sickle-cell molecule carried too few electric charges. He thus examined the blood of people who had heterozygotes, who had inherited the disease from only one parent, and it contained two haemoglobins, one with the normal charge and one with the abnormal charge.

I managed to obtain some sickle-cell blood after that and Murdoch Mitchinson and I looked at it with a high-powered polarised microscope and found that this transition is accompanied by a crystallisation of the haemoglobin in the red cell. The haemoglobin is a solution; it crystallises; and it has the typical refringence of crystals of human blood, the oxyhaemoglobin. This observation showed me for the first time that protein aggregates are not tolerated by living cells and this was very important.

Shortly after that there was a young chemist in northern England, John Irovolsci, and Francis Crick and I needed to him to try and look into this chemically. He developed a very good new method of chromatography that he called fingerprinting, which involved spreading the haemoglobin. He digested the haemoglobin with an enzyme, with trypsin, and then spread out the peptides in one direction by their electric charge and in the other direction by their mobility, and found that the difference appears in those from peptides. The year after that he showed that the change in charge that Pauling and his colleagues had observed was due to the displacement of a single pair of amino acids among the more than 500 other pairs: the normal haemoglobin carries glutamic acid, which has a carboxylic group, which carries a negative charge, and in sickle-cell heamoglobin this is replaced by another amino acid, valine, which differs from the carboxyl

group in that two oxygen molecules are replaced by three hydrogens. Thus it was discovered that chemical exchange in a protein can give rise to a deadly disease. This discovery was also crucial because this was the first time that we actually realised what genetic mutation does.

We did not as yet know anything about the genetic code, but we realised that this had to be due to a change in the genetic code which replaces one amino acid in a protein by another. This posed in acute form the question of the genetic code, and really stimulated Crick and Brenner to work on the discovery that the triplet code was the three basis code for one amino acid. Since then I think literally thousands of other such amino acid replacements have been discovered as the causes of a variety of genetic diseases: most are caused by amino acid replacement, some are caused by extensions, and others by deletions, of amino acids.

However, in 1993 an entirely new and different form was found as a result of a tremendous cooperative effort, an unprecedented cooperative effort by about a hundred medical people and scientists to discover the gene for Huntington's disease, a terrible neurodegenerative disease, one of the worst diseases there is. It is dominantly inherited and leads first to uncontrolled movements, which is why it is also called Huntington's chorea, then to movement disturbances, then to progressive dementia, and finally to death. It is a late onset disease. It starts in middle age, and people are healthy and unsuspecting; they have children not knowing that they actually have the disease and will transmit it to an average half of their children. It is deadly and sinister. There was this tremendous international effort and in 1993 a paper appeared in Cell by 61 authors from six American and two British universities reporting the discovery of the gene. The gene was huge, it consisted of over 60 axons spread over thousands of kilobases and it coded for a protein of a single chain of almost 3.140 amino acid residues. But near one end, near the amino end, was an extraordinary feature, a repeat of a single amino acid, glutamine, as can be seen from the diagram.

The glutamine is a little different from glutamic acid, and as can be observed instead of there being two oxygens, one oxygen is replaced by nitrogen, and when this oxygen carries a negative charge the nitrogen carries a positive one. There is a dipole here: the oxygen carries a partial negative charge and the nitrogen a partial positive charge, and this is crucial for the mechanism of the disease. What they found was that in healthy people the length of this repeat varies anything from half a dozen to 37 glutamines, but in people with the disease the length is over forty, and that the longer the repeat, the earlier the disease sets in and the more

severe it is. There were some juvenile cases, but the repeat was over a hundred glutamines long.

I read this one night, coming home on the train and became very excited because a few weeks earlier my attention had been drawn to a repeat of glutamines in certain proteins of drosophila, mainly transcription factors, and I wondered what they could possibly do, what the meaning of this was. What does a crystallographer do? He builds a model. So I built an atomic model of two glutamine chains and I found that they act as polar zippers, that is to say, as can be seen from the diagram, there are two chains: one protein chain on one side and another chain on the other, and I found that there were protein chains known to stick together by hydrogen bonds between their main chain CO and H groups. That was well known but I found when I built this model that actually these dipoles at the end of the glutamine can also form hydrogen bonds, so that this chain is held together not only by hydrogen bonds in the main chain, but by pairs of hydrogen bonds all along on the side chain. In the diagram can be seen in blue the bonds that that herd together. And in fact I had a paper in press on polar zippers in proteins.

When I read this paper about Huntington's disease it seemed to me that this might be the clue to the molecular mechanism of the disease: that these longer repeats of glutamines might lead to aggregation of the protein, that it adheres at two appendices. How could I test this? I asked the chemist in our lab to make me a synthetic chain of polyglutamine which would have been in solubles and I asked him to put two astatines at one end of the chain and two lysines at the other end, and the optical properties of this showed that it does indeed form aggregates in solution with this sort of structure. When the astatines at the end were discharged at neutral PH it actually formed little fibres and the fibres actually grew and gave the typical X-ray refraction picture of this sort of structure, and so I published this and sent it to *Nature* who rejected it as of no interest without sending it to referees. I then published it in the proceedings of the National Academy.

In the accompanying diagram the atomic model can be seen rather better than it can be seen in real life: there are hydrogen bonds in the main chain and H and CO groups, and there are the hydrogens between bonds in the side chains. The next diagram contains my suggestion in that paper that the extension of the glutamine repeat may cause the affected proteins to agglomerate and precipitate in neurons. Symptoms may set in when the precipitates have reached a critical size or have resulted in a critical number of neural blocks. I published this in 1994 but there was at

that time no evidence whatever that such aggregates existed. The immunostamine showed the protein to be in sort of little dots, isolated molecules in the cytoplasm, with no aggregates. But in August 1997 a paper by Bates and Davies appeared in Cell which really put an entirely new complexion on the disease.

Gillian Bates for the first time succeeded in reproducing the disease in mice, in an animal, and she did it by a piece of genetic engineering. She just used the first axon of the very large gene which expressed the first hundred or so amino acids, which included the glutamine repeat, and she made transgenic mice and one set of mice with a repeat of 20 glutamines, and another set of mice with a repeat of 150 glutamines. Among these the ones with 20 glutamines remained healthy and the ones with 150 developed the typical symptoms of Huntington's disease. This was a great success and her mice have since been used the world over for experiments on Huntington's disease.

In the same paper there was a report by Steven Davies, a lecturer at University College, London. Gillian Bates and Davies worked at the university hospital in London and Davies cut up and then examined sections of the brains of these mice under a electromicroscope. One morning he stormed into my room in Cambridge in a very excited state. He started telling me about his experiments even before he had shut the door and showed me a series of electromicrographs demonstrating the aggregates that I had predicted three years earlier, that is to say, he found that in the cell nucleoid of the neurons in these mice there were aggregates. There were clumps of protein which staved with antibodies against pectide and also with antibodies against ubiquitin and other proteins.

They published this discovery and in the same issue of *Cell* there was a paper by Henry Shwanken and his group at the Max-Planck Institute for Molecular Genetics in Berlin, who used Gillian Bates' axon to introduce it into colibacteria. They introduced it with a series of different lengths of glutamine repeats and found that with repeats shorter than 40 the protein remained soluble and with repeats longer than 40 glutamines it produced aggregates. The important fact in this discovery was that the length of glutamine repeat that produced aggregates in vitro was exactly the same as the length that produced this disease in the patients.

After this, Marian Di Figlia's group at the Harvard Medical School looked again at the sections of the post-mortem brains of Huntington's patients and found the same aggregates there. They had been overlooked before but since then very many groups have been found in patients. The

largest structures of neurons blocked by these aggregates were found in juvenile patients, and there is a distinct relationship: the fraction of neurons carrying these aggregates depends on the length of the repeat – the shorter the repeat, the rarer they are, and vice versa.

Since that time seven other neural degenerative diseases have been discovered caused by the extension of glutamine repeats in totally different proteins of different molecular weights. In some the glutamine repeat is near the amino end, and in others near the carboxyl end, and in others in the middle, but it does not seem to make any difference, and the striking thing is that in six of these diseases the length of repeat that produces the disease is the same as in Huntington's disease. Thus the rule is: with fewer than 37 repeats people are healthy, if they have more than 40 glutamines in the repeat they are likely to experience this neural degeneration.

I wondered what this meant and, thinking about the question, it seemed to me that there would be entropic reasons why longer repeats form ordered hydrogen bondage sheets, while shorter repeats would form random ones. The following diagram shows this. If the roots are free they attract water. The group depicted in it attracts the water molecules which attach themselves because they also carry partial electric charges. The water molecules become immobilised at the cost of entropy within the system. When they come together and form hydrogen bondage with each other, these water molecules become liberated and the entropy of the system increases. At the same time, when the chain is a random coil it has a large entropy, but when it forms a hydrogen bondage structure of this kind it loses rotational and translation entropy.

The chain becomes longer and more and more hydrogen bonds are formed. The loss of rotational and translational entropy becomes smaller and smaller until it becomes negligible, but the gain in entropy through the liberation of water molecules in that group remains constant, so eventually the gain in entropy wins and the two come together.

This is expressed in the next diagram. There are a number of glutamine repeats against the entropy changes in the system and there is a gain of entropy; there are liberated water molecules which remain constant; a loss of rotational and translational entropy which decreases; and eventually the cross over in between. Due to this entropic effect, a length of more than 40 would form a hydrogen bondage structure which would need the aggregation. Certainly I have not been able to test this idea because anything you make with long repeats is insoluble in the first place, and you cannot obtain any spectral information in solution.

The situation now is that many people have tried to reproduce the disease in other animals and cells. Nancy Bonini has reproduced it in Philadelphia, has produced it in flies, in drosophilae, and other people have introduced the Huntington gene into all kinds of different cell lines and done experiments on it. As a result readers become absolutely confused: reports of aggregates without cell death, neural cell death, and of cell death without aggregates. Whereas workers in England are convinced that it is the aggregates that cause the disease, in the United States it is now widely believed that the aggregates are what they call an epiphenomenon and that the true cause of the disease is still unknown and has yet to be sought.

I was disturbed about this controversy and was wondering what to do about it when I came across two interesting observations. Last July a paper appeared in Nature reporting the probability of neural death in twelve neural degenerative diseases, and they found that in Huntington's disease the probability of neural death remained constant with time. In other words, it behaved like a random event that was due to singeing heat rather like radioactivity, so that each neuron in Huntington's disease has a constant half-life. But they offered no interpretation for this result.

Last September a paper appeared in a new journal, Nature Reviews, by Guzella and McDonald of the Harvard Medical School. They plotted the age of onset of the disease against the lacks of glutamine repeats in all eight diseases due to extension of glutamine, and found an extraordinary thing: in every case the age of onset, or rather its inverse, was an exponential function of the length of glutamine repeat.

In the next diagram are to be found Steven Davies' aggregates and in the following one are Marian DiFiglia's aggregates in the human neuron. These aggregates form in the cell nucleoid and they are composed not of the entire Huntington's disease molecule but of a fraction of it. So the first step in the disease seems to me to be a proteolytic severance of the molecule and then a fragment in the nucleus about one tenth the size of the whole molecule, and the nucleus forms this granular of aggregates.

The next diagram shows this remarkable fact. In it you have a number of CAG glutamines and proteins against the age of onset of the disease, and the larger the number of CAG glutamines the faster the age of onset: hence these remarkable logarithmic curves.

These two observations made me wonder. I said to myself: surely they are telling me something? And if you look at the cell as a collection of large molecules, what sort of event could give rise to this phenomenon, to this radioactive life decay, and to this exponential dependence on the moment

of onset of the disease? And it struck me that there was only one phenomenon, and that was nucleation.

The classical system of nucleation is clouds; water droplets in a cloud. Supersaturate cloud water molecules come together and form aggregates, but as long as the aggregates are small the surface tension is very high and the weight of loss of water molecules from the aggregates exceeds the weight of aggregation. But at some moments in time the aggregate reaches a critical size where the rate of loss is less than the rate of aggregation, and then it forms into a drop, and this is a totally random event which is also obtained in crystallisation, which is another example, and in solutions of large molecules.

It struck me that this was probably due to the same thing, but I found that I did not really know enough about the theory of nucleation, so I got together with a physicist, Alan Wildon, who is a Professor of Material Science in Cambridge and works on the aggregation of synthetic polymers and is an expert on the nucleation of such polymers into aggregates. He taught me a great deal and together we produced a paper on this subject in which we showed that nucleation is exactly such a random event and that in high polymers the rate of nucleation is indeed an exponential function of the length of the polymer, exactly as you see in the curves in the diagram. This paper has been submitted to Nature. I hope that this will end the controversy, but I am not sure that it will because in the case of sickle-cell haemoglobin it took twenty years for the paper by Mitchinson and myself to be accepted, and I fear it may take just as long again. But I will no longer be here.