

IO

3128
94

Comets

The Artificial Kidney

Biochemistry To-day

The Physics of Low Temperatures

X-Ray Studies of the Kidney Circulation

Ultra-Microscopic Surface Topography

How Old is the Earth?

Research Report

Dating the Past



Ampl. N46
N4701C
168362

One shilling and sixpence

SCIENCE NEWS

10



10-12-11
12



SCIENCE NEWS

EDITED BY J.L. CRAMMER

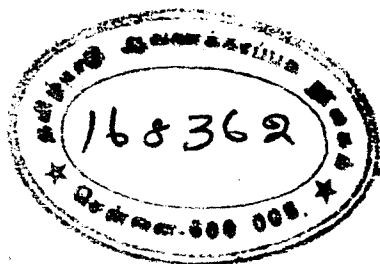
10

PENGUIN BOOKS

1949

First published January 1949

DL
Am 1, 246
249-10



Photogravure plates printed
by Eric Bemrose Ltd., Liverpool

Made and printed in Great Britain for
Penguin Books Ltd., Harmondsworth, Middlesex
by C. Nicholls & Company Limited
London Manchester Reading

CONTENTS

Editorial	7
Comets by Martin Davidson	9
Dating the Past	28
How old is the Earth? by J. J. Kipling	32
Biochemistry To-day by J. L. Crammer	41
Physics of Low Temperatures by K. Mendelssohn	71
The Artificial Kidney by John McMichael	92
X-ray studies of the Kidney Circulation by K. J. Franklin	101
Research Report by A. W. Haslett	124
<i>Six-thousand-mile waves – ocean sound – radar and meteorology – thunder cloud exploration – tracer research and industry – crystal amplifier – island telephony – rats and heredity</i>	
Ultra-microscopic Surface Topography by S. Tolansky	140
Glossary	156
Our Contributors	158
Index	160

Editorial

THE present century is increasingly an age of biological engineering. It began with the processing in large quantities of the sera and vaccines required for inoculating against tetanus, typhoid, cholera. It went on to the design of machinery for large-scale fermentations, the production of commercial citric acid by the growth of moulds, vinegar manufacture, and now the mass-production of penicillin and streptomycin. It also developed on another line in the mechanisation of blood transfusion. Machines were devised for separating the plasma from whole blood and for drying it ready for future use at any time. As off-shoots of this process come the globulin which protects against measles when injected; and fibrinogen, which clots over a raw wound such as a burn and seals it up, or can be used for joining cut nerves.

In these processes the engineer is called upon to make machines for the handling of products of living activity. But another big field is now thrown open to him. The film *Experiments in the Revival of Organisms* was a reminder of the work Russian engineers are doing in developing pumps to simulate or replace the heart's action. They are following in the steps of the American engineer, Lindbergh, who devised a complete and very complicated method of keeping single organs of the body alive with the help of artificial lungs and artificial heart. Now in Holland and in Sweden, work is going ahead on an artificial kidney. Professor McMichael, a British exponent of this latest invention, describes in this issue of *Science News* the medical value of this machine, crude though it yet is. There is no doubt that ultimately, when the engineers and the doctors have together got down to the job, great practical advantages will accrue for the treatment of the sick, and research will also be aided by this latest mechanisation.

Another worker with an important contribution to make to the processes of biological engineering is the biochemist. It is clear from Professor McMichael's article to what an extent the artificial kidney depends on biochemical knowledge. From time to time in the medical press there appear accounts of attempts to use the blood plasma of other animals for human blood transfusion, or even to substitute synthetic plastics, and here again the work of the biochemist is eventually going to clarify the problems, and in the end revolutionise blood transfusion by the introduction of substitutes of these kinds. The long article *Biochemistry To-day* is an attempt to compress a statement of the main ideas and methods of the science into the space of an essay. With such compression the choice of actual subject matter becomes in some degree a personal foible, and the condensation may at times demand thought for its understanding. This article is, however, an introduction only, a glance over the whole field. In subsequent issues it is planned that experts shall discuss their own particular branches of biochemistry in greater detail and clarity, and by giving many views, help the constant reader to a true picture of the whole.

Just now the kidney is a main focus of scientific interest. Professor Franklin's article deals with important new techniques of studying it, and the fascinating results of its working thereby obtained. This research has aroused fresh interest in the anatomy of its blood supply, so demonstrating that morphological anatomy is still a living subject, still a field for basic research under the stimulus of new questions.

Fruitful results from another newly applied technique are reported by Professor Tolansky, who by simple and elegant methods can photograph a hill only one molecule high on the surface of a crystal or a metal sheet.

Dr Davidson's report on *Comets* is interesting not least because it reveals the vast number of unsolved problems in this branch of astronomy. *How Old is the Earth?* describes the latest isotope studies of the problem; apparently iso-

topes are also now becoming important in Archæology, as an interesting correspondence testifies.

Finally, in the course of this quarter's *Research Report* A. W. Haslett describes a neo-Lamarckian experiment on rats which is an object-lesson to all concerned with animal experiments on heredity and evolution.

Comets

DR MARTIN DAVIDSON

COMETS are amongst the most lawless members of the solar system; while all the planets and also all the known minor planets or asteroids, as they are frequently called – about 2,000 in number – revolve round the sun in the same direction, comets are quite indifferent about observing the rule of the road. About half the known comets revolve round the sun in the same direction as the planets and the other half in the opposite direction, but it must not be assumed that the paths in which they move have any resemblance to planetary orbits. This will be made clearer by referring to Fig. 1.

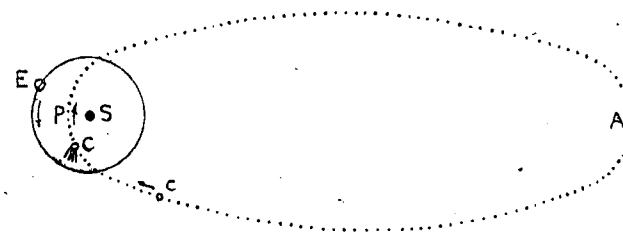


Fig. 1 – The paths followed by a comet and a planet are shown in the diagram, the broken line denoting the comet's path. The planet moves nearly in a circle, while the comet follows an elongated ellipse, coming closest to the sun at P and going off to its greatest distance at A. Near the sun a comet develops a tail which points away from the sun, as shown. At other parts of its orbit the tail disappears. The comet is moving in a direction opposite to that of the planet. Less than half the comets have this retrograde motion. (Reproduced from *An Easy Outline of Astronomy* (Davidson) by courtesy of Watts and Co.)

In this Figure we see the typical path of a planet denoted by E which may, if we wish, represent the earth, one of the

nine planets. The path is shown as a circle, though none of the planets revolves round the sun in an exact circle, but on the scale of the diagram it is quite impossible to distinguish any deviation from a circle. This applies to all the planets, though Pluto, the last to be discovered in 1930, is about 4,600 million miles from the sun when at its greatest distance, and about 2,700 million miles when at its least distance. The difference between the two distances may seem very great, but actually it is extremely small when we compare it with that between the greatest and least distances of a comet from the sun.

As the diagram shows, the comet *C* is moving in a path which is oval-shaped, and approaches the sun *S* at *P* closer than does the earth, but recedes from the sun at *A* to an enormous distance *SA* in comparison with *SP*. It will be seen from the arrows that the comet is moving in a direction opposite to that of the earth – a characteristic, as has just been pointed out, of about half the comets. The paths of the earth and the comet are represented as lying in the same plane – the plane of the ecliptic – but it is very seldom that a comet's path lies in this plane though there are several cases where the paths are close to the ecliptic plane. Unlike the planets, however, comets move at all inclinations to the ecliptic, from about 0° to 90° , the great majority of the orbits being at fairly steep inclinations compared with those of the planets. It may be pointed out at this stage that many minor planets also move at steep inclinations to the ecliptic, but none of them moves in a direction opposite to that of the planets. Fig. 2 shows a comet whose path is inclined at a small angle to the plane of the ecliptic, denoted by 'orbit of the earth'. Later on it will be necessary to refer again to this figure in connection with meteors.

The orbits of comets (we shall use the word 'orbit' in future to denote the nature of the curve which a comet describes in its motion round the sun) may be ellipses, parabolas or hyperbolas. No comet has ever been discovered the orbit of which is a circle, nor do the orbits of many

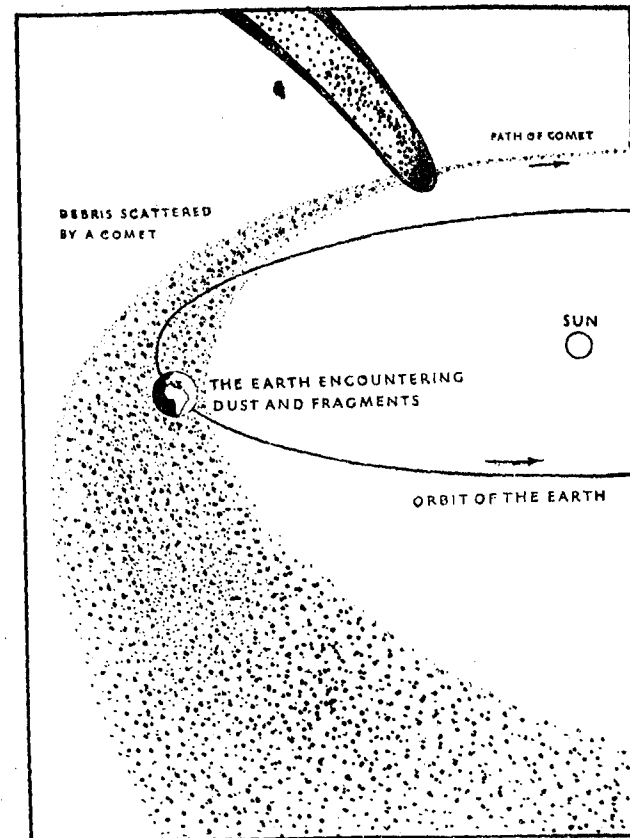


Fig. 2. – A comet discards a certain amount of its matter as it moves around the sun, and meteoric debris is spread out along its orbit, this debris continuing to revolve round the sun just as Saturn's rings revolve round the planet. The earth encounters this debris on certain occasions and a shower of meteors results from the encounter. The debris cannot be seen unless it encounters the earth's atmosphere, and then it appears only as flashes of light from the particles becoming incandescent. (Reproduced from *Splendour of the Heavens* by courtesy of Hutchinson & Co. (Publishers) Ltd.)

comets approximate to circles to the same extent as do those of the planets. An ellipse is a well-known figure and it is unnecessary to spend much time in describing it. A simple method for drawing an ellipse is shown in Fig. 3 and the caption explains this method. It will be seen that the ellipse

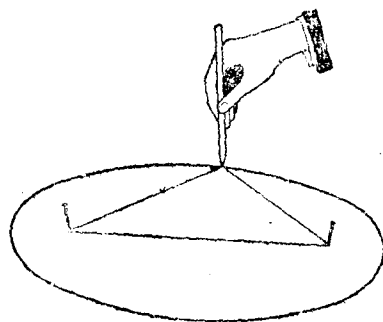


Fig. 3. — An ellipse can be drawn by inserting two pins in a piece of paper, passing a loop over them, and then moving a pencil round the paper, the point of the pencil keeping the string tight. The two points made by the pins are the foci of the ellipse. The sun is in one focus of all the planetary orbits and of cometary orbits also. The other focus of these orbits is known as the 'empty focus'. (Reproduced from *Elements of Mathematical Astronomy* (Davidson) by courtesy of Hutchinson & Co. (Publishers) Ltd.)

is a closed curve, so that comets whose orbits are ellipses return in the same path or nearly the same path for a long time: in the case of Halley's Comet this is more than 2,000 years, but very often comets become worn out with frequent returns. The reasons for this will be given later.

Fig. 4 shows an ellipse, a parabola and a hyperbola, and without dealing with the properties of these curves, which anyone can look up in an elementary work on conic sections, it may be said that neither of the last two is a closed curve, so if a comet should move in a parabola or a hyperbola, it will go off into the remote depths of space, the sun losing control of it. Notice that in the neighbourhood of the sun there is very little to distinguish between these three curves

and that it is only when the comet has receded some distance from the sun that the difference is apparent. At this point it is necessary to correct a popular error regarding the orbits of comets, and this error occasionally occurs in the works of those who are experts in other branches of science but have not studied this branch of astronomy.

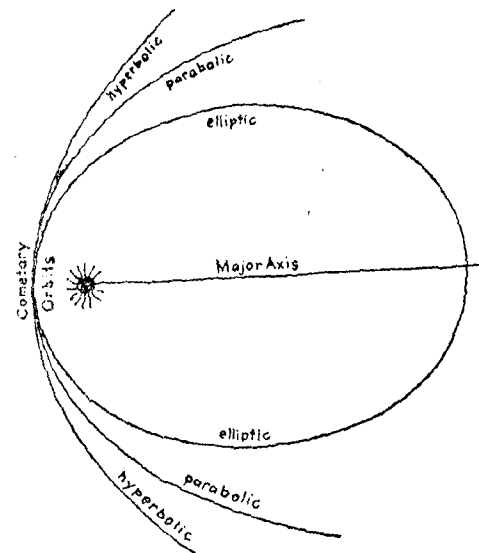


Fig. 4. — Several cases of all of the above forms of orbit are known, though the hyperbolic ones are almost indistinguishable from the parabolic. Near the sun the shapes of all three forms are similar, so observation over a fairly long arc is necessary to decide the character of an orbit: (Reproduced from *Splendour of the Heavens* by courtesy of Hutchinson & Co. (Publishers) Ltd.)

It is sometimes assumed that many comets have come into our solar system from outside its bounds, and hence that they are strangers in our system, merely paying a short visit and then going off again into the remote depths of space, from which they will never return. It is quite true that some comets do go off from the solar system, the sun losing gravitational control of them, in which case their

orbits are either parabolas or hyperbolas. The probability of an orbit which is absolutely parabolic is, however, so small that we need scarcely consider it, and for all practical purposes we can divide the orbits of comets into those that are elliptical, and those that are hyperbolic. But though some comets are lost to the solar system, this does not necessarily imply that they came into it from the remote depths of space. The reasons why those which were once members of the solar system become lost will now be explained.

Every body in the universe attracts every other body, the attraction being greater the greater the masses of the bodies under consideration, and decreasing with increasing distance between the bodies. For example, the planet Mercury is attracted by Venus and the earth, and by the other planets, but we will confine our attention to the case of Venus and the earth. Now Mercury comes much closer to Venus than it ever does to the earth, and on this account is attracted more by Venus. On the other hand, Venus is not quite so massive as the earth, and in this respect the earth should attract Mercury more than does Venus. But the greater distance of the earth from Mercury more than offsets the greater mass of our planet, and hence the attraction of Venus is finally the greater on Mercury. This attraction pulls Mercury out of the path that it would otherwise describe, and so it does not always follow exactly the same path in its revolutions. The same applies to all the planets, their mutual attractions causing 'perturbations', that is, small displacements from their normal places.

A number of comets make comparatively close approaches to different planets and in such cases they are perturbed by the planets, these perturbations being specially noticeable in the case of the larger planets, Jupiter, Saturn, Uranus and Neptune. Hence it is quite possible for a comet which is moving in an elliptic orbit round the sun, if it comes fairly close to one or more of these planets, to be so perturbed that its orbit is changed from an ellipse into a hyperbola. If this occurs the comet goes off in an orbit like

the outside curve in Fig. 4, and as this curve – a hyperbola – does not close in on itself, the comet will simply move on and on away from the solar system into the depths of space, perhaps some day to be captured by some star and made to move in an orbit round it as it once did round the sun. This is not an imaginary story; it has actually happened that comets have been compelled by the attractions of some of the larger planets to move in hyperbolic orbits, being thus expelled from the solar system. But notice that in every case where a comet has been detected in this act, it has been proved that it was formerly a member of the solar system, pursuing an elliptic orbit round the sun, and its expulsion from our system happened because it ventured too close to a massive planet.

In the past, comets have always made a strong appeal to the public, and the same would apply to-day if only we got a few very spectacular ones. Unfortunately really spectacular comets are rather rare, though they do occur from time to time and in some cases cause a scare, especially among the superstitious. In ancient times they were looked upon as omens of evil and were also regarded as dangerous, especially if they had long tails, as some comets have. In medieval times they were believed to be luminous appearances in our atmosphere, and in such circumstances it is not surprising that people thought they were responsible for pestilences and various other evils.

It was not till the sixteenth century that this delusion about their atmospheric origin was dispelled by the work of Tycho Brahe. Soon after he had settled at his observatory in the island of Hveen in the Sound a bright comet appeared, and Tycho kept it under observation for about two months. Although his instruments were very crude he was able to decide quite definitely that this comet was further off than the moon. It was necessary to make simultaneous observations from two places on the surface of the earth (and one of these was made by a collaborator of Tycho) to find the distance of the comet, just as a surveyor uses a base line and two

angles to find the distance of an object. The comet was so far away and the base line so short that no parallax could be detected, so Tycho knew that the object could not have been sublunary. While the great majority of comets are far away from the earth – many millions of miles – there have been a few cases where they have come fairly close to us and on occasions the earth has passed through the tails of comets, but without suffering any evil effects. On rare occasions we have even encountered a portion of the head.

We must now deal with the question, What are comets and of what are they made? Probably some who have seen the glowing tail of a comet – Halley's, for instance, in 1910, or more recently Comet Bester,* especially in the southern hemisphere – would conjecture that the tail at least consisted of an enormous volume of hot gas, which might prove detrimental to every form of life if we encountered it. This conjecture would not be quite correct, and we shall now consider the three main parts which are usually recognised as the components of comets.

First of all there is the *nucleus*, which is not always easily detected with the telescope. The nucleus is not a solid body like a planet but a loose agglomeration of particles of varying sizes; some are merely like grains of sand and others are fairly large, occupying many cubic yards in volume. The particles composing the nucleus are similar to many of the elements and compounds that we find on the earth – iron, nickel, magnesium, sodium, etc., which suggests a common origin for the different bodies of our system. Surrounding the nucleus is the *coma* which develops as the comet approaches the sun. At a certain distance, depending on the nature of the particles in the nucleus, the heat of the sun causes gases to exude from these particles, and the gases often carry fine dust with them which, with the gases, diffuses outwards from the nucleus. This reflects the light of the sun so that the coma becomes visible. The tail may be regarded as a development of the coma.

*This was the eleventh comet discovered in 1948.

In most cases the *tail* of a comet is the most spectacular part, but it does not always develop according to expectations and often it is disappointing. When it develops it always points away from the sun – in apparent defiance of the law of gravitation – and the cause of this phenomenon will now be briefly explained.

In the early part of the seventeenth century Kepler suggested that the tails of comets were possibly deflected away from the sun owing to the pressure of sunlight. We now know that this is a correct view and that in the case of very small bodies the pressure of the sunlight constitutes a repulsive force exceeding the gravitational attraction of the sun. Many years ago Arrhenius showed that if this light repulsion is to exceed the sun's attraction, the diameters of the particles must not exceed 0.001 mm. The ratio of the repulsive to the attractive force increases inversely as the diameter of the particle, but only over a certain range, and the computations become a little complicated as the diameters of the particles approach the wave-length of light, but it is unnecessary to deal with this or with the densities of the particles, which also have an effect. It will be sufficient to say that the repulsive force, acting on the finer particles in the tail, which is formed from the coma as described, drives these away from the sun. In some cases two or more tails are seen and these are probably due to the sorting action of the light-pressure on the particles of different sizes. Many sudden changes are sometimes observed in the tails of comets; limits of space prevent us discussing these, but before leaving this subject it may be remarked that electric repulsion as well as light-repulsion cannot be ignored in dealing with cometary phenomena. Morehouse's Comet is shown in Pl. 1 and Delavan's Comet in Pl. 2.

It is not surprising that the tails of comets have sometimes caused alarm when we remember the enormous length to which some tails extend. Although in most cases the tails are not very impressive, in others they may be from 50 to

150 million miles long and perhaps more than one-twentieth of these figures in width. At the last return of Halley's Comet, its tail was shown to be nearly 9 million miles long in the early part of February, 1910, but by May 5 it had increased its length to nearly 28 million miles. The earth passed through the outer edges of the tail and although various phenomena, such as auroral glares, seemed to indicate that it had certain meteorological effects, nevertheless, as such phenomena are observed at other times, it is difficult to say that they were actually associated with the tail of Halley's Comet. Bright moonlight masked any faint effects that might have been otherwise seen. Owing to the extremely rarefied condition of the gases in the tail of any comet there is practically no danger, even from the deadly cyanogen which the spectroscope has shown to be present. So rarefied are the gases that stars seen through the tails of comets are not dimmed to any appreciable extent, and in many cases not at all (see Pl. 1). The molecules composing the tail are obviously some distance apart.

Where do comets come from if they are not visitors from outward space? This is a question that has not yet been satisfactorily answered, but we shall deal with some of the theories that have been advanced to explain their origin.

It has been conjectured that comets have been ejected by the sun, and some support is lent to this view from the fact that a few comets pass very close to the sun - in fact they practically graze his surface. The orbit of the Great Comet of 1882 is an example of this. The perihelion distance (the least distance from the centre of the sun) of this orbit is about 720,000 miles, and as the radius of the sun is 433,000 miles, it will be seen that this comet, known as 1882 ii, passes within a distance of about 290,000 miles from the surface of the sun. Although irrelevant for our present purpose, it may be pointed out that it takes about 760 years to complete a revolution around the sun, or about ten times as long as Halley's Comet, and that its greatest distance from the sun, known as its aphelion distance, is

about 16,000 million miles, which is more than four times the mean distance of Pluto.

It may be objected that the sun could not eject matter with sufficient velocity to carry it off from his surface and cause it to revolve round him afterwards, but this view is not correct. The sun is always erupting matter with a high speed, as the phenomenon of solar prominences shows, and if this matter attained a velocity of 270 miles per second it could move round the sun in a circular orbit. If it just exceeded 380 miles per second it could describe a very elongated ellipse such as the Great Comet of 1882 does, and for speeds between these two it would describe ellipses not so elongated as this comet's orbit. For speeds a little greater than 270 miles a second the orbit would be an ellipse but resembling the orbits described by most of the planets - more or less like a circle. Unless these orbits were modified by the perturbations of some of the planets, they would intersect the sun and hence the cometary matter ejected would be re-absorbed by the sun at some time. Planetary perturbations could, however, modify the orbits sufficiently to make the matter merely graze the surface instead, though they could also modify them in another way - make them strike the sun at a point deeper than that from which ejection took place, if we can imagine such a thing possible. In the latter case, however, as the comets would return into the sun we should not have much opportunity of seeing them, and most or all of them might be undiscovered.

There is one insuperable difficulty about this view. We know the composition of the meteoric particles that compose the nucleus of a comet because we sometimes encounter these, and the larger portions of the nucleus, instead of being dissipated in gases in the upper regions of the atmosphere as ordinary meteors, strike the ground as meteorites. Now it is certain that the sun could not have ejected these meteorites just as they are now because the surface temperature of the sun - about 6,000° C. - would vaporize them. They must have been ejected in the gaseous condition

and have condensed later into solid bodies. But the physicist objects to such a view. Comparatively small gaseous masses at a high temperature would not solidify into bodies like meteorites; on the contrary, they would rapidly dissipate into space, their own gravitational attraction on the molecules of their gases being too small to retain these. It has been calculated that no planet with a radius less than 1,200 miles could have been formed as the result of gradual condensation from a gaseous mass, and a planet with such dimensions would be incomparably more massive than any comet. The radius of the planet Mercury is 1,500 miles, and certain calculations based on its mass and other factors show that if this planet condensed from a gaseous mass the density of this mass must have been about 0.8 that of water, which, of course, is quite impossible. It may be added that some calculations made on the mass of Halley's Comet showed that this could not have exceeded ten thousand million million tons (written for convenience in the form 10^{16} tons) and may have been only a thousand millionth of this. The mass of Mercury is about 2×10^{20} tons, which is very much greater than the maximum value given above for Halley's Comet. The actual mass of this comet, practically all of which lies in the nucleus, as with all comets, is almost certainly much nearer the lower limit given above, which is only a minute fraction of the mass of Mercury.

From what has just been said it is obvious that it is difficult to account for sun-grazing comets as the condensations of gaseous material ejected by the sun, but even if this were allowed we have still to account for the origin of the others - running into many hundreds - which do not graze the sun's surface.

Another theory is that certain comets were either ejected by the major planets or that they were so perturbed by these planets that they were forced into the orbits in which we find them to-day. We have purposely used the word 'certain' in dealing with comets because they are roughly divided into two classes: (a) those with long periods like

Comet 1882 ii, just mentioned; (b) those with short periods which vary from a few years to about 76 years or even more; the latter is the average period of Halley's Comet, while Encke's Comet has a period of only $3\frac{1}{2}$ years. Yet both are included under the category of 'short-period comets'.

Short-period comets, to which we shall confine our attention for the present, are usually associated with one of the planets. For instance Comets Halley, Pons, Brooks and others are associated with Neptune, and the reason for this will be seen from the following considerations.

The aphelion distance of Halley's Comet is about 3,340 million miles; the mean distance of Neptune from the sun is 2,797 million miles, and this approximate correspondence in distances suggests that at some time Halley's Comet may have been ejected by Neptune. The same thing applies to the other major planets, each of which has a 'family' of comets, Jupiter having the largest family.

The late Richard Proctor postulated the planetary origin of the short-period comets and assumed that their ejection took place in the remote past, when the planets were much hotter than they are to-day. It was necessary to make this assumption as otherwise it is difficult to explain the high speed of ejection necessary to produce the comets. Take the case of Jupiter, whose large family of comets have all direct motion, that is, they move round the sun in the same direction as the planets; this applies to most, but not to all, the short-period comets. Could Jupiter eject comets to-day?

Each planet has what is known as its 'velocity of escape', which means the velocity with which a body must be projected from its surface to make it go off into space. In the case of the earth this is 7 miles per second and in the case of Jupiter it is about 38 miles per second. If, therefore, Jupiter ejected matter which afterwards became a comet, its velocity of ejection should be about 38 miles per second. It may be objected that a gaseous mass might describe a circular orbit around Jupiter with a less velocity than this, which is true, and in fact a velocity of ejection of

27 miles a second would suffice for this. But in this case the ejected matter would merely revolve around Jupiter, whereas all comets revolve around the sun, and for the latter a velocity of about 38 miles a second would be necessary. Small modifications, arising from the fact that Jupiter's orbital speed is 8 miles a second, need not be considered.

This velocity seems very high when we remember that the gaseous surface of Jupiter is very cold – about -140°C . – and that any activity such as we see on the surface of the sun is out of the question. On the other hand, as planetary observers know, there is a considerable amount of activity on Jupiter, especially on the great Red Spot, and it is possible that underneath the cold exterior there may be hot layers. Here we enter the realm of speculation, and one can only say that it is possible that Jupiter and other major planets may eject matter which later becomes comets.

If we reject this view, there is the other theory that the short-period comets were once long-period comets which made close approaches to the major planets, and were annexed as parts of their 'families'. At first sight this theory seems attractive, but a careful examination of the question shows the extreme improbability that this could have happened. The theory requires that comets must pass very close to one of the major planets at some time and, taking into consideration the space through which comets can possibly move, and the relatively small volume of space surrounding each planet in which capture could occur, the probabilities against such captures are enormous. It is true that if we are given an extremely long period of time such captures could occur, but our time is limited for the following reasons.

A comet does not last indefinitely; some wear out fairly quickly with repeated revolutions, though Halley's Comet is one of the few exceptions. Even in the case of Halley's Comet, however, its splendour in recent times is less than it was many centuries ago, and like other comets it is losing

its material. Comets sometimes disappear completely, as happened with Biela's Comet, which split in two in 1845, and on its next return in 1852 the two parts were about a million miles apart. In September of that year the parts disappeared from view and the comet has not been seen since, but its previous existence still makes itself known in the shower of meteors which takes place towards the end of November. These meteors are known as the Bielids, and also as the Andromedids because their radiant point is in the constellation of Andromeda. The shower is now very faint and is evidently almost worn out, or perhaps the earth misses the main portion of the debris of the comet, but in 1872, when the comet was due to return, the shower of meteors was a wonderful sight. On that occasion the earth encountered concentrated debris of the disintegrated comet.

Considering that the short-period comets are wearing out, some renewal would be necessary, but renewal by capture would be inadequate. It has been estimated that many thousands of years would be required for Jupiter to add one comet to his family, and as practically all his existing family would be defunct in that time, renewal would not keep pace with dissolution.

Even if we admit that short-period comets have either been ejected by the major planets or annexed by them, this does not explain the larger number of long-period comets. Where did they come from? Their existence, and indeed that of the short-period comets as well, still remains one of the riddles of our solar system. Until we know how the solar system came into being (and we seem a long way yet from a solution of this problem) we can merely speculate about the origin of these strange wanderers which, in all ages, have been responsible for arousing more superstitious fears than have all the other heavenly bodies. Before leaving this subject it may be added that there is some doubt about the reality of the families of comets which are supposed to be associated with the major planets, in particular about the families of Saturn, Uranus and Nep-

tune, but the same scepticism does not exist about Jupiter's family. The reason for this is briefly as follows.

Astronomers have a certain criterion that they can use in ascertaining the effects on a comet of a comparatively close approach to a planet. Now it is a simple thing to determine the minimum distances of the short-period comets from the orbits of the major planets, and, assuming that a major planet was at this minimum distance at the critical time, that is, when the comet was nearest the planet's orbit, the effect can be determined. Prof. H. N. Russell, as a result of his investigations, concluded that Jupiter is the only one of the four major planets that plays an important role in the capture of comets, that of Saturn and Uranus being of little importance, while that of Neptune is quite negligible.

There are over forty short-period comets that have been observed to make more than one return, that is, they have been observed on at least two occasions separated by many years, from 151 in the case of Comet Herschel-Rigollet, 76 in that of Halley, nearly 73 in that of Olbers, and so on down to Encke with its short period of $3\frac{1}{2}$ years. There are nearly as many (actually thirty-eight) whose periods have been computed and found to be comparatively short - all under 65 years - which have been observed only once. Of course small errors in computation would be responsible for preventing astronomers from finding them on their return, because in these circumstances the predicted places would not be exact. It must not be assumed, therefore, that these comets have disintegrated like Biela's: many of them may be rediscovered in the future.

The known number of long-period comets, those with periods running into many hundreds and in some cases into thousands of years, is several hundred, and the number is increasing every year with new discoveries. Prof. Elis Strömberg, who made a special study of comets, believed that there were millions of them in the solar system. We have already seen that some known comets are wearing out

or have already worn out, and it is impossible to believe that these millions of comets (if we accept this high figure, and perhaps we should exercise care in our judgment) have existed since the solar system was formed about 3,000 million years ago. Their presence constitutes one of the major problems of astronomy.

The speeds of comets can be easily calculated from the following simple rule which applies approximately to those whose periods are not very short, like Encke, when they are visible from the earth. They are then comparatively close to the sun, comparatively close at least for comets, which, as we have seen, frequently move out a long way from the sun.

A comet which is as far from the sun as is the earth has a speed of about 26 miles per second. If it is four times as far off as this, its speed is 26 divided by the square root of 4, that is, it is 13 miles per second, and so on. Thus a comet half as far from the sun as is the earth has a speed of 26 divided by the square root of $\frac{1}{2}$, which is nearly 37 miles per second. The rule, though not exact, is useful for approximate results, but should not be used if the comet is far away from the sun.

We shall now say something about the connection of comets with meteors. It will be necessary to refer to Fig. 2 to explain why certain showers of meteors are closely associated with well-known comets - about ten altogether.

As a comet moves round the sun its nucleus is held together by its own gravitational attraction, which is very feeble in comparison with that of the planets. Hence it would not require a great force to produce disintegration of the nucleus; the attraction of the sun, more especially when the comet is at or close to perihelion, is one factor responsible for disintegrating the nucleus. Another factor is light pressure, in particular on the smaller particles of the nucleus, as already explained. It will be seen that as a comet revolves round the sun it will leave a lot of debris in its wake, but this debris will follow on more or less in

the track of the nucleus, though it will become spread out to some extent in the plane of the comet's orbit. It is possible, therefore, that in the course of centuries the debris of a comet will be spread along its path, though it may be more concentrated at some places than at others. It is certain that this has happened with Halley's Comet because each year the earth encounters some of the debris in the early part of May, the result of which is the shower of meteors known as the Aquarids. It is not surprising, therefore, that Halley's Comet is gradually losing its material, like many others, and even if the earth did not encounter the debris, the nucleus must be continually decreasing in mass because of the spread of debris along the track of the comet.

Fig. 2 shows the debris scattered along the track of a comet which has approached the sun at one point to a distance nearly the same as the earth's distance from the sun. The debris is following on after the comet and the earth is seen passing through it. In these circumstances the particles will strike the earth with speeds which depend on three factors; (a) the speed of the particles, which is practically that of the comet, (b) the speed of the earth, (c) the angle between the direction of the earth's motion and the direction of motion of the debris. Another factor which plays a minor part is the increased velocity of the debris owing to the earth's attraction. Taking this latter and also the other three factors into consideration, the actual speeds with which the debris strikes the upper regions of our atmosphere vary from 10 to 45 miles per second. In the figure, which shows that the earth and the debris are moving in nearly opposite directions, it is easily seen that the conditions are favourable for the higher velocity just mentioned, whereas if the comet were moving in the opposite direction, the speed of encounter would be about 10 miles per second. At such high speeds the smaller portions of the debris are quickly vaporised owing to the intense heat generated by friction with the atmosphere, but larger por-

tions are not so easily dissipated into gas and strike the earth. As this occurrence is comparatively rare we must conclude that by far the greater portion of the nucleus of a comet consists of the finer particles.

In this brief account we have dwelt on the more important features of bodies which still puzzle the astronomer. It is quite impossible to deal adequately with the subject in an article of this length, but probably sufficient has been said to convince the reader that in spite of the extraordinary progress in astronomical science during the last few decades, and not least in our knowledge of some of those stellar systems separated from us by millions of light-years, our ignorance concerning these parochial problems of our own puny solar system is still extraordinarily profound.

BIBLIOGRAPHY

There are very few books on comets, but there is a large amount of literature on the subject, which can be found in various astronomical publications, English and foreign. Those who desire further information on the subject could consult some of the following books:

Amédée Guillemin, *The World of Comets*, Translated and Edited by James Glaisher, F.R.S. - (Sampson Low, Marston, Searle and Rivington, 1877).

George F. Chambers, *The Story of the Comets* - (Oxford at the Clarendon Press, 1909).

Charles P. Olivier, *Comets* - (Baillière, Tindall and Cox, 1930).

Mary Proctor and A. C. D. Crommelin, *Comets, Their Nature, Origin, and Place in the Science of Astronomy* - (The Technical Press, Ltd., London, 1937).

Dating the Past

A Comment -

READERS of I. W. Cornwall's excellent article on 'Scientific Prehistory' which appeared in *Science News VIII* may be interested to learn of a new method for dating the past which has only recently been developed. It can be applied to any organic remains which are up to about 50,000 years old, and will therefore obviously prove of the very greatest value in future archaeological studies. The method is similar to that used in estimating the ages of some ancient rocks by measurement of their radioactivity, as mentioned by I. W. Cornwall; but is applicable only to materials which contain the element carbon - i.e. all organic matter.

Ordinary carbon consists of stable atoms, mostly having an atomic weight of 12; but it has been found that the carbon dioxide of the earth's atmosphere contains a very small proportion of an unstable radioactive form of carbon, having an atomic weight of 14. Atoms of this unstable 'isotope' are being continually generated in the atmosphere by cosmic rays, and are also being continually converted into a stable atomic species (nitrogen with an atomic weight of 14) by radioactive decay. In the free atmosphere there is a balance between these two changes, and the actual number of unstable carbon atoms remains constant. Any piece of living organic matter is continually exchanging carbon-dioxide with the atmosphere, so it follows that there is the same equilibrium proportion of radioactive carbon in all living organisms. However, as soon as death occurs, the organism no longer exchanges carbon dioxide with the atmosphere, and the process of radioactive decay will immediately begin to diminish the number of unstable carbon atoms in the dead organism. Virtually no new

unstable atoms will be formed within the organism because the number of cosmic rays striking it is very small.

It so happens that the unstable carbon atoms change into stable nitrogen atoms at such a rate that *half* of them disappear in approximately 5,000 years. One of the characteristics of a radioactive change such as this is that there is a definite 'half-life' period, and no external influence will alter the rate at which this change takes place. Therefore, 5,000 years after death only *half* the unstable carbon atoms originally present will remain, after 10,000 years only a *quarter*, after 15,000 years an *eighth*, and so on.

The small quantity of unstable carbon-14 atoms present in a recently dead specimen is just detectable by means of a sensitive Geiger-Muller counter. (The carbon-14 is present to the extent of about one in a million-million, and β -rays of very low energy are emitted). To carry out any accurate measurements it has so far been found necessary to concentrate the small proportion of carbon-14 by means of thermal diffusion, or some similar process, and then to employ either a mass-spectrograph or a Geiger-Muller counter to determine its amount.

Although very great accuracy has not yet been obtained, there seems little doubt that this method is capable of further development and will provide an absolute time scale for archaeological remains, especially for those with ages within the range 1,000 to 50,000 years, i.e., periods which are comparable with the 5,000-year half-life of carbon-14.

The method is being actively developed in the Chemistry Department of the University of Chicago, and readers who are interested in further details are referred to the following publications: -

- Physical Review*, Vol. 69, p. 671 (1946).
- Physical Review*, Vol. 72, p. 931 (1947).
- Physical Review*, Vol. 73, p. 350 (1948).
- Science*, Vol. 105, p. 576 (1947).

J. T. KENDALL

- and the reply

I was most interested to read Dr Kendall's comments on radioactivity-dating of organic archaeological objects.

I was, of course, aware that radioactive isotopes, and in particular, carbon-14, were being used as 'tracers' in physiological and biochemical researches, but the gradual breakdown of C^{14} in organic remains and the application of this fact to prehistoric dating were unknown to me, as, I think, to most other prehistorians. We, therefore, owe Dr Kendall a debt of gratitude for introducing this new tool to us, and to *Science News* for bringing us together.

The method is so new as to have been hitherto untried in this country, but it would seem to present tremendous possibilities. The half-life period of 5,000 years fits it particularly for dating the last few millennia where, as far as Western Europe is concerned, dated points in time are scarce and not by any means always reliable.

It would be interesting to know whether or not the estimation of C^{14} in any individual sample is a matter of great difficulty and expense. Most prehistorians are notably short in the purse, and determinations, each occupying much time of a skilled investigator, are expensive. Perhaps Dr Kendall could also indicate who would be prepared to undertake C^{14} estimations in this country. The archaeological fraternity can supply unlimited quantities of material for investigations!

This limitation of expense is shared by pollen-analysis, which is extremely tedious.

There is a further new method of dating, scarcely tried as yet, which I did not mention in my article. It depends on a progressive accumulation in fossil bones of traces of fluorine from percolating water. By micro-estimation of fluorine in a specimen, and comparison with the fluorine-content of other specimens of known age, found in the same locality, or in surroundings with a comparable natural fluorine-concentration, a relative dating of the specimens

seems possible. The method and its limitations have been briefly described by Dr K. P. Oakley in *The Advancement of Science* (Quarterly Journal of the British Association) IV, 1948, p. 336.

I. W. CORNWALL

How old is the Earth?

J. J. KIPLING

How old is the Earth – the planet we are living on? Civilised man appears always to have had an answer to that question, though he has not always formulated it very consciously. The answers which he has given form a fascinating sequence, the last part of which shows very strikingly the way in which such a problem is tackled by the scientific method.

The date of the Creation – 4004 B.C. – found on the first page of many Bibles, is due to Bishop Ussher who, in the seventeenth century, deduced the age of the earth from a study of the Hebrew chronological tables and the account in Genesis. Much earlier, the Persian Zoroaster thought the world was then twelve thousand years old, while Babylonian astrologers gave the age of mankind (a rather different thing, as we shall see later) as half a million years. When we consider the modern estimate, we may be inclined to sympathise rather with the ancient Brahmins who thought the world had always existed.

The scientific approach to this question was first made as recently as the nineteenth century. It was made possible by the coming of the Uniformitarian Theory in geology. Formerly all geological processes had been thought to be sudden and catastrophic, a belief which subsequently came to be called Catastrophism. The early geologists suggested however that such natural features of the landscape as river gorges were not suddenly formed overnight, but were the result of a gradual process of erosion lasting over a very long period of time. The process was much more uniform than was suggested by Catastrophism.

Geologists could actually measure the rate at which

river valleys were being eroded. They then estimated the amount of rock which the rivers must have carried away to leave the present valleys. Assuming erosion to be a uniform process, they calculated that many rivers must have taken a million years to wear down their valleys to the present level. It follows that the rocks into which these rivers have carved their valleys must be considerably older than a million years.

These estimates were only very rough. In 1898 Joly tried to form a more accurate estimate. He observed that the rivers carry down to the sea small but measurable amounts of salts (of which ordinary common salt, sodium chloride, is one). Careful measurements of the world's principal rivers showed that 156 million tons of sodium, in the form of various salts, are added to the oceans annually. The oceans are estimated to contain 12,600 million million tons of sodium as its salts. Hence if the addition of salts has been going on uniformly, it has taken 81 million years to make the oceans as salty as they are now.

This estimate looked very accurate, but further reflection showed a number of ways in which it might be in error. Thus there might have been some salt in the original oceans before the rivers added to it, in which case the estimate is too high. On the other hand, large quantities of salt are blown back on to the land from the sea in spray; this makes the estimate too low. But quite apart from such errors, many reasons were discovered for thinking that although the removal of salts from the land has been a gradual process, it has not been uniform. For example, during the Ice Ages, large areas of land were protected from the action of rivers and the rate at which salts were carried to the sea must have been much smaller than it is now. Again, the land has not always been raised to the heights of the mountains we now know. Lower mountains would give rise to slower-flowing and therefore less vigorous rivers which, again, would carry less salt to the sea. For reasons such as these, the estimate had to be revised, though

a revision could only be done by inspired guesswork. A value of 330 million years emerged.

Meanwhile the physicists were making a very different approach. The earth is believed to have been formed from the sun. It has therefore cooled from the very high temperature of the sun to its present relatively low temperature. Early in this process, a solid crust must have formed and the rate at which the earth has cooled must therefore have depended on the rate at which heat could be conducted through the crust. Lord Kelvin measured the heat conductivity of some of the rocks forming the earth's surface. These varied somewhat and, when coupled with different estimates of the thickness of the earth's crust at different stages of its history, gave a range of values between 25 and 40 million years for the time of cooling.

These values were in marked contrast with those of Joly and other workers. The discrepancy was at once recognised as important because developments in both biological and geological theories favoured the longer time scale. By this time the theory of evolution had taken shape in quantitative terms, that is, biologists had built up a time-scale for the various stages of evolution. This, however, required a much longer time than was allowed by Lord Kelvin's estimate, based on physical theory. The biologists, moreover, were supported by the geologists, who had worked out a time sequence for the formation of successive groups of rocks. The time required for this sequence was again longer than was allowed by Lord Kelvin's calculations. Apart from minor modifications, however, no error was found in Kelvin's method and a serious controversy was carried on for many years between those scientists who thought in terms of a relatively young earth and those who considered it to be much older.

The disagreement was explained early in the twentieth century. Becquerel discovered the phenomenon of radioactivity. The Curies isolated radium, an element which spontaneously disintegrates. They showed that this process

of disintegration (radioactivity) is accompanied by the evolution of large quantities of heat. Lord Rayleigh then found that radium is widely distributed throughout the earth's crust and is generating large quantities of heat there. So much heat is being generated in this way that at present the earth produces just about as much heat from radioactive materials as it loses by radiation from its surface. Thus its temperature is at present being maintained practically constant, instead of falling progressively as Kelvin assumed. Kelvin was right in assuming that the earth is radiating heat, but he did not know that this heat is being replaced as the radioactive materials disintegrate. In the past, of course, the loss of heat from the earth's surface was greater than it is now, and the supply of heat from the radioactive sources was not sufficient to prevent the earth from cooling. But the total time during which the earth has been cooling is much greater than was assumed by Lord Kelvin. The earth's store of heat is not limited to what it had on leaving the sun. Fresh additions are being made continuously and help to replace what is lost by radiation.

With the discovery of the reason for Kelvin's short life for the earth came a further method of estimating geological age. It was discovered that radium is only one of a long chain of chemical elements all of which are radioactive, i.e., undergo spontaneous disintegration. Radium is in fact formed from uranium, though there are several other members of the chain in between these two. Each is formed when the previous member disintegrates; each in turn disintegrates and gives rise to the succeeding member and also to either an alpha particle (the nucleus of a helium atom) or a beta particle (an electron). The chain is continued beyond radium and ends with lead, which is not radioactive.

The net result of the process is that an atom of uranium disintegrates to give (ultimately) one atom of lead and eight atoms of the gas, helium. (In addition, as we have seen,

large quantities of heat are liberated.) If, therefore, we can find the rate at which uranium disintegrates, and if we can also find how much helium or lead has been produced from a given piece of uranium ore, we can deduce the age of the ore and of the rock formation in which it occurs. Now the rate at which uranium disintegrates can be measured very accurately in the laboratory. This is the first advantage of this radioactive method over previous methods which all, to a greater or less extent, relied on an estimate rather than an accurate measurement.

The method based on radioactivity has one further advantage which makes it vastly more important than previous methods. Our theory of radioactivity makes it clear that the rate of disintegration of uranium must have been the same throughout the earth's history as it is now. This has been confirmed by two different kinds of experiment. In the laboratory uranium has been subjected to very widely varying conditions of temperature, pressure, etc., such as might have obtained during the course of the earth's history; no alteration in its rate of disintegration has been brought about. Secondly, we know that the helium nuclei (alpha-particles) which are emitted in the disintegration of uranium can travel only a limited distance which is dependent on the rate of disintegration. In certain minerals (e.g., micas) which contain crystals of a uranium ore, the alpha particles cause a darkening where they come to rest. This darkening takes the form of a ring whose radius is the range of the alpha particle. It is found that the oldest rings are as sharply defined as the youngest and have the same radius; we can therefore conclude that the range of the alpha particles and hence the rate of disintegration of the uranium has been the same throughout the earth's history, as far as we can discover.

In point of fact, each member of the disintegration series which emits an alpha particle gives rise to a ring of characteristic radius. The collection of rings forms a pleochroic halo, examples of which are shown in Plate 8.

For those who are more mathematically minded, it might be explained that the disintegration of uranium takes place according to a sort of reverse 'compound interest' law. The number of atoms which disintegrate over a short period is proportional to the number of uranium atoms present *at that time*, just as the interest added to a savings bank deposit in a given year depends on the total amount in the account for that year and not on the amount originally deposited, perhaps some years previously. For uranium, it is the 'interest rate' (or rather the rate of *removal* of 'capital') which remains constant.

It follows that we can calculate (by an equation rather similar to the one used in calculations of compound interest) how long the disintegration of a particular piece of uranium has been going on to give an observed amount of either lead or helium. Both lead and helium have been used in these investigations, each giving rise to a special method. It must be stated at this point that the determinations are a little more complicated than has been described so far. There are actually three radioactive chains. Readers of *Science News II* will remember that uranium consists of three isotopes, the atoms of which are chemically alike, but have different masses. Uranium 238 is the first member of one chain, uranium 235 is the parent of a second chain, and another element, thorium, gives rise to the third. (The third isotope of uranium, of mass 234, is a member of the chain which starts with uranium 238.) All three chains produce both helium and lead, each at its own rate. These facts have to be taken into account in making the calculations. What results are obtained?

We have said that helium is a gas. During the many upheavals which have taken place in the earth's crust, therefore, the helium which had accumulated in the rocks affected must have been lost. The helium which we now find has therefore accumulated only during the lifetime of the rock (in its present form) in which it is found. This means that this method does not give a direct value for the

age of the earth. It does, however, give the age of each individual rock formation. This in itself is very valuable to geologists. Moreover, it enables us to date the fossils found in different rocks and thus to give at least an approximate time-scale to evolutionary processes - see Table 1. Together with the results of the lead method, it gives a continuous record of rock formation from as far back as 1850 million years ago to the present time. Thus the 'long' time-scale is confirmed, and geologists and biologists are now in agreement with physicists and chemists.

TABLE 1. GEOLOGICAL AGES

<i>Period</i>	<i>Stage of the earth's history</i>	<i>No. of million years ago</i>
—	formation of the earth	3,350
Archaean	no remains of life found	1,800
Proterozoic	invertebrates appear; plant remains form graphite	900
Cambrian	primitive invertebrate marine life; no land life	500
Ordovician	marine life - crustaceans	400
Silurian	earliest fishes; simple land plants	350
Devonian	some fresh-water fish appear	300
Carboniferous	advance of fish and development of amphibians; dense forest of club-moss etc., insects; land animals develop with appearance of flora	280
Permo-Triassic	reptiles appear	200
Jurassic	giant reptiles	130
Cretaceous	decline of reptiles; advance of mammals, birds, flowering plants	100
Tertiary	mammals predominate	60
Quaternary	appearance of man	20

(after Shand, with additions)

Calculations based on lead are similar to those based on helium, but there is one complication. Most rocks contain small but appreciable amounts of lead which have not been formed by the disintegration of uranium. In any particular rock, how much of the lead has been formed from uranium and how much was there originally? This question was

difficult to answer until some valuable experiments were done by Aston in this country and Nier in America. They showed that lead, like uranium, and many other elements, consists of more than one isotope. In fact there are four isotopes of lead, having masses of 204, 206, 207 and 208. The last three are, respectively, the end-products of the three radio-active series which begin with uranium 238, uranium 235 and thorium. Isotope 204, however, is not known to be formed in any natural radioactive process. It is therefore an index of the amount of 'common' lead (i.e., lead not originating from a radioactive source) in any rock. It remained to examine samples of lead not contaminated with uranium and thorium to find what proportions of the isotopes 206, 207 and 208 are associated with lead 204 in common lead. From an analysis of the isotopes of lead in a uranium ore the amount of common lead can be determined and hence the amount of lead from radioactive sources (called, for convenience, radiogenic lead).

There have been difficulties in developing both the helium and the lead methods. Recent work has shown that there were many errors and inaccuracies in earlier work, but the latest values have been examined carefully and seem to be trustworthy. They are, however, ages for rock formations and not for the earth itself. At this stage we could nevertheless still say that the greatest age recorded for a geological formation is a minimum value for the age of the earth. In 1941 the greatest age recorded for any rock was 1850 million years. There is at least one rock formation older than this, of which, unfortunately, only traces remain. Hence it was concluded that the earth is at least 2,000 million years old.

In 1946, a further development was made in the radio-active method which seems at last to give us the real age of the earth, and with a fair degree of accuracy. When Professor Nier examined several samples of common lead, he found small but significant differences in their isotopic compositions. These were not big enough to make appreci-

able differences to calculations of the ages of the rocks, but were more than could be accounted for by experimental error. Nier concluded that the 'primal' lead (present at the formation of the earth) was contaminated by radiogenic lead before becoming the 'common' lead found in any particular rock. (On becoming incorporated in a rock, of course, contamination would cease in the absence of a radioactive material.)

The variations in the isotopic proportions of common lead are systematic and bear a definite relation to the age of the rock in which the lead occurs. Professor Holmes of Edinburgh has used this variation as the basis of a very elegant method for determining the age of the earth. Any two samples of common lead must once have had the same isotopic composition, that of primal lead. From the time of formation of the earth, uranium 238 added lead 206 and uranium 235 added lead 207 at their respective rates. These additions continued until the rocks we now know were formed at times of (to take two examples) 200 and 600 million years ago. From these data, Professor Holmes has calculated the time when the two samples had the same isotopic composition, i.e., when the contamination of primal lead began. He concludes that it was 3,350 million years ago.

Thus we have at last a real and reliable value for the age of the earth, though it will, like all scientific results, be subjected to careful scrutiny by other scientists and, if need be, to revision in the light of later research. It remains only to say that this value finds considerable support from the researches of astrophysicists who are interested in the age of various celestial bodies (including the sun and the earth) and of the whole universe. But that is another story.

Biochemistry To-day

J. L. CRAMMER

What is Biochemistry?

THE division of scientific knowledge into distinct and separate sciences is in the last resort arbitrary, but convenient, and often hallowed by long convention. Everybody knows that geology is the subject that deals with the nature of the Earth, and the stratification, history and composition of the rocks that make it up. Yet there exist also specialities such as mineralogical chemistry, which studies the detailed chemical nature of each rock; palaeobotany and palaeozoology, concentrating on the fossils found in them; and geophysics, a name given to the investigation of their magnetic, gravitational, seismic, and other physical properties. These subjects span scientific frontiers: they are part of geology, but also parts of chemistry, biology or physics, and how they are regarded is purely a matter of convenience.

Biochemistry is a subject in similar case, a meeting-place of many boundaries. Unlike geology, however, it is a new development, a new delimitation of subject matter in the field of natural knowledge, unheard of a hundred years ago; and because of this newness of its creation, its purposes and potentialities are still largely unrecognised. Few people are aware of its true frontiers, or of the particular advantages to be gained in making it a separate study. To the laboratory chemist, biochemistry is primarily a matter of the chemistry of natural products – the dyes which colour flower petals, the scent of musk, the myriad sugars and starches of vegetables, the crystalline substances to be obtained after boiling hair for twelve hours or so with strong hydrochloric acid. To the practising doctor, on the other hand, bio-

chemistry is the measurement in a hospital laboratory of the amount of sugar in a patient's blood to help in the diagnosis of diabetes, or the amount of hydrochloric acid in his stomach, to throw light on the possible existence of an ulcer or a cancer. The public and the politician equate it with nutrition, the study of vitamins, calories, and a balanced diet. The physiologist thinks of it as the study of digestion, excretion, and metabolism. To the orthodox bacteriologist, it is a valuable but unacknowledged help in distinguishing different kinds of bacteria, which appear confusingly alike beneath the microscope but call attention to their individuality by their differing chemical powers. For instance, the common bacilli inhabiting the intestine are capable of fermenting lactose (milk sugar), whereas the very similar-looking bacilli of dysentery cannot attack this sugar, and are grouped as 'non-lactose fermenters'.

The list of viewpoints grows and grows. It is not surprising that reports of biochemical discoveries appear in many different kinds of scientific journals, and come from many different kinds of laboratories besides the regular biochemical institutes. This scatter in turn explains the variety of opinions as to what biochemistry is really all about. If you are busy identifying bacteria by their power to make a smell or a pigment, you are apt to forget that this smell or colouring is important also to the organisms themselves, a part of their natural lives and their normal activities. The diphtheria bacillus does not make its poison out of sheer caprice, or to damage its host if it can, but because toxin-production plays some role in its own internal life economy. The toxin may be an enzyme essential to its own living machinery but accidentally shed, or a waste product cast away in daily life, but it is not primarily merely a weapon, it is an integral part of the organism's functioning as an individual.

Man, like all other animals, uses up the oxygen of the air and makes carbon dioxide. His pure saliva, spat into a starch solution, slowly changes it chemically so that it no

longer colours blue with iodine. Bacteria turn sweet milk sour and acid, yeast produces alcohol from the grape. The realisation that living things work chemically, carrying out chemical reactions with substances, making them and changing them, sprang from a discovery made in 1828, which started biochemistry. When in that year Wöhler found urea, a simple organic chemical, in human urine and found also that he could synthesise it in his laboratory, he was introducing the idea of the human body as a chemical factory.

The power to make chemical changes is fundamental to all living things. But the converse is also proved true by observation. Chemical substances profoundly influence all organisms. Some substances, like carbolic (phenol) or lead, are poisonous, and kill. Others modify the body's working, as aspirin lowers the temperature in fever, or epsom salts (magnesium sulphate) relieve constipation, or chloroform produces anaesthesia. Modern pharmacology (to-day a branch of biochemistry) has twin roots, the 16th-century teachings of Paracelsus and the iatrochemists, that simple minerals such as mercury or gold are good treatments for diseased bodies, and the older knowledge of the herbalists who discovered so many plants which chemists have since shown to contain active drugs. How very significant and important it is that many of the drugs in the pharmacopeia are chemicals produced by living things. The peculiar chemistry practised by plants, including moulds and bacteria, results in substances particularly powerful in affecting the functions of other living things. To-day we can go some way in explaining why this should be (see later). To-day, also, pharmacologists appreciate the fact, and in the search for new synthetic drugs, model their new molecules on this peculiar chemistry. Paludrine, against malaria, is one fruit of such an approach.

Living things produce chemical changes: they are strongly influenced by chemicals: it is not surprising, therefore, to find that they are themselves made up of chemical sub-

stances: these are the three foundation stones of biochemistry. The realisation of the truth of the third statement grew only slowly, however, and met theoretical opposition from the vitalists, who were impressed by the uniqueness of life and its apparent total distinctness from the inorganic material world. At first only seemingly non-living organs, the skeleton for instance, were conceded a chemical status, and for some centuries, phosphorus for matches was made by distilling bones. Various chemical stores were also recognised – fats and starches. Then pigments were allowed into the chemical category, and it was shown that the red colouring of human blood could be ascribed to a brown iron-containing substance called haem, a part of haemoglobin. The green colouring of leaves could be extracted and shown to be due to a chemical relative of haem, but containing magnesium instead of iron, and called chlorophyll.

Eventually, continued research revealed that tissues and organs are a complex of chemical substances. Whether they are anything more than this, as vitalists maintain, will be discussed later. For the moment, it is enough to appreciate the *complex chemical nature* of living things, for this discovery gives the clue to the true scope of biochemistry as a science. It is the winning of all possible knowledge of the chemical nature of living things in order to understand what constitutes the breath of life. How muscles move and nerves send messages, why children resemble their parents, what species of miracle directs one tiny egg to grow into a worm, another into a fern, and a third into a man, these are questions of natural philosophy to which biochemistry will give an answer. In doing so, it will also provide us with the knowledge to wipe out diseases and produce supermen, the power to *invent* new and valuable kinds of animals and plants, a complete mastery of the living world.

The sections which follow develop this theme. They proceed in order from simple to complex, beginning with an introduction to some of the chemicals in cells, and the

ways they are studied, passing on to the functioning of substances in the cell, then to the chemistry of organised systems and whole tissues, and finishing with a general discussion of the problem of growth, development, individuality, and inheritance.

Porphyrins

When blood is drawn off into a bottle containing a little sodium citrate to prevent it clotting and allowed to stand, the red cells slowly settle out to the bottom, falling under the influence of gravity. The same separation of the red cells is obtained much more quickly by spinning the bottle in a centrifuge, a machine in which bottles are held clamped at the periphery of a rotating wheel. When a stone is whirled round and round on the end of a piece of string, it tries to fly off at a tangent and the centrifugal force developed is felt as the pull of the stone on the string; the actual strength of this pull depending on the mass of the stone, and the speed of whirling. Now the bottles in the centrifuge are clamped, but their contents are free to move, and the effect of spinning blood is to throw down the red cells in a packed mass at the bottom of the bottle, leaving the clear plasma above. Centrifugal separations of this kind are very common in biochemical technique because they are rapid, can be done with small quantities without losing any material, and apply equally to gelatinous substances which would only clog an orthodox filtration apparatus.

Using the centrifuge, then, the separated red cells can be washed free of all trace of plasma by mixing the packed mass up with some 0.9 per cent salt solution (a harmless strength, neither shrinking nor swelling the cells) and spinning them out again. Two repetitions of this are enough. Now the clean red cells are mixed with distilled water, which causes them to burst open. The red colouring matter, some thirty per cent of the whole cell, is set free in solution, and cell debris begins to rise to the top. This process again is speeded by centrifugation, but first it is convenient to add

a quantity of ammonium sulphate (about half the total amount which would saturate the solution) which makes various unwanted cell proteins less soluble, precipitating some of them, and also assists the final stage, crystallisation of the haemoglobin, which occurs on cooling. In this way crystals of purified red blood pigment are available for study.

What kind of a substance is it? If we add acetone containing hydrochloric acid to a strong solution, it goes a clear dark brown, and deposits a crumbly white precipitate like fragments of hard-boiled egg, which can be easily filtered off. This experiment has split the brown haem from the protein globin. The haem can be crystallised and studied by the methods of orthodox organic chemistry; it contains an atom of iron, and has the carbon-nitrogen network structure characteristic of a porphyrin (see Figs. 5 and 7).

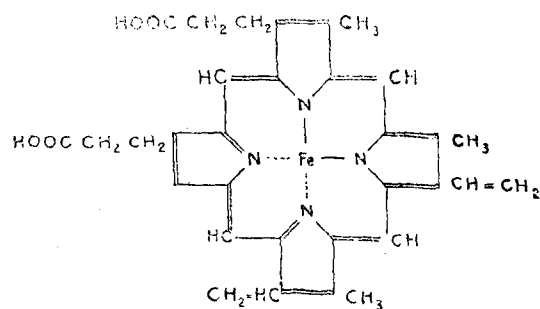


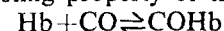
Fig. 5. - Haemin.

The globin is a protein - the implications of this are discussed fully in the following section - which has been denatured by the technical manipulation adopted. That is to say, it has been made insoluble in water and also incapable of recombining with haem to form haemoglobin, so that a good deal of its individuality as a substance has been destroyed. But not irrevocably so, because if the white insoluble precipitate is very cautiously dissolved in dilute

alkali, a good deal of the protein renatures, becomes 'native', and now if a little haem is added to the solution, haemoglobin is produced.

It is to the atom of iron (in the ferrous state) that haemoglobin owes its important property of combining with oxygen in a reversible manner. This is its physiological function in the blood, to add on molecules of oxygen as it circulates through the lungs, and to yield them up to the oxygen-hungry tissue cells as it circulates through the brain and muscles. If blood were like water, each 100 ml. would take up 0.7 ml. of oxygen; the presence of haemoglobin permits it instead to carry nearly 20 ml., a thirty-fold increase. Note that the iron atom remains throughout in the ferrous state. If it is *oxidised* to the ferric state, the molecule (now called methaemoglobin) will no longer take up oxygen, in fact, has a different set of chemical properties altogether.

Haemoglobin will combine with other gases beside oxygen, in fact, it is much more eager to take up carbon monoxide (by over two hundred times) which explains why carbon monoxide is so poisonous, and so popular for suicides. An interesting property of this reaction



is that the back-reaction (the dissociation back into blood and free gas) is greatly encouraged by shining a light on the flask in which the experiment is being performed. Naturally enough the colour of the light used is rather important, because some colours, some parts of the spectrum, certain definite wavelengths, are absorbed by a coloured substance much more strongly than others - in fact that is why it appears coloured. (Haemoglobin appears red when seen by ordinary white light because it absorbs particularly the green, but rejects to a large extent the red part of the spectrum, which consequently alone enters the eye. And, of course, only light waves actually absorbed by the molecule are capable of producing chemical changes in it).

Warburg turned this matter to good account in an

interesting and classical experiment on living yeast cells. These cells, like very many other living things, give off carbon dioxide and take in oxygen. He showed that the intake of oxygen could be stopped by adding carbon monoxide to the atmosphere in which the yeast was living (e.g., to the air in a closed flask, fitted with a manometer to indicate the rate of oxygen uptake), and that this inhibition could be abolished by shining a bright light. The relative effectiveness of each light wavelength, each part of the spectrum, in producing this abolition was studied, and in this way he was able to plot the absorption spectrum of the carbon monoxy-compound of a substance in yeast cells which must play a vital role in their respiration. By this ingenious means he could determine in great detail the colour, so to speak, of the compound, and could compare

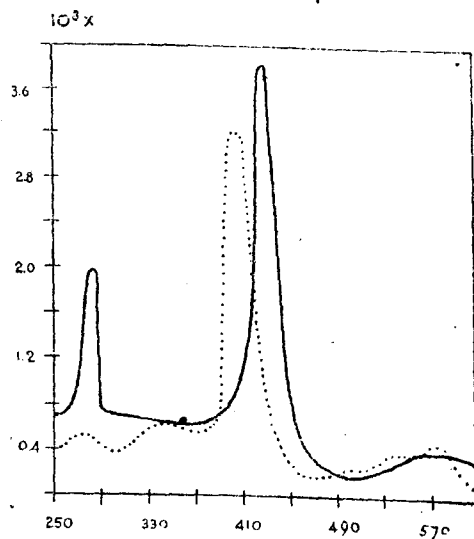


Fig. 6. - The absorption spectra of haemin (dotted) and of Warburg's substance (smooth) compared. From left to right, increasing wavelengths from ultraviolet to red; the peaks of the curves indicate the colours of light most strongly absorbed.

it with a known carbon-monoxo-compound, that of haemoglobin, which is also light-sensitive. On the basis of their close resemblance (see Fig. 6) he tentatively identified the unknown in the yeast cells as also an iron porphyrin compound, and the research of the succeeding twenty-five years has done everything to confirm this view.

The single biggest confirmation was Keilin's discovery of the cytochromes in 1926. These are porphyrin-protein substances present in many animal and plant tissues, and in the case of cytochrome C from heart muscle, obtained in the pure state, the porphyrin is known to be very similar to that of haemoglobin. Probably that of the others is also similar. The iron in them is reduced to the ferrous state by the chemical activity of the cell, and reoxidised by Warburg's porphyrin and oxygen, but not by oxygen alone. This is known because the cytochromes can be watched at work in the living cell, in yeast and bacteria, in a bee's wing muscle or a slice of rat's brain; although invisible in the oxidised state, they become coloured much more strongly when reduced, and produce dark absorption bands, visible in a spectroscope attached to the microscope. Thus if carbon monoxide, or cyanide, or sodium azide are added to yeast or brain, the colour bands of reduced cytochrome slowly appear, though the chemicals are not reducing agents, but merely poisons for the cytochrome oxidase or

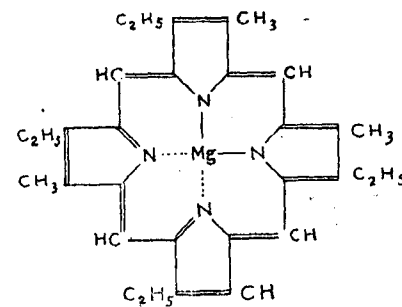


Fig. 7. - Chlorophyll.

Warburg porphyrin. The same bands appear if the cells are merely deprived of all air, and disappear at once when air is readmitted. This does not happen, however, if a poison is added shortly before letting in the air, because then the cytochrome oxidase is out of action, and the cytochromes not being autoxidisable (directly oxygen-sensitive) remain in the reduced state.

So porphyrins have at least two physiological roles: haemoglobin, which merely adds on and gives up oxygen: and cytochrome, which goes through a shuttle process, ferrous to ferric and back again, with oxygen indirectly promoting the oxidation. But this is not the whole of the porphyrin story. The green pigment of plants, called chlorophyll, can be extracted and on chemical analysis proves to be a protein attached to a porphyrin not unlike haem (see Fig. 7), but containing magnesium in place of iron. This pigment is associated with the power plants have to build up their substance with the help of sunlight – photosynthesis – in which carbon dioxide is absorbed by them, and oxygen given off into the atmosphere, but only while the illumination lasts.

Thus porphyrins are found to be well-nigh universal constituents of living matter, and it is no longer so surprising to learn that haemoglobin occurs not only in human blood, and vertebrate blood generally, but also in a few worms, a few insects, and in the intestinal wall and liver of the ordinary snail; and even in the nodules on the roots of plants of the pea, bean and clover family, where bacteria lodge as guests and turn the nitrogen of the air into ammonia and nitrate for the host-plant. More generally, porphyrins are sometimes noticed as colouring matters, in birds' feathers (a copper-porphyrin Turacin, in S. America), or in the skin of the earthworm, or in the shell of hens' eggs, making them brown, their presence being inexplicable except in terms of the process which produced the shell, the feathers, etc.; they are to be regarded as the stains of manufacture.

But to return to the subject of haemoglobin. Strictly

speaking, one should write 'of haemoglobins', because they are a group. Human haemoglobin differs from that of rat and rabbit and dog and camel, not to speak of cod or mackerel, worm or plant root-nodule. Each species has its own variety of haemoglobin, which crystallises in a different way, has a slightly different solubility in water, a slightly different absorption spectrum, a different hunger for oxygen. All the haemoglobins are red, all pick up oxygen when there is enough about, and give it up again when there is none, and prefer carbon monoxide (especially in the dark), and all contain the identical porphyrin haem. The difference lies then in the globin, the protein of the molecule. Each species of animal produces a different kind of globin. Even simple chemical analysis, for instance, of the amount of sulphur in a globin, shows at once the difference between sheep and horse. Yet though there are individual differences between globins, they also have one characteristic in common, the absolutely vital one of combining with haem to form an oxygen-carrying pigment, haemoglobin. Other proteins cannot do this. Either they cannot combine with haem at all, or at the best they form brown compounds (called haemochromogens) which are quite inert. A globin itself, once it is denatured (by heat, organic solvent, etc.) can do no better, its individuality destroyed.

So the curious fact emerges that proteins are molecules with strange modifying powers: Haem alone is inactive, haem with the proteins of milk (casein, lactalbumin), of white of egg (ovalbumin), or of blood plasma, is equally inactive, but haem with a certain group of molecules called globins combines and becomes capable of oxygen combination. If the ferrous iron of haemoglobin is oxidised to ferric, the brown methaemoglobin no longer carries oxygen, or has a physiological function. But if the methaem is attached to a different special protein, as in the enzyme catalase, a very powerful catalyst for the decomposition of hydrogen peroxide is produced. A drop of catalase (methaem-protein) in a beaker of hydrogen peroxide, and there is a whoosh!

as it splits into water and oxygen almost explosively. Methaemoglobin does not behave like this: it is the protein, acting again as a modifier, bringing out the chemical potentialities of the porphyrin structure, which is responsible. It is to the chemistry of proteins, therefore, that we must now turn.

Proteins

The fact that protein can profoundly modify the chemical reactivity of haem, introduces this section. The same discussion would have been forced on us, however, if instead of porphyrins, we had started talking about flavines, or nucleic acids, or metals, or almost any topic of biochemistry. Whichever way we look, proteins are found to be at the core of the matter. They form a major part of what the biologist calls protoplasm, the sticky viscous jelly-like material of cells. Nearly all the pieces of cellular mechanism we can identify are found to be protein – either simply and solely so, or combined with a non-protein part or ‘prosthetic group’ (like haem in haemoglobin or in catalase) which determines the kind of function the molecule as a whole will play. Thus in studying the chemistry of proteins we come closer to grips with the fundamental questions, the nature of muscular contractility, and growth, and inheritance.

Some of the properties of the great majority of globular proteins can be simply demonstrated with raw white of egg. This stuff dissolves in or mixes with water, is coagulated (i.e., denatured) by heating above 50° – 60° C., or by strong acids and alkalis, or by adding alcohol or other organic solvent, but is precipitated from solution in water unchanged (and therefore re-soluble) by the addition of a certain precise quantity of a salt such as ammonium sulphate. This last property is particularly important because it is the basis of an early method of sorting out mixtures of proteins. Some proteins are very easily precipitated by small quantities of salts, while others require high concentrations, and the classification of proteins into albumins and globulins,

at one time regarded as important, depended on this fact, that complete saturation of the solution with ammonium sulphate is necessary to precipitate an albumin, whereas a globulin would come out at only the half-saturation point. To-day, the sensitivity of proteins to varying salt concentrations continues the basis of the standard methods for their separation and purification, but with many added refinements, such as carefully graded denaturations by heating for a few minutes to 50° C. or so (taking advantage of slight differences in heat sensitivity), or precipitations with ice-cold alcohol (which will not denature while it is ice-cold), or adsorptions on alumina or calcium phosphate jellies (to which only some proteins will adhere, under given conditions). The application of these techniques is a process of delicate cookery which can only be learned in the laboratory, partly by personal experience, partly by being apprenticed to one of the experts, who are guided by an uncanny intuition as to the right thing to do, the result of many years’ experience on their part.

Purification is important because it is only by the detailed study of one substance at once that its chemical structure and hence the explanation of its various peculiarities can be successfully attacked. But the purification of proteins presents very great problems. All proteins are very sensitive to physical conditions and easily upset by a little too much heat or acid or alkali, or even to standing about a little too long at room temperature. And all proteins are physically and chemically rather alike. Therefore every method of separation must be mild, and very sensitive, taking advantage of what small differences in properties there are. Even more important, there must be some test to distinguish a given protein when it has been separated. One is little the wiser if one merely separates a single colourless watery jelly into two colourless watery jellies; that sort of game can go on for ever. And yet all proteins are so similar, that to find a test is difficult. Imagine a European unfamiliar with the East asked to pick out the individual

members of one Chinese family from a great crowd of Chinamen; he would be completely at a loss, unless he had some peculiar bit of information to guide him, such as that the members of the family all had webbed feet, and then it would be easy enough.

In the case of a protein the test must be *quantitative*, it must not merely pick out the individual protein, but indicate with suitable measurement how much of that protein is there. There are at present two ways of doing this. One is to take advantage of some property of the prosthetic group. For instance, we might separate haemoglobin from a protein mixture by measuring the intensity of the red colour given by the haem, for this intensity is proportional, within wide limits, to the amount of combined porphyrin, and therefore of globin present. But clearly this would be a useless test for separating different kinds of haemoglobins; it works only within limits, therefore, as a way of following the progress of a purification.

The other method, also not perfect, is to follow an activity, an ability to do something. For instance, the gastric juice (stomach digestive juice) is capable of digesting protein, for instance of dissolving hard boiled white of egg. The agent in the juice which brings this about is called pepsin, and is a protein, and that is about all that is known about it chemically at the start of the investigation. So its function is made a test. A unit as arbitrary as the inch or the pint is invented, something convenient to measure, for instance how long it takes to digest a standard amount of egg white under standard conditions. The unit will then be defined, for instance, as that amount of pepsin present per milligramme of whole dried gastric juice which will digest the egg in one minute. If, then, later on, one milligramme of purified dried juice will digest the same amount of egg in one second, we shall know that the purified dried juice contains sixty times as much pepsin as our standard, i.e., 60 units. With an activity test of this nature we can follow the purification of the digestive agent a long way – but again the method

breaks down, if more than one kind of protein molecule can show a similar activity, and this has actually been claimed for pepsin. It is certainly the case that the crystalline protein from the pig can be fractionated into two different pepsins of somewhat different activities; while salmon pepsin is known to be chemically distinct from that of the pig. It is the discovery about haemoglobins all over again. Pepsins are a group of closely allied proteins, not a single entity. Consequently even the activity test fails at the last, or at least, may do so.

But the activity test is usually good enough to lead to crystals of protein. To the orthodox chemist this implies a high grade of purity. In biochemistry, unfortunately, it is otherwise. Proteins are large and complex molecules, and all rather alike. Consequently the active crystals may be ten per cent of one protein and ninety per cent of another, and there is no knowing which of the two is the one we want. Sometimes too the crystals may be a more complicated mixture than this. All the same, a great deal of important work can be done with materials of this grade of purity, which may be the end result of a thousandfold purification. Beyond this point, physical measurements become important, two tests of homogeneity which are of theoretical as well as practical importance.

The early twentieth century biologists thought of protoplasm as colloidal, and this concept was taken over by the protein chemists. It expressed the gluey, viscous nature of the substances, which were not exactly solids, nor yet liquids, in fact sometimes described as the fourth state of matter. A colloidal substance was imagined to be a mass of particles of many shapes and sizes, and expected to have all sorts of odd properties not displayed by mere conventional solids and liquids. In fact the term 'colloid' became more and more unprecise and eventually a barrier to progress. Any unusual phenomenon had only to be labelled 'colloidal' for it to be assumed that all was explained and understood about it; in reality the word was a cloak

of complete ignorance. There have been, and still are, other magic words of the same character, which delay research for a time, 'permeability' for instance, and 'protoplasmic poison', as we shall see later.

As long as proteins were imagined as colloidal, any attempt to study them in detailed physico-chemical terms seemed hopeless. Svedberg's invention of the ultracentrifuge marks the beginning of a change of view, because with this machine he was able to show a regularity in the properties of proteins, where diversity had been assumed the rule. The idea behind it was simple. In the ordinary centrifuge solids in suspension (such as red cells in blood) are whirled round and forced to the periphery or bottom of the container, the heavier ones being driven down the faster. What Svedberg did was merely to increase the centrifugal force by making the machine run very much faster (a skilled problem in metallurgy and engineering), and also to arrange a sort of camera which could record from time to time how the suspended particles were settling down. With the final versions of the ultra-centrifuge he could generate sufficient centrifugal force to begin to spin down glucose molecules in water and obtain them in higher concentration at one end of the centrifuge tube than at the other. This was a terrific feat. Protein molecules are very much larger than those of the sugar, and consequently he had ample power for studying particles of colloidal dimensions. His machine enabled him to examine them in two ways. By making various measurements he could determine the actual weight of a single suspended particle (e.g., the glucose molecule in the extreme case just cited); and he could tell whether all the particles had the same weight, or whether there was a mixture of differing weights – in other words, he could test for homogeneity.

When he applied his machine to the study of haemoglobins, he was able to show that a good preparation from a single sort of blood contained particles (molecules) all alike as regards the rate at which they sedimented; in other

words, a single sort of haemoglobin, say that from sheep blood, was a definite physical entity, homogeneous, and constant in its behaviour whenever prepared: its molecular weight could be measured, making certain reasonable assumptions, and was about 68,000. Further, the other haemoglobins also mostly had a weight about 68,000 times that of the hydrogen atom, except for a few which were only a quarter of this (i.e., 17,000). Extending the observations to other proteins, he confirmed this experience of homogeneity of particle size for each sort of protein, and showed further that values around 70,000, and around 35,000 (and also 140,000) were quite common for molecular weights. This probably has no theoretical significance, and even the values have more recently been questioned as experience has brought greater accuracy. But at the time, they served to emphasise the idea of regularity in the physics of proteins. A true colloid (a gold sol, for instance) was markedly heterogeneous with many different-sized particles. A protein was therefore not a colloid, after all, but a macromolecule. Since starches and many modern plastics are also in this category, the physical chemistry of macromolecules has since grown to be almost a speciality on its own, and we must now say a little about the other important technique in this field.

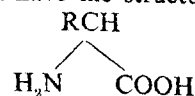
Electrophoresis, for which Svedberg's colleague Tiselius developed the necessary apparatus, depends on the elementary fact that most macromolecules are electrically charged, either positively or negatively, and will move to one electrode or the other when an electric current flows through their solution. In electroplating, the chromium or nickel ions (charged atoms) in the solution move swiftly to the electrode when the current is switched on, and deposit on it, and the deposition is the part we find interesting. The bulky macromolecules on the other hand are slow movers, their speed depending on the degree to which they are charged (and this varies according to the acidity of the solution) and the interesting thing about them is this

speed, and not the final deposition. Thus electrophoresis is another test of homogeneity, not this time for likeness of weight, but for likeness of electric charge under standard conditions. Proteins of different charge are separated, since they move at different rates and can sometimes be removed separately from the machine, so that experiments can be done on the different fractions. Electrophoresis is thus also a method of protein purification.*

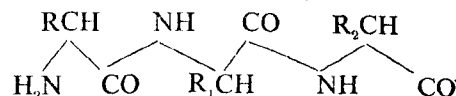
The outcome again is that study of the best protein preparations (e.g., of haemoglobin) displays homogeneity. The argument grows therefore: the macromolecules of a single protein are all built up in the same pattern of constituent atoms. This view is supported by X-ray study of protein crystals, which give well-defined diffraction patterns, indicating fundamental regularity throughout the crystal.

These methods thus give rise to the idea of the protein molecule, as distinct from the colloidal particle. At the same time they offer purity tests for preparations of single proteins, and they lead to estimates of molecular size.

But why are these molecules electrically charged? To understand this, it is necessary to indicate briefly the results of analytical chemistry. Boiling with hydrochloric acid eventually produces from any protein a mixture of twenty-odd different amino-acids in various proportions. These amino-acids all have the structure



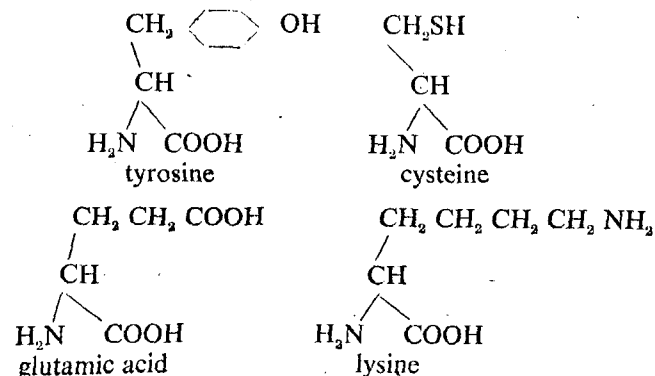
where R is different in each one, and they are bound together in a chain, called a peptide, thus



In the macromolecule the chain is characteristically coiled

*More recently, the ultracentrifuge has also been used for preparations on a micro-scale.

up, presumably in a ball. The R groups or side chains are extremely important because at their tips they carry reactive atoms or groupings. Thus R may be a phenol (as in tyrosine), or end in -SH (cysteine), or in the acidic -COOH (glutamic acid) or basic -NH₂ (lysine):



— there are, of course, other possibilities too. Now there are strong reasons for thinking that under normal conditions most of the glutamic acid and lysine R side chains are ionised, and really end in -COO⁻ (with a negative charge) and -NH₃⁺ (with a positive). Tyrosine and cysteine are also acidic and capable of forming salts, but they are weak acids and only behave as such in alkaline solutions.

So we must picture a protein macromolecule as covered with a forest of charged side chains, like a burr, or a rolled-up hedgehog, all spikes, a mass of upthrust negative and positive charges and reactive chemical groupings. And we know from various lines of evidence, including those already discussed, that these charges (or polar groups, as they may be called) are not haphazard in space, but arranged in significant order, in a *pattern*. Only on this basis can we explain why globins will combine with haem to make haemoglobins, whereas other proteins at the best form only a haemochromogen. Only globins share the polar group pattern for combination: denature a globin,

disorder the groups, and it can henceforth only make a haemochromogen like other proteins. Perhaps the protein macromolecule can be pictured as an assortment of electrical plugs and sockets arranged side by side over the surface of a sphere with their fitting ends outwards: its combination with other molecules, including other proteins, depends on the other molecule fitting the right combination of sockets.

Thus one of the great problems facing biochemistry to-day is how to discover the precise pattern of groups each protein possesses. A molecule like pepsin contains 3,000 atoms or so, and haemoglobin nearly twice as many. Ideally the chemist would like to place every single atom in the three-dimensional protein network. But this is a very distant goal. The immediate task is the development of a stereochemistry which shall enable him to, map out the 'surface' groups, on which the specific properties of each protein molecule depend. For the pattern of these reactive groupings in the amino and side-chains is the most significant aspect of protein chemistry for biology.

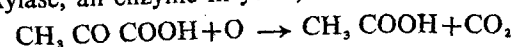
Enzymes

An enzyme, or ferment, is a protein which possesses the ability to promote some chemical reaction. Pepsin, for example, digests white of egg, which is to say that it hydrolyses or breaks up the protein into smaller soluble components – amino-acids, or short peptide chains. It does this in water at the temperature of body heat (say 37° C.). To do a similar thing in the laboratory requires some hours boiling with strong hydrochloric acid, much more drastic procedure.

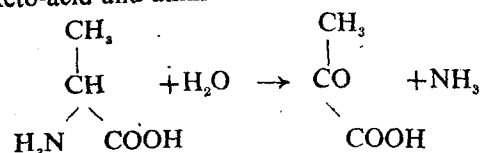
All the chemical changes living organisms perform are made through the agency of these active proteins, which conduct the reactions by easy stages, single steps each with its own enzyme. When yeast makes alcohol from glucose, for instance, the first step is to add on a phosphate group to the glucose, i.e., to phosphorylate it, the next step is to

transfer the phosphate to a different position in the glucose, the third step is to add on a second phosphate group, the fourth is to split the hexose (six-carbon-atom) sugar diphosphate into two molecules of triose (3-carbon) phosphate – and so on, each step catalysed by a different enzyme. The introduction of polar phosphate groups alters the strength with which the different carbon atoms of the sugar cling together, and eventually results in a smooth split between them at room temperature: without these groups, ordinary laboratory chemistry is forced to use drastic and complicated means to gain the same end.

In the last ten years a great many enzymes have been purified from minced muscle, or kidney, or smashed-up yeast cells, and so on, by the methods outlined in the preceding section, and their chemistry has been much studied, ultimately to discover what peculiarity of structure enables them to bring about chemical reactions. This goal is still distant, but certain important facts have come to light. One is that many enzymes will not function except in the presence of coenzymes, chemical assistants which in many cases turn out to be vitamins. Thus the conversion of pyruvic acid to acetic acid and carbon dioxide by carboxylase, an enzyme in yeast,

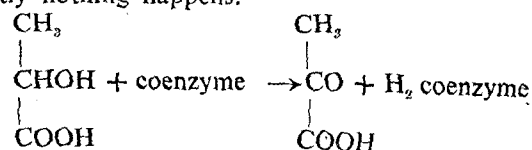


in the presence of oxygen only takes place if co-carboxylase, vitamin B₁ pyrophosphate is also present. Amino-acid oxidase of mammalian kidneys converts an α -amino-acid to a keto-acid and ammonia:



and when purified, this enzyme is found to contain a prosthetic group, flavine adenine dinucleotide, which is built up from riboflavin (vitamin B₂). Lactic dehydrogenase

of muscle converts lactic acid to pyruvic acid, an oxidation which involves the removal of two hydrogen atoms (i.e., two hydrogen ions and two electrons) and requires their addition to the coenzyme diphosphopyridine nucleotide (coenzyme I) which is built up from nicotinamide, another B vitamin, the cure of pellagra. Without the coenzyme apparently nothing happens.

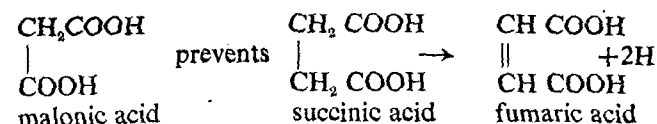


Sometimes the coenzyme or prosthetic group is merely a metal, a manganese or magnesium ion, copper or zinc, etc. The important point is that without these metals the enzyme does not work. Thus the fact that animals and plants become diseased in the absence of traces of metals in their diets or soils begins to find an explanation. Like the vitamins, these metals are required as a part of normal nutrition to put the finishing touches to some piece of cell machinery, to complete some enzyme ready for function.

Another much-studied matter is enzyme specificity. Clearly amino-acid oxidase will not convert lactic acid to pyruvic. But will it act on any other amino-acid than the particular one (alanine) instanced above? The answer is that it will act on quite a number but not all: lysine for example is untouched. But it shows definite tastes even with the amino-acids it will attack. They must all be *alpha*, i.e., with the amino-group attached close to the carboxyl; and they must all have the right spatial atomic layout. Two versions of alanine are possible, one the mirror-image of the other, and differing in the way it affects polarised light. One enzyme only attacks the *d*-form, whereas all proteins consist of *l*-amino-acids. Another enzyme is also known which, contrariwise, attacks the *l*- but not the *d*-.

The importance of stereochemistry in biochemical processes is thus again emphasised. It is involved also in the

existence of competitive inhibition, when a substance similar in spatial layout to the true substrate of an enzyme may combine with that enzyme and so prevent substrate-combination, and hence chemical reaction. Thus succinic dehydrogenase converts succinic acid to fumaric acid, and has no action whatsoever on malonic acid. Nevertheless, if malonate is added to the solution of enzyme and its substrate, the rate at which substrate becomes fumaric drops proportionately; and this is explained as enzyme combining with malonate at the same point on its surface as is normally reserved for succinic, which is consequently elbowed out.

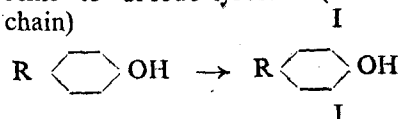


The reactive groups, some electrically charged, positively or negatively, on the surface of the enzyme molecule are thus involved in its activity. Alter the pattern by denaturation, and the enzyme no longer works. Cover up a corner of the pattern with a competitive inhibitor, or some other reagent (see below) and it ceases to function. Clearly the enzyme does its work in virtue of its three-dimensional patterned structure, which enables it to combine with certain substances, which are then much more liable to change than they were before. Perhaps the concentration of electrical charges on the protein surface 'strains' and 'distorts' the substrate. At present no-one knows, because the nature of its pattern as yet eludes us.

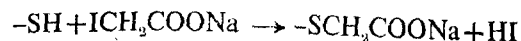
The use of poisons

Some of the reactive groups involved in enzyme reactions can be tested by means of special reagents which combine with a given group without spoiling or denaturing the protein as a whole. For example, the amino-groups of pepsin can be acetylated ($-\text{NH}_2$ becomes $-\text{NH}.\text{COCH}_3$)

without affecting the enzyme's activity. But if pepsin is carefully treated with iodine under standard conditions, so as to convert tyrosine to di-iodo-tyrosine (iodination of the phenol side chain)



it becomes completely inactive. The same is true of the hormone, insulin, a protein used for treating diabetics; by further careful manipulation the iodine can also be removed from the tyrosine, and the hormone then recovers its activity. Many enzymes, of which succinic dehydrogenase is one, depend on the integrity of their -SH side chains for full activity; treatment with iodo-acetate poisons them at once



by adding the acetate to the sulphur and so prolonging the chain. The antiseptic actions of mercury and silver are due in the first place to this protein reaction, compounds such as -SHg and -SAg being irreversibly formed, so that to call them 'protoplasmic poisons' though correct is unnecessarily vague.

Very few substances are found on detailed enquiry to be *general* poisons: research shows them to take part in specific reactions with just one or two enzymes. Mustard gas, for instance, combines with and thereby inactivates hexokinase, the protein which converts glucose to glucose 1-phosphate, and leaves a myriad other enzymes unharmed and unaffected. In fact, the explanation of the actions of all drugs is to be sought in interference with the functioning of single enzymes; and in seeking new synthetic drugs, the aim is to invent molecules which will interfere, as a competitive inhibitor or a pseudo-prosthetic group, for example,* or best of all by paralleling the enzyme 'surface' in

* See *Antivitamins* by Professor Yudkin in *Science News VII*, 56.

COMETS

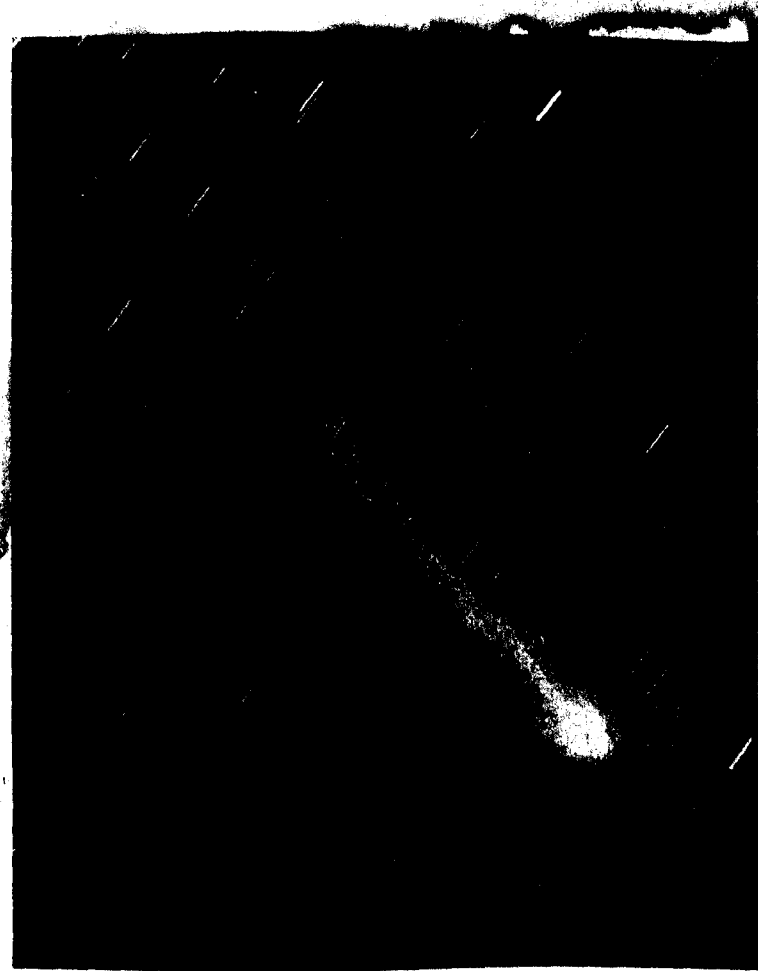


Plate 1. Morehouse's Comet, photographed at Greenwich, 1908 Sept. 29. This comet approached the sun to within 88 million miles and is included under the 'long-period' comets. Notice the different sheaves in the tail. The transparency of the tail is shown by the visibility of star images through it (trails in the photograph).—By courtesy of the Astronomer Royal.



Plate 2. Delavan's Comet, photographed at Greenwich, 1914 Sept. 26. This comet was discovered nearly a year before it reached perihelion. It would have been a spectacular object if it had made a closer approach to the sun; its nearest distance to the sun was 102 million miles and at this distance comets do not always become very spectacular. Its orbit was slightly hyperbolic, an explanation of which is given in the text.—By courtesy of the Astronomer Royal.

LOW TEMPERATURE PHYSICS



Plate 3 This is part of the auxiliary equipment required in modern low-temperature research. Compressors at the Clarendon Laboratory, Oxford.

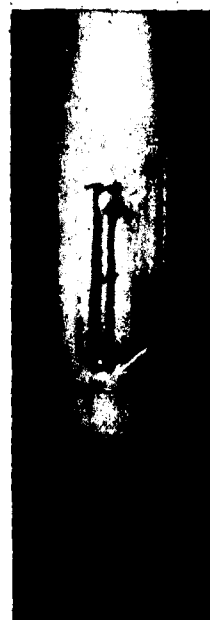


Plate 4. Film Transport in Liquid Helium II. A drop of helium can be seen at the bottom of the beaker.

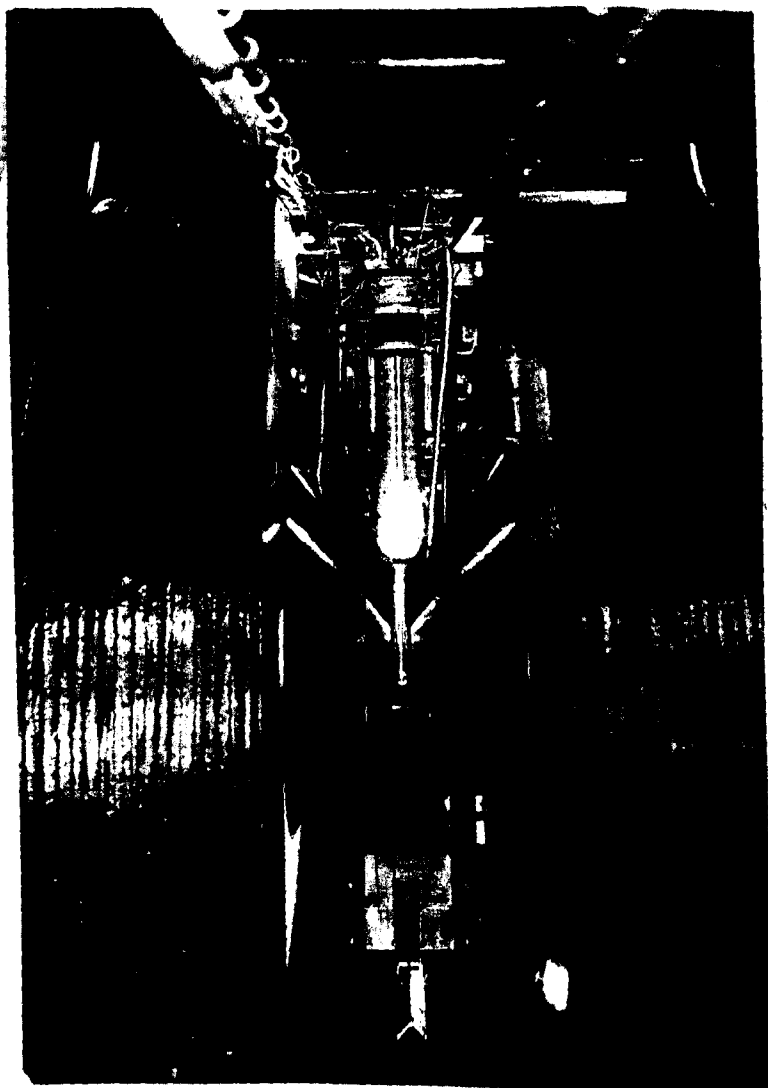


Plate 5. Demagnetisation apparatus. The narrow part of the vacuum flask seen in the centre between the two poles of the powerful electromagnet contains the paramagnetic salt at about 1°K, ready for cooling by demagnetisation.

ARTIFICIAL KIDNEY

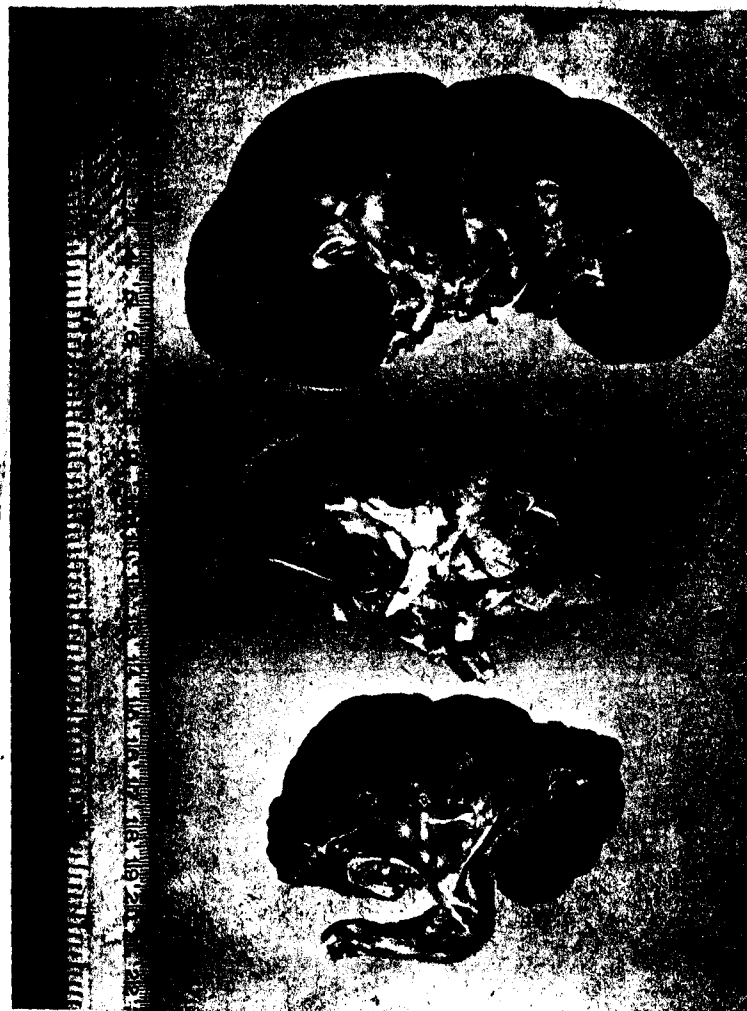


Plate 6. The kidney in the centre is a normal human kidney. The kidney at the bottom is the shrunken kidney of a patient with nephritis. This is an irrecoverable disease, and treatment with the artificial kidney may prolong, but cannot save, life. The kidney on the top is the swollen kidney which results from a mismatched blood transfusion. If the patient can be kept alive with the artificial kidney, this lesion will recover. (These kidneys have been sliced in half lengthways and these are their cut surfaces).

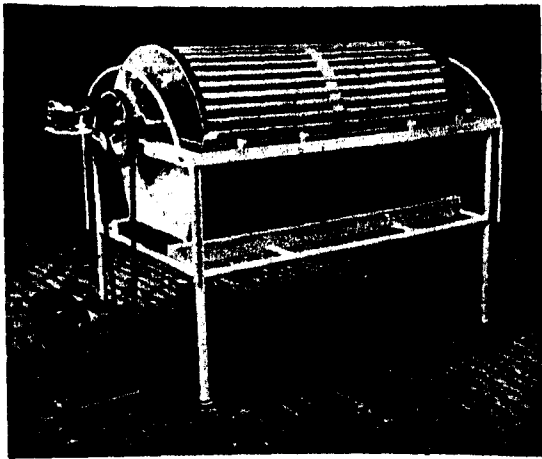


Plate 7. The artificial kidney. Cellophane sausage skin is wound round the wooden cylinder which rotates in a saline bath.

HOW OLD IS THE EARTH?



Plate 8. Pleochroic haloes surrounding a speck of uranium ore; the significance of this is explained in the article. By courtesy of the Oxford University Scientific Club, reproduced from *The Geological Age of the Earth* by Dr John Joly, O.U.P., 1926.

X-RAY STUDIES IN KIDNEY CIRCULATION

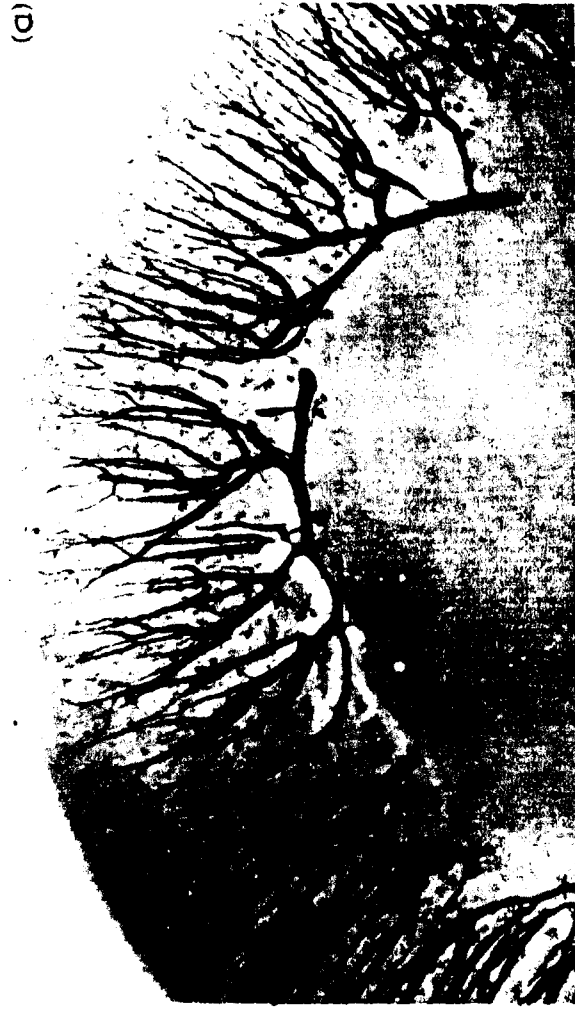


Plate 9. After an old print of the Radcliffe Observatory, since 1935 the Nuffield Institute for Medical Research, Oxford.



Plate 10. Photographs of half of the right kidney (a) and (c) and of half of the left kidney (b) and (d) of an anaesthetized rabbit, after injections of 5 per cent methylene blue solution had been made into the respective renal arteries. These show the reflex effects upon the left kidney of stimulation of the central end of the divided left sciatic nerve (for further description see text). Note the considerable staining of the outer surface of the right kidney (a) as compared with the virtual absence of such staining in the case of the left kidney (b). The cut surface of the right kidney (c) shows diffuse staining throughout cortex and medulla, but the cut surface of the left kidney (d) shows a staining confined to the deepest layer of the cortex and to the subcortical zone of the medulla. (Reproduced by courtesy of Blackwell Scientific Publications, Ltd., from *Studies of the Renal Circulation*).

Plate 11
(a)



(a)

Radiomicrographs of sections of right (a) and left (b) kidneys of an anaesthetized rabbit, showing the effect of stimulation of the nerves running along the artery to the left kidney. Radiopaque medium (colloidal bismuth) was injected during life via the aorta. In the unstimulated kidney (a) the medium fills the interlobular arteries to their terminations and also great numbers of cortical glomeruli; a few basal glomeruli are seen, but the vasa recta are poorly filled. In the stimulated kidney (b) the outer segments of the interlobular arteries are strongly constricted, but the proximal segments are still patent, and many basal glomeruli are well filled: so are the vasa recta in the medulla. The diversion of blood from cortex to medulla is also shown by the relative over-all pallor of the cortex in (b) as compared with the normal cortical density in (a). (Reproduced by courtesy of Blackwell Scientific Publications, Ltd. from *Studies of the Renal Circulation*).

Plate 12
(b)



ULTRA-MICROSCOPIC SURFACE TOPOGRAPHY

Plates 14/31 illustrating this subject are best understood by direct reference to the text of the article.

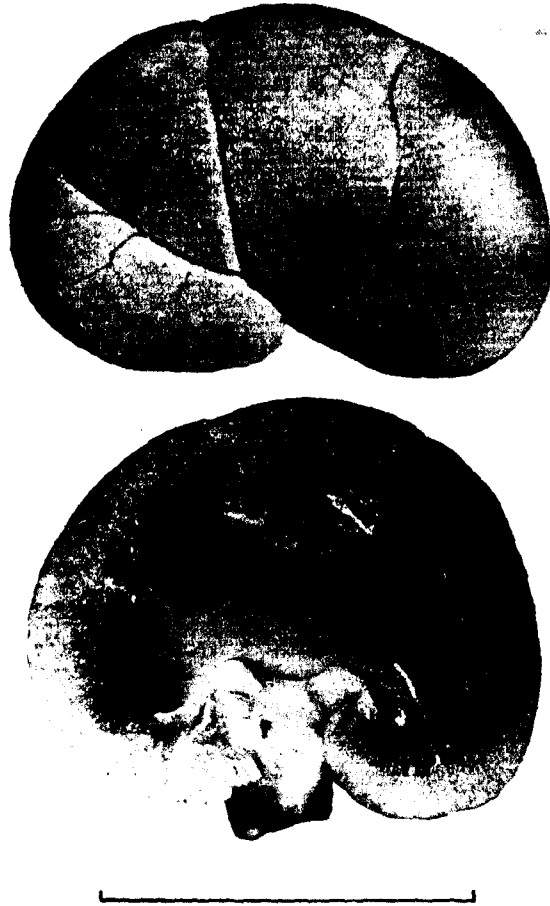


Plate 13. Photographs of a kidney of an animal which was anaesthetized and thereafter breathed air containing carbon monoxide. The first photograph shows the blanched external surface; the second (which depicts one half of the organ after it had been divided lengthwise) shows a bloodless cortex contrasting with a medulla which is rich in blood, particularly in its outer part. In other words, severe carbon monoxide poisoning has brought into operation the renal shunt. (Inch scale.)

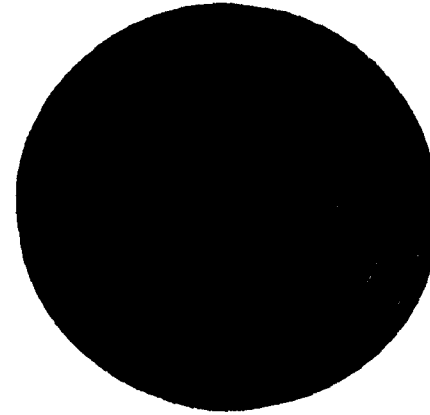


Plate 14. Multiple beam rings from interference with parallel plates

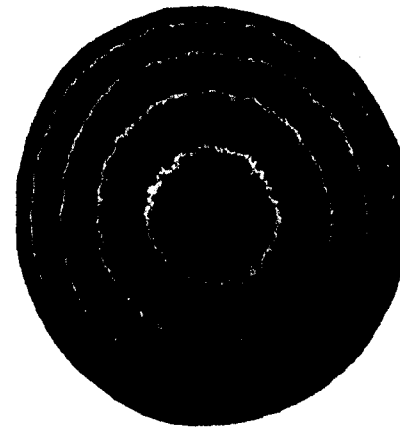


Plate 15. Interference fringes from a spherical lens in contact with a glass plate.

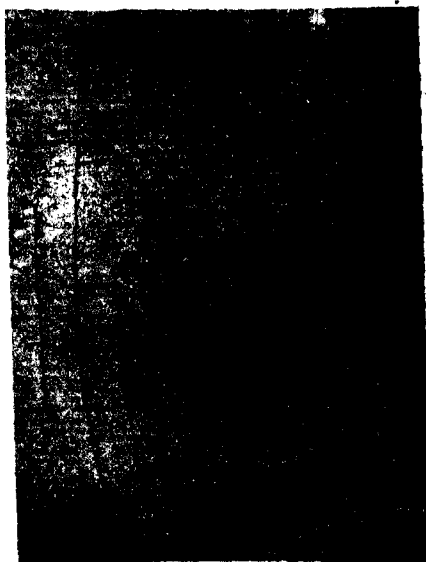
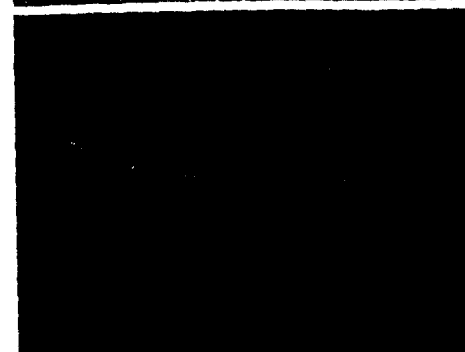
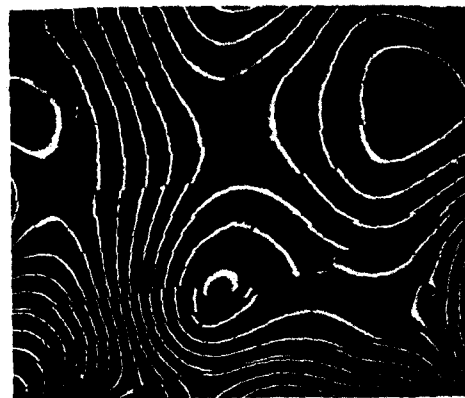


Plate 16. Part of the natural unpolished face of a quartz crystal--about 0.3 inch long.



Plate 17. The same surface as plate 16, by a different method. Surface ruts only, 20 molecules deep are seen.



Plates 18 and 19. Surface of a flake of mica, showing cleavage steps, multiples of 20 A.U.



Plate 20. Surface of a flake of selenite.

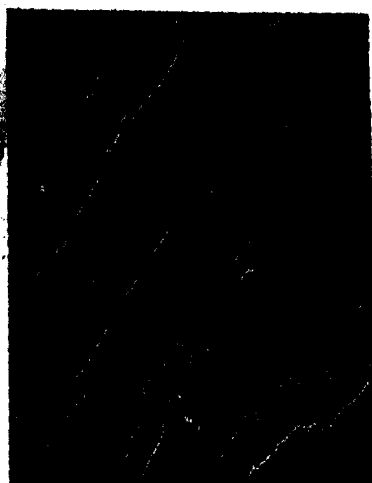


Plate 21. The face of a natural unpolished diamond. Sharp line fringes.

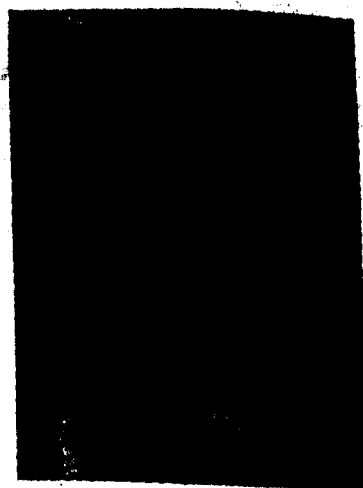


Plate 22. The same diamond surface. Broad high dispersion fringes.



Plate 23. Crossed fringes. Superposition of systems in plates 21 and 22.



Plate 24. A small crater made in polished tungsten by a single electric spark: the depression is about $1/25,000$ inch deep and $1/250$ inch across.



Plates 25 and 26. Two further craters of comparable dimension produced by sparks.

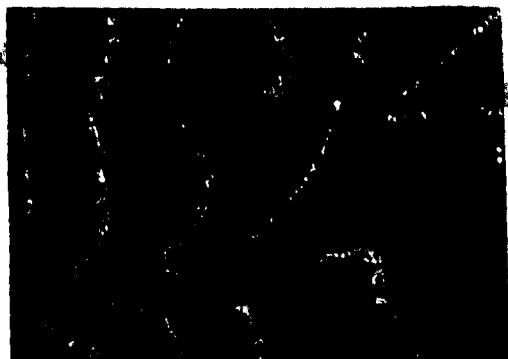


Plate 27. A scratch $1/100,000$ inch deep on buffed steel. Shown by the replica technique.

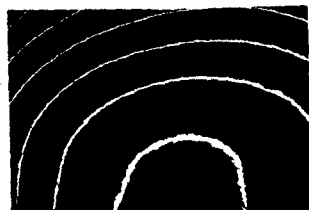


Plate 28. Fringes for the surface of a quartz crystal. At rest.



Plate 29. The same, oscillating under the influence of an alternating electric field.



Plate 30. Superposition of fringes during oscillation on those observed while the crystal is at rest shows there is no change in separation between the quartz face and the glass surface on which it rests.

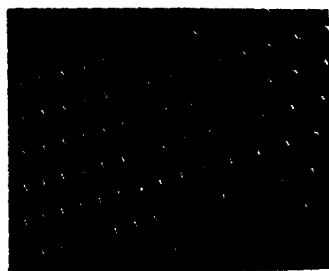


Plate 31. A crystal vibrating in high harmonics.

some way, and blocking it, for this clearly gives the most specific of all specifics.

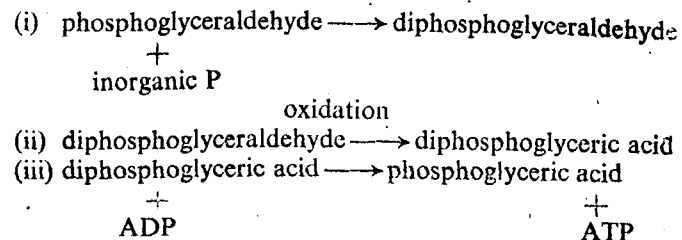
This idea of discrete poisons affecting only one step in a series, blocking only one enzyme while the others continue functioning normally, has been turned to great advantage in studying how the cell carries on its metabolism. Yeast converts glucose to alcohol: the intermediate steps might never have been discovered if it had not been found that addition of iodoacetate to the fermentation causes the accumulation of hexosediphosphate in place of alcohol, whereas the addition of fluoride permits the disappearance of hexosediphosphate as soon as formed, but blocks a later step, phosphoglyceric acid (a 3-carbon compound) now appearing instead of alcohol.

Another approach to the chemical reactions going on in whole living cells, whole organs, even whole animals, is provided by so-called detoxications. Many of the strange substances made by organic chemists, or produced industrially, when fed to animals are excreted in the urine in an altered form. Thus benzoic acid, which is poisonous, appears in the urine combined with the amino-acid glycine or with glutamic acid, and naphthalene (e.g., moth balls) appears in combination with cysteine; this has been interpreted teleologically as the animal doing its best to render poisonous substances harmless. But it is better looked on as an interference or sidetracking of normal biological chemistry. Thus we know that the animal liver is very busy synthesising glutathione, a peptide of cysteine, glutamic acid, and glycine. (Why it carries out this synthesis is a mystery.) A whole series of enzymes must be working together to make this substance. Is it surprising, then, if some of them are unable to help themselves and turn from their normal job of fixing glutamic acid to cysteeyl-glycine, and attach it to benzoic acid instead? Offering the tissues organic chemicals becomes a way of exploring their chemical capabilities, and is proving a valuable extension of the early use of poisons as dissectors of metabolism.

Living machinery

We must now turn from the analytical side of the subject, and see what the study of purified enzymes eventually leads back to. Three topics demand attention, phosphorylation, respiration, photosynthesis. They are all aspects of the energetics of life. Examination of them will show how a living organism exists, like a small co-ordinated chemical factory, going its way through the storms, the chaos, the resistances of its inorganic environment, and manufacturing more and more of itself. In one word, the answer is phosphorylation.

All cells contain adenosine triphosphate while they are alive. This substance is like a bank, an accumulator, it is a reserve of stored chemical energy. When glucose is converted into glucose 1-phosphate the reaction is promoted by an enzyme, it is true, but it also needs the putting in of energy, and merely mixing glucose and inorganic phosphate will do nothing. But if adenosine triphosphate is offered in place of the inorganic ion then all goes well, for this stuff, colloquially ATP, contains phosphate and energy in one: it donates one of its phosphate groups, and becomes adenosine diphosphate (ADP). If necessary it can yield another and become the monophosphate, adenylic acid (AA), but it has then reached rock bottom, its chemical energy reserve is exhausted and must be replenished. This is done by coupling its re-phosphorylation with an oxidation. In other words, inorganic phosphate is (indirectly) empowered to couple with AA or ADP and reform ATP. For example, the breakdown of glucose in a muscle without oxygen follows the lines previously described of formation of hexosediphosphate and splitting into two triosephosphates, one of which is phosphoglyceraldehyde, which can be oxidised by enzymes to phosphoglyceric acid. But this oxidation is associated with the uptake of inorganic phosphate. In detail, the process runs:



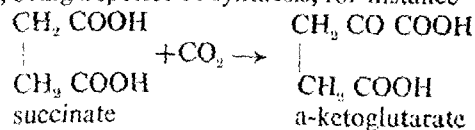
In muscle or kidney or other tissue respiring in oxygen, the fact that the respiration is coupled with the resynthesis of ATP can be amply demonstrated, but the details of the mechanism remain to be worked out. It is only in the details of the anaerobic Cori-Meyerhof cycle just quoted that we can as yet see the basic chemical mechanism at work: even here the mechanism of the reaction in atomic and electronic terms awaits full exploration.

But the essential is established. ATP is the fuel for muscular activity. The contractile protein of muscle, myosin, changes its shape (probably the molecules in a myosin fibril reorientate themselves to one another, rather than an individual molecule actually contracting) in the presence of ATP, and at the same time breaks it down to ADP. Thus movement, a sign of life, can be evoked in a test-tube, with purely chemical substances. ATP is also the energy behind synthesis. The polymerisation of glucose to form starch (which consists of many glucoses bound together) requires energy, and is accomplished in living plant and animal tissues, by making glucose 1-phosphate which polymerises in the presence of the enzyme phosphorylase, forming starch, and liberating inorganic phosphate. When the time comes to utilise the starch, the process reverses: inorganic phosphate and starch yield a flood of glucose 1-phosphate again, capable if necessary of rephosphorylating AA to ATP, or doing other valuable service. Biochemists to-day believe that all synthetic processes involve ATP as a source of the necessary chemical energy. But so far it has been demonstrated in only a few reactions, though the

list grows monthly. There seems little doubt that phosphorylation is part of the basis of life, from the bacteria to man.

In the reaction sequence briefly described above, oxidation (i.e., the removal of two hydrogen atoms) provided the power to run the phosphorylation machine. Now these two hydrogens will ultimately meet and combine with an oxygen atom, and so fade away as H_2O – water. But the cell in its respiration is very economical. It transfers the hydrogens in a series of simple steps (through diphosphopyridine nucleotide, flavoprotein, cytochrome) to cytochrome oxidase where it finally meets its oxygen, and each step is linked in some way with more rephosphorylation, more ATP production. Strictly speaking, it is not hydrogen atoms which are transferred, but electrons (and hydrogen ion which wander about in the solution), so that living matter runs off electricity, or the flow of electrons, in the final analysis.

Finally, a word about photosynthesis. All animals depend ultimately on plants for their existence. Man, for instance, eats plant starches in potatoes and bread, and with the inorganic phosphate in his body or in his drinking water or his other food, makes glucose 1-phosphate and hence charges up his ATP. But the potato has only sunlight to help it convert its ADP to ATP, and so to store chemical energy. Now as every botanist knows, photosynthesis means the uptake of carbon dioxide and the liberation of oxygen in sunlight by the green leaf, which contains the porphyrin pigment chlorophyll. Biochemists have discovered that taking up carbon dioxide, even in complete darkness, is a widespread habit amongst living things. Bacteria do it, pigeon liver does it, even plants do it in the dark, as the use of labelled carbon has made abundantly clear. This dark uptake requires ATP (of course), being a species of synthesis; for instance



So the uptake of carbon dioxide is unlikely to be directly the

sunlight-promoted reaction. We know porphyrins only as agents of oxygen metabolism, and chlorophyll is no exception. Hill has in fact been able to break the metabolism of the leaf in half and show that the production of oxygen in sunlight will take place independently of the rest of the processes. Presumably this oxygen comes from water, and the hydrogens also produced in that case must undergo gradual stepwise reoxidation, coupled with rephosphorylations. But this is a research frontier: We await fresh intelligence.

Conclusion

Modern biochemistry thus has a considerable amount to say about the chemical composition of living matter, and the way an organism gets its energy and expends it. Its daily life is seen to be already largely describable in mechanistic terms, in physics and chemistry, and every month of research is a triumphant march further along this road. The biochemist cannot be a vitalist, oppressed by the mystery of life and doubting the possibility of its mechanistic interpretation, otherwise he will fail to prosecute his investigations with full fervour. Mechanism is sweeping the board as he sees to what extent already the discovery of order in living processes has led to understanding. All living things from the lowest to the highest in the evolutionary scale are made of similar materials and work in similar chemical ways, using the same vitamins and prosthetic groups, and producing similar kinds of proteins. Who can doubt that just as pattern, spatial order, determines the behaviour of the individual protein molecule, so order also is present in the individual cell determining its behaviour, with enzymes positioned, not haphazard in a jumbled protoplasm? And this protoplasmic order must spring from a regularity and pattern in the seed and the egg, so handed on from individual to individual, from generation to generation.

The year 1935 was the red-letter year in which W. M. Stanley in America, and F. P. Bawden and N. W. Pirie in Britain, first showed that certain plant viruses are nucleo-

protein molecules, and can be crystallised. Yet a few molecules introduced into a healthy plant will induce the production there of millions more like themselves. In other words, the plant can be induced to manufacture something totally foreign to it, but identical with the 'parasite' molecules, and which are thus reproducing themselves. It is facts of this kind which make nonsense of terms like 'living' and 'lifeless'. Before 1935, a virus was a living organism, now it is a self-reproducing molecule.

The language of everyday speech is unsuited to scientific discussion, since it is based on a different, narrower, range of experience. We talk of a man dying: but at what stage does death occur? Is it when his breathing becomes shallower and stops, refusing to begin again? Or is it when his heart halts through lack of oxygen? Or is it when his brain cells begin to get disorganised through failure of the circulation? Even then, this decay is a process of many stages, for hours after 'death', brain cells are still capable of carrying out some chemical changes, and if the brain cells were removed early on, they could be persuaded into immortality by careful tissue culture. But gradually the number of their activities shrinks as the protoplasmic pattern collapses.

The virus discovery brings us to the newest, most intriguing front of biochemical research to-day, the nature and functioning of nucleic acids and nucleoproteins. Like viruses, the chromosomes and genes, the immortal hereditary molecular patterns, are composed of nucleoprotein, and are self-producing. It is they which determine the growth of an egg into a worm or a man, exerting the ultimate control over enzyme functioning. If we can elucidate their physics and chemistry, we shall have the key to a scientific understanding and control of life.

READING LIST

- Harrow, *Textbook of Biochemistry* (4th Ed., Saunders, 1946).
 Baldwin, E., *Dynamic Aspects of Biochemistry* (Camb. Univ. Press, 1947).
 Needham, J., and Green, D. E., *Perspectives in Biochemistry* (Camb. Univ. Press, 1937).

Physics of Low Temperatures

DR K. MENDELSSOHN

Introduction

MAN lives in a peculiar and, as compared with the universe generally, in a somewhat unrepresentative place. The crust of the earth enveloped by its atmosphere forms a kind of greenhouse, the temperature of which is kept tolerably constant at a level of about 300 degrees centigrade above absolute zero. That is very much cooler than the average star, whose central temperature is of the order of several million degrees, and very much warmer than the space between the stars, which is only a few degrees above absolute zero. On the other hand, the temperature region of our greenhouse is of special interest, not merely because it affects our daily life but because it is unique in the variety of forms in which it permits matter to exist. Just in this particular temperature region, the elementary building bricks of matter will combine in innumerable different ways to form the immense multitude of structures which surround us. At lower temperatures this combination will not take place, while at higher temperatures the complex structures will again be broken down into their simple constituents. Life, embracing the most complex of all structures, thus appears to be limited to this narrow range of temperatures.

The complexity of aspect which the terrestrial scene thus presents, while it is an indispensable adjunct to life, does not necessarily make it easy to understand the fundamental facts. In his long endeavour to unravel the nature of things, man has found it profitable to subject matter to different temperatures and to observe the effect. This line of investigation began with the mastery of fire, which in due course yielded the breakdown of all structures short of the atomic

nucleus, and has in our time reached a final crescendo in the release of atomic energy. The temperature generated in the explosion of an atomic bomb approaches that of a star, and the high-speed particles employed by the nuclear physicist also correspond in energy terms to these high temperatures. Thus, extremely high temperatures are used to break up the ultimate structures of matter. But what about the other end of the temperature scale?

Absolute zero

About 150 years ago scientists began to realise that the temperature scale which they had established has a lower limit. While the scale appears to stretch into infinity as far as high temperatures are concerned, the lower end is located at 273 degrees centigrade below the temperature of melting ice. This is the so-called 'absolute zero', and for many scientific purposes, particularly for those of low temperature research, it is convenient to make absolute zero rather than the ice point the origin of the temperature scale. Instead of talking about -270°C. , we call this temperature 3°K. and instead of 0°C. , we say 273°K. This is the absolute or Kelvin scale of temperatures, so named after Lord Kelvin, its originator.

Once it was realised that there exists an ultimate lowest temperature, attempts were made to reach this rock bottom of the physical world. However, not much progress was made at first. Somehow, it turned out much easier to make high than to make low temperatures, and the reason for this curious fact was not well understood for a long time. Then, on Christmas Eve of 1877, before the French Academy of Sciences were read two communications which mark the first great step in the approach to absolute zero. In one of these letters, which was dated December 2nd, Louis Cailletet working at Châtillon-sur-Seine announced the liquefaction of oxygen. The second communication, a telegram, showed that the same result had been achieved on December 22nd by Raoul Pictet in Geneva. Cailletet and

Pictet had used quite different methods, but as so often in the history of science, the same step forward, which was due to come, had been made at the same time independently at two places. By the liquefaction of oxygen a temperature of only 90°K. had been reached.

It is significant that the chief aim of these early workers was not to make low temperatures but to force oxygen gas into the liquid state. The low temperature achieved was at first regarded as an incidental result, of which relatively little notice was taken. The unique range of terrestrial temperatures, to which we have referred above, allows us to observe without difficulty the change of many substances from one state of aggregation to another; water being, of course, the most striking example. To see whether *all* substances can be obtained in the gaseous, the liquid and the solid state was one of the first important questions experimental science had to settle. A few substances, oxygen, nitrogen, and hydrogen, seemed to resist all efforts at liquefaction, and many physicists began to believe that these were 'permanent' gases. This explains the great consequence of the success achieved by Cailletet and Pictet.

Disorder and temperature

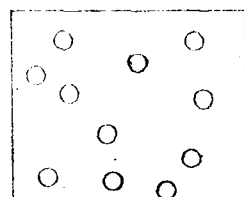
The liquefaction of oxygen gave strong support to the hypothesis that a sufficient degree of cold will turn any gas into a liquid, and this fundamental relation between temperature and change of state holds the key for the importance of low temperature research. The progress in atomic and molecular physics in the last three-quarters of a century has made the phenomena of liquefaction and freezing appear in a new light. What does really happen when ice melts into water, and when water in turn evaporates?

The molecules in a bit of frozen water, say in a snow flake, are all arranged in a definite and orderly pattern. It is this regular arrangement on the atomic scale which is responsible for the beautiful star-shaped little crystals of which the snow flake is made up. When caught in the hand,

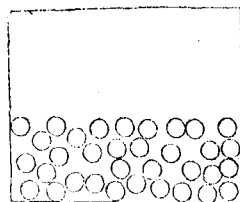
the little stars will persist for a short while, then they collapse and the pretty pattern changes into a droplet of water. What happens is that heat from the skin is transmitted to the water molecules in the crystal and causes them to vibrate more and more violently. Finally the stage will be reached when the heat motion becomes stronger than the cohesive forces holding the molecules in their positions. The swing of the vibration becomes so large that an individual molecule will not return to its place in the crystal, the regular pattern of the crystal is lost and the substance melts.

In the liquid state, the heat motion is sufficient to prevent the molecules from sticking to their individual positions, but it is not strong enough to detach them completely from each other. Thus, the molecules of liquid water, while not retaining a rigid pattern, still cling together. Only when the temperature is raised further will the cohesive forces be entirely overcome. Then the water molecules break away from each other completely. Each individual molecule travels through space in solitude with such speed that whenever it collides with one of its fellows the force of the impact will make them bounce apart again. The liquid has evaporated (See Fig. 8).

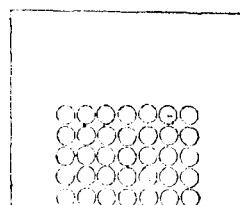
The change in the atomic pattern produced by the application of heat at once reveals the true significance of low temperatures as a physical phenomenon. As energy is



in a gas



in a liquid



in a solid

Fig. 8.
Distribution of Atoms

supplied, the regular arrangement of the crystal is transformed into the complete chaos of the gas; the liquid state forming a curious kind of halfway house. More than any other aspect, this *destruction of order* with rising temperature has now become recognised as the relevant physical change brought about by the application of heat.

The importance for physical research of this relation between temperature and disorder is at once apparent. We have seen that it is the complexity of our terrestrial scene which makes it so difficult to recognise the basic principles of physical phenomena. In spite of the impressive and sometimes terrifying mastery of the physicist over the forces of nature, our understanding of the fundamental processes is still in its infancy and any means of simplifying the complex aspect presented by the things around us must become a powerful method of research. One such method is provided by the 'high temperature' of nuclear bombardment which breaks up matter into its elementary building bricks. The other method is refrigeration, which allows us to study the pattern of matter in a state of ever-increasing order as we descend on the scale of temperatures.

Calorimetry

In the case of the melting snow flake we were able to observe directly the change of state and the breakdown of the orderly pattern, but more often the changes in the state of order are hidden and may pass unnoticed. For instance, when melting an ordinary lump of ice we still observe the change from a solid to a liquid, but the obvious transition from a degree of higher to one of lower orderliness is obscured because the crystal pattern is not as readily seen as in snow. Even worse, at a certain temperature well below the melting point, the crystal pattern of ice undergoes a complete change which does not, however, alter the shape or the aspect of the substance. It is clear that our recognition of the importance of atomic order is of doubtful value if we are not sure that changes in this state of order can be detected.

Fortunately this problem was solved even before anything about the atomic structure of crystals was known. If we were to measure the heat passing from the hand into the snowflake, we would find that for each degree the flake is warmed, a certain amount of energy (measured in calories) has to be supplied; this is of course the *specific heat*.^{*} Then, on melting, a relatively large amount of heat, the latent heat, is taken up without change of temperature, and finally the specific heat of water must be furnished to warm up the droplet to the temperature of the skin. In this way, by merely measuring energies supplied and temperatures produced, a complete record of the changes in the atomic pattern, the increasing disorder of the crystal, the removal of the molecules from their positions and the weakening in the cohesion of the water molecules, is obtained. The great beauty of the method, and as we shall see later, its weakness, is that it does not require observations on the atomic scale. A physical quantity (heat divided by temperature) derived from these calorimetric determinations is a direct measure of the degree of disorder. Its importance was realised long before its statistical significance was known; it is the entropy.

According to the third law of thermodynamics all substances are in a state of complete order at absolute zero; their entropy is zero. By adding heat and observing the attendant increase in temperature step by step, we can plot the thermal life history of the substance. Such an entropy chart, reaching from the neighbourhood of absolute zero to room temperature, then presents a full record of all physical changes which the substance has undergone. The process is analogous to the exploration of a shaft in rock without descending into it. All one has to do is to pour down measured quantities of water and observe the rise in level with a plumb-line; the rate of rise providing a measure of the diameter. If, for instance, at a certain depth the level

^{*}Called *specific*, since it is characteristic of the particular substance and differs from one sort to the next.—Ed.

does not rise while water is poured in, we must conclude that at this place there exists an underground passage or reservoir which has first to be filled. Similarly the low temperature physicist carrying out a calorimetric determination will know by a temporary halt in his thermometer readings that at this temperature the substance undergoes some physical change.

Knowledge of the exact shape of the entropy curve often provides a valuable guide to the type of rearrangement taking place. A sudden discontinuous rise in the entropy (corresponding to a complete stop in the temperature rise) always indicates the occurrence of a phase-change like melting, or a transformation of the crystal structure; it means that at this temperature two modifications of the substance exist side by side. A steep rise in the entropy, which, however, is not discontinuous, is the sign of a 'co-operative phenomenon' like ferro-magnetism or the spontaneous electric polarisation of substances such as KH_2PO_4 . Our knowledge concerning these changes is still very limited and more often than not we are unable to attach an atomistic interpretation to the anomalies charted by calorimetric exploration.

Refrigeration

However, before any of these experiments can be carried out the substance has to be cooled; the shaft has to be emptied before we can fill it with water. It now becomes clear why the production of cold is so much more difficult than the generation of heat; it is always easier to create chaos than to make order. In fact, the second law of thermodynamics postulates that, irrespective of the process involved, the total entropy of a physical system *can never decrease*. At first sight this offers a bleak prospect for the production of cold. Fortunately, even the laws of nature have loopholes and we can score against this one on a technical point. We can guide a physical process in such a way that in one part of our apparatus order is actually

created, provided we keep within the law by seeing to it that the *total* entropy of the system increases. Such a mechanism, which produces cold at one spot at the expense of more heat created all round, is called a refrigeration machine. Its realisation depends on finding a suitable working substance whose entropy varies with some factor other than the temperature. The close connection between gas liquefaction and refrigeration is due to the fact that the entropy of a gas decreases on compression. When a bicycle tyre is pumped up, the 'squeezing out' of the entropy can be noticed by the liberation of heat.

The simplest type of refrigeration cycle, though not necessarily the most common, consists of a compressor (see Fig. 9) from which heat is carried away by the cooling water, and an expansion engine. In the latter, the compressed gas pushes out a piston until it has regained its original pressure. The expanded gas cannot, however, regain its original temperature unless the heat lost in the compressor is allowed to return to it. By thermally isolating the expansion engine, the return of this heat is prevented and the gas

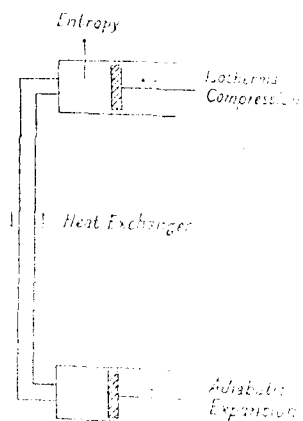


Fig. 9. - A Refrigeration Cycle

cools. The gas cooled in the first stroke can now be used to lower the temperature of the compressed gas coming in for the second stroke, and so on. The heat exchanger, consisting usually of two concentric tubes, in which the expanded gas pre-cools the compressed, thus forming an essential link between the 'isothermal' compression and the 'adiabatic' expansion. It combines the two processes into a continuous cycle and is a characteristic feature of all types of refrigeration machine.

Refrigeration by means of the Joule-Thomson effect is another important method of gas liquefaction. Here the gas is expanded through a nozzle and the work is not done against a piston but against the molecular forces. This process is less efficient than an expansion engine, but it can be worked in a continuous cycle without the use of moving parts at the cold end. In view of the lack of suitable lubricants at low temperatures this is an important consideration. In a cryogenic (low temperature) laboratory, simplicity and ease of operation will usually outweigh the factor of efficiency economy, but in modern plants for oxygen liquefaction, which has become an industrial process, both types of expansion are combined to produce the maximum yield for a given power input. Domestic refrigerators, and also some stages in large scale liquefaction installations, make use of the large change in entropy when a substance passes from one state of aggregation into another (compare the latent heat of the snow flake). Since for practical reasons the working substance should remain 'fluid' throughout the cycle, the change of state commonly used is that between gas and liquid, e.g., liquid to gaseous ammonia.

The liquefaction of oxygen in 1877 was followed in 1898 by the liquefaction of hydrogen and in 1908 by that of helium. Sir James Dewar besides liquefying hydrogen made the most important invention in cryogenic technique; the vacuum flask which bears his name. However, with the beginning of the 20th century the centre of low temperature research shifted to Holland, where Kamerlingh Onnes had

founded the famous cryogenic laboratory which held a position of unchallenged supremacy for nearly 50 years.

Helium which, as the temperature is lowered, is the last gas to condense, boils at 4.2° K. By pumping off the vapour above the liquid and thereby abstracting from it the latent heat of vaporisation, Kamerlingh Onnes could reach a temperature of slightly below 1° K. The final limit of refrigeration by gas liquefaction had thus been reached.

Temperatures below 1° K

The progress towards absolute zero, which by then had become synonymous with gas liquefaction, now came to a temporary standstill. It was clear that refrigeration below the helium range demanded a novel type of entropy removal. A physical system had to be found which is still in appreciable disorder at 1° K. and which can be ordered by an external change similar to the application of pressure to a gas.

In 1926 Giauque and Debye independently suggested the same solution to the problem. They proposed cooling by magnetic means. The magnetism of individual atoms is caused by the rotation of electrons in their orbits or around their own axis; the so-called electron spin. In most substances these 'elementary magnets' are in a state of order, neighbouring magnets pointing in opposite directions and thereby cancelling out each others' magnetism (i.e., there are electrons with opposing spins within a single atom or molecule, which is consequently neutral). Some salts, however (for instance, iron ammonium alum), have unsaturated electron spins, which means that in each molecule or ion there is one rotating electron whose magnetism is not cancelled by the rest. At very low temperatures these spins, too, will become ordered by mutual interaction, but at 1° K. they are still pointing at random in all directions. However, when a strong magnetic field is applied to the salt, it will make the little electronic magnets all point in the direction of the lines of force, and thereby produce order.

Like the compression of a gas, the magnetization squeezes out entropy from the substance and heat is liberated. This heat is carried away by the liquid helium, which corresponds to the cooling water in the compressor. Then the salt is thermally isolated and the magnetic field is removed. Since the heat of magnetization has been abstracted from the spins and cannot be regained, the substance cannot return to its former state of disorder, and must cool. Its temperature will fall to that level which corresponds to the degree of order produced by the magnetic field.

Seven years after it was suggested, the magnetic cooling method was first employed at Berkeley, Leyden, and Oxford (see Plate 5). The experimental difficulties were considerable. High magnetic fields had to be provided, and the influx of heat into the salt had to be reduced to a minimum. Nevertheless, the technique of work with temperatures as low as 0.01° K. has since been developed to such extent that these degrees of cold can be maintained for many hours, and physicists are at present preparing experiments which may yield temperatures of 0.0001° K. or even 0.00001° K.

These new attempts also will employ the magnetic method. It is to be expected that the usefulness of electron spins will be exhausted in the range of 0.01° K. and 0.001° K., because by then the electronic magnets will be in a high state of order. There exist, however, other elementary magnets which remain disordered even at these very low temperatures: the atomic nuclei. The magnetic interaction between one atomic nucleus and its neighbour is extremely weak, and calculation suggests that they will become ordered only at about one millionth of a degree above absolute zero. While thus theoretically a method is open to us to extend the range of refrigeration a hundred or a thousandfold beyond the present limit, the obstacles are very great indeed. All arrangements designed to produce cooling require the making and breaking of heat contacts. In the temperature region covered up to now, this has been accomplished by the use of gases, and the lack of a suitable medium pro-

viding heat exchange at 0.01°K . may prove one of the most serious difficulties in realising the method.

Another more fundamental problem is that of heat exchange between the nuclear magnets and the substance as a whole. The temperature of a body is measured by the vibration of the atoms and, unless the energy given off by the nuclei can be taken up by the whole atom and vice versa, the action of the magnetic field will produce no thermal changes. Theoretical calculations have indeed indicated that this energy exchange may be so slow as to make nuclear magnetic cooling impracticable, but recent experiments raise the hope that nature may be kinder to the experimentalist than theoretical prediction warrants.

Thermometry

The question is often asked whether, when the low temperature is reached, it can also be measured. In other words, if we have found a new method of reducing the temperature, will this be useless until we are lucky enough to discover a means of *measuring* temperatures in the newly opened range? Fortunately the two problems are not independent ones. The fact that a cooling can be obtained means that the entropy of the working substance changes in the region of refrigeration, and this also means that we can employ this entropy change for a thermometric process. Thus, in the region of gas liquefaction use is made of the variation of the pressure or the volume of a gas with temperature and one uses a gas thermometer. In the range covered by magnetic cooling the magnetic properties (the susceptibility) of the salt also serve as an indicator for the temperature. Similarly in the proposed nuclear cooling stage the effect of the nuclear magnets will provide a suitable thermometer.

While therefore no difficulty in temperature measurement will arise in principle, the experimental procedure involved is likely to be laborious. The magnetic susceptibility, which is not very hard to measure, can only serve as a temperature

indicator and not as a true measure of the absolute temperature, unless we can account theoretically exactly for the manner in which this quantity will change with temperature. Theoretical prediction of this kind becomes increasingly difficult as the mutual interaction between the elementary magnets grows in importance, i.e., just in the region of the lowest temperatures which can be reached with the method. The only alternative is to carry out very accurate determinations of thermodynamic quantities from which the true 'thermodynamic' temperature can be deduced rigorously.

In addition to the thermometric processes based on the same entropy changes which provide refrigeration, other physical quantities can be employed as 'secondary' thermometers. Indicators of this kind may be simple in operation and have other desirable features, but they all have first to be calibrated against some primary thermometer which measures the thermodynamic temperature directly. The most important secondary thermometers are electric ones; thermo-couples, and the variation of the electrical resistance of some metals (in particular platinum) with temperature.

Quantum effects

In the discussion of refrigeration methods for very low temperatures, the reader will have noticed the curious fact that as the temperature decreases, the steps in cooling also appear to decrease. Helium liquefaction brings us from the lowest temperature obtainable with hydrogen (about 10°K .) to a little below 1°K . Demagnetization of electron spins reduces the temperature from about 1°K . to about 0.005°K . and the proposed nuclear stage may end at a millionth of a degree. These facts would have been quite incomprehensible to a physicist at the turn of the century. He would have expected either of the last two methods, or possibly even the helium stage, to reach absolute zero.

In 1906 Walter Nernst formulated an empirical law of

physics which soon took its place among equals with the first and second law of thermodynamics. This third law of thermodynamics states that at absolute zero the entropy of all substances is zero, and it expresses the fact that, while absolute zero can be approached to any degree, it can never be reached. Whereas, therefore, absolute zero is a finite limit to our temperature scale, it has slipped into infinity as far as its experimental attainability is concerned. Dealing with low temperatures we must revise the old accepted view that degrees of temperature, like centimetres or inches, are equal measures of a physical quantity. One degree centigrade only represents a third of a per cent of the absolute scale when measured at room temperature (300°K.) but it amounts to 50 per cent when a cooling from 2°K. to 1°K. is considered. A reduction from 0.01°K. to 0.0001°K. may only be a hundredth of a degree, but it represents a cooling to 1 per cent of the initial temperature, corresponding in range to the step from ordinary temperatures to those of liquid helium.

The third law is a direct consequence of the fact that, in any physical process, energy is emitted or received not as a continuous stream but in small indivisible parcels, the quanta. The quantum theory is the basis of modern physics, without which the structures of atoms and molecules and the processes taking place within them would remain utterly incomprehensible. The reason why the discontinuous nature of energy had escaped the notice of physicists until less than fifty years ago is that the fundamental constant involved is very small (6.6×10^{-27} erg-sec.), which normally limits the influence of the quantum principle to atomic dimensions. At sufficiently low temperatures, however, quantum effects become noticeable in large-scale physical determinations. In a way, the appearance of the liquid and the solid state when a gas is cooled can already be considered as a quantum effect, because the cohesive forces between the atoms and molecules are quantum-mechanical ones. Nevertheless, in most substances, even if they are solid, the

energy is shared out equally at normal temperatures. This, however, changes completely when the temperature is lowered sufficiently and the available energy quanta become accordingly smaller. Then an energy economy of plenty is gradually replaced by one of scarcity. The quantum phenomena begin to be noticeable in large-scale measurements.

In fact, it was through the experimental work at low temperatures that the quantum principle, which up to then had been regarded as a mathematical oddity rather than a physical reality, entered modern physics. Early in the nineteenth century it had been discovered that the specific heats of solid elements had nearly all the same value (6.3 cal.) when the same number of atoms was considered. Returning for a moment to our analogy of a shaft, the width of which can be explored by filling with water, this discovery means that all the shafts representing elements are equally wide. However, there are a few exceptions, as for instance carbon in the form of diamond, the 'shaft' of which has less than a quarter of the normal width. Since the specific heat of diamond comes up to the value of 6.3 cal. at high temperatures, it was decided to investigate the specific heat of the other elements at low temperatures. These determinations yielded the remarkable result that sooner or later all the shafts get narrow, becoming infinitely small as absolute zero is approached. This striking effect, which is caused by the variation of quantum size with temperature, could not be explained at all on the basis of the 'classical' conception that energy is a structureless fluid.

The anomalies in the specific heat mentioned earlier (page 77, the underground reservoirs), acquire new meaning when viewed in the light of the quantum theory. The large increase in entropy accompanying the destruction of electric polarisation in the salt KH_2PO_4 , for instance, occurs at about 120°K. This means that the molecular change taking place swallows up energy quanta of a size corresponding to this temperature. A calorimetric measurement will therefore not only detect physical changes, it

also reveals, by the temperature level at which the particular change occurs, how big is the energy involved in the change.

Superconductivity

Among the great number of quantum effects which make their appearance at low temperatures there are two which have claimed special attention. This is partly due to the extraordinary phenomena which are connected with them, and partly to the fact that so far they have found no adequate explanation. The first of these effects is the super-conductivity of metals, and the other the superfluidity of liquid helium.

The peculiar properties which distinguish a metal from an insulator, say sulphur, are due to the free movement of electrons within the metal. Inside the metal these electrons behave in many ways like the molecules of a gas and, being electrically charged, they can be made to move by attraction and repulsion. Hence the application of an electric field to a metal produces a current. This flow experiences a resistance, because the electrons collide all the time with the vibrating atoms and are deflected from their path. As the temperature is lowered this atomic vibration decreases, the electrons are not scattered so much, and the electrical resistance of the metal becomes smaller. Theory suggests that, in a perfectly pure metal, at absolute zero, the resistance should entirely disappear.

Of course, neither can absolute zero be reached, nor can we hope to make a perfectly pure metal. It was therefore hardly credible when in 1911 Kamerlingh Onnes reported the observation that at 4°K. mercury completely loses its resistance. Nevertheless the discovery proved to be correct, and in the following years this new phenomenon of *superconductivity* was found in a considerable number of metals. The superconductors all have medium atomic volume; metals with very small or very large atomic volume do not seem to show the phenomenon. They are in certain ways similar in physical and chemical properties; monovalent and ferromagnetic metals do not become superconductive.

The loss of electrical resistivity appears to be confined to temperatures below 15°K., and a recent American claim to have discovered superconductivity in a solid solution of sodium in ammonia at more than 100°K. has not been borne out by subsequent experiments.

The electrical conductivity seems to be infinite, no trace of resistance having ever been found in any experiment. A current which has once been set going in a closed ring of superconductive metal continues to run for hours and days, as long as the ring is kept cold. When a superconductor is cooled in a magnetic field, the metal will eject the magnetic lines of force; it becomes perfectly diamagnetic. A small bar magnet brought close to a shallow trough of superconductive metal will float above it, because the magnet is repelled by its own magnetic image which it has induced in the surface of the metal. Since an electric current is always associated with a magnetic field, the complete diamagnetism of a superconductor confines the flow of electricity to a thin surface layer. These surface currents are limited in strength to a critical value, which increases as the temperature is lowered. This restriction in the current which can be passed without resistance is the reason why superconductivity has found no practical application, except for recent attempts to use superconductive wires as sensitive detectors for infra-red radiation.

It is probably wrong to regard superconductivity as an extreme case of ordinary electric conduction. The electric field, which normally by accelerating electrons produces a current, is absent in a superconductor. Its place is taken instead by the magnetic field, with which the super-current is connected in a peculiar way unknown in ordinary electro-dynamics.

Superfluidity

On the same day when Kamerlingh Onnes first liquefied helium he tried by further cooling through evaporation to produce solid helium. The attempt proved unsuccessful, and we know to-day that helium will remain liquid down to absolute zero. Solid helium can only be obtained by subjecting

the liquid to external pressure. This astonishing observation was only the beginning of a long line of strange discoveries. First it was found that below a certain temperature (2.19°K.) the liquid which up to then on cooling had contracted, began again to expand. Calorimetric experiments showed that this temperature is also marked by a large anomaly in the specific heat, and it became clear that the substance must undergo some kind of a change when cooled below 2.19°K. The remarkable feature of this transformation is that helium I, above this temperature, and helium II, below it, are *both in the liquid state*. In a solid substance various forms of change, which are due to the mutual rearrangement of the atoms, may occur, but in a liquid where the atoms are not arranged in an ordered pattern, transformations of this nature cannot be visualized.

However, it soon became apparent that the two liquid forms differ very much. Whereas helium I is a fairly normal liquid, helium II exhibits properties unlike those of any other substance. Its heat conductivity is enormous, many times exceeding that of copper or silver, the best heat conductors known so far. Its internal friction is negligible, and it can flow almost without resistance through narrow

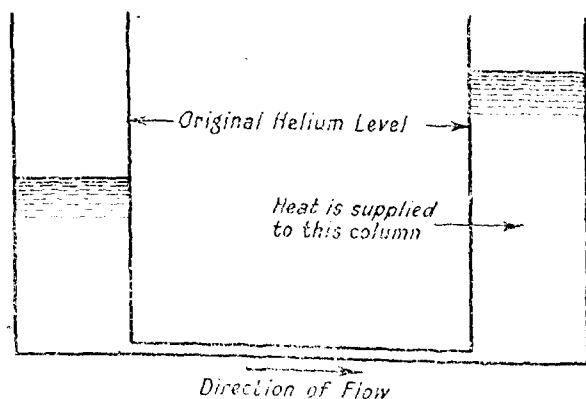


Fig. 10a. - Heat effects connected with the flow of liquid helium II.

capillaries. When a heat impulse is supplied to the liquid, instead of being dissipated, it will travel along like a sound wave in air, and the phenomenon has been called 'second sound'. All these effects seem to suggest that helium II is a mixture of a normal liquid and a 'superfluid' one. Any such description has, of course, to be treated with caution, and provides possibly nothing more than a convenient way of looking at things. As such it has had considerable success. If, for instance, heat is supplied to one of two volumes of helium II which are connected by a narrow capillary, a flow of liquid in the direction of the higher temperature will occur. This can be interpreted as a flow of the superfluid part towards the warmer end through the narrow connecting tube which, however, presents an impassable obstacle to the normal part. Conversely, when helium II is forced through a capillary, the emerging liquid cools while that which remains behind warms up, indicating that the superfluid part of the liquid has a low entropy.

In addition to all these phenomena, helium II forms a film of about 100 atoms thickness on all solid surfaces with which it comes into contact, and it is in this film that the most striking effects occur. An empty glass beaker which is

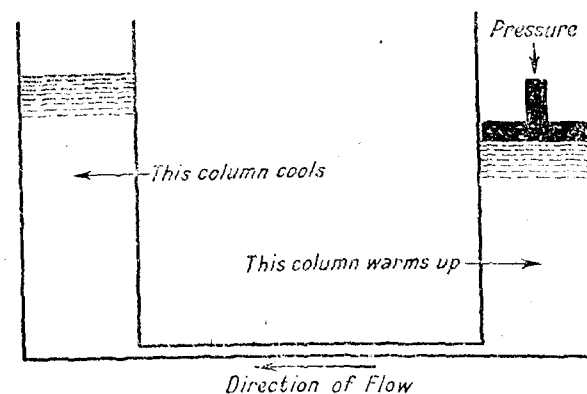


Fig. 10b. - Heat effects connected with the flow of liquid helium II.

lowered into the liquid only so far that its rim stands well above the level, can be seen to fill up gradually until the liquid inside has reached the same level as the bath of helium outside. A flow in the reverse direction will take place when

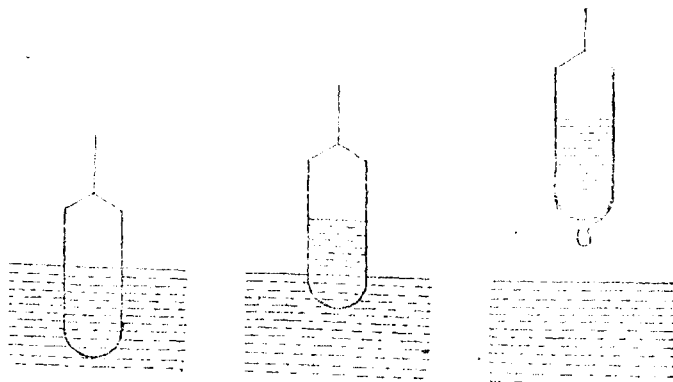


Fig. 11. — Film transport in liquid helium II.

the beaker is now raised, and when it is lifted completely out of the bath little drops of liquid helium can be seen forming on the outside bottom of the beaker. These drops fall back into the bath at a steady rate until the beaker is empty. This transport, which takes place through the film, is uninfluenced by the level difference or the height of the glass wall above the liquid. The fact that the rate of the film flow does not depend on these factors strongly indicates that the transport is completely without friction. Indeed the films offer a unique possibility of separating the super-flow from the ordinary viscous flow (Fig. 11, and Pl. 4).

The Fourth State of Aggregation

As already indicated, there exists as yet no satisfactory theory for either superconductivity or superfluidity; however, the beginnings of an explanation are clearly

indicated by the theoretical attempts made recently. It appears that we are confronted here with phenomena in which the quantum effects completely overshadow the processes of classical physics. As to what exactly happens, physicists are by no means agreed yet. In the author's opinion the similarity between superconductivity and superfluidity is so strong that in spite of the difference of the particles involved (electrons in one case and helium atoms in the other) we should look for a common explanation. In both cases there is frictionless transport which breaks down when a critical rate is exceeded, and in both cases there is a rapid drop in the entropy when the substance is cooled into the frictionless state. This decrease in entropy clearly reflects some sort of ordering process, but in neither case is there any indication of the particles taking up an ordered position as, for instance, the water molecules do in the snowflake.

However, when describing an assembly of particles it is not their mutual position alone which matters. The atoms and electrons also move, and we must consider their velocities as well. Normally, as for example in a gas, there will be particles with all kinds of velocity, small, medium and large, the average speed depending on the temperature. It is conceivable that in a superconductor and in helium II, more and more particles will all have the same velocity as the substance is cooled and that the phenomena we observe are due to such a condensation, not according to position but to velocity. In this case we would have to add to the three known states of aggregation, solid, liquid and gas, a fourth one; the frictionless state of aggregation. At our present state of knowledge, considerations of this kind border on the realm of somewhat somewhat wild speculation, but it is likely that in a few years' time the experiments which are now under way will enable us to form a more balanced view on these questions.

The Artificial Kidney

PROFESSOR JOHN MCMICHAEL

1. How the Kidney works

THE normal human and mammalian kidney consists of a bundle of many thousands of functioning units called *nephrons*. Each of these is a microscopic coiled tubule into the blind expanded upper end of which is projected a

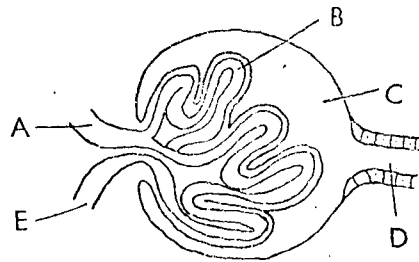


Fig. 12 - Diagram of a kidney glomerulus. A is the afferent artery leading blood into the glomerulus capillary loops (B) which are inserted into Bowman's capsule (C) like fingers into a glove. D is the upper end of the tubule lined by cubical cells. E is the efferent blood vessel leaving the glomerulus.

special filtering apparatus called the *glomerulus* (Fig. 12) while the lower end joins, with a multitude of other nephrons, a duct leading from the kidney to the urinary bladder. The glomerulus or filter has a relatively simple structure, and consists of a coiled loop of thin-walled blood vessels (capillaries, Fig. 12, B) covered over by a thin membrane, the whole projecting into the upper end of the funnel-like tubule which is known as Bowman's capsule (C).

Blood is carried through the glomerular capillaries at a high pressure, equal to 60 mm. mercury (nearly one-twelfth of an atmosphere). At this pressure the watery part

of the blood, containing dissolved substances, is filtered through the membrane, into Bowman's capsule. This glomerular filtrate has been directly examined by an ingenious micro-technique by Professor A. N. Richards and his school in Philadelphia, who proved that it consists of a protein-free solution, otherwise identical in composition with the blood plasma. This watery solution then passes down the tubule, where a process of selective reabsorption takes place. The total glomerular filtrate of both kidneys in man is about 100 ml. per minute, or six litres an hour, or 148 litres in 24 hours. As the amount of urine passed is about 1-2 litres each day, this means that 98 per cent of the filtered water must be reabsorbed. Not only is water reabsorbed, but substances necessary to the body-economy, such as sodium chloride and glucose, are also selectively reabsorbed by the tubule cells back into the bloodstream.

2. The problem of Kidney Failure

When the kidneys are destroyed or thrown out of action by disease, the elimination of waste products no longer keeps pace with their production by the processes of metabolism necessary for continued life. Urea steadily accumulates in the blood-stream, together with other substances which are much more poisonous. Patients become drowsy, wasted and thirsty. Twitching of the muscles sets in, and death is never long delayed. There are certain conditions in which kidney failure comes on suddenly, and, when the kidneys are abruptly thrown out of action, the patients seldom survive for more than a week. There are many circumstances in which, about the end of this critical period, urine secretion may begin to recover, and the patient turns a precarious corner towards recovery from the acute *uraemia*, as the condition of kidney failure is called.

3. Past efforts at treatment

In the conditions of acute *uraemia* described above, there was often little that the physician could do, except keep his

fingers crossed and pray. Old-fashioned methods such as purging, bleeding and sweating could only remove a tiny amount of toxic substances and were, for the most part, nearly useless, in addition to being uncomfortable and miserable for the patient. Efforts were made to make the kidneys work again, by forcing the patients to drink large quantities of water, and even infusing salt solutions into the veins. As the kidneys could not get rid of this water, it often constituted an excessive load on the circulation, precipitating heart-failure. Various drugs which promote the formation of larger volumes of urine do so by preventing reabsorption by the tubules; that is, they tend to depress the normal function of this important part of the nephron, and their use might be accompanied by further impairment of the kidneys.

It was often thought that the blood-flow through the non-functioning kidneys in uraemic patients was reduced much below the normal. This view has, in fact, been substantiated by modern research.* Efforts were therefore made by surgeons to improve the blood-supply to the kidney, by various operations on the blood-vessels supplying it, and by such measures as releasing the swollen kidney from pressure, imposed on a swollen organ by the tough resistant and inelastic capsule of fibrous tissue which encloses the organ. This latter operation was called decapsulation.

Needless to say, the results of all these measures were disheartening, and even though an occasional recovery followed, it was always difficult to say whether the patient had recovered because of, or in spite of, treatment.

4. *The development of the Artificial Kidney*

It has been emphasised and proven that the glomerulus of the kidney works like a filtering or dialysing membrane. A membrane of collodion or thin sausage-skin will allow water and dissolved substances to pass through. If the membrane,

*See *X-ray Studies of the Kidney Circulation*, page 101.

constructed like a bag, is filled with a salt and water solution, and the bag is immersed in a bath of water only, the salt will pass out into the water until an equal concentration is reached on both sides of the membrane. If, however, the membrane bag and the bath both contain a salt solution of the same concentration a state of equilibrium would exist, in which no significant interchange takes place. It has long been realised that the composition of the blood might thus be controlled by causing blood to flow along cellophane tubes, which were bathed outside by a solution of known composition. Efforts were made as long ago as 1912 to eliminate toxic substances from the blood of experimental animals by this means. Immense technical difficulties arose, however, mainly because of clotting. Clotting troubles prevented any further development of these techniques, until large-scale manufacture of heparin was achieved in Canada and Sweden a few years before the war. Heparin is a substance which, on injection into the blood-stream, will prevent blood from clotting. Otherwise, it has no important action and is quite harmless. This technical advance made possible another attack on the problem of uraemia. Circumstances, curiously enough, were such that the first real practical solution was reached in Holland, during the years of German occupation. Work in the Dutch Universities was virtually suspended by the Germans, but in the small country town of Kampen, Dr W. J. Kolff found conditions favourable for a certain amount of clandestine technical work. He gained the co-operation of the staff of a local enamelware factory, who constructed a large bath of suitable shape and dimensions. In this bath was placed a rotating drum, round which was wound 40-50 yards of cellophane sausage-skin. The next problem was to convey blood from the patient through this cellophane tubing, round the rotating drum. Blood was led in through the axle of the drum, the rotating coupling of the early model being, in fact, a coupling from the water-jacket of a Ford engine. The rest of the technical development followed lines

now well-known in experimental physiology. The patient was given an injection of heparin to prevent his blood from clotting, and, by means of a small incision over the radial artery (which is the pulse at the wrist) the artery was connected to the tubing, through which blood would flow, the blood being subsequently led back into the patient's veins at the bend of the elbow. By this means Kolff was able to "wash" 30-120 litres of blood in ten or twelve hours. This means that all the blood in the body (about 5 litres) was taken through the machine at least six times and sometimes as much as twenty-four times (see Plate 7).

The washing fluid which filled the bath had to be very carefully made up. Glucose was added in sufficient quantity to prevent loss of glucose from the body. Sodium bicarbonate was added, to neutralise some of the acid products which accumulate in the blood of uraemics. By means of various chemical adjustments, the substances escaping from the blood can be reasonably controlled. As much as 250 grammes of urea have been removed from a patient by a single 12 hour dialysis through the artificial kidney. This is equivalent to the amount of urea eliminated in the urine of a normal person in ten days.

5. *The clinical use of the Artificial Kidney*

The conditions which arise in clinical medicine and which lead to suppression of kidney function fall into two groups.

(i) Those in which the kidneys have been practically completely destroyed by disease, and in whom, therefore, there can be no hope of recovery, even although the patient may be kept alive a few days longer by means of the artificial kidney.

(ii) Those in whom uraemia develops as a result of an acute illness affecting the kidney, and where suppression of urine formation is a temporary matter and in whom recovery may be hoped for, provided the patient can be tided over the critical days which follow suppression of urine formation at the end of the first week (see Plate 6).

The irrecoverable cases were naturally the first patients chosen for trial of the artificial kidney. Although prolongation of life was only achieved for a few days, the experience, none-the-less, proved that the method was safely applicable. Thereafter, the artificial kidney was applied in the treatment of the second group, in which the kidneys were subjected to some sudden severe damage, known, however, not to be fatal, if the patient could only be kept alive sufficiently long to allow normal healing to take place. Very quickly the artificial kidney began to justify the hopes of its inventor, and a small but steadily increasing number of patients are alive today, who would certainly have died without this life-saving measure.

Needless to say, much new practical information has come to hand as the result of prolongation of life in patients who would otherwise have died. The clearance of excess waste products from the blood-stream often seems to be followed by the sudden reappearance of a satisfactory urine flow. The more old-fashioned methods of treatment, such as infusion of salt solutions in large quantities, have been found to be harmful influences, should the patient subsequently require treatment by means of the artificial kidney. The circulation is often overloaded by such measures, making the condition of the patient precarious for the application of the modern technical procedures. With steadily increasing knowledge and new developments the method has undoubtedly come to stay.

6. *Some further useful ideas in the treatment of Uraemia*

The area of the filtering surface of the glomeruli in the human kidney is estimated at 20,000 sq. cms., while the filtering surface of the artificial kidney is 24,000 sq. cms. The volume of blood flowing through the normal human kidney is about 1200 ml. per minute, while that flowing through the artificial kidney is at least 100-200 mls. per minute. The artificial kidney, therefore, although having a useful filtering surface, cannot be supplied with blood at a

rate comparable to the normal human kidney. For this reason other methods have been tried, of which the most promising is *peritoneal dialysis*.

The peritoneum is the thin, shiny membrane which lines the inside of the abdomen, and which covers over the stomach, liver, and intestines. It has a good blood-supply and a reasonably large area, and it was realised that if this surface could be used for dialysis, something worth-while might be achieved. The problem was the preparation of a large volume (50-100 litres) of sterile, non-irritant saline solution, which could be washed through the abdominal cavity, so as to take away urea and other substances which would diffuse into it from the blood-stream.

The practical application of this method has now been achieved. Under a local anaesthetic a small puncture wound is made on each side of the belly wall, and through each puncture is passed a short length of perforated rubber tube. The sterile saline solution runs into the abdomen at one side and drips out into a receiver at the other. In practice it is found that this procedure has to be carried on for twenty-four to thirty-six hours, to achieve the same purification of the blood as is produced by a twelve-hour dialysis in the artificial kidney. None the less, it has the advantage that it requires less continuous watching than the artificial kidney, and, from the patient's point of view, there is little difference in the relative comfort of the two methods. Peritoneal dialysis, however, cannot be repeated, and it carries a slight risk of the introduction of infection to the peritoneum (*peritonitis*) when the rubber tubes have to be left in position for a long time.

The artificial kidney of Kolff's design is only at present available in a limited number of centres in Britain. It has the slight disadvantage of being cumbersome, and it requires a special room, properly equipped, to which the patient is brought. To overcome some of these features, a new dialyser has been devised in Stockholm, by Dr Nils Alwall. This consists of a cellophane tube wound round a

cylinder, which is enclosed in a larger cylinder. The space between the cylinders is small, so that the cellophane tube is kept flat, an advantage which increases the surface exposure of the blood for the same volume, and hence without excessive 'bleeding' of the patient into the tubes. The rinsing fluid runs down vertical grooves on the opposing surfaces of the enclosing cylinders. The patient's blood is forced through the cellophane tube by his own arterial pressure. This avoids the trouble with rotating couplings and pumps which had to be incorporated in Kolff's original machine. This new development in method is already proving itself of value, and it has the great advantage of being portable, so that, with a little extra equipment, it may be brought to the bedside of the patient.

7. *Dietetic measures*

It has long been known that urea and most of the other important waste substances are the product of the breakdown of protein, either that which is ingested in the food, or that which results from the wear and tear of the living cells of the body. Most of us can remember the days when patients with kidney disease were advised not to eat meat. While this is reasonable advice, the trouble was that, even if the patient starved himself, he would simply break down the proteins of his own body instead. It was thought that the break-down of body proteins could not be reduced below a certain minimum level, and therefore, in more recent times, opinion was developing that limitation of protein intake was useless and unnecessary. The problem, however, was taken up by another distinguished Dutchman, Professor J. C. G. Borst, of Amsterdam. Borst recognised that a certain amount of body-cell protein breakdown was inevitable, if the patient was starving or practically starving. He therefore tried the effect of giving a diet containing no protein at all, but which would, none the less, supply 2,000 calories per day, i.e., enough energy to meet the needs of the average patient lying in bed. These calories were

supplied by a syrup-like mixture of butter and sugar. Borst's Dutch patients seemed to take to this diet, and while on it, breakdown of body protein was reduced considerably and the amounts of urea formed each day fell from 15 grammes to under 5 grammes. By this means it was shown that the rate of formation of urea, and therefore the development of severe uraemia, could be slowed up. Other and more palatable diets have now been devised by Borst which make the treatment easier. By dietetic measures alone many uraemic patients can be kept alive beyond the critical days. The more dramatic procedures of the artificial kidney and peritoneal dialysis can be reserved for the most intractable and severe cases only. By the use of these methods the outlook for many patients with severe kidney disease has now undergone a dramatic and hopeful change.

X-Ray Studies of the Kidney Circulation

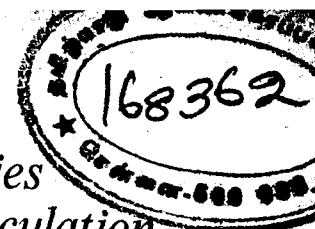
PROFESSOR K. J. FRANKLIN

Introductory

THIS article is in the main a résumé and appraisal of work done at the Nuffield Institute for Medical Research, Oxford, from the autumn of 1945 to 1947; it culminated in the production of a book describing and picturing the results,* and in a large-scale demonstration to the 17th International Congress of Physiologists, assembled in Oxford in the summer of 1947. The work was a cooperative effort by a surgeon, a physiologist, a radiologist, a graduate assistant in research, and – after the initial three months – a pathologist. It was, therefore, a fulfilment, in a particular sphere of investigation, of ideas which Lord Nuffield (then Sir William Morris, Bart.), had in mind when he purchased the eighteenth-century Radcliffe Observatory Building (Plate 9) and offered it to the Oxford Medical School as an Institute for Medical Research. For he saw how often clinicians and preclinicians are artificially separated from each other when, for the solution of many medical problems, they should be working together in one and the same research centre.

From the outset, for reasons into which we need not here enter, part of the building was allocated to research by means of X-ray cinematography (or cineradiography, as it is now more commonly styled) and allied techniques, and it is of interest to trace briefly the history of the application of X-rays to the investigation of the blood circulation.

*J. Trueta, A. E. Barclay, P. M. Daniel, K. J. Franklin, and Marjorie M. L. Pritchard (1947). *Studies of the Renal Circulation*. Oxford: Blackwell Scientific Publications, Ltd.



Franklin, K. J.
1947, 246
1949.10

Wilhelm Conrad Röntgen (1845-1923) discovered his new kind of rays on November 8, 1895, and on December 28 of the same year the Secretary of the Institute for Physics of the University of Würzburg received from him the preliminary note about his great discovery. As the necessary apparatus for producing the *x*-rays or Röntgen rays was available in most Physics Departments, radiographs were soon being made throughout the civilized world; the first human one to be taken in this country was made by Messrs A. A. Campbell Swinton and G. Stanton as early as the second week of January, 1896.

Under such circumstances, despite difficulties and hazards which surviving pioneer radiologists can still vividly recall, radiography of certain bodily structures became very rapidly an effective technique even if, as we see in retrospect, too much immediate attention was devoted to the pathological side before the normal had been adequately studied.

In contrast to this rapid application of *X*-rays to the examination of the bones, etc., of the body was their delayed application to the examination of the blood circulatory system. For normally blood vessels do not show up in a radiograph, and to make them visible it is necessary to inject into some part of the system a solution which is opaque to *X*-rays. So far so good, but the trouble in the earlier days was that all the radiopaque substances available were toxic to the living body in the concentrations in which they had to be used. But about 1930, media more suitable for the purpose, such as thorotrast (a colloidal thorium dioxide compound), diodrast or perabrodil (an organic iodine compound), and the like began to be obtainable and, while they rendered the inside of the heart and of the blood vessels visible in the radiographs, they were more or less innocuous to the living animal.

For a full study of the circulation, however, ordinary radiography, even if it is combined with the intravascular injection of radiopaque media, is insufficient. This is because the blood flow is a *movement* of appreciable rapidity, and a

mental picture of movement cannot be obtained from the examination of still radiographs. Some idea of it is given if one watches the fluorescent *X*-ray screen during the passage of injected contrast medium, but in ordinary screening the intensity of the fluorescence is not very great. Nor can the fleeting impressions be analysed as an objective record can; the most they can do is to give a hint to the skilled observer as to what is happening, i.e., he may get a glimpse of the truth, but he cannot effectively demonstrate it so as to produce conviction in others less skilled than himself. The upshot of all this is that, since changes occur in the circulatory picture in small fractions of a second, their full demonstration and analysis require an equally rapid rate of recording. In other words, for the full *X*-ray recording of the blood flow cineradiography is an essential technique.

Its general principles were formulated by McIntyre of Glasgow as early as 1897, but its evolution as a practicable proposition was more or less contemporaneous with that of innocuous contrast media for intravascular injection. It has two forms, the direct and the indirect. With the former of these, relatively small radiographs can be produced at a maximum rate of about 8 per second upon a continuous band of film. The direct apparatus of this kind which has been most used for circulatory research is that devised by Dr A. E. Barclay and Mr W. S. Farren; it takes serial radiographs, 5 inches square, at a maximum rate of 200 frames per minute, and it was used in the kidney circulation studies shortly to be described. The chief advantages of the direct technique are that the records are more or less life-size, and have the detail necessary for fine analysis. The chief disadvantages are that the field is small, that one cannot see with any ease what is happening during the recording, that the record is incomplete because the sequence is too slow, and that projection of the record, when reduced on to ordinary cine-film, is too jerky if any appreciable changes have taken place in thirds of seconds. For cine-projection a rate of not less than sixteen frames per

second is essential in order to avoid flicker, and normally the rate of taking should be not lower than this. So direct cineradiography at 200 frames per minute would be more correctly styled rapid serial radiography.

Indirect cineradiography involves photographing the fluorescent screen with a cine-camera, and to do this the highest quality is required in the X-ray tube, screen, camera lens, and film. The best results to date are those obtained by Professor Robert Janker of Bonn, and by Dr Werner Böhme, who studied under him and thereafter worked in Rostock. The 35 mm. film records produced by these two men can be projected as cine-films, for the rate of taking has reached, on occasion, over 200 frames per second. Analysis, though laborious (it has to be done through drawings from enlarged stills), yields very satisfactory results even if the detail is inevitably less than with the direct method. Finally, one can follow the course of an injection on the very brilliant fluorescent screen during a recording and so see the answer at once, or else get some very valuable first impressions. It is the present writer's belief that the renal short-circuiting (see below) which was discovered after some months' hard work at Oxford could have been revealed, and recorded in a cine-film suitable for projection, during a relatively small number of experiments with Professor Janker's apparatus.⁴

One last development of radiographic technique, initiated in large measure by F. Bohatyrtshuk in 1943-4 but since then very considerably extended and improved by Dr Barclay, is microarteriography, by means of which one can get details of very minute blood vessels. Inasmuch as it was of subsidiary influence upon the main results during 1945-7, and in addition Dr Barclay is producing a monograph on the technique,² this passing reference is probably sufficient here.

¹ Tentative arrangements for the making of such a film were upset by Professor Janker falling ill. We hope, however, that records will have been secured by the end of 1948 - K.J.F.

² To be published at Oxford by Blackwell Scientific Publications, Ltd.

With that necessary introduction, we can pass to the year 1941, by which time the Nuffield Institute had a considerable record of achievement in the fields of radiography and cineradiography and a policy, in line with Lord Nuffield's original ideas, of co-operating as fully as possible with outside researchers who sought the advantages accruing from the use of those techniques.

To the Institute, in 1941, came the distinguished Catalan surgeon, Dr J. Trueta, and his colleague, Dr J. M. Barnes, with ideas about the mechanism of a post-traumatic (injury) condition which Bywaters and Beall, the first to recognize and describe it, had named the 'crush syndrome'. They had reported four cases of air-raid injuries in which the victims had died of kidney failure (i.e., failure to continue making urine) after having had a limb or limbs crushed for several hours under fallen masonry, stones, or heavy beams. Dr Trueta believed, in opposition to almost all contemporary opinion, that the observed falling-off in the power of the kidneys to produce urine might have a nervous and not a toxic basis, i.e., that the crushing of the limb might cause a spasm* of the limb arteries and that this spasm might extend to the renal (kidney) arteries, with impairment of kidney function as a result. He and Dr Barnes asked, therefore, if they might investigate the problem experimentally, and with Miss Prichard, graduate assistant in research at the Institute, they made a beginning which showed that the spasm produced by a tourniquet tightly applied over the root of an anaesthetized rabbit's thigh (an experimental imitation of a crush injury) could be observed radiographically not only in the femoral artery on the tourniquet side but also in that on the opposite side; further, that the spasm on the tourniquet side on occasion spread centrally to involve the iliac artery just inside the abdomen. Unfortunately, the contrast medium was sodium iodide in high concentration (which is toxic); it had to be

*A constriction of the artery, due to the active contraction of muscle fibres in its walls squeezing the hollow channel of the blood vessel (see also p.116) - Ed.

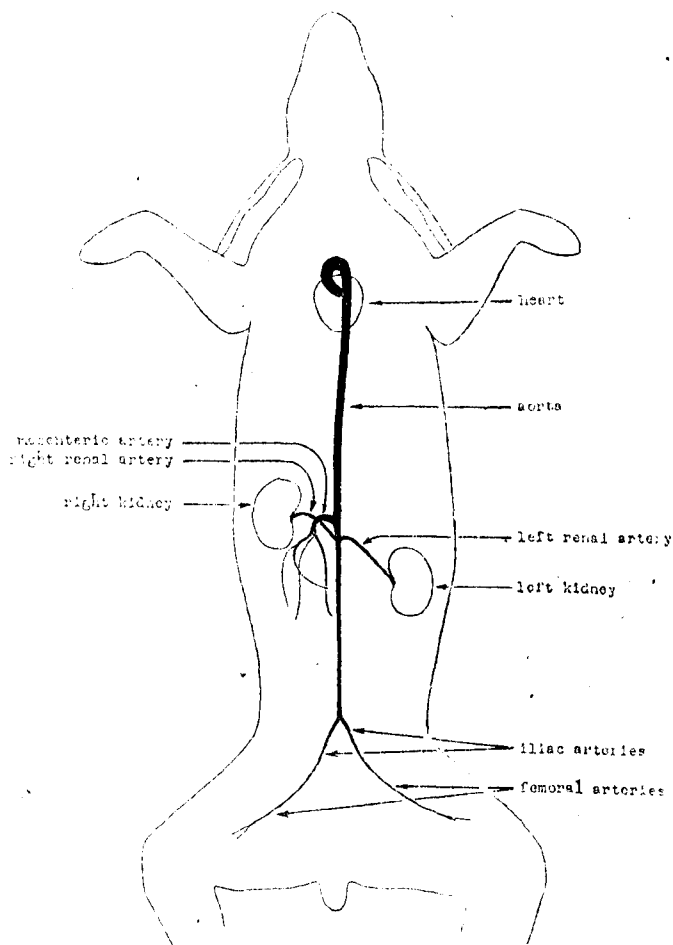


Fig. 13. - Diagram, based on an angiograph of an anaesthetized rabbit, to show the relative positions of the heart, aorta, and other blood vessels mentioned in the text.

X-RAY STUDIES OF THE KIDNEY CIRCULATION 107

injected into the terminal part of the aorta (the chief artery carrying blood from the heart to the main arteries of the body), and the radiographic field did not include the renal arteries; finally, Dr Trueta could spare no more time from his general war surgery. So the continuance of the research had to be postponed to a more opportune time.

This came in the autumn of 1945 when Drs Trueta, Franklin, and Barclay and Miss Prichard held a conference and planned a full-scale research to determine the part played by the nervous system in the distribution of the circulating blood volume,¹ one of the more immediate objects being to see if the renal arteries could be reflexly constricted through appropriate stimulation of afferent² nerves in the hind-limbs. It was decided to do the work on anaesthetized rabbits, partly because this species had been the one used in 1941, but mainly because rabbits were at the time the only animals obtainable in sufficient number. As it turned out, this was a fortunate limitation of choice, for the ordinary laboratory carnivores are not nearly so suitable as rabbits. It was necessary after a certain stage to study the naked-eye changes in the kidney surface, and in the rabbit this is a straightforward matter, whereas in the carnivore the changes are in part obscured by a network of superficial vessels. As a contrast medium, diodrast or perabrodil was found to be unsuitable because of its effects, in the strength required, upon the blood pressure. Thorotrast, on the other hand, appeared to be without untoward side-effects and it was, therefore, the medium chosen.

¹ It is important to realize that the normal total volume of blood in the body is not enough to fill the whole circulatory system, if all the blood vessels are wide open at once. In the healthy person the actual volume is, however, adequate since the circulation is controlled, i.e., some arteries are more or less shut when others are opened, so that the main quantity of blood is switched as occasion demands first to one organ, then to another. The details of the control are still incompletely known (see also pages 118, 121).

² Afferent nerves carry messages into the brain and spinal cord; efferent or effector nerves carry the instructions of the central nervous system to the organs.

The first step after this was to determine, by means of direct cineradiography, the times taken by a small amount of thorotrast, injected into the right external jugular vein (in the neck), to appear in the renal arteries and veins. It was found that these times were respectively 3.5 and 6 seconds, so most of the subsequent recording was done on large-sized X-ray films at more or less these intervals from the time of injection, the advantages being that the films covered a much larger part of the animal (thorax, abdomen, and the beginnings of the thighs), could be developed and fixed in a matter of minutes for examination, and were less expensive. The term angiograph, probably first introduced in 1934 by Stör, was applied to these radiographs of media-containing blood vessels. Control angiographs were made before each experiment so that the normal calibres of various vessels could be measured for comparison with their calibres as affected by experimental procedures. Measurements were made very accurately by means of specially adapted calipers fitted with a vernier scale, and the experimental calibres were expressed as percentages of the controls. The main vessels (see Fig. 13) reported on were the aorta, mesenteric artery (to the intestines), renal artery and vein, and iliac and femoral arteries (to the legs).

At the outset, anaesthesia of several hours' duration was found to be without effect on the calibres, so this factor was eliminated from possible complications. Application of a tourniquet to the left* thigh was then tried, and its usual effect both during and after application was to produce constriction of the left femoral and renal arteries, no change or perhaps slight dilatation in the case of the mesenteric artery, and marked dilatation of the left renal vein. Various procedures (nerve stimulation, injections of drugs, haemorrhage) were thereafter tried to see if they would cause corresponding change in the renal vessels, and it was found

*As the effects were normally all on the same side (ipsilateral), and the left kidney is the one more easily radiographed and also the one more readily observable after opening of the abdomen, most of the experimental procedures were carried out on the left side.

X-RAY STUDIES OF THE KIDNEY CIRCULATION 109

that the only ones which similarly produced constriction of the renal artery and dilatation of the renal vein were stimulation of the central end of the divided sciatic nerve and stimulation of the distal end of the divided splanchnic nerve. As in the tourniquet experiments, the degree of response varied from individual to individual, but the general concept was emerging that the application of a tourniquet to the hind limb produced a constriction of the renal artery of the same side through the intermediation of the nervous system: i.e., that impulses passed along a reflex arc, the effector fibres of which were included in the splanchnic nerve.

At this point, however, certain other observations shifted the focus of attention from the reactions of the extra-renal

TOURNIQUET EXPERIMENT

RABBIT

RENAL CIRCUIT TIMES				
BEFORE	DURING: T. ON 4½ HRS		AFTER	
	BEGINNING	END	T. OFF 15 MIN	OFF 1½ HRS
				
2.7 SEC.	1.8 SEC.	2.1 SEC.	1.5 SEC.	1.65 SEC.
100%	67%	78%	55%	61%

CROSS SECTION OF RENAL VESSELS				
ARTERY 100%	87%	75%	75%	81%
VEIN 100%	46%	150%	176%	160%

DETERMINED FROM DIRECT CINERADIOGRAPHIC RECORDS

Fig. 14. - Table based on measurements taken from the direct cineradiographic record of an anaesthetized rabbit which had had a tourniquet applied tightly to the root of its left thigh (T, tourniquet). To show the reduction in the renal blood circuit time observed during and after application of the tourniquet. (Reproduced by courtesy of Blackwell Scientific Publications Ltd., from *Studies of the Renal Circulation*).

vessels to what was happening within the kidney itself. It was noted that on occasion contrast medium appeared in the renal vein nearly simultaneously with its appearance in the renal artery, and this suggestion of a more rapid passage through the kidney was borne out by direct cineradiographic records of a tourniquet experiment (see Fig. 14). As the artery was constricted while the renal circuit time was reduced, it was thought that there was probably short-circuiting of the blood flow within the kidney, and in view of the probable importance of such an effect the main attention was thereafter directed towards the *intra-renal* blood circulation.

The first convincing evidence of the by-passing of part of the kidney by the blood flow was provided by an anaesthetized rabbit in which the central (spinal-cord-connected) end of the divided left sciatic nerve had been stimulated at intervals over a period of some hours in order that the neural conditions of a crush injury in the human subject (i.e., intermittent stimulation of afferent nerves in the crushed limb) might be more closely imitated. The angiographs taken from time to time had given no conclusive indication as to what was happening, and it was therefore decided to inspect the kidneys themselves during further stimulation of the sciatic nerve. But this procedure proved equally uninformative. When, however, as a last resort, 0.5 ml. of 5 per cent methylene blue solution was injected into each artery in turn, the results were dramatic. For the dye reached the surface of the right kidney in appreciable amount, but hardly any reached the surface of the left kidney. Further, when the organs in question were divided by transverse cuts, the right kidney was seen to be diffusely stained all over, whereas the cortex and the inner part of the medulla (see Pl. 10) of the left kidney were completely unstained, i.e., they had been reflexly deprived of their blood supply through stimulation of the sciatic nerve.

The rest of the researches can be regarded as deriving from this experiment, since it not only showed that the renal

blood flow had a hitherto unsuspected changeability in its intrarenal distribution, but also gave a valuable first indication of the parts which were rendered ischaemic (bloodless) and hyperaemic (full-blooded) respectively during such redistribution. We see, then, that a clinical surgical problem gave rise to an hypothesis, that the testing of the hypothesis by radiographic methods in the *intact* animal gave some inkling of the working of a previously unknown mechanism, but that the first striking proof came from the application of a non-radiographic technique. The subsequent researches were conducted by means of a great variety of techniques, with radiography and microradiography included but not dominant, indeed, with the hitherto somewhat puzzling radiographic appearances gradually becoming more fully explicable as the result of work with the other techniques.

The further researches had three main objects:—

(1) To determine the way in which various circumstances affect the blood flow to the kidney, its distribution within that organ, and its return from the same.

(2) To ascertain the pathways which are used by the blood when it is diverted, in part or in whole, from the cortex or outer part of the kidney.

(3) To discover if the functional and structural pictures so obtained of the rabbit's renal circulation and vascular system have their counterparts in other mammalian species.

To understand the answer to the first problem, the reader must have some idea of the mammalian kidney and of its blood-vascular and urine-producing systems (see Figs. 15-21). The organ itself is divided into an outer part, called the cortex, and an inner part, called the medulla. In a rabbit weighing about five pounds, each kidney is an inch or so long, and it is supplied with blood at fairly high pressure by the renal artery; a filtrate formed from the blood during its passage through the kidney is changed into urine and passes down through the ureter to the bladder; the modified, purified, blood which leaves the kidney does so by the renal vein.

The individual urine-producing units, or nephrons, are

present to the number of about 200,000 in each rabbit kidney, and each nephron is composed of a Malpighian corpuscle and a long uriniferous tubule. The corpuscle consists of a collection of capillary-sized blood vessels,

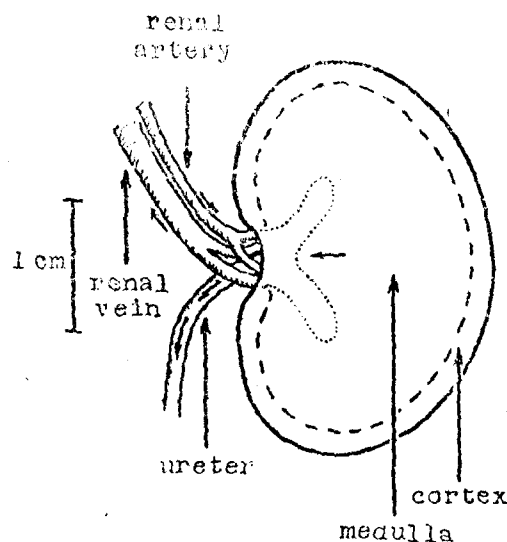


Fig. 15. - Diagram showing the left kidney, renal artery, renal vein and ureter of a rabbit. It also indicates the relative widths of the cortex and medulla at the mid-kidney level.

called the glomerulus, which is interposed between an afferent vessel and an efferent vessel and is enveloped by a double layer of very thin membrane, the inner of the two layers being wrapped closely round the individual capillary vessels. Between the two membranous layers is a space, or potential space, into which a protein-free filtrate is separated off from the blood in the capillaries; on leaving the space the filtrate passes in succession through the various parts (proximal convoluted tubule, descending and ascending loops of Henle, distal convoluted tubule) of the long urini-

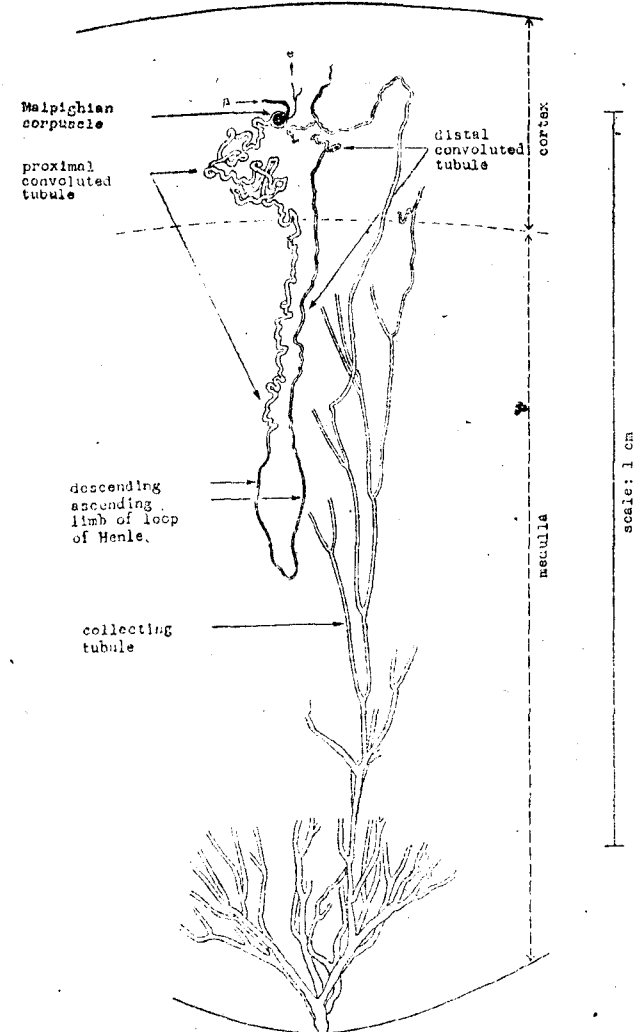


Fig. 16. - Rabbit's kidney. Scale-diagram (after Kenneth C. Strong) of a single urine-producing unit, or nephron, together with the larger urinary vessels with which the nephron is connected. In preparing his specimen, Dr Strong had necessarily to disturb somewhat the normal close relationship of the distal convoluted tubule and the afferent vessel of the glomerulus, which are naturally in contact with each other over a limited area.

ferous tubule. During this course, water and some substances are absorbed from the filtrate and returned to the blood, other substances are secreted by the tubule cells into the filtrate, and often its reaction is greatly changed; at the end it is urine.

With that description of the nephron or urine-producing unit of the kidney out of the way, we can describe the general vascular arrangements of the organ. The renal

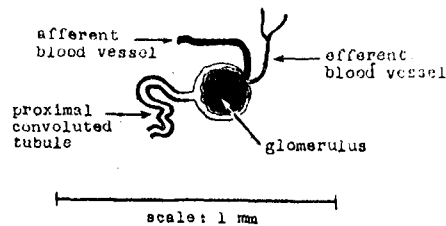


Fig. 17. - Rabbit's kidney. Scale-diagram of cortical glomerulus, to show its relations to the afferent and efferent blood vessels and to the first parts of the nephron.

artery, very shortly after its entry into the kidney substance, subdivides into interlobar arteries and these in their turn give rise to so-called arcuate arteries, from which interlobular arteries pass outwards through the cortex towards the surface of the kidney. The interlobular arteries give off afferent vessels to the Malpighian corpuscles and these vessels, as noted above, break up into collections of capillaries, styled glomeruli, which are thereafter drained by efferent vessels; usually there is one afferent and one efferent vessel connected with each glomerulus. The rest of the story varies according to the type of Malpighian corpuscle with which one is dealing.

In the case of a 'juxtamedullary' or 'basal' glomerulus, i.e., one of those lying deepest in the kidney substance, the efferent vessel is as wide as the afferent one, gives off but a few capillaries, and those near its origin, to uriniferous tubules, and then breaks up into a leash of straight vessels

(vasa recta), each about as wide as the parent one. After passing inwards for a variable distance, each of these vasa recta bends back and finally unites with an interlobular vein. From the bend of such a loop another vessel often arises, and passes inwards for a further distance before it too bends back towards the interlobular vein. So far as the evidence goes, it is by this vascular route, through the glomeruli of the relatively large juxtamedullary Malpighian

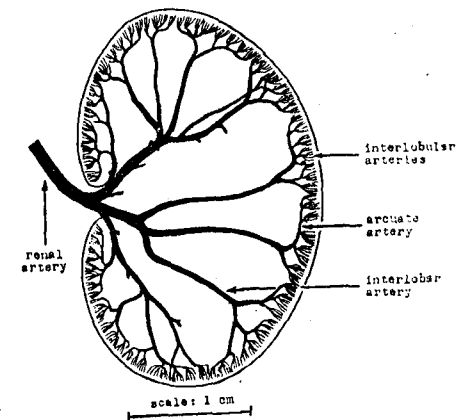


Fig. 18. - Diagram to show the distribution of the renal arterial system in a transverse section through the middle of the rabbit's kidney.

corpuscles, their efferent vessels, and the leashes of vasa recta, that a greater amount of blood than usual passes when the cortex is partly or wholly deprived of blood.

The efferent vessels of the remainder of the Malpighian corpuscles, styled 'cortical,' which are slightly smaller but very many times more numerous than the juxtamedullary or basal ones, are narrower than the corresponding afferent vessels, and break up into capillary plexuses round the uriniferous tubules; the plexuses are drained by small veins which open into interlobular veins. The rest of the course of the venous blood is the same for both vascular routes,

namely, into arcuate veins, interlobar veins, and finally the renal vein.

Much of the above account of the intrarenal vascular system derives from the work carried out at the Nuffield Institute, and it is true to state that, previous to that work, the proper synthesis of available anatomical knowledge (e.g., the findings of Lee-Brown, 1924) was very incomplete, especially in respect of the vasa recta and their drainage; this was mainly because a definite physiological directive was lacking until 1946. Doubtless there is further work to be done before all the details are known, and the account could even now be amplified. We shall, however, content ourselves with one functional addition before summarizing the general physiological conclusions of the work as a whole. The functional addition is as follows. Arteries are able to reduce their calibre as a result of the contraction of more or less circularly arranged muscle fibres in their coats. In the case of the renal vessels, reflexly stimulated by a tourniquet applied to the thigh or by other means, the main artery itself is seen to constrict, but in relation to the shift of the blood flow from the cortex to the medulla the vessels most concerned are the fine interlobular arteries and their branches to the glomeruli. These arteries do not constrict simultaneously throughout their length, but from the

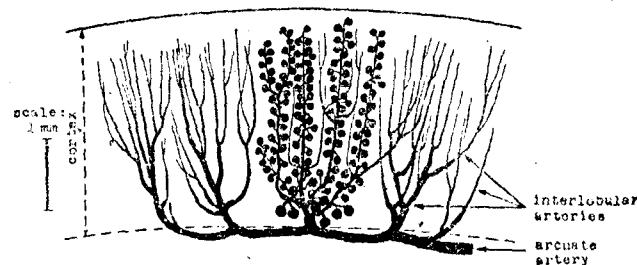


Fig. 19 - Diagram to show the relations of the arcuate and interlobular arterial supply to the glomeruli of the rabbit's kidney.

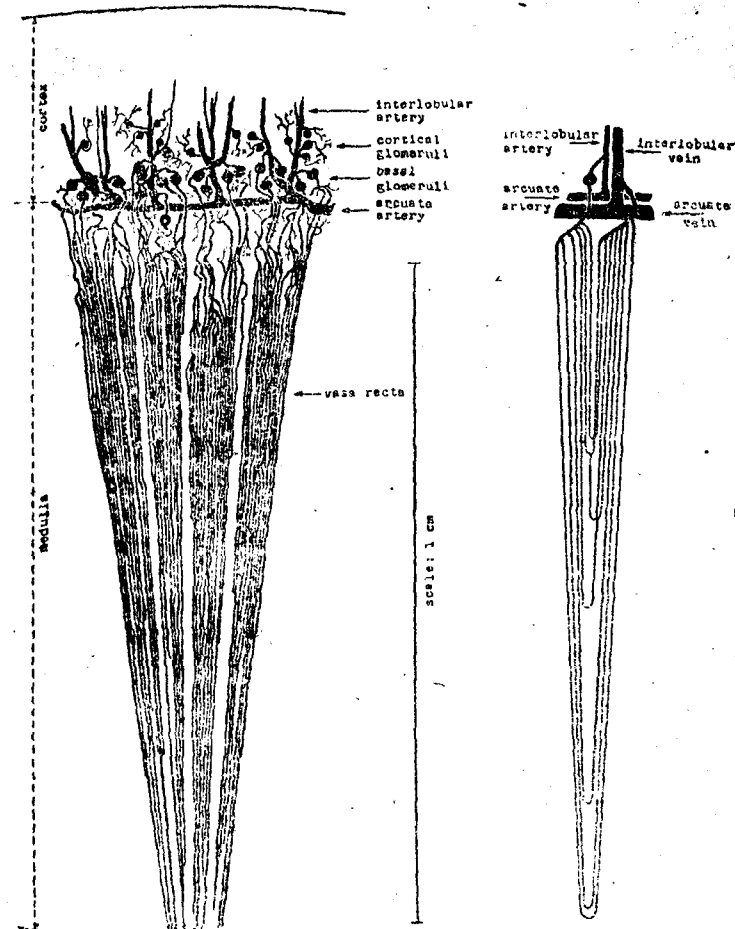


Fig. 20. - Rabbit's kidney. Diagrams to show the leashes of vasa recta which arise from the efferent vessels of the basal glomeruli (*left-hand diagram*) and the way in which the contained blood of one such leash passes into interlobular and arcuate veins (*right-hand diagram*.)

periphery, i.e., the part nearest the surface of the kidney, inwards (see Plates 11 and 12). As the process develops, more and more layers of glomeruli are deprived of their blood supply, the surface of the kidney becomes ever paler, and an increasingly greater proportion of the total renal blood flow passes through the juxtamedullary or basal glomeruli and the vasa recta associated with them.

With that essential point about the diversion explained, we can pass to the general physiological conclusions, which have been expressed in two forms. The first states that 'The course taken by the blood in its passage through the kidney does not show that degree of constancy which has commonly been imagined. In fact, the kidney has two potential circulations, a greater and a lesser, and in extreme conditions the blood may pass either almost exclusively through one or other of two pathways, or in less abnormal circumstances, to a varying degree through both. The vessels making up the pathway of the greater circulation are those associated with the cortical glomeruli; the channels of the lesser circulation are those associated with the juxtamedullary glomeruli.'

The second version is that 'The blood reaching the kidney has two potential routes through that organ and, according to circumstances, it may pass almost exclusively by one or other of these routes, or in varying proportions through each of them.'

'The two routes diverge where the afferent arterioles of the juxtamedullary glomeruli leave the interlobular arteries. One route, the medullary, continues through the juxtamedullary glomeruli, the efferent vessels of these glomeruli and their derivative vasa recta, to the interlobular veins. The other route, the cortical, continues through the interlobular arteries, the afferent arterioles of the remaining glomeruli, these glomeruli themselves, their efferent vessels and the cortical intertubular capillary network into which these break up, and finally through the veins draining this network into the interlobular veins. The rest of both

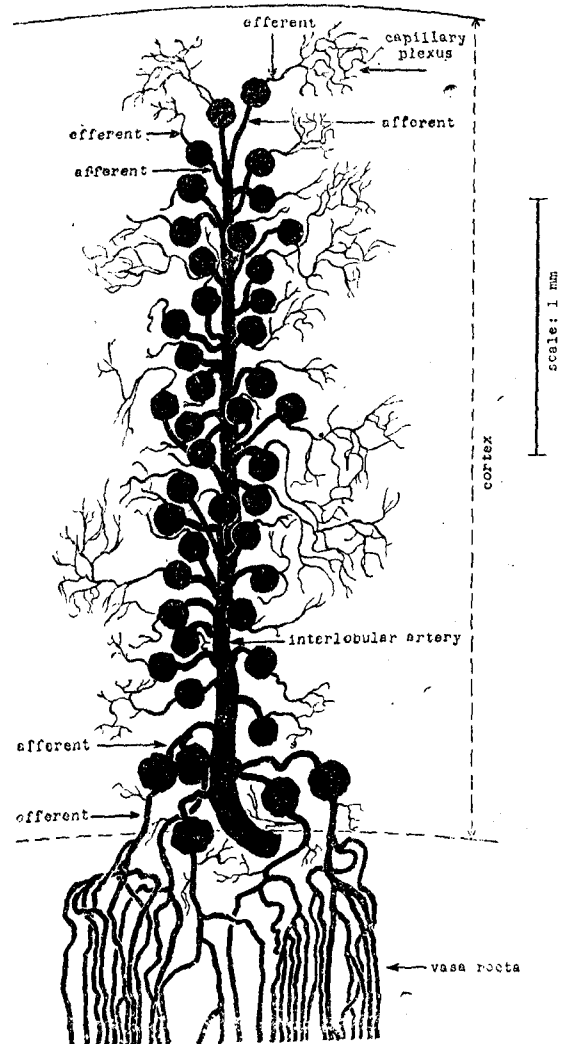


Fig. 21. - Rabbit's kidney. Diagram showing the efferent vessels of the cortical glomeruli breaking up into capillaries (round the uriniferous tubules), while the efferent vessels of the larger, basal glomeruli connect mainly with leashes of vasa recta. *N.B.* - For the sake of clarity, not all the glomeruli present are included in the diagram.

routes, like their beginning, is identical and is through ever larger venous trunks to the main renal vein.'

The rabbit, as we have seen, was the animal which was primarily used for the Nuffield Institute researches, but the physiology was extended to the rat and, to some extent, the cat, while the anatomy was checked not only in these extra species but also in the dog and in the human subject. Since then, American observers have extended the physiological findings to the dog and to the monkey, and there is even some evidence accumulating in respect of man. It may, therefore, be believed that in the upper mammals generally, under various circumstances, the renal blood flow may be diverted, in whole or in part, from the cortex and so render that portion of the kidney temporarily, and in varying degree, bloodless.

As it is somewhat cumbersome to describe the diversion in the above sort of way, American workers have invented for it the phrase, 'the renal shunt' or 'the Oxford shunt.' A valuable extra discovery in relation to it has been made by a Yale student, Mr J. H. Cort, working with Professor D. H. Barron in Professor John F. Fulton's laboratory at New Haven. For Mr Cort found that the shunt is controlled or influenced from area 13 of the cerebrum, i.e., from that supraorbital part of the frontal lobe (a part of the brain just behind the forehead) already known to be concerned in several visceral functionings.

A very simple way, discovered by the present writer, of demonstrating the shunt is to stop the air entry into the wind-pipe of an anaesthetized animal, the abdomen of which has been opened to expose the surfaces of the kidneys for inspection. With the rapid occurrence of oxygen lack ('suffocation'), the surface of the kidney is seen to pale and the blood is found to be diverted from the cortex to the medulla. With free entry of air into the lungs restored, recovery of colour occurs in a matter of seconds. Many experimental methods for bringing the shunt into operation cannot be used again too soon after the first occasion; this

X-RAY STUDIES OF THE KIDNEY CIRCULATION 121

method, however, can be employed several times in succession at quite short intervals. Administration of carbon monoxide to an anaesthetized animal (Pl. 13) also brings the shunt into operation, but in this case there is not the same reversibility.

The importance of the renal shunt

In the adult human being, it has been estimated, about 1,200 millilitres of blood (nearly four times the weight of the two kidneys together) pass each minute through the glomerular capillaries, and this amount represents about a quarter of the total output of the left side of the heart. If, therefore, the renal arteries contract down and the shunt comes into full play on both sides, a considerable addition is made to the blood supply available to other organs. Certain viscera from which the blood can be expelled for the supply of other parts were named by the late Sir Joseph Barcroft blood-reservoirs or blood-depots, but the renal cortex does not come into this category, for Sir Joseph stated that an organ can only be regarded as a blood-depot if it is clear 'that the blood is not in the organ because it is being used there,' and this postulate is obviously not fulfilled in the case we are discussing. The renal cortex can instead be styled a potential blood-donor and it seems, from the work of the present writer and of others, that the skin, skeletal muscles, non-pregnant or pregnant uterus, alimentary tract or parts of it, and possibly or probably other parts of the body, fall from time to time into the same category. Under certain conditions, there is a definite order in which the parts named give up a proportion of their blood flow, and it is clear that each such blood-donor organ must be able to exist for appreciable periods without oxygen.

In the case of the renal cortex, such lack of oxygen provides, according to some workers, the conditions necessary for the formation of a precursor of *hypertensin*, a substance which raises the blood pressure. So there is a possible direct link between the operation of the shunt mechanism,

if it is excessive in length of time, and hypertension (high blood pressure). In the present state of knowledge it would, perhaps, be unwise to lay too much stress upon this particular possibility, but the involvement of the kidney in the general story of hypertension does lend support to the hope, held by some at least of the members of the Nuffield Institute team, that the discovery of the renal shunt may assist in the elucidation of the problem presented by this distressing complaint. For the killing rate of high blood pressure, apart from its untoward effects in other directions, makes its prevention the outstanding objective of medical research; in the United States, for instance, the condition is said to be responsible for as many as one-third of all deaths.

Even if hypertension is for the moment excluded from consideration, there are still many physiological and pathological conditions in which the renal shunt is involved. There is no need to list all these here, for a preliminary survey has been published in *Studies of the Renal Circulation*. It is, however, of interest that letters continue to reach the authors of the book, stating that this or that observation, made by the particular writer and previously puzzling to him, is explicable when the shunt is taken into consideration. The field covered includes veterinary as well as human medicine.

When the cortex is rendered partly or wholly ischaemic, the formation of urine may be reduced or even, in the very severe cases, altogether suppressed. As many tests of kidney function were elaborated before the shunt was discovered, research has been begun to discover how the operation of the latter affects the results of the former.

The above are some indications of the possible importance of the Nuffield Institute results. We may hope that, in addition, they will assist in clarifying the functions of the various parts of the mammalian nephron, or urine-producing unit. For many of the current ideas about the kidney as a producer of urine are in large measure derived from research on amphibia (frogs particularly), and 'the physio-

logical analysis of the particular functions associated with the different portions of the tubule is still incomplete in mammals.* So possible new lines of approach are eminently desirable, and controlled variation of the blood supply, with concurrent observations of other changes, may well prove a useful procedure.

For reasons of space, the account given in this article is necessarily incomplete, and it certainly does not keep to the Editor's suggested title, namely, '*X-ray studies of the renal circulation*.' Perhaps, however, it is of greater interest for that very reason. For it shows how radiographic studies of the intact animal provided a starting-point for a variety of functional and structural studies. So it stresses the stimulating power of radiographic and cineradiographic procedures upon developments in other fields, and it is this vivifying action of X-ray techniques to which all who have worked in the West Wing of the Nuffield Institute feel bound to testify.

*F. R. Winton and L. E. Bayliss. *Human Physiology*. 3rd Edn., 1948, p. 244. London: J. & A. Churchill, Ltd.

recording instrument was laid at a depth of 110 feet off Pendeen on the Atlantic coast of Cornwall. Two others of different type are at 70 and 35 feet off Perranporth. Pressure changes as recorded naturally represent the sum of many different wave movements, some of near and others of more distant origin, and analysis is necessary to sort out the contributory wave systems. At any moment, waves of a number of different periods will be arriving. These will come, in general, from different storm centres — or, if from the same storm, then at different phases in its development. The problem in identifying any particular storm centre is to pick out those waves which, arriving at different times in Cornwall, could earlier have been in the same place at the same time, and so could have had a common origin.

The most precise check was provided by a tropical storm off Florida at a distance of about 2,800 miles. It was selected for examination because, being of compact structure, it provided a more accurate check of the calculated rates of travel of the waves. This has confirmed the theory already outlined, within the probable accuracy of the measurements made. It showed also that, given enough meteorological information, any wave system can be traced back with some confidence to the storm which produced it. Also, since the longest waves travel fastest, storms with the highest wind speeds should produce waves with the longest periods. Thirteen storms, for which many reports of wind strength were available, have been examined in this connection. It has been found that up to a wind strength of 9 on the Beaufort scale (41-47 knots) wind strength and greatest wave period increased together. But in the case of those storms which produced the longest waves, no correspondingly greater wind speeds were reported. The most likely explanation is that, whether from fear of exaggeration or otherwise, there is a tendency on the part of ships' masters to 'write down' the highest wind speeds. If this is the correct interpretation, it follows that the analysis of wave periods in Cornwall may give more reliable informa-

tion about the frequency of the most severe storms in the Atlantic than is otherwise available.

Barber and Ursell. *Phil. Trans. A.*, 1948, 240, 527

Ocean Sound

A further example of oceanographical research, this time from the United States, is the discovery that at a depth of about three-quarters of a mile beneath the surface of the oceans there is a natural sound-transmitting layer within which sound waves are indefinitely confined, and so travel with comparatively small loss of energy for great distances. As a result, the explosion of a 4-lb. bomb or one even smaller at the right depth can be detected at a distance of up to two thousand miles. The effect was demonstrated during the war by the Woods Hole Marine Biological Station, Massachusetts, and is now being applied by the United States Navy to provide an experimental distress-signal service on the ocean air-route between San Francisco and Hawaii.

The existence of this sound-transmitting layer is due to a variation in the speed at which sound travels at different depths in sea water. This is affected by both temperature and by pressure, but the temperature effect is the greater. Measurements over a twelve-miles range in the English Channel in 1923 showed that sound travelled 3 per cent faster in summer water at a temperature of 16.95 degrees centigrade than in winter water at a temperature of 7 degrees C. (1,510 metres per second, as against 1,471).

Since temperatures in the oceans decrease with depth, except near the Polar regions, the first effect is a reduction in the speed of sound as depth is increased. But pressure, naturally, also increases with depth and this produces an opposite but smaller effect. Adding the two together, there is a comparatively sharp increase in the speed of sound near the surface, where the temperature effect is the more important; but at greater depths the pressure effect takes control and the speed of sound begins, more gradually, to increase.

In between, there is a layer in which sound travels most slowly, and it is within this layer that long-distance transmission can take place. An advancing wave disturbance, directed slightly upwards from this layer, would be bent down again into it, because its upper edge would be travelling faster than its lower edge; while a wave disturbance which tried to escape downwards would be similarly bent up again into the critical layer. Waves originating in an explosion within the layer are thus spread out only horizontally, and not vertically as well, and can therefore be detected at very much greater distances than would otherwise be possible. A further and convenient peculiarity, which here can only be stated as a fact, is that the initial signal at a distance is comparatively weak, but builds up into a 'kettledrum' finale which can be used for timing measurements.

The use of these effects as the basis of a warning system for the location of aircraft lost at sea is due to B. W. Stifler of Engineering Research Associates and W. F. Saars of the U.S. Navy Underwater Sound Laboratory, Connecticut. One such station has already been installed in Hawaii by the U.S. Navy Electronics Laboratory, San Diego, and tested against explosions at the required depth—in this latitude about 700 fathoms—off the coast of California. Two more stations are to be built on the west coast of the United States and the three together should fix the position of any underwater explosion which takes place at the right depth between San Francisco and Hawaii. The name 'Sofar' has been given to the system, and under the conditions proposed the overall time to obtain a 'fix' should be between 30 and 45 minutes.

It should perhaps be added that the route selected is favourable to the system in several respects. In relation to the need for distress-warning, it is exceptional both in the length of ocean crossing and the comparative scarcity of sea traffic: in relation to operation, the fact that the route is sub-tropical and predominantly east-west in direction

means that the depth of the critical layer should be more nearly uniform than would otherwise be the case.

Stifler and Saars, *Electronics*, 1948, 21, 98

Radar and Meteorology

From the sea it is natural to pass to the air, and this brings us to the Ministry of Supply's Telecommunications Research Establishment at Great Malvern where a demonstration was lately given of the measurement of raindrop size and cloud height by the application of radar methods. Both have progressed a long way beyond the simple wartime discovery that tropical rainstorms could give rise to 'spurious echoes' on radar display tubes; and that they could be detected and plotted in much the same way as could aircraft and ships.

In the case of raindrop size, a theory has been worked out connecting the strength of the echoes produced with both the size of the reflecting particles and also the radio frequency used, and to test this theory a simple drop impactor has been designed, to be carried in the nose of the aircraft. The speed of the aircraft being known, the impact of individual drops on a microphone gives a measure of their size: and from a comparison between these aircraft observations and simultaneous radar echoes from the same rainstorm, a direct test of the theory will become possible. Finally, when the correct relationship has been established, it should be possible to measure the size of drops by radar methods only.

Even with the shortest radio wavelengths which can be used, only particles of raindrop size give appreciable echoes, and for the charting of cloud heights in general it has been found necessary to make use of the shorter wavelengths of visible light. Cloud heights, up to now, have been measured by two different methods, neither of which is capable of continuous operation. The procedure by day has been to release pilot balloons, watch the disappearance of the balloons into the cloud from not less than three points

and then calculate the height of the cloud by the same sort of methods that are used by surveyors. By night, the method of converging searchlights is used and, again, the co-operation of at least three parties is necessary. The new equipment has been designed to do both jobs, and to give a direct measurement of cloud height above the observer. It consists, essentially, of a pulse-operated searchlight working on the same principle as that of a radar station. A succession of high-intensity light-flashes is directed vertically upwards through a telescope, and the time-interval is measured before the 'reflection' of these flashes from the cloud. Apart from the particular application, the method is also of interest as an example of the combination of techniques. The idea of pulse-transmission and echo-timing comes from radar and, earlier, from sounding. The high-intensity light flashes used were greatly developed during the war as an aid to night photography. And 'photomultipliers' are used in reception, of the same kind as those which have lately been employed by astronomers to measure the light-intensity of stars.

Thunder-cloud Exploration

Thunderclouds were investigated from a somewhat different point of view during the war by a joint U.S. Army-Navy research team and an analysis of the results has lately been presented by Dr Ross Gunn of the U.S. Weather Bureau. The observations which were made involved aircraft flights through nine different thunderstorms, successive traverses at different heights being made on a number of occasions, and during one flight the investigating aircraft was itself struck by lightning. The figures obtained show that considerably stronger electric fields are common inside thunder-clouds than had been suggested by earlier measurements with sounding balloons, carried out from Kew Observatory.

The investigation was known as the Precipitation Static Project and had its headquarters at Minneapolis. It had at

its disposal three aircraft, seventeen physicists and engineers, a hangar with sufficient clearance to enable high-voltage experiments to be carried out on full-size aircraft suspended within it, and recording instruments designed for the project by the U.S. Naval Research Laboratory. The motive was to discover ways of reducing the radio interference commonly experienced on aircraft sets when flying through cloud, and the thunderstorm investigation was incidental.

Radio interference was found to be due to the building up of high electrical voltages on the aircraft, often reaching more than half a million volts, and the development of the luminous or corona type of discharge from its surface. Charge was most rapidly built up when flying through extended layers of dry snow or ice. Under these conditions, the building up of charge was due to friction between the snow or ice and the aircraft. Alternatively, charge could be given to the aircraft by the impact on it of water droplets which themselves carried an electric charge. It was this second process which led to the thunderstorm part of the investigation. Previous ideas of the rate of change of electrical pressure within a thunder-cloud were based, as already mentioned, on experiments carried out at Kew. These led to the conclusion that electric pressures more than about one-hundredth of that at which a lightning stroke would develop were rare within thunderclouds 'except in relatively small regions in which discharges originate.' By contrast, these American measurements suggest that electrical fields approaching the flash discharge level are common, and not the exception, within thunder-clouds. On the occasion when the aircraft was struck, the rate of change of electrical pressure in a vertical direction was 3,400 volts per centimetre (or about 8,500 volts per inch) just before striking. In the same cloud almost equally strong electrical fields in the reverse direction were twice encountered, and in nine thunderstorms the greatest vertical field averaged 1,300 volts per centimetre. A good deal of evidence was also accumulated suggesting that a considerable fraction of

the total electrical gradient was often in a horizontal direction, so that the figures quoted should be underestimates of the total gradients encountered during these aircraft flights.

Gunn, *Journal of Applied Physics*, 1948, 19, 481

Tracer Research and Industry

The chief impression formed by those who visited the Ministry of Supply's second major research station - the Atomic Energy Research Establishment at Harwell near Didcot - was of the large amount which could be seen, and the large amount of information which could be given, without that 'disclosure of secrets' which Sir John Cockcroft, director of the establishment, assured his visitors would not take place. As to the main object of the station, the position can be simply enough stated. The principal facts about the release of atomic energy have long since been published, the problems raised by its proposed peacetime application as a source of power are technological, and what progress may have been made towards their solution would be unlikely to be of general interest at the present stage, even if no security ban on its publication were in force. There is a personal interest in a first sight of an atomic pile in operation, although as one of the Harwell physicists suggested 'it would be more impressive if it made some noise': but little to be gained, at second hand, by describing it as a concrete-surrounded cube suggestive in height of a three-story building. Yet apart from the control-room, which looks much like any other control-room, this is all that is to be seen.

There remains the work of Harwell as a research laboratory in nuclear physics, on which good progress has been made in the installation of equipment; the elaborate health services for the protection of staff and public; and the arrangements for the production of radioactive materials for research use in universities and in industry and the plans which have been made for the servicing of such research. There is an obvious and genuine enthusiasm for

this side of Harwell's activity, not only because it can be freely discussed without restriction, but even more because it is the one immediate and certain benefit which atomic energy research can contribute. Power is for the future: the use of radioactive materials as tracers in research, chiefly biological and chemical, is an achievement already realised, and awaiting only the increased supply of materials for its great expansion.

Harwell's proposed industrial service is perhaps its most novel contribution. As well as producing the radioactive materials needed, and advising on their use and the type of problems which they may help to solve, the establishment will also be prepared, when asked, to send out its own research teams and undertake whatever industrial investigations may be required for any firm which lacks the specialised staff necessary. This should be a most helpful procedure, for only through mutual discussion are the full applications of the tracer method likely to be appreciated.

Its basis, as already described in *Science News VI*, is that radioactive atoms take part in the same chemical and physical changes as do normal stable atoms of the same chemical element; but because of their radioactivity, can be followed, with no room for argument, wherever they may go. With every chemical element potentially available in radioactive form, and the majority capable of ready preparation, it will be evident that the method is of very wide scope indeed, and it has been suggested that, with adequate supplies available, it offers enough work for a full generation of research on problems which can be immediately visualised. This can scarcely be an exaggeration, and the danger, if any, is that it may be employed uncritically and for its own sake instead of being treated rather as one more addition, though extremely useful to research methods. University laboratories are already aware of what can be done and have been held up only by limitations of supplies, and the same position is also true of industrial firms which possess adequately staffed laboratories. The value of the visiting

service to be provided will lie in the more rapid extension of the method.

Most of the radioactive materials required are made by irradiation in aluminium containers in one or other of Harwell's two atomic piles. In the 'gleep' – the low-energy nuclear pile first installed at Harwell – these are small cylindrical tubes of about the same cross-section as a half-penny, and they are inserted in graphite blocks which are drilled with holes of the same size as the containers. Usually the radioactive atoms are of the same chemical element as that irradiated, and no process of chemical separation is therefore necessary. In these cases no further processing is required, and after removal with long forceps from the graphite blocks, the aluminium cans are transferred as they stand to lead containers for transport.

A further development is the separation, also for research use, of 'heavy carbon' or carbon-13. This is not radioactive, and differs only in its slightly greater mass from normal carbon-12 with which it is mixed in any natural source in the proportion of about one part in a hundred. Several different methods of separation are possible. The one used at Harwell depends on a slight difference of less than one-tenth of a degree centigrade between the boiling-points of carbon monoxide – which is a gas at normal temperatures – according to whether the carbon atom in this compound is 'heavy' or normal. The plant should be complete by the time this note is published, and is designed to produce carbon-13 at a rate of about 0.3 grammes per day. This amount, small though it is, will probably be sufficient to meet immediate research needs, and work on other such separations is proceeding. Carbon-13 fulfils the same purpose of acting as a tracer as do the radioactive forms of normal chemical elements; but special apparatus is required for its effective use, and it is for this reason that the present limited output is judged sufficient.

Crystal Amplifier

A possible substitute for the radio valve, in many applications for which there is no present alternative, has been described by Dr J. Bardeen and Dr W. H. Brattain of the Bell Telephone Laboratories. It is a development of the metal rectifier or crystal and 'cat's whisker' of early radio receivers, so modified that it will act also as amplifier or oscillator. The extension of use is of the same kind as that achieved in the case of the radio valve by Lee de Forest's introduction of the triode in 1907. The earlier diode, invented by Sir Ambrose Fleming, acts as a valve in the literal sense of the word, allowing electrons to pass across it in one direction but not in the other; but without de Forest's third electrode, a normal radio valve cannot be used either for the amplification of weak signals or as a source of oscillations. Similarly, in the 'cat's whisker' type of rectifier, electrons can pass more readily from the crystal to the 'whisker' than in the reverse direction, and like the diode it has therefore a valve-like action.

There was a demand for improved rectifiers of this type during the war for the rectification of currents of very high frequency, and it seems to have been from this development that the new crystal amplifier has originated. In these improved rectifiers, the crystals used were of the metal germanium, and the 'whisker' was a tungsten wire. Germanium is usually described as 'rare', but is found in the flue dusts from certain coals in the proportion of about 0.1 per cent, so that if the demand arises effective supplies could be readily increased. In the new crystal amplifier, to which the name 'transistor' has been given, the same materials are used, but in place of one 'whisker' there are two. These are placed close together – at a separation of between 0.005 and 0.025 centimetres – on the same side of the crystal, and the amplifying action depends on the interaction between them. This in turn depends on the existence of a thin surface layer with different physical properties from that of the main body of the crystal.

Practically, the arrangement has been shown to give a 100-fold amplification, with a maximum power output at present of about 50 milliwatts. This is well below the level required in the final stages of a domestic radio receiver. On the other hand, being all-metal, unheated, non-vacuum and of simple construction, the transistor should have a long working life and, provided that stability of performance can be ensured, it should be capable of many uses. As an example, there is an obvious demand for an amplifier which can be installed in a telephone circuit and there left for long periods without need for the inspection or replacement of components; and a similar requirement arises in the case of many of the electrical control devices which are possible in industry. For some such purposes the transistor may prove the best solution.

Alternatively, its existence may act as a further stimulus to research on longer-life valves, on which work is already proceeding; and in the end it may be expected that both the conventional valve and the transistor will have their uses. In Great Britain, research on long-life valves is being carried out at the Post Office's Engineering Research Station at Dollis Hill, and in the United States a conference on 'electron tubes for instrumentation and industrial use' was held lately in Philadelphia by the American Institution of Electrical Engineers.

Bardeen and Brattain, *Physical Review*, 1948, 74, 230
Electrical Engineering, 1948, 67, 740

Island Telephony

A Post Office development of more immediate application is the production of a battery-operated radio-telephone. It has been designed by the Post Office Engineering Research Station and is intended for installation as an individual-subscriber equipment to meet the needs of those living on the western islands of Scotland or other remote sites, from which a direct view to a site accessible to telephone cables can be obtained. There has been considerable development

already in the installation of normal radio-telephone links. These either act as substitutes for extra cables or serve to connect outlying groups of subscribers to the main system. What is new in the present equipment is the idea of extending the same facility to an individual subscriber, and without the need for any special procedure at the mainland end. Normal dialling is provided, and with the equivalent of twelve 3-minute calls a day, the dry batteries provided should enable the equipment to operate without maintenance for six months. Receiver and transmitter are continuously energised only when a call is in progress, but every thirty seconds the receiver listens for incoming calls. The set operates on a wavelength of between 3.75 and 4.3 metres, and there must be many more remote parts of the world in which similar equipment could be of service.

Rats and Heredity

The controversy over the suggested inheritance of acquired characteristics is an old one in the history of science, and for nearly twenty years an experiment on the training of successive generations of rats, which was carried out by Prof. W. M'Dougall at Duke University, North Carolina, has remained as a wraith disturbing the accepted view that acquired skills cannot in fact be inherited. Prof. M'Dougall's experiment was continued effectively through thirty-two generations of rats and his conclusion was that learning capacity was increased by training. It was repeated with negative results by Prof. F. A. E. Crew at the Institute of Animal Genetics in Edinburgh, but with no more satisfactory explanation than that in the case of Prof. M'Dougall's rats some unsuspected form of selection must have been at work. An even more protracted repetition has now been carried out by Prof. W. E. Agar and others in the Zoology Department of the University of Melbourne, and the results serve as a general warning of a difficulty which may arise in any form of experiment which depends on the maintenance of a 'pure' line of animals through any

large number of generations. His conclusion, in brief, is that however closely inbred the original stock may have been, it cannot be relied upon to remain unaltered.

The essence of the original experiment, which was retained in both repetitions, was the introduction of rats at regular intervals into a water tank from which two alternative ways out were provided. Each exit was up a wire ladder, but one of these administered an electric shock to a rat ascending it, whereas the other did not. The electrified ladder might be either on the right or the left in any particular test; but, whichever side was dangerous, that side was always illuminated and the safe ladder remained dim. The learning accomplishment required of the rats was accordingly to associate light with liability to shock and absence of light with a more comfortable exit from the tank.

The essential addition to the experiment in the Melbourne repetition was the maintenance of two independent lines of rats - breeding in one line being from rats which had been trained, and in the other from rats which had not. A number of rats from the untrained line were tested at each generation, like the others, for their learning capacity, but these rats were not used for breeding. There was thus a comparison at each generation between rats, whose ancestors on both sides had been trained, and control rats, none of whose ancestors had been trained. The interest is in the contrasting results obtained at different stages of the experiment. Up to the twelfth generation, the untrained as well as the trained rats showed an improvement, and at about the same rate. From that point up to the 25th-28th generations, both lines continued to improve, but the trained line did progressively better. At this stage, therefore, Prof. M'Dougall's results could have been regarded as confirmed. Finally, from the 29th to the 36th generation, both lines learnt more slowly and the difference between them disappeared.

Various other differences were also noted. The most obvious was a difference in body size, rats of the trained

line being both bigger and heavier. There was also a concealed difference in colour-pattern. This was discovered through the use of a number of the rats bred from these lines in other experiments. It showed itself in crossing with piebald rats and, like the divergence in weight, has maintained itself, once established, through successive generations. Further, these differences in size and colour-pattern have been shown to be independent, and no connection has been found between size and learning rate. The interpretation, therefore, is that three independent differences of a hereditary kind have appeared, one temporary and two permanent; and the general conclusion is that any experiments based on the supposed fixity of inbred lines of rats, mice or any other laboratory animals must be treated with caution. Since many medical and biological experiments are carried out which involve an assumption of this kind, it will be obvious that Prof. Agar's laborious investigation must be of wider interest than its original purpose, and he has stated in his last report that it is being continued.

Agar, Drummond and Tiegs, *Journal of Experimental Biology*, 1948, 25, 103

Ultra-Microscopic Surface Topography

(Multiple-Beam Interferometry)

PROFESSOR S. TOLANSKY

Introduction

SINCE 1862, when the French physicist Fizeau introduced his celebrated interferometric procedure, the use of light-waves for the study of the sub-microscopic shapes of surfaces has had considerable application, and in fact, Fizeau's simple method is to-day widely used in numerous precision tool and optical workshops, both as a means of gauging distances and examining surfaces. The reason for this lies in the smallness of the length of a light-wave. The distance between the crests of visible light-waves (say green light) is some $1/50,000$ th of an inch, and it is possible to arrange to utilise this small distance as a unit of measurement.

It is clear that the unit which is to be adopted in any measurement depends upon the length to be measured. Thus, for example, distances between towns are measured in miles, the length of a football field in yards, a line drawn on a sheet of paper in inches, the diameter of a piece of wire in thousandths of an inch; but after that, what? For many technological purposes it is necessary to have very smooth or flat surfaces on which 'hills' and 'valleys' should not exceed, say $1/200,000$ th of an inch, a quantity too small to be measured even by a microscope. It is here that the light-wave is used as the yard-stick, for the small quantity just mentioned is a quarter of a light-wave, and in any form of practical measurement with light-waves, one can measure to at least a twentieth of the unit adopted. The manner in which this is achieved will be described in detail later, but

ULTRA-MICROSCOPIC SURFACE TOPOGRAPHY 141

the limitations even of the light-wave as a unit will first be considered.

How far are we likely to wish to push the limit of mechanical measurement? It is known that the major number of solids have an ultimate crystalline structure, and the separations between the atoms comprising the crystal units have been found by the use of X-rays. These distances are very small and it is convenient to use a small unit to describe them. This is called the angstrom unit (written A.U.) and is one two-hundred-and-fifty-millionth part of an inch ($1/250,000,000$). On this reckoning, a green light-wave has a wavelength of 5,000 A.U. The separations of atoms in crystals vary from perhaps 1 to 5 A.U., and it might thus be expected that there can be irregularities on surfaces as small as about $1/1,000$ th part of a light wave (a ridge one atom high). What, then, is to be the way to enable one to measure such small surface irregularities?

One indirect method is that called *electron diffraction*, in which electrons are reflected from the surface, and from the manner of reflection information as to surface structure can be derived. Another indirect method, which gives great magnification across the surface, is to use the electron microscope, which is a costly and elaborate technique. A third method, that developed by the author, is called multiple-beam interferometry, in which light-waves are so modified that an irregularity of $1/1,000$ th of a wave can be detected. This has the merit of being simple, direct, inexpensive and accurate. The magnifications easily attainable are of the order of half a million, and height changes of the order of 5 or 10 A.U. can be evaluated. The procedure is complementary to that of the electron microscope in that the interferometric method gives measurements of *heights and depths* whilst the former gives measurements in *extension* across the surface.

Interference of Light

When a beam of light strikes a transparent thin film (e.g.

a soap bubble, or a thin sheet of cellophane) part of the light is reflected from the front face and part from the back face of the thin film or sheet. If the film is thin so that the faces are very close together, the effects due to interference between the two reflected light-beams are easily seen. The light reaching the eye from the two faces has travelled different distances, and if the crests of the waves from the one fall on the troughs of those from the other, the two rays 'interfere' and annul each other. It is this interference which leads to the coloration of soap bubbles, since according to the thickness, different colours are reflected or annulled.

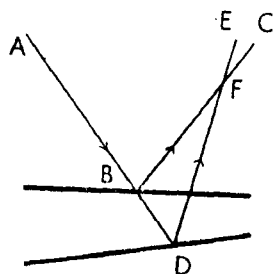


Fig. 22

The way in which this simple principle is applied is shown by Fig. 22. Here we have two flat glass plates inclined at a very slight angle, and these are illuminated from above by monochromatic light (i.e. light of one colour or wavelength, say, 5,000 A.U.) The ray AB partly reflects to C and part carries on to D and thence to E. The two rays are reunited at F, and when the difference in optical retardation between the two waves is half a wavelength (or $1\frac{1}{2}$, $2\frac{1}{2}$, $3\frac{1}{2}$, $4\frac{1}{2}$ etc. light waves) the two beams annul. When it is a whole number of wavelengths they reinforce. As one moves along the air wedge between the plates from the apex, one successively passes positions of light and darkness, and a change from one dark to the next dark position means a

ULTRA-MICROSCOPIC SURFACE TOPOGRAPHY 143

change in air film thickness of half a light wave if the light be directed perpendicularly on the surface of the glass (half, since the light has gone to the bottom and returned to make the whole wave difference.)

In general, the light and dark regions appear as bands,

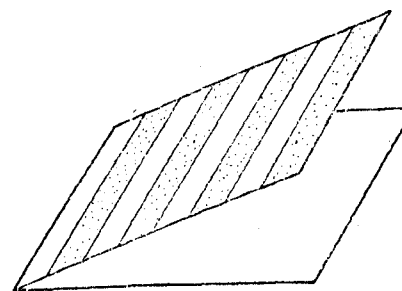


Fig. 23

such as are shown in Fig. 23, and these are called *interference fringes*. In this figure the wedge apex has been set opposite to that in Fig. 22. Owing to the nature of light vibrations, they are not sharp-edged but fall off in intensity,

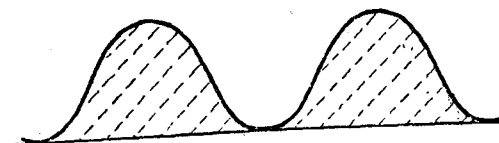


Fig. 24

as shown schematically in Fig. 24, in which an attempt is made to show how the light and dark spaces are equal. This intensity-shape considerably restricts the accuracy with which measurements can be made, and an evaluation to about one-tenth (at best, one-twentieth) of a fringe separation is often all that can be achieved.

The fringes produced by two inclined flat surfaces are straight lines, but if one surface is flat and the other has

some distinctive topographical shape, then the fringes become contour lines of this shape, exactly as in the familiar contour map of geography. Suppose one surface (the reference plane) is flat and the other spherical, then the trace of all the points which are the same distance from the flat surface forms a circle. Thus a succession of circular ring fringes appears. They are in fact the contour lines of a sphere, and their distribution can be visualised if we picture those successive rings which are obtained by cutting the sphere in parallel equidistant slices which are separated from each other by half a wavelength of light. Those familiar with the interpretation of geographical contour maps will easily be able to interpret contour interference fringes. It is only necessary to remember that the contour lines correspond each to a height change of half a light wave, i.e. $1/100,000$ th inch (or alternatively 2,500 A.U.) if a wavelength of light of 5,000 A.U. is employed.

Multiple-beam interference

As already indicated, the use of simple interference for

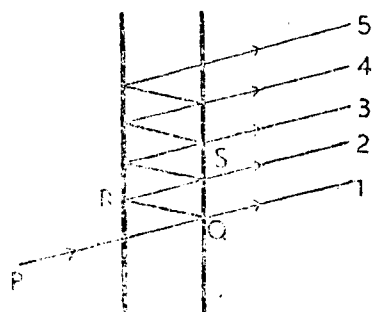


Fig. 25

topographical studies is over 80 years old, but recent developments of this classical technique have considerably extended the possibilities and power of the method, and

ULTRA-MICROSCOPIC SURFACE TOPOGRAPHY 145

many advances have resulted. These new advances are based upon the correct application of what is called *multiple-beam* interferometry, as distinct from the older method which used *two beams* only. It is possible to deposit a thin layer of silver on to a surface such that it can reflect about 90 per cent of the light falling on it and transmit about 5 per cent, the remainder being absorbed. If two such surfaces are brought *very close together* (less than $1/2,000$ th of an inch) and illuminated with monochromatic light (light of one wavelength or colour) the interference fringes formed by the plates are profoundly modified by the high reflectivity. Consider what happens if the plates are parallel as in Fig. 25. A beam from P on reaching Q transmits a small fraction but most is reflected to R, and from R to S at which the phenomenon repeats. Clearly there emerges a set of beams 1, 2, 3, 4, etc. of diminishing intensity. If these are collected with a lens, interference results at the lens focus, where all the beams meet.

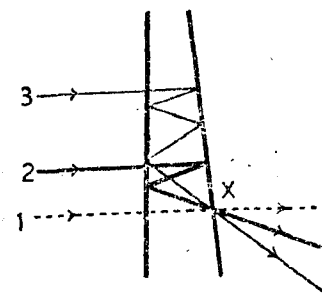


Fig. 26

If the two surfaces are inclined as in Fig. 26, then in like manner at the point X there will arrive beam 1, then beam 2, after two reflections, beam 3 after 4 reflections, etc., and at this point X, interference takes place, this time also with multiple beams.

If the reflectivity is high, then a very large number of beams of comparable intensity arrives at the point X. Calculation shows that the result of this is to produce highly sharpened fringes which are so fine and narrow that it is possible to use them to measure surface alterations of even only $1/1,000$ th of a light wave, i.e. 5 A.U. This condition is only arrived at if the reflectivity is high, the separation of the two surfaces very small, and the light illumination critically directed perpendicularly on to the surfaces. The necessary high reflecting layer is obtained by evaporating a thin film of silver on to the surfaces in a high vacuum. The conditions of cleanliness are critically important, as is the thickness of the silver layer, which should be about 500 A.U. thick. A technique has been developed in which the silver exactly contours the underlying surface and this silver deposition introduces no errors in deriving the surface topography. That this is indeed the case has been the subject of exhaustive tests.

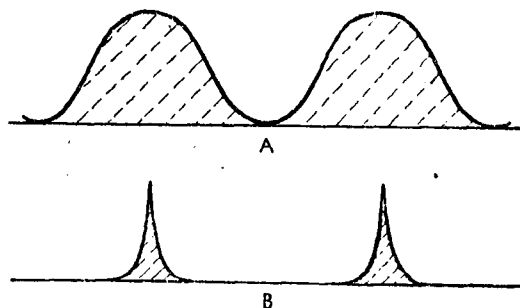


Fig. 27

The profound modification in fringe shape produced by the multiple beams is indicated schematically in Fig. 27, which compares (A) two-beam with (B) multiple-beam fringe shape.

Plate 14 shows the multiple-beam rings which can be produced by an arrangement such as in Fig. 25. These are

ULTRA-MICROSCOPIC SURFACE TOPOGRAPHY 147

included to illustrate simply what is meant by fringe sharpening, and it will be recognised that the fringe width is only a small fraction of the separation between fringes. Strictly speaking these rings are not exactly contour lines of the type under discussion, but their fringe shape is the same, and by no other means can perfect circles be obtained to illustrate the fringe width.

Topographical applications

Newton's rings

It has long been known that if a spherical lens is placed on a flat glass surface, interference rings (so-called Newton's rings) appear, and indeed such circular rings are the contours of the spherical lens. Now if the lens and glass plate are suitably coated with silver, held close together and correctly illuminated, multiple-beam fringes form as shown in Plate 15. [The light used is the green light filtered from a mercury lamp.] This reveals two features. The rings are sharp and are almost circular contours, but a closer inspection shows too that the contour lines exhibit small irregularities. These are in fact the slight defects of polishing on the glass surfaces, leading to small alterations in height, which the contour lines reveal. Now in moving from one fringe to the next, the surface height has changed by half a light wave, i.e. 2,500 A.U. The fringes are so inherently sharp that a wriggle of less than $1/100$ th of a fringe separation can readily be measured. Such a wriggle thus represents a height change of less than 25 A.U. One is immediately struck by the fact that this is approaching molecular dimensions; in fact many common molecules are larger than this, and it is indeed surprising that with such ridiculously simple means one can virtually 'see' height changes which are of molecular depth. It is clear that this simple device opens up a means for the precision inspection of surfaces, and a representative group of some of the results recently obtained by extensions of this procedure will now be considered below.

Quartz crystal

An interesting case of surface shape is illustrated in Plate 16. This shows the topographical features of a natural *unpolished* face of a quartz crystal. The length of the surface reproduced is about $\frac{1}{4}$ of an inch. In this picture only one kind of light-wave has been used and the fringes appear single. It is at times advantageous to use several wavelengths since this assists in the interpretation of the contours. Those familiar with contour map reading will recognise that the feature on the left corresponds to a ridge with very slightly curved sides, the central region is flat and the upper right region cylindrical. It should be noticed that the whole complex topography of this natural crystal surface is encompassed within a height range of only four light-waves, i.e. $1/12,500$ inch.

Attention is drawn to the fact that these interference fringes are produced in *transmission*, and this necessarily implies that the two silver surfaces be deposited on transparent bases. It will be shown later that the technique of multiple-beam interferometry can be extended to the study of opaque substances by placing a silvered glass surface on a silvered opaque surface, and this opens up a very important field, the examination of metal surfaces. To do this, the light must be *reflected* from the surfaces, not transmitted. Calculation shows that whilst transmitted fringes are sharp *bright* lines on a dark background, reflected fringes on the contrary are sharp *black* lines on a bright background. Such reflection fringes can also be obtained with transparent bodies (they are, of course, the only kind possible with a metal) and Plate 17 shows such reflection fringes for the same quartz crystal surface as given by Plate 16. At this point it should be noted that the fringe patterns in Plates 16 and 17 are not identical. This difference arises from the particular angle which has been set between the two interfering surfaces, and by adjusting this angle, fringe patterns are seen which superficially appear to be very different. However, the ultimate analytical interpretation in such cases always leads

ULTRA-MICROSCOPIC SURFACE TOPOGRAPHY 149

to an identical result. Attention is drawn to the sharpness of the fringes in Plate 17 which easily permits the recognition of surface ruts which are only 20 silica molecules deep. Ruts 2 molecules deep could have been identified, if existing, providing their *widths* were not less than $1/5,000$ th inch.

Cleavage of mica

It has been known for many years that mica crystals can be cloven into large thin flakes which have what appear to be highly perfect surfaces. The fringes given by such a surface are shown in Plate 19 and are of much interest. Two main features may be noticed, namely, (a) the surface is covered with 'hills' and 'dales,' for the contour looks like that of a mountainous country, (b) sharp discontinuities are clearly visible. These are cleavage 'steps' and from the slight displacements of the fringes the step value can be measured. When this is done a surprising fact emerges, for it is found that the steps are exact simple multiples of 20 A.U., and in fact many steps of only 20 A.U. have been observed. It is known from X-ray studies that the sheets of atoms which form mica crystals are in fact separated by 20 A.U. This is called the crystal lattice separation and can be considered to be the thickness of a mica 'molecule.' Thus the observations reveal that when mica cleaves it shows steps which are small simple whole number multiples of the molecular size. As a by-product of the observation, this lattice separation can thus now be independently evaluated with light-waves. This is relatively an optical achievement. The cleavage step shown in Plate 18, with higher magnification than Plate 19, shows how very sharp are the individual fringes, and justifies the claim that one can confidently measure a step of 20 A.U. which is $1/125$ th of a fringe separation.

Cleavage of selenite

Selenite (gypsum) is a mineral which resembles mica in its perfect cleavage and transparency, and indeed in the very

early history of crystallography the two minerals were often confused. However, when a cleavage face of selenite is examined by multiple-beam interferometry an immediate distinction is apparent. Plate 20 shows such fringes and it is evident that the characteristic hills and dales of mica are absent. There are numerous analogous cleavage steps (which incidentally are found to be multiples of 15 A.U., which X-rays give as the lattice spacing of selenite), but there is, in addition, a secondary 'mosaic' type of cleavage which leads to the fringes having a very ragged appearance. Space prevents further discussion of this interesting feature.

Natural unpolished diamonds

A powerful extension of the technique, which has been called the method of 'crossed fringes', will be illustrated by its application to the examination of the natural faces of diamond crystals. Plate 21 shows the highly sharpened contour fringes given by a natural diamond face, of length about 1/10th inch. Three light radiations have been used, namely, the green and two yellow spectrum lines given out by a mercury lamp. In the picture, the single lines are the green fringes and roughly half-way between are the close pair of yellow fringes. It is clear that the crystal face has a complex topography, and if one takes a large number of such pictures, it is possible to build up a model of the surface from them. Such a procedure is difficult and laborious, and furthermore smaller details may easily be missed.

The 'crossed fringe' method simplifies and improves the procedure. Plates 15 to 21 have all been taken with wedge conditions as in Fig. 26. If instead the surfaces are set as near parallel as possible (as Fig. 25) and a *small group* of radiations is used instead of *one* radiation, then an interference picture with extremely broad fringes results as in Plate 22. It is emphasised that these are sensitive interference fringes. This picture shows up surface structure with a remarkable wealth of detail not visible by ordinary means. The reason is as follows. Owing to the multiple-beam interference, a

ULTRA-MICROSCOPIC SURFACE TOPOGRAPHY 151

surface feature corresponding to a change in height of only a few angstrom units causes a big change in the *intensity* of the interference fringe when the surfaces are nearly parallel. Calculation shows in fact that a feature only 3 A.U. high can cause a 10 per cent change in intensity, and a 20 A.U. feature causes a 50 per cent change in brightness. Clearly, such a picture is of great sensitivity for revealing sub-microscopic detail. Yet it cannot be *numerically* interpreted since one has really no easy means of measuring the clearly visible features which are *qualitatively* shown.

The solution to the difficulty lies in crossing the fringe types of Plates 21 and 22. This is accomplished by photographing both systems on top of each other on the same photographic plate, with the result shown in Plate 23. It will be recognised at once what a considerable advance has been made, for the broad revealing pictorial fringes act as a background enabling the meaning of the sharp contour line fringes to be immediately appreciated, whilst the sharp contour line fringes permit of the exact numerical evaluation of the broad pictorial fringes. Small triangular pits can be seen and measured, growth sheets on the surface can readily be identified and in fact a highly complex topography is revealed, and can be measured up with the highest of precision; and all from one photograph. The interpretation becomes simple.

Observations on Metals

The application of the multiple-beam methods to the study of metallic surfaces has only just been begun, yet the results so far obtained are of such promise that a determined attack now being made in this direction will certainly soon have valuable results. The opacity of the metal base leads necessarily to the use of reflection methods. A technical consequence is the exclusion of the use of the crossed fringe procedure. There are many difficulties associated with the examination of metals. To begin with the various constituents of an alloy may have distinctly different optical

reflecting properties. Again difficulties are often encountered in depositing silver on a metal surface, although these difficulties are now in the process of being overcome. Metal surfaces are inherently rough, even when highly polished; and this considerably limits the accuracy of observation; finally, when a high-power microscope is used to magnify detail, grave optical difficulties arise. However, these have been overcome, and in the following two sections some examples of studies on metals will be described.

Surface pits

Plate 24 is a most interesting picture. It shows multiple beam Fizeau fringes in reflection for a small crater formed by the passage of a single strong electric spark, into the surface of a disc of highly polished tungsten metal. The linear magnification is $\times 750$ of the original, which is the highest yet reached with a metal. The photograph shows a depression of diameter about $1/250$ th inch and with depth of about two light waves, i.e., closely enough $1/25,000$ th inch. Not only can the amount of metal flashed off be accurately calculated but, in addition, the actual quality of the remaining surface can be appraised from the fringe shape.

Another such pit produced by a different spark on the same surface is shown under identical magnification conditions in Plate 26. The shape of the hollow is distinctly different from that of Plate 24, and the quality of the residual surface very markedly different. Measurements indicate that there is little difference in the amount of metal removed in the two cases. Attention may be drawn to the particularly fine definition in Plate 26. The fringes are very sharp, when the high degree of magnification is taken into consideration. A further spark pit is shown in Plate 25. The differences in surface quality in the three pictures are notable.

A replica technique

Many of the difficulties associated with the study of metals have now been overcome by means of a replica casting

ULTRA-MICROSCOPIC SURFACE TOPOGRAPHY 153

technique which has just been developed for the purpose. This has limited applications, but is likely to prove useful in many directions. If the metal surface to be studied is heated to 140° centigrade a perfect cast of the surface can be obtained by lightly pressing on to the warm surface a moderately thick sheet of either Perspex or Transpex synthetic transparent plastic. By making replicas of known surfaces of pieces of mica it has been unambiguously established that the replica is a very close reproduction of the surface almost to within molecular dimensions. Having once secured a transparency, this can be silvered and the surface topography of the metal obtained by the relatively simpler methods open to transparent materials, including the use of crossed fringes.

Plate 27 shows the fringes given by the replica of a piece of buffed steel which has on it a deep scratch. Much is revealed by this photograph. Thus the exact depth of the scratch is measurable (it is half a light wave); it can be distinctly noted how the metal is ploughed up on the sides of the scratch mark. The surface polish and surface quality can also be appraised from the fringe width and character. With such a replica the crossed fringe technique can once again be adopted, with consequent improvement in ease of interpretation. Many such problems in metal physics are now being studied by interferometry.

Oscillations of quartz crystals

As a final example, the most recent application of multiple-beam interferometry will now be described because of its quite novel character. Radio transmitters and modern high precision astronomical clocks are controlled by what are called quartz oscillators. A thin section cut from a quartz crystal in a suitable manner sets itself into very rapid mechanical oscillation if an oscillating electric field is applied to its surfaces. A radio transmitter is in fact a device for producing oscillations, say about a million oscillations per second, the particular number depending

on the wavelength. Now the crystal has a natural frequency of oscillation which it takes up most easily, the particular frequency depending on the size and shape. Once set into oscillation it can impose its period on that of the transmitter, and thus the latter gives a radiation of very stable and constant frequency.

It has been found possible to study these extremely rapid oscillations of the crystal surface by means of multiple-beam Fizeau fringes. The crystal surface under examination is heavily silvered and this rests on a silvered flat. The opposing face of the crystal is lightly silvered and electrical contact is made between the two silvered faces of the cut crystal. Monochromatic interference fringes between the crystal face and the glass flat are first set up with the system stationary. Plate 28 illustrates Fizeau fringes for a particular crystal which has a slightly curved surface and shows fringes in ring form, the fringes being sharp. Plate 29 shows the appearance when a slight amplitude of oscillation is set up and it is seen that on some parts of the surface the fringes broaden out because of the amplitude of movement perpendicular to the surface on these regions. It will be noticed that there are certain true nodal regions, where the surface is at rest although neighbouring parts are moving at about 1,000,000 times a second. From the fringe broadening at any point the exact local amplitude of vibration can be calculated. In the example given this is but a small fraction of a light wave. Clearly much should be learnt by the intensive application of this procedure. A further picture, Plate 30 is shown. In this a form of 'crossing' has been adopted. Superposed are the fringes produced both with and without oscillation. The former are broad and the latter very sharp. The sharp lines can be seen running down the centres of the broadened fringes, showing that in this specific instance the vibration did not produce any alteration in the separation between the glass surface and the quartz interface, but only an oscillation about this position.

The most striking interference picture yet obtained is

shown in Plate 31. This shows the remarkable shape taken on by a crystal which is giving a high harmonic of its fundamental frequency. It is oscillating at over 3,000,000 times a second and multiple nodes are clearly visible.

It will be seen from the foregoing brief descriptions that the field of investigation is considerable and extensions are being made. Only one branch of the subject has been described. There exists a further complete branch which makes use of white light fringes.

BIBLIOGRAPHY

The details of the technique, with more complete discussion of results obtained, have appeared chiefly in the Proceedings of the Royal Society, and the Physical Society of London, and a complete survey is given in *Multiple-beam Interferometry of Surfaces and Films*, published (1948) for the author by the Clarendon Press, Oxford, who have kindly permitted the reprinting of Plates 2, 5, 10, 16 here.

GLOSSARY

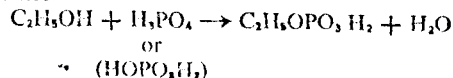
BLOOD: Human blood is a suspension of cells or corpuscles in a fluid called *plasma*, which is a solution of various proteins in water together with salts, sugar, and traces of other substances (urea, vitamins, etc.). One of the proteins in plasma is called fibrinogen and forms a part of the clot when blood coagulates: plasma minus fibrinogen (e.g. after clotting) is termed *serum*. The red corpuscles, which are the majority of the suspended cells, contain haemoglobin as their principal (30%) constituent.

DENATURATION is a term used in protein chemistry to indicate the loss of characteristic properties, a change away from the natural or native state of a protein. Thus raw egg-white coagulates on boiling, and this is one example of a denaturation. The term is somewhat imprecise; but it is not usually used to describe the loss of single properties, only that of a number together. Thus the inactivation of insulin or pepsin by iodine does not come under this head, probably because no dramatic change in solubility results.

ECLIPTIC, plane of the, is an imaginary sheet through the heavens, with the sun at about its centre, and the earth's orbit round the sun as a more peripheral milestone on its extension. The purpose of choosing this plane is to provide a base-line or reference plane from which to start, in describing the various orientations of the movements of heavenly bodies one to another.

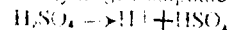
MANOMETER. An instrument for measuring a change in gas pressure, and hence the production or removal of gas within a closed system.

PHOSPHORYLATION. The process of adding on a phosphate radicle to an organic molecule, any process of phosphate esterification: for instance



This type of reaction has a particular significance for biochemistry, since it is the prelude to many chemical breakdowns of molecules into smaller fragments, and also to some synthetic reactions, the phosphate appearing to facilitate the chemical changes which follow.

POLAR GROUPS. It is well known that an acid such as sulphuric acid when in water dissociates or ionises, for instance into a hydrogen ion and a hydrogen sulphate ion



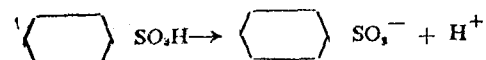
This is due to a peculiarity of the nucleus of the hydrogen atom, which is prone to wander off by itself leaving its solitary orbital electron behind with the rest of the molecule. Both thus become

GLOSSARY

157

electrically unbalanced, the hydrogen nucleus (*proton, or ion*) being positively charged, the sulphate carrying the *single excess electron* and therefore being negative.

Now consider the case where the sulphuric acid has reacted with a longer more inert molecule, for instance a benzene sulphonic acid. When this ionises, the negative charge is localised in space



at one end of the molecule and it is convenient to refer to it as a polar or charged group within the molecule. Further polar groups could of course be added to the type of ring here shown, with the proviso that a negative charge at one end will influence the reactivity of the rest of the ring, and repel the approach of other negative ions. In addition, in a ring system such as that of benzene, the introduction of, for instance, a non-ionising chlorine atom into the molecule influences the even distribution of electrons throughout it, moving them preponderantly to one end. Thus an asymmetrically charged or polar molecule is produced, and the disturbing atom or group may be termed a polar group.

SPECTROSCOPY. Ordinary daylight passed through a glass prism is split into the well-known red, yellow, green, blue, violet of the spectrum. When light from a sodium lamp is similarly treated only two very narrow yellow lines, closely side by side, appear as a spectrum: they do not even fill the whole yellow band of the full spectrum. Similarly, light from a mercury lamp gives a sharp line in the orange, one in the green, and a number in the blue-violet. These are known as line spectra or emission spectra.

But if daylight is passed through sodium vapour or mercury vapour, these compounds absorb light of just the wavelengths they emit when electrically excited, and when the full spectrum is formed by a prism, dark vertical lines are seen across it where light of certain colours is missing through the absorption. (Absorption spectra.) Each absorption is characteristic for the particular substance. Careful study of sunlight shows two narrow dark lines in the yellow at the wavelengths characteristic of sodium. It is accordingly concluded that there is sodium in the outer part of the sun. It also shows other absorption bands, indicative of other elements.

Similarly when ordinary white light is passed through 'coloured' solutions, certain wavelengths are absorbed, and some more than others, so that an absorption spectrum is formed. Conclusions can be drawn from the wave-lengths absorbed, and the degree of absorption at each, as to the chemical nature of the absorbing substance.

About Our Contributors

J. L. Crammer has been Editor of *Science News* from the first issue, originally under a pseudonym. He has had experience of Service life and of journalism, is medically qualified, and has practised medicine. He is a science graduate of Cambridge, where he specialised in biochemistry and later did research.

Rev. Martin Davidson is Vicar of Holy Trinity, Canning Town, E.16. He is a graduate in science, London University, and is also a M.Sc. and a D.Sc. of Queen's University, Belfast. He was President of the British Astronomical Association, 1936/8, and Director of the Comet Section for many years. He is a member of the Council of the Royal Astronomical Society of which he has been Vice-President, and has contributed a large number of papers to various scientific journals, chiefly on astronomical subjects. In addition to his astronomical interests he has written extensively on other subjects.

Kenneth James Franklin was Assistant Director of the Nuffield Institute for Medical Research from 1935 to 1947, Fellow of Oriel College for twenty-three years, and Dean or Acting Dean of the Medical School, Oxford, for over ten years. He is now Professor of Physiology at the Medical College of St. Bartholomew's Hospital. His work has been mainly on the circulation, especially in the veins, the foetus, and the kidney; much of it has been done by the use of X-ray techniques.

A. W. Haslett is a Cambridge graduate, who has made scientific journalism his career. He is Editor of *Science Today*, and for some months past has been a regular speaker in the B.B.C. Forces Educational Programme.

James T. Kendall is a Science graduate of Cambridge University. Since 1939 he has been working in the research laboratories of Metropolitan-Vickers Electrical Co. Ltd., at Manchester.

J. J. Kipling read chemistry at Oxford University and carried out research in physical chemistry for D.Phil. Spent some time in industry as managerial chemist. Now a Lecturer in Chemistry at University College, Hull.

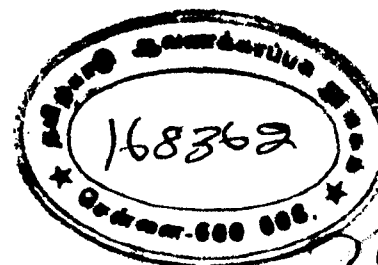
John McMichael, born 1904, Professor of Medicine in the Postgraduate Medical School of London. Graduated in medicine in University of Edinburgh, 1927. Held various university teaching appointments in medicine and physiology, and was a Beit Research Fellow. Was director of the Department of Medicine at the Postgraduate Medical

ABOUT OUR CONTRIBUTORS

School of London during the war years, and carried out research on circulatory failure in wound shock and allied conditions. Is especially interested in the application of physiological methods to the problems of clinical medicine.

K. Mendelssohn, University Demonstrator in Physics, Oxford University. Born and educated in Berlin. Since 1933 in Oxford, where he installed the first helium liquefier in this country. Chief fields of research: superconductivity, superfluidity, and application of physics to medicine.

S. Tolansky, now Professor of Physics at Royal Holloway College, London University, was born at Newcastle-on-Tyne in 1907. He studied physics at Newcastle, Berlin and London, and was on the staff of Manchester University 1934/46. His research work has been mainly carried out in two distinct fields, namely high resolution nuclear spectroscopy, and multiple-beam interferometry, and on these topics he has published over fifty research papers and four books. He has recently been awarded the Boys Prize by the Physical Society of London for his contributions to interferometry.



Handwritten: April, 1946
168362

INDEX

- Activity tests for proteins, 54
- adenosine triphosphate, role in the cell, 66
- amino-acids, chemistry of, 59
- angiographs, 108
- angstrom unit, 141
- Blood pressure, high, 122
- Carbon-13, production, 134
- carbon-14, natural, 28
- catalase, nature of, 51
- centrifuge, 45, 56
- chlorophyll, formula and function, 49, 68
- coenzymes and metals, 62
 - and vitamins, 61
- coma, nature of (astronomy), 16
- comets, origin of, 19
 - speeds of, 25
- crush syndrome (air-raids), 105
- cytochromes, role in the cell, 49
- Diet and kidney disease, 99
- Earth, sea salt of, 33
 - cooling of, 34
- electrophoresis, 57
- enzymes, 60
- escape from planets, velocities of, 21
- Fluorine and age of fossils, 30
- Haem; chemistry of, 46
- Helium II, properties of, 88
- Heparin, use of, 95
- Hypertensin, 121
- Jupiter, conditions on planet, 22
- Kidney, anatomy of, 92, 111 *et seq.*
 - circulation in the, 114-116
 - disease and diet, 99
- Liquefaction of oxygen, 72
- Magnetic cooling, 80 *et seq.*
- metals as coenzymes, 62
- metal surfaces, 151
- metal, replica casting, 152
- metals, origin of, 11, 26
- mica, cleavage of, 149
- Newton's rings, 148
- nucleus of comets, 16
- Ocean waves, analysis of, 124-126
- orbits of comets, 9 *et seq.*
- Peritoneal dialysis (medicine), 98
- perturbations of planets, 14
- photosynthesis, 68, 49
- phosphorylation, 60 *et seq.*, 66
 - et seq.*, *Glossary*
- poisons, specific actions of, 64
- porphyrins, chemistry and function, 45, 50
- pressure of sunlight, 17
- protein structure, 52 *et seq.*
- Quantum principle discussed, 83 *et seq.*
- quartz crystal, surface of, 148
 - oscillations, 153
- Radioactivity 28, 35-40, 133
- refrigeration cycle, how works, 78
- Salt of the sea, 33
- selenite cleavage, 149
- specific heats, significance of, 76, 77, 85
- stereochemistry, importance of in biochemistry, 62
- sugar, breakdown of (*see also* phosphorylation), 61
- superconductivity, 86
- superfluidity, 87
- Tails of comets, 18
- thermometry, 82
- transistor (semi-conducting crystal), 135
- Uremia, 93
- Vitamins as coenzymes, 61
- viruses, nature of, 69
- Zero, absolute, 72