Inaugural Lecture Series 50

PROTEIN,

NATURE'S VERSATILE

DEVICES

by A. ABODERIN

UNIVERSITY OF IFE PRESS

PROTEINS, NATURE'S VERSATILE DEVICES

by

A. Aboderin
Professor of Biochemistry

An Inaugural Lecture delivered at the University of Ife,
on 17th February, 1981 .

Inaugural Lecture Series 50

UNIVERSITY OF IFE PRESS . ILE- IFE . NIGERIA.

Modern Biochemistry, the chair of which is being inaugurated in this University for the first time today (and which it is my proviledge to do), is largely a product of this century. Indeed, not a few of the principal architects and builders of the wonderful edifice of the knowledge which constitute the subject are still alive today. Biochemistry, in spite of its name, is less a syncretic notion as it is a derivative science. It has been fathered by biology. It has, at the same time, been fruitfully nurtured by chemistry. Its efflorescence in this century can be correctly regarded as the culmination of the philosophical commitment of natural science to the cartesian principle of objectivity and of the application of this principle with its full consequence and in all its rigour to the living state. Biochemistry could not have been possible but for the contributions to natural science of those men of the 19th century, men such as Lavoisier, Berzelius, Liebig, van't Hoff, Ostwald, Gibbs Arrhenius and Emil Fischer, who laid the foundations of the new chemistry, and the scientific achievements of some of which involved quint-essentially biochemical problems.

Biochemistry at the University of Ife has been regarded and treated logically. Due and correct attention has been paid by the founding fathers of the Institution to both its paternity and its nuture. To have classified it as a subject in the natural sciences in the company of physics, chemistry, biology and mathematics and as consequence, to have included it within a Faculty of Science are actions which, apart from making it possible for a Professor of Biochemistry to be a Dean of a Faculty of Science, provide for the enrichment of science education in general and of biological education in particular.

Our last point immediately introduces a dilemma, however. Although a natural science, biochemistry, like biology from which it is intellectually derived, cannot be said to share in the universality of either physics or chemistry. It is more parochial.

1

Barring an unexpected discovery today by the space craft Voyager in its odyssey to the outer limits of the solar system, the biosphere, which is the habitat of those elements that constitute the subject matter of biology and its daughter subject is known to be a thin crust of about 5-10 kilometres above and below sea level of this planet, no doubt an almost negligibly thin crust compared with the expanse of the expanding universe. What the biosphere lacks in physical extension however, it makes up for in the almost endless variety and complexity of life forms which it supports. It is known, as a conversative estimate, that man inhabits the biosphere in the company of some 2000 other animal species, about a million plant species plus and unknown number of species of bacteria, other prokaryotes and viruses which are only the successors of probably an unknown but, in all probability, astronomical number of species that have existed since the beginning of biological time. In spite of this knowledge however, it is clear that the scientific laws and principles to be derived from the study of the biosphere and its associated phenomena cannot have the universality of physical laws. What can be expected, and that which has been achieved in the application of the principle of objectivity to the biosphere, is that those biospheric principles and laws will be in full consonance with, and be explainable by, the universal laws of physics and chemistry.

Principle of Biogenesis

There is a quasi-religious element in the scientific enterprise which is self-evident but denied by all. And this is the underlying motive power of Faith. In face of the immense phenomenological vastness of natural phenomena, there is always the certainty that there exists or will exist unifying principles and laws. Scientific experience suggests that this faith is not unfounded. It is our intentions in this lecture to explore some of the unities that characterise life from

the intellectual niche of the protein molecule. To the extent that it can be demonstrated that the choice is not a parochial one, but one that belongs, if not at the centre, close to the centre of the phenomenon of life, to that extent will the intention of this discourse have been fulfilled. As is usual in this type of setting however it is fit and proper for me to seek the understanding of both the initiates and the laity, obviously for very different but understandable reasons.

Proteins, as nature's versatile devices, function within comparatively macroscopic entities known as cells. As all students of biology and the history of biology are aware, the cell is the basic unit of biological life. As everyone again is aware there are different kinds of cells. In multicellular organisms such as the present lecture, there are brain cells, liver cells, kidney cells, all of which, within the proper observational grid, not only look different but also have different properties and functions.

One of the basic unities of nature as expressed in the bio-sphere is the almost tautological but powerful principle of biogenesis which states very simply that behind each living organism today there is an unbroken lineage of descent going back to the beginnings of biological time: *Omnia cellula ex cellula,* as expressed by Schwann and Schleiden in 1838, the same year coincidentally that saw the beginning of the systematic study of proteins by the Dutchman, Gerardus Johannes Mulder.

From the vantage point of post-Darwinian biology the biogenetic principle can be rephrased for completeness. We may say that the existing diversity of life forms has arisen by the progressive diversification during the course of biogenesis.

There are certain implications of this principle of biogenesis, the consideration of which leads to some of the defining notions that distinguish living things from the non-living.

In the first place the principle admits of the possibility of a self-reproducing system; a system, that is, which is capa-

3

able of producing a replica of itself. This is a feature which is normally referred to as reproductive invariance. Secondly, such a self-reproducing system if it is to ultimately yield two of its type must be capable of such self controlled growth as to make this reproductive invariance possible. This characteristic is that known as autonomous morphogenesis. In the third place, a biological self-reproducing system can be subject to changes in its properties. Such changes as it is subject to can be transmitted faithfully by the self-reproducing machinery, to the descendants of such a changed cell. This is the origin of diversifications.

These three principal characteristics of living things, to repeat; reproductive invariance, autonomous morphogenesis and diversification, added on to the observable success story of biosphere — the persistence of life in all its multifarious forms — compellingly suggest the notion of goal-directedness or purposiveness in the performance of those activities which guarantee not only reproductive invariance and autonomous morphogenesis, but also this biospheric success. Modern science arose in part from a revolution against the tyranny of scholasticism which derived not only the structure but also the content of knowledge from the ancient Greeks — in particular from Aristotle. The knowledge of final causes or teleology is one of such elements of knowledge in the Aristotelian corpus. For modern biology which has to live within the shadows of the antecedent revolution, the term teleonomy, a descriptive category, rather than the Aristotelian teleology, an epistemological category, has been adopted to describe those purposive or goal-directed behaviour and performances of living things be they macroscopic or microscopic or molecular which make both reproductive invariance and autonomous morphogenesis possible.

This notion of teleonomy can be employed in biology in the form of two useful even though slightly circular and tautologous axioms. We may say in the first place that there exists in living things, whether uni or multi cellular, teleonomic activities that guarantee orderly growth and faithful

4

reproduction as already summarised. The second, following from the first, will be the postulation of the existence within all cells of an apparatus whose performances within the living system result in the teleonomic activities.

This brings us a little closer to our chosen topic since it is known that in all cells proteins constitute the teleonomic apparatus. Let it however be said at the outset that our discourse will have little to say about the origin and preservation of diversification in general or of the teleonomic apparatus in particular, central though it is to biology. Our concern, on the contrary is with the complex order that is life and the management of that complexity by the protein components of the teleonomic apparatus. The community is lucky that a full consideration of the evolution of this complexity and that of the teleonomic apparatus will be the concern of another in this year's series of lectures.

Simplicity, Subtlety and Versatility of the Teleonomic Apparatus.

The central position of proteins in biology, the realisation of which has served as one of the primary impetuses for a great deal of the phenomenal and revolutionary advances in the study of life that has characterised the science of the last fifty years, and which is still expected to lead to greater but yet unforseeable triumphs, has been aptly summarised by Francis Crick who twenty-three years ago, in his contribution to the discussion of protein biosynthesis at the XIIth Symposium of the Soceity for Experimental Biology, a symposium which was devoted to the central issue of the biological replication of Macromolecules, observed *inter alia* that (i) :

> In Biology proteins are uniquely important ...
> Biologists should not deceive themselves with the
> thought that some new class of biological mole-
> cules of comparable importance to the proteins

5

remains to be discovered. This seems highly un-
likely. In the protein molecule Nature has devised
a unique instrument in which an underlying
simplicity is used to express great subtlety and
versatility: it is impossible to see molecular biology
in proper perspective until this peculiar combina-
tion of virtues has been clearly grasped.

Simplicity, subtlety and versatility, the three Crickian
attributes succintly summarize what biochemistry in general
and protein chemistry in particular — in an interval of history
of about a century and a half which can be conveniently
bounded by the isolation, and identification of the simplest
amino acid, glycine through the cummulative work of Henri
Braconnot, Henri Dessaignes and Charles Gerhardt between
1820 and 1846 and by the completion of the determination,
by John Kendrew, of the three dimensional structure of the
oxygen storage protein, myoglobin in 1960 — has been able
to learn about the teleonomic apparatus. An apparatus which
with typical gallic grandiloquence has been described by
Monod (2) as being:

entirely logical, wonderfully rational and per-
fectly adapted to its purpose which is to preserve
and reproduce the structural norm.

It is this simply designed and functionally subtle and versa-
tile device of nature which is perfectly adapted to its purpose
that is the subject of our discourse.

Teleonomic Activities, Functional Diversity and Management
of Complex Order

The consideration of the structure function and mecha-
nism of action of the teleonomic apparatus is an intricately
inter-woven subject. For one is meaningless without the
other. In a lecture such as this one, therefore, a certain

measure of deliberate didactic ordering will have to be assumed.

But perhaps this is not a particularly necessary step since, if we accept the principle of the goal-directedness of all biological activities as defined earlier, the functions for the sake of which the structures exist, must take a biologically-sanctioned precedence. To distinguish a living cell from non living matter it has been necessary and sufficient to empha- size the two principal properties of reproductive invariance and autonomous morphogenesis. On the other hand life, even to our laymen's sense of observation and even intuition displays a welter of other features, Without enumerating, we shall attempt to subsume all of these other features under the general rubric of Complex Order, a complex which manifests itself both in space and time.

The notion of complex order must be construed in a dynamic sense for one very important reason. Living systems are in the final analysis physico-chemical systems, and they are therefore naturally subject to the strictures of the second law of thermodynamics. They are at the same time open systems which are, by elementary definition, capable of exchanging both matter and energy with their environments. Given such constraints, it should be clear that for dynamic order, reproductive invariance and coherent growth to be possible in an open system, there must be a proper manage- ment of the fluxes of energy, matter and also biological information. In the same vein, it must be expected that an apparatus which has been designed specifically to maintain and manage this complex order isothermally and in open system must be capable of performing some minimal number of activities. These we shall enumerate, and we shall take the opportunity of enumeration to introduce ourselves to elements or parts of the apparatus.

(a) The first of these activities, starting from the simplest is transportation: the transportation, that is, of cell com- ponents from sites where they are produced or available to sites where they are required. A simple example which is

7

familiar to all and sundry is the transport of oxygen to the tissues from the organ which is in direct contact with the source of the element. For such transport requirements nature has evolved specific proteins. If attention was restricted to the simple example of oxygen — an indispensable element to all aerobic life — there are at least four different solutions (apart from free diffusion in unicellular organism) that have been found workable in evolution, each of which is a specific protein that can carry oxygen and drop it wherever it is found necessary. Thus there is the familiar haemoglobin which is responsible for the colour of blood found in all vertebrates. There are also the related molecules: chlorocruorin and hemerythrin in certain invertebrate classes as well as the radically different haemocyanins found only in some arthropods and gastropods. (3)

(b) A second activity that the aparatus is expected to perform, and one of great importance is the transformation, or the reshuffling, of the covalent chemical bonds in the various macro and small molecules present in and brought into the living system. During these transformations, permissible transmutation of covalent bonds to yield usable free energy also takes place. These transformations, and transmutations of covalent bonds take place continuously in all living cells and constitute as a group what is referred to as metabolism. What is important for our present purpose is that if these chemical transformations and transmutations must take place in an isothermal setting and at speeds which are to be compatible with the maintenance of life, catalysis will normally be required. It is a characteristic of living things, and one which takes on the nature of a universal bisopheric law, that every single, biochemical reaction is catalysed by an enzyme, which is, following upon the epochal work of Summer in 1926, of Northrop and Kunitz in the period 1930-6, and the cimulative experience of all biochemistry since, always a protein (wholly or in very large measure).

This is a convenient point to draw attention to an important consequence of this law. This is the principle of biolo-

gical specificity. The notion is one of fundamental mechanistic importance to which we shall return later. Suffice it to say at this juncture that with respect to enzymes, the principle says that a given enzyme can catalyze only one type of reaction, involving one specific molecule (or at worst a family of similar molecules).

That enzymes catalyse biochemical reactions very efficiently is news, but it is not all the news. A cursory glance at any of those metabolic charts which adorn the walls of biochemistry offices and laboratories (and which epitomize the scourge that biochemistry constitutes for students with only an ephemeral but compulsory interest in the subject), demonstrates a situation of bewildering complexity. There is a maze of converging degradative pathways by which exogenously supplied food stuffs and endogenous reserves are converted into a small number of simple key intermediates. On the other hand, there is a network of diverging and inter-locking pathways and cycles through which such key intermediates are used for the construction of other cellular components and for the generation of metabolically usable free energy.

The coordination of all these metabolic circuits, or in our earlier phraseology the management of this metabolic complexity that will lead to the control of energy and material fluxes, a goal that must be achieved if the functional coherence of the cell or organism is to be guaranteed, cannot be possible if enzymes were only the efficient and specific catalysts that they are. This point has been made quite vividly in contemporary technological idiom that (4):

> to define an enzyme as a protein catalyst is as inadequate as it would be to define an electron tube (valve) as a conductor of electricity. A tube with a short circuit between the plate and cathode is a better conductor than an intact tube, but it is useless. A cell in which the enzymes had been replaced by mere catalysts, no matter how efficient.

9

would resemble a living cell no more than a radio
with each tube replaced by a piece of copper wire.

Within the boundaries that are imposed by thermo-
dynamic and physical factors, it is known that a large amount
of metabolic control is achieved because enzyme proteins are
designed to serve as molecular cybernetic elements that
respond in many specific ways to appropriate chemical
signals in such a manner as to modulate their catalytic
activity. Since the initial reports of Umbarger (5) and of
Pardee (6) in 1956, numerous examples of the modulation
of enzyme activity by feed-back control, either of activa-
tion or more commonly, of inhibition have been uncovered,
with such controls exerted usually on the enzyme which
catalyzes the reaction that constitutes the first committed
step in a sequence of reactions leading to the synthesis of a
given metabolite

(c) Chemical free energy arising from the transformation
and transmutation of covalent bonds is all that most cells
can call upon to perform energy-requiring processes such as
motion. Transduction of such energy is thus necessary. To
continue with the technological motif, and to emphasise once
again the versatility and efficiency of the teleonomic appa-
ratus, let us quote *in extenso* an advertisement to a lecture
delivered almost 12 years ago to the date by a Professor
D.R. Wilkie to the Institution of Electrical Engineers in
London:

Available now, linear motor, rugged and depen-
dable: design optimized by world wide field testing
over an extended period. All models offer the
economy of "fuel cell" type energy conversion and
will run on a wide range of commonly available
fuels. Low standby power but can be switched
within m.sec to as much as I kw mech/kg (peak,
dry). Modular construction and wide range of
available subunits, permit tailor-made solutions to
otherwise intractable mechanical problems.

10

Choice of two control systems

(1) *Externally riggered mode.* Versatile, general-purpose units. Digitally controlled by pico-joule pulses. Despite low input energy level, very high signal to noise ratio. Energy amplification 10^6 Approx. Mechanical characteristics (1 cm module) max speed; optional between 0.1 and 100 mm/sec. Stress generated 2 to 5 x 10^{-5} newtons m^{-2}

(2) *Autonomous mode with integral oscillators.*

Especially suitable for pumping applications . . . . Many optional extras e.g. built-in servo (length and velocity) where fine control is required. Direct piping of oxygen. Thermal generation etc. Good to eat".

That, in engineering nut-shell, is a description of nature's mechanical device *par excellence* — the muscle — which is built basically around three important and especially designed contractile protein molecules, myosin, actin and tropomyosin, found not only in muscle, but generally in cells as elements of the cytoskeleton, which is responsible both for cell shape and cell motion.

(d)   To the foregoing teleonomic activities of transportation, transformation, transmutation and transduction must be added that of defence. Just as in the Nigerian budget, defence also takes a large (but perhaps more justifiable) part of the activity of a living cell. This is to be expected if, as we have emphasized all along, the cell is indeed an open system. The success story of biological evolution testifies to the existence of extremely efficient stratagemes for biological defence. Specialized protein molecules to be found either in circulating body fluids on security patrols or on the surfaces of specialized cells on sentry duty constitute important elements for this defence system.

11

Before leaving this section in which we have attempted to see and appreciate the versatility of proteins in the functional sense it is not inconsequential to reacquaint ourselves with an almost routine but nonetheless important observation about the success of the biospheric experiment.

The experiment in life forms has been successful not only because there has been, but that there will always be, an endless variety of living things. It is just as remarkable that there is hardly any type of environment — be it in the heights of the Himalayas or the depths of the Mindanao gorge; in the frigid conditions of the Antaretica or in the warm Ikogosi waters of Ondo State — that has not at one time or the other, been successfully colonized by one form of life or another, in the long history of this biospheric experiment. We suggest that this has been possible primarily because of the adoption by nature of a simple design and an equally simple structural motif for the construction of the teleonomic apparatus which admit of an almost infinite number of different structures characterized by subtle but real differences in similar properties. In other words nature effectively can solve the problems of flexibility and diversity by the proper manipulation of a few simple unities. The success has also been possible because, in spite of this creation of diversity there is a given and rigidly maintained principle of mechanistic unity within structural and functional diversity.

Unity in the Face of Structural and Functional Diversity

Both themes, diversity in unity and unity in diversity are again necessarily and intricately interwoven. A prior consideration of the mechanistic peculiarities of protein molecules, in addition to that of the functionally diversified roles which proteins are called upon to fulfil in biological systems (to which attention has already been drawn), will, however, make it easier to appreciate the design problems that were encountered for the fabrication of these versatile and flexible devices during the course of evolution and the perfectly logi-

cal solutions which were found and which are consistent with the continuing success story of life.

What do we mean when we refer to mechanism with respect to protein action? No more than one of the ordinary meanings of the word, which is simply the way in which a protein works. Defined more formally, mechanism describes the order in time and space of fundamental processes involved in an action or reaction (7). To invoke a principle of mechanistic unity in the face of structural and functional diversity is to say that there exists a fundamental process common to all the actions and reactions — be they those of transport, transformation and transmutation of covalent bonds, transduction of energy or defence — that are mediated by proteins as components of the teleonomic apparatus.

This is the principle which was always implicit in the cumulative experience of enzymologists with respect to the remarkable specificity of enzyme action (see for example the pioneering work of Bergman and Fruton in 1937 on the substrate specificities of the proteases trypsin and chymo- tryspin (8) and the more remarkable but essentially similar specificity of serological reactions (9), and which can be explained simply as follows: All proteins whatever their individual teleonomic role or performance have one charac- teristic in common. They are able to recognise specific topological patterns, or, simply put, shape. The recognition of shape, is the special priviledge of proteins; and it is the basis of all manifestations of biological specificity. A specific shape may be provided by the chemical substance whose chemical reactions are catalyzed by a specific enzyme pro- tein, or it may be an epitope (or antigenic determinant) which provides a recognition site for a specific and appro- priate antibody, etc.

Because of the shape-recognition characteristic of proteins and because all the teleonomic performances discussed earlier mark out proteins as devices that function in a physical milieu, the obligatory first step in the mechanism of action of all proteins therefore involves physical binding between

protein and that which is bound by it, its specific ligand.

Physical binding is a simple and relatively non-specific way of describing the fact that the change of free energy accompanying the process is less than zero.

$$P + L \; \rightleftharpoons \; PL$$
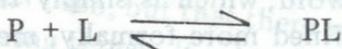
In many cases particularly with respect to antibody-antigen interactions binding constants of the order of $10^{10} M^{-1}$ corresponding to changes in free energies of the order of -15 Kcal/mole have been encountered. This non-covalent and specific interaction, sufficiently strong in some cases to lead to the isolation of the formed non-covalent complex, is found to be mediated by non-specific weak forces of short range which, because of their importance to our structural concerns, will be summarized in the Table below:

Type	Binding Energy (Kcal/mole)	ΔF(kcal/mole) $H_2O \rightarrow EtOH$
1. Dispersion forces	-0.03	
2. Electrostatic Interaction		
(i) Salt Bridge	-5.0	
(ii) Dipole-dipole	-0.3	
(iii) Hydrogen Bond	-0.4	
3. Entropic Forces		-2.4
4. (Covalent-C-C)	58.6	
" C-H	87.3	
" C-N	48.6	
" C-O	70.0	

Attraction energies in all these cases, except for the case of salt bridges involving monopoles, are proportional to multiple powers of the inverse distance between charges or dipoles as the situation demands.

We can therefore arrive at two points of important structu-

ral consequences with respect to the specific ligand binding characteristics of proteins which we have identified as the common mechanistic denominator of protein action. Both the strength of binding, and the identity and the distance dependence of the mediating forces demand (i) that the physical binding of ligand involves the close physical apposition of both protein and ligand which will require (ii) the additive cooperatively of more than element of interaction.

The theoretical formulations, Emil Fischer's Lock and Key hypothesis, first advanced in 1894 (10) and Daniel Koshland's Induced fit hypothesis (11) proposed sixty-five years later in 1958 have been advanced in an attempt to rationalize this shape recognition and specific binding characteristic of proteins. In contradistinction to the static notion of recognition inherent in Fischer's approach, the induced fit idea, which envisages not only a quasi-instructive, dynamic and cooperative affair between protein and ligand, but also a specific structural transformation (a conformational change), in the protein molecule as a result of the specific binding of ligand mediated by non-specific and non-covalent forces, captures the only other common denominatorial element in our general consideration of the mechanism of protein action.

On the Need for a Complex Structure: Diversity in Unity

The discussions of the last two sections emphasizing, as they have done the basic mechanistic unity of all protein molecules in the face of diverse teleonomic activities that have had to be, and still are mediated under different possible environmental conditions, permit us to consider the next and final issue which is a survey of those structural specifications that needed to be catered for in any design process for such a functionally diverse and versatile group of molecules. This is the second issue, which is one of those factors to which have been ascribed the success of the biospheric experiment and which has been labelled Diversity in Unity.

Nature's solution which, from the advantage of hindsight cast over a two billion year period, can be declared a veritable success and which for this same reason can be considered as the only feasible and logical one that could have been adopted and selected for, is the adoption of the macromolecular option.

Speaking in A.D. 1981 at a time when the all-pervading presence and influence of macromolecules in daily life are not news it might be hard to believe that a scant half a century ago, a period which had already seen the concretization of the new organic chemistry of Van't Hoff and Fischer, this comment of Staudinger in 1929 (12):

> I speak of large molecules and of intermolecular forces and thereby assume that macromolecular substances have the same structural principle as those of low molecular weight

was not as self-evident then and was certainly a highly debatable proposition that was viewed with considerable scepticism by a large number of his contemporaries. That he was awarded the Nobel prize a few years later was a recognition of the conceptual revolution that he caused to be within chemistry (not to talk about the practical implications).

This slight detour into chemical structural history is instructive in some respects. Even though it was not completely appreciated at the time, it was the physical data on the size of protein molecules, starting with the pioneering studies of Ludwig Thudicum in 1872, the subsequent work of Gilbert Adair half a century later in 1925 on the size of the haemoglobin and most particularly the introduction of the ultracentrifuge and its application to the study of the size of protein molecules by The Svedberg in 1925, that provided the first hard facts on the existence of these "giant" molecules.

Contemporaneously with the experimental contributions of Adair and Swedberg and the conceptual re-ordering of Staudinger was the demonstration by Summer, in 1926,

of the crystallizability of proteins. Macromolecules in general and proteins in particular, then, not only had the same structural principle as those of low molecular weight in the words of Staudinger, they also had definite structures, which, like those of low molecular weight compounds, were determinable.

Why then have we described the choice by nature of the macromolecular option as the one feasibly logical choice for the construction of the teleonomic apparatus? The reasons are many: but is convenient once again to classify them under two general rubrics which are defined by structural and mechanistic imperatives.

Macromolecules, though very large are simple compounds from the synthetic point of view. Using the word addition in its most general sense, these large molecules may be looked at as being built by the simple successive addition of smaller units usually called monomers. The process of addition is in the trade given the name polymerization and the product of the process is simply a polymer, such as polyethylene, no doubt a useful compound but one which is chemically very simple and monotonous though large.

A simple view such as this is sufficient to lead us to three useful generalizations. Firstly it should be obvious that in principle, polymerization can be made to produce large compounds of the same type but of different sizes depending on the nature of control imposed on the polymerization process. Secondly the nature of the bond between two successive monomeric residue will depend on the chemical characteristics of the monomer. And thirdly the physical and chemical properties of the resultant polymer will depend ultimately on the characteristics of the monomer.

By adhering rigidly to only one type of bonding between successive monomeric units which, on the surface at least, have a purely random arrangement from one end to the other, nature has allowed herself the possibility of generating an infinite number of different protein structures. To guarantee an equally infinite range of chemical and physical

properties in the resultant polymers for current and unforseen teleonomic activities it is only necessary and sufficient to call upon monomers which will belong in the same chemical family but which will have different properties. The monomeric units are the familiar alpha amino acids representable by the general formula:

$$R$$

$$NH_2 \quad H \quad COOH$$
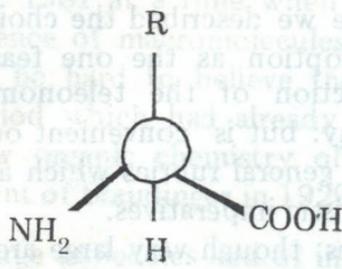
Nature makes use of twenty different amino acids, compounds with the general structure but with different R (or side chains) for the building of proteins, joining these by what is called the peptide bond, which also has a fixed and invariant transoid configuration

$$C \quad\quad H$$
$$C - N$$
$$O \quad\quad C$$
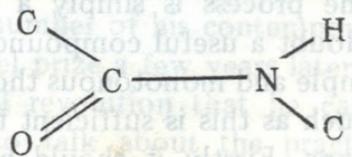
To drive home the point about the enormous possibilities available to nature by the adoption of this simple synthetic scheme, it is a matter of simple arithmetic to show that for a chain of 150 amino acid residues with MW of 17000 daltons (a relatively small protein as they come) there are about 10^{200} unique possibilities; a number that is greater than the estimate of the total number of protons and neutrons in the universe. Each unique possibility is different from the other of the amino acid residues in the chain. This order is what has come to be known as the primary structure of the protein. The deciphering of the primary structure of pancreatic insulin in 1953 by Frederic Sanger is one of the few major events in the history of the study of life and in the advancement of the science of biochemistry.

Crystallizability is only the most direct characteristic of proteins which permit the inference of a definite structure as can be seen in the structure of a simple protein such as hen egg white lysozyme, an enzyme which catalyses an equally simple hydrolytic reaction involving a bacterial polysaccharide, a nice compact structure, almost a sphere. Stripped down by the removal of the side chains the course of the peptide chain can be observed. There are regularities hydrogen bond stabilized helix structures sheet structures and beta turns — examples of what are collectively called secondary structures, the elucidation of which bears the pioneering imprint and the genius of one of the greatest minds of the age, if not of all time — Linus Pauling. And there are irregularities seemingly without a definite purpose but which allow for the type of local structural flexibility within a global definiteness that permits the propagation of chemical signals in the form of specific conformational changes over a large molecular distance as has been demonstrated for example in our work on the linked reactivities of functional groups is this same protein (13,14).

Reclothed with its side chain vestments we observe in the core of the two noticeable structural domains the packing of basically non polar or oily side chains of the appropriate amino acids and the dispensing of the more hydrophilic groups on the outside surface. The structure of such a molecule is defined only in an appropriate aqueous environment which is characteristic of life.

This architectural tour of a protein molecule points to the determinant role of the side chains in the fashioning of the three dimensional structure of such a molecule. This is because of the possibilities offered by the variety of structures in the side chains for a variety of short range interactions of the types that have been spoken of previously. Of primary importance in this context is the concept of hydrophobicity of the amino acid residues, a simple and heuristic scale for which has been proposed by us (15), and to which reference will be made shortly.

That the choice of the macromolecular option makes functional sense is also without doubt. The priviledge of proteins, it has been emphasized, is to recognise shape. Other things being equal, large flexible molecules are intrinsically better discriminators of different topologies. Or simply put, a large flexible structure is easier to wrap around any shape. Thus, do proteins recognize, and bind specifically to their ligands, as the example of the binding of substrate to lysozyme through the intermediary of various short range forces acting cooperatively shows.

A large flexible molecule, capable of the type of directed conformational change described above in response to a chemical signal, is in addition, the best molecular candidate to serve as device for cybernetic control. Such devices would normally be expected to have more than one recognition site for similar (homotropic) or different (heterotropic) ligands with the binding of one influencing, through a protein structural change, the binding of the other, either for potentiation or inhibition, as it is necessary. It is one of the instrinsic qualities of proteins to act as multi-input devices principally because of their size, which allows for the existence on the same molecule of multiple sites with different topological specificities. A good example of known tertiary structure is muscle phosphorylase (16).

The Current Scene and Conclusions

As managers of complexity, proteins are complex structures. That such structures can be determined has been amply demonstrated clearly from the limited examples given, thanks to the existence of adequate and powerful analytical procedures and instrumentation whose introduction over the past fifty years may be said to have been responsible in large measure for the tremendous amount of knowledge on proteins and their role in biology that have been acquired.

It is a different matter, and a much more difficult one however, to predict a protein three-dimensional structure

given a primary structure (or more primordially, the amino acid composition). This is an area of current and very intense research activity here in our own laboratories and in similar places abroad. For many reasons which do not need to be elaborated upon, nature as usual and hopefully only temporarily, is guarding her secrets. This is not however, to suggest that the front is entirely bleak. There are well established guiding principles derived either from a general consideration of the intrinsic nature of complex structures or extrapolated from hitherto accumulated experience (in true Baconian fashion) which must be elements in the final solution of the riddle of protein chain folding. Thus, it has been correctly observed that in the primary structure alone is encoded all the information required for the specification of the three-dimensional structure of any polypeptide (given, that is, the correct solvent conditions) (17).

Secondly, by defining structural relatedness amongst protein molecules with respect to non-zero homology indices (a criterion that is easily derivable from the Hydrophobicity scale described earlier) and studying the permissible types of amino acid replacements in a family of such structurally related protein families, it is possible to infer, through the construction of simple amino acid replacement matrices, that protein chains must fold in such a manner as to conserve and maximize the hydrophobicity of the interior of the resultant globulor structure (15). The principle of the maximization of hydrophobicity provides not only a theoretical rationalization for observed structures of native globular proteins, it also helps to emphasize the primary role of entropic forces in the fashioning of those structures.

In the third place it can also be shown, by appropriate analysis of proteins with known tertiary structures, that certain amino acids have statistically proven predilections for one of the standard secondary structures.

With respect to all complex systems in general the observation has been made by Simpson (1) that they:

21

will evolve (or be formed) from simple systems
much more rapidly if there are stable intermediate
forms than if there are not. The resulting complex
forms in the former case will be hierarchic (or in
the terminology of Arthur Koestler holarchic) .

Accumulating evidence will tend to suggest that this holarchic view applied to protein structures is entirely realistic and heuristic. The determination of the three-dimensional structures of some fifty different proteins has led to the growing appreciation of the probable existence of holarchic motifs found within structural domains which are composed of varying combinations of the elementary secondary polypeptide structures. Examples of these are the coiled coil helix, β meanders the $\beta \epsilon \beta$ unit or the Rossman fold. Fifty structures constitute but a very small sample when compared with the very large number of different protein molecules extant. An average cell, it may be remarked, contains at least 10,000 different protein molecules.

We do not yet know if it is in the solution of such complex structures that the ultimate vindication of the inductive approach of Bacon will be found; or if, as is normally the case, there will be that intuitive quantum leap. Whatever the situation, the stakes are very high. To be able to predict the three-dimensional structures of protein molecules *a priori* will not only be to better understand one of the more beautiful aspects of nature (the study of biology, we may observe, is an exercise in natural aesthetics). Such knowledge, which will surely lead to the design and synthesis of protein molecules with desired and specific biological and chemical properties, is bound to be of incalculable economic value to man, whose positive distinguishing characteristic, apart from language, is his ability to utilize his knowledge of nature and natural phenomena for the improvement of the quality of his own existence on this planet. Let it be noted once again that these. the elements of the teleonomic apparatus, are the ultimate not only in miniaturization, (even in this age of the

microprocessor) but also in the matching of versatile function to structure.

There is yet another lesson to be drawn from our consideration of the interwoven themes of Unity and Diversity. It is to remind all of us that we share in the partimony of nature for whom the seemingly contrary themes of unity and diversity constitute no dialectical problem that calls for a dialectical resolution. Nature in the final analysis is Janus. It is for us therefore to accept our patrimony, to learn to recognise unity in diversity and to cope with diversity in unity.

REFERENCES

1. Crick, P.H.C. (1958). S.E. B. Symposia XII 138.
2. Monod, J. (1972). *Chance and Necessity Vintage Books Edn.* Page. 20.
3. Antonini, E. and Brunori, M. (1971). Hemoglobin and Myoglobin in their Reaction with Ligands. Elsevier NY.
4. Atkinson, E.E. (1970). The Enzymes Vol 1 (Chap. 9).
5. Umbarger, H.E. (1956). Science 145, 674.
6. Yates, R.A. and Pasdee, A.B. (1956). J. Biol. Chem. 221, 757.
7. Schulz, G.E. and Schirmer, R.H. (1979) Principles of Protein Structure (Chap. 11).
8. Bergman M. and Fruton, J.S. (1937). J. Biol. Chem. 117, 193.
9 Landsteiner, K. (1945). The specificity of Serological Reactions (2nd Ed.) Cambridge Harvard Univ. Press.
10. Fischer, E. (1894). Chem. Ber. 27, 2985.
11. Koshland E.E. (1958). Proc. Natl. Acad. Sci. U.S. 44, 98.
12. Staudinger, H. (1929). Z. angew. Chem. 42, 37.
13. Aboderin, A. Boedefeld, E. and Lusi, P.L. (1973). Biochem. Biophys. Acta 328, 20.
14. Aboderin, A and Boedefeld, E. (1975). Biochem. Biophys Acta 420, 177.
15. Aboderin, A.A. (1971). Int. J. Biochem. 2, 537.
16. Fletterick, R.J. and Madsen, N.B. (1980) Ann. Rev. Biochem. 49, 31.

17 Anfinsen, C.B. Haber, E., Sela, M. and White, P.H. (1961). Proc. Natl. Acad. Sci. 47, 1309.

18. Chou, P.Y. and Fasman, G.D. (1974). Biochemistry 13, 222.

19. Simpson, H.G. (1962). Proc. Am. Phil. Soc. 106, 6.