J. theor. Biol. (2002) **219,** 325–342
doi:10.1006/yjtbi.3128, available online at http://www.idealibrary.com on IDEAL®

The Protein Folds as Platonic Forms: New Support for the Pre-Darwinian Conception of Evolution by Natural Law

MICHAEL J. DENTON*†, CRAIG J. MARSHALL† AND MICHAEL LEGGE†

†*Department of Biochemistry, University of Otago, P.O. Box* 56, *Dunedin, New Zealand*

(*Received on* 27 *November* 2001, *Accepted in revised form on* 8 *July* 2002)

Before the Darwinian revolution many biologists considered organic forms to be determined by natural law like atoms or crystals and therefore necessary, intrinsic and immutable features of the world order, which will occur throughout the cosmos wherever there is life. The search for the natural determinants of organic form—the celebrated "Laws of Form"—was seen as one of the major tasks of biology. After Darwin, this Platonic conception of form was abandoned and natural selection, not natural law, was increasingly seen to be the main, if not the exclusive, determinant of organic form. However, in the case of one class of very important organic forms—the basic protein folds—advances in protein chemistry since the early 1970s have revealed that they represent a finite set of natural forms, determined by a number of generative constructional rules, like those which govern the formation of atoms or crystals, in which functional adaptations are clearly secondary modifications of primary "givens of physics." The folds are evidently determined by natural law, not natural selection, and are "lawful forms" in the Platonic and pre-Darwinian sense of the word, which are bound to occur everywhere in the universe where the same 20 amino acids are used for their construction. We argue that this is a major discovery which has many important implications regarding the origin of proteins, the origin of life and the fundamental nature of organic form. We speculate that it is unlikely that the folds will prove to be the only case in nature where a set of complex organic forms is determined by natural law, and suggest that natural law may have played a far greater role in the origin and evolution of life than is currently assumed.

Introduction

Before Darwin, the majority of leading biologists adhered to a Platonic model of nature, referred to by Owen (1849) in his classic monograph *On the Nature of Limbs* as "the Platonic cosmogony." According to this conception, all the basic recurrent forms of the organic world, such as the pentadactyl design of the vertebrate limb, the body plans of the major phyla, the forms of leaves and so forth, as well as the recurrent forms of the inorganic realm, such as atoms, crystals, etc., represent the material manifestations of a finite set of immutable immaterial archetypes or "ideas" termed by Owen (1849) "predetermined or primal patterns." These pre-existing abstract types or ideas are materialized, or to cite Owen (1849) again, "clothed in material garb," by the agency of

*Corresponding author. Tel.: +64-3-479-7863; fax: +64-3-479-7866.
E-mail address: mikedenton30@hotmail.com (M.J. Denton).

0022-5193/02/$35.00/0

natural law, or more precisely in the case of organic forms by a special class of natural laws which applied uniquely to the vital realm—the celebrated "Laws of Form."

And just as today we account rationally for the diversity of inorganic forms such as atoms, crystals, chemical compounds, and even subatomic particles by various sets of laws or constructional rules—atom building rules, laws of crystallography, laws of chemistry and so forth—which allow for a rational deductive derivation of all possible atoms, crystals, chemical compounds, subatomic particles, etc., so pre-Darwinian biologists hoped to provide a rational and lawful account of the diversity of organic forms via the "Laws of Biological Form" (Driesch, 1929; Webster & Goodwin, 1982).

The fundamental goal of biologists in the pre-Darwinian era to seek rational and lawful explanations for biological form is reflected in Goeffroy St Hillaire's attempt to derive all the basic body plans of the major biological types from a basic fundamental plan by a system of simple natural transformations—his most famous being the derivation of the vertebrates from the invertebrates by simply turning the invertebrate body plan on its back. The attempt of Carl Gustave Carus to provide a rational and lawful account of all skeletal forms in terms of a set of simple geometric transformations is indicative of the same tendency. According to Russell (1916): "He was seeking to identify the inner law which presides over the formation of the skeleton throughout the animal kingdom. His system was … an attempt to work out a geometry of the skeleton … his thesis is that all forms of skeleton … can be deduced from a hollow sphere … every skeleton can be represented schematically by a number of hollow spheres suitably modified in shape and suitably arranged. We may expect then all skeletons to be composed of spheres, cylinders and dicones in diverse arrangements."

Conceiving of organic forms to be natural kinds, determined by natural law like inorganic forms, it is easy to see why the crystal was one of the most popular metaphors for organic form in the early 19th century. It was used widely by Carus (Russell, 1916) Theodore Schwann, Owen and Robert Chambers. Schwann, the co-founder of the cell theory considered "cytogenesis as a form of organic crystallization" (Rupke, 1994) and in the last chapter of his *Microscopical Researches* he draws extensive parallels between cells and crystals: "The process of crystallization in inorganic nature … is … the nearest analogue to the formation of cells … should we not therefore be justified in putting forward the proposition that the formation of the elementary parts of organisms is nothing but a crystallization and the organism nothing but an aggregate of such crystals … if a number of crystals capable of imbibition [absorption] are formed, they must combine according to certain laws so as to form a systematic whole, similar to an organism" (Schwann, 1847). Owen (1866) used the analogy unambiguously in the final chapter of his *Anatomy of Vertebrates* in the context of a discussion of the causes of segmentation: "the repetition of similar segments in a vertebral column and of similar elements in a vertebral segment, is analogous to the repetition of similar crystals." And Chambers (1969) used the crystal as an analogy of organic form throughout his popular *Vestiges of the Natural History of Creation*, first published in 1840: "In some crystallizations the mimicry [of biological form] is beautiful and complete; for example, the *Arbor Dianae* is a crystallization resembling a shrub." The fact that many different crystal forms can be generated from a small number of basic patterns added to the attraction of the analogy. In the case of calcite, for example, the rules permit the construction of about 600 different molecular arrangements which can be combined to build over 2000 different combinations (Lima de Faria, 1988).

The widespread belief that organic forms are lawful "givens of nature" explains why it was that throughout the pre-Darwinian period from the *naturphilosophie* of the late 18th century, right up to the period just before the publication of the Origin, although it was universally accepted that organisms exhibited functional adaptations, for Goethe, Carus Goeffroy and Owen, it was always form which was of primary concern. Form came first and function was viewed as a secondary and derived adaptive feature (Russell, 1916; Richards, 1992). Owen (1866) believed that during evolution: "change

of structure would precede that of use and habit.'' Goethe, one of the foremost exponents of the *naturphilosophie*, took the ''form first position'' to extraordinary extremes. He even went so far as to claim: ''We are not to explain … the tusks of the *Babirussa* by their possible use, but we must first ask how it comes to have tusks. In the same way we must not suppose that a bull has horns in order to gore, but we must investigate the process by which it comes to have horns in the first place'' (Russell, 1916). In other words it is form which is primary and ''given,'' and adaptive functions are merely secondary modifications.

It followed from this Platonic conception of organic forms as ''built-in'' lawful givens of physics, like molecules or crystals, that the whole pattern of evolution was itself in a sense already pre-determined or pre-specified by natural law—being the material manifestation of a pre-existing and eternal plan. As Owen (1866) put it in the concluding chapter of his *Anatomy of Vertebrates*, the path of evolution was ''preordained … due to an innate tendency … by which nomogenously created [generated by law] protozoa have risen to the higher forms.'' The idea that the paths of evolution were all pre-ordained by natural law was vigorously defended by Chambers (1969) in his *Vestiges*. Discussing life in the cosmos he comments: ''Thus as one set of laws produced all the orbs and their motions and gnognostic arrangements, so one set of laws overspread them with life.'' And of course it followed that if biological forms were indeed the inevitable ends of natural law—''givens of physics''—life throughout the cosmos wherever it exists should be based on the same set of organic forms. And this radical possibility was raised by Owen (1849) in the concluding section of *On the Nature of Limbs* where he raises the possibility of the vertebrate body plan having been modified in different ways on different planets: ''The laws of light as of gravitation being the same [on other planets] … the inference as to the possibility of the vertebrate type being the basis of organization of some of the inhabitants of other planets will not appear so hazardous.'' And Chambers (1969) following the same logic considered it highly likely: ''the inhabitants of all other globes of space bear not only a general, but a particular resemblance to those of our own.''

The conception of evolution by natural law also raised the obvious possibility—again by analogy with inorganic phenomena—that the evolution of organic forms might have occurred *per saltum*. Chambers (1969) for example puts forward the idea that the origin of life might have come about *per saltum* as the result of a process analogous to that of crystallization, a view which echoes some modern thinking in this area (Kauffman, 1993). The possibility of saltational evolution was also raised by Owen (1866) in *Anatomy of Vertebrates* in the discussion on the evolution of the horse. Of course the conception of organic forms as lawful ''crystal-like'' features of the natural order did not rule out alternative more gradual modes of origin via built-in ''pre-determined constructional or evolutionary paths.''

Holding organic forms to be the result of a set of natural laws which applied uniquely to the organic realm, nearly all the leading biologists of the pre-Darwinian era might be described as ''vitalists,'' according to the definition of the term given by Driesch (1929) in his *Science and Philosophy of the Organism*, as the belief that life: ''has its elemental laws, laws of its own,'' which are additional to those which are known to operate in the inanimate world. This type of vitalism is however a very weak form of the doctrine, and is obviously compatible with a quite materialistic conception of the nature of life (McDougall, 1938). It amounts to little more than the claim that the properties and structures of organic forms which comprise the vital realm arise like the properties and structures of all other classes of natural forms—like those of atoms and molecules for example—from the intrinsic natural properties of their constituents. [Expressed in the jargon of today we might say that they believed organic forms to be the result of the self-organizing properties of the unique bio-materials which comprise living systems.] As Russell (1916) points out, Schwann's interpretation of cells as ''crystals'' was for example quite materialistic. According to Schwann (1847), if there were vital laws, unique to the organic realm, these were no less immaterial, no less part of the physical world than the laws of chemistry,

or the laws of crystallography—which were also unique to the realms of chemistry and crystallography respectively. Owen was also accused at times of veering toward materialism (Rupke, 1994). Indeed it seems he even considered the possibility that the vertebrate body plan might be the result of the action of purely physical forces such as electromagnetism, which implied that the laws of organic form are just a special subset of ordinary physical laws. As Rupke (1994) recounts, William Conybeare argues in a letter to Owen: "although you do excellently distinguish organic laws from the mere mechanical laws of inorganic nature—yet you are inclined to assign a very considerably larger influence to the inorganic forces in the organic kingdom than I can persuade myself to do." And as Rupke continues: "Galvanism brought us too near to the materialistic belief in the spontaneous generation of life advocated by such foreign fiends as Lamark and Oken. Such fears proved fully founded when at a later stage Owen indeed attributed the origin of life in its most primitive and elementary form, not to any separate life force … but to a constellation of organic particles, like elementary dipoles concurring to produce a magnet." [Owen had more to fear from such charges than his continental fellow travelers stationed as he was in England where creationism was so prevalent. For Owen was quite aware that according to the Platonic/naturalistic conception of nature, the order of biology was an in-built necessary order and not a contingent artifactual order as the creationist doctrine implied. This clash is nowhere clearer than in *On the Nature of Limbs* where Owen has to explain that because adaptations are *ad hoc* functional modifications of the primal patterns given by natural law and not contrivances designed specifically by God to serve particular functions, we should not expect organisms to exhibit *perfect* adaptations.]

The Platonic biology of the pre-Darwinian era with its emphasis on evolution by natural law and its conception of a rational order underlying the diversity of life, represented a grand scientific vision, whose heroic goal was nothing less than the unification of biology and physics. It collapsed primarily because it failed to identify the elusive laws of form which might have provided a rational account of organic form and explained how the evolution of the basic invariant forms or types, from cell forms to the body plans of the major phyla, and deep homologies such as the pentadactyl limb, might have come about as a result of natural law. That they had no convincing explanation was explicitly conceded by Owen (1849) in the final paragraph of "*On the Nature of Limbs*": "To what natural laws or secondary causes the succession and progression of such organic phenomena may have been committed we as yet are ignorant."

POST-DARWINIAN BIOLOGY: THE ARTIFACT AS METAPHOR OF LIFE

After 1859, the whole Platonic typological scheme was overthrown. Indeed the very concept of organic forms as real natural existents, as necessary parts of the eternal fabric of the world order, like atoms or crystals was abandoned. Instead a new model of organic form — that of the machine or artifact took its place. Necessity was replaced by contingency and natural law was replaced by natural selection. Organic forms were now viewed as contingent mutable assemblages of matter, like the constructs of a child's erector set such as Lego, put together during the course of evolution piece by piece by natural selection for various biological functions. Such a model implied in Driesch's (1914) words that life " is distinctive as a [functional] combination and not because of its own laws."

And because according to the new Darwinian framework, it is selection for function which generates organic form—an inversion of the previous Platonic "form first function second conception"—the great majority of biologists since 1859 have come to see selection, in Kauffman's (1993) words "as the overwhelming, even the sole source of order in organisms." Even the deep homologous patterns which underlie the major body plans such as the vertebrate body plan or the pentadactyl limb, for which no convincing selectionist explanation has ever been provided, are now assumed to represent ancient adaptations entrapped by genetic inertia and perpetuated as non-adaptive features in their present-day descendents.

[We note in passing that there is some irony in the fact that in adopting the metaphor of the artifact or machine, the Darwinists had adopted the same metaphor as their creationist opponents. For both creationists and Darwinists, life's order is contingent and artifactual, like the order of a machine, like the order of Paley's watch. Significantly, Darwin himself admitted how impressed he was with Paley's *Evidences* (Darwin, 1958). For creationists it was God who had contrived life's contingent order for the Darwinists it was a "Blind Watchmaker"(Dawkins, 1986) who relied on time and chance.]

The adoption of the "contingent mutable artifact"as the metaphor of organic form ushered in the modern era of biology and changed the whole explanatory framework of biological science. It was truly a change of revolutionary import. The very naturalness of life—the idea of organic forms as necessary parts of nature was abandoned. The metaphor of the crystal was replaced by that of the watch! Where before Darwin the order of life and its unique properties had been an "element of nature" (Driesch, 1929) "freely given from within" by natural law, the same order and properties now arose like those of an artifact from special functional contingent combinations of matter which had to be instructed or specified from outside of nature, like an artifact, from information in a blueprint or program (Webster & Goodwin, 1983; Yockey, 1992; Keller, 2000; Davies, 2001). And because the order and properties of machines are determined from the bottom up by their parts, the focus of biology shifted away from the study of organisms and higher forms to the study of life's most elementary components.

And it is implicit in the new Darwinian view of organic forms as "artifact-like assemblages," that the actual biological forms that make up the organic realm of earth—molecular forms, cell forms, body plans, etc.—represent *a tiny finite set of all possible forms which have been drawn by selection during the evolution of life on earth from a potentially infinite set*. Organic forms on other planets should be, on this theory, quite different from those on earth—if the tape of life were to be replayed (Gould, 1989) we would not expect to see vertebrates on extraterrestrial planets, as Owen inferred.

However the pre-Darwinian conception of organic forms as intrinsic features of the natural order was never completely laid to rest after 1859. It survived well into the 20th century particularly on the continent of Europe (Gould & Lewontin, 1979). Much of Driesch's (1929) *Science and Philosophy of the Organism* was not so much an argument for an indwelling nonmaterial vital force as what Driesch terms the "autonomy of life" by which he meant the existence of natural laws peculiar to or autonomous to the biological realm. D'Arcy Thompson's *On Growth and Form* (1942), first published in 1917, represents a great classic in the Platonic tradition and contains more than an echo in many sections of the thinking of Goethe, Carus and Goeffroy. In the final chapter of *On Growth and Form*, where Thompson shows how the forms of many animals can be related by simple mathematical transformations, one is instantly reminded of the attempt of Carus to create a universal geometry of all skeletal forms. Two recent authors who must also count as belonging to the Platonic tradition are Brian Goodwin and Stuart Kauffman (Webster & Goodwin, 1983; Kauffman, 1993; Goodwin, 1994).

Of course no serious biologist doubts that *some* biological forms may be given by natural law and arise spontaneously out of the intrinsic self-organizing properties of their constituents and may not need any genetic program for their specification. The spherical form of the cell and the flat form of the cell membrane are two well-known examples. Other more complex examples cited by Waddington (1962) are the various cytoplasmic structures made up of multiple layers of membranes such as the grana and intergrana regions of chloroplasts, the hexagonal arrangement of the rhabdomeres in the eyes of insects and the many forms described by Thompson (1942) in *Growth and Form*, including radiolarian skeletons, the shapes of mollusk shells, the curved shape of animal horns. But on the whole, natural law is considered to play a very trivial role in the generation of biological form and particularly in the generation of complex seemingly asymmetric biological forms such as protein folds, cell forms, body plans, etc.

The only area of modern biology where a strong deterministic and naturalistic element is

still evident is the "origin of life" with many researchers viewing life's origin as an inevitable and determined end of planetary and cosmic evolution (Kenyon & Steinman, 1969; Lehninger, 1982; De Duve, 1991; Morowitz *et al.*, 2000; Sowerby *et al.*, 2001). Here we argue that in another important area of modern biology, one related to the origin of life, that involves the evolution and origin of one of the most important classes of complex biological forms—the basic protein folds — the pre-Darwinian concept of organic forms as "built-in" intrinsic features of nature determined by natural law provides a more powerful explanatory framework than its selectionist successor (Denton & Marshall, 2001).

The Platonic Nature of the Protein Folds

The protein folds are the basic building blocks of proteins and therefore of the cell and indeed of all life on earth. Each is a polymer between 80 and 200 amino acids long consisting of from about 1000 to 3000 atoms folded up into a complex intricate three-dimensional shape. Most folds exhibit a hierarchical structure composed of basic *secondary structural* elements such as α helices and β sheet conformations which are often arranged into more complex motifs which are in turn combined together to make up the native conformation of the fold. The elucidation of the atomic structure of the protein folds has been one of the triumphs of 20th century science. Many different folds have now been identified and together they form an exotic and beautiful set of organic forms. The strangeness and beauty of these enigmatic abstract forms (see Fig. 1) is apparent to anyone on even a cursory examination.

It is important at this stage to note that the great majority of functional proteins in the cell consist of two or more basic folds linked together into multidomain or multifold complexes. In this paper we are considering only the fundamental nature and evolutionary origin of the folds and not of the higher order adaptive structures into which they are combined. These higher order complexes resemble—"Lego-like"—contingent assemblages put together by natural selection for various biological functions during the course of evolution by gene duplication and fusion (Brandon & Tooze, 1999).

FOLD TYPOLOGY

For most of the first half of the 20th century the structure of proteins was a subject of great speculation. It had been known since 1910, after the pioneering work of the great German chemist Emile Fischer, that proteins were long polymers consisting of amino acids. But it was only in 1957 when the first 3D structure of a protein, whale myoglobin, was finally determined by X-ray crystallography that the 3D arrangement of the polypeptide backbone of at least one protein was finally revealed. By the late 1960s several other proteins had been determined including hemoglobin and lysozyme. Despite these early successes the lack of any apparent regularity in protein structures, and the great dissimilarity among those that had been determined, provided no basis for a rational classification (Ptitsyn & Finkelstein, 1980; Richardson, 1981). The picture was still in those early days compatible with the Lego model—that the folds in living organisms on earth might be individual members of a near infinite set of contingent material assemblages put together by natural selection over millions of years of evolution.

It was only during the 1970s, as the number of 3D structures began to grow significantly, that it first became apparent that there might not be an unlimited number of protein folds—that the folds might not belong to a potentially infinite set of artifactual Lego-like constructs. On the contrary, it became increasingly obvious as more structures were determined that the protein folds could be classified into a finite number of distinct structural families containing a number of related but variant forms, i.e. that the classification system of fold structures was typological (Ptitsyn & Finkelstein, 1980; Richardson, 1981; Orengo *et al.*, 1997). This was an important finding as the very fact that protein folds can be grouped in such a way was itself significant, for it provided the first line of evidence that the folds *might be* natural forms determined by physical law.

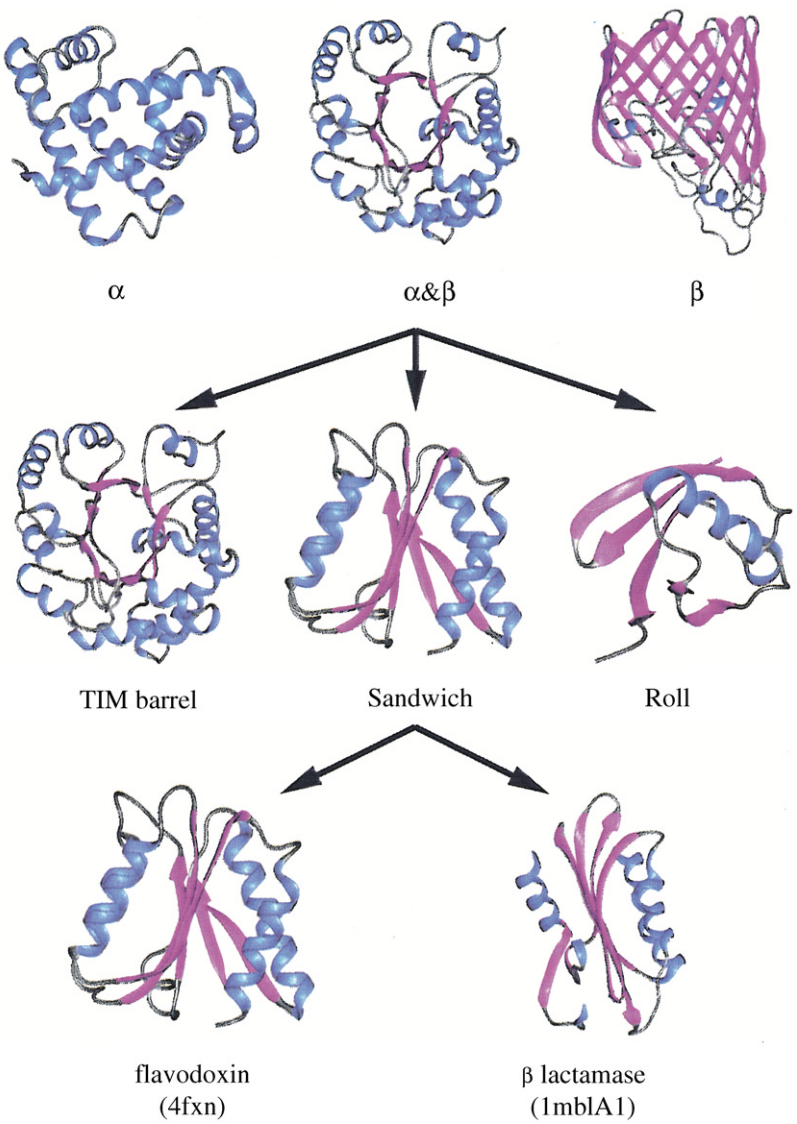

FIG. 1. Structural classes of protein folds. (Top row) Three basic fold classes—α, containing only α helices; α and β, containing α helices and β sheets; and β, containing only β sheets. (Middle row) Three different architectural subclasses of the α and β class—TIM barrel, three-layer sandwich and roll. (Bottom row) Two different arrangements of the three-layer sandwich. The spiral conformations are the α helices, the broad arrows are the β sheets. (Reprinted with permission from Orengo *et al*. (1997). Copyright 1997, with permission from Elsevier Science.)

It also became apparent that the 3D structures of individual folds were essentially invariant—some such as the Globin fold and the Rossman fold for example, having remained essentially unchanged for thousands of millions of years. Both their invariance and the typological classification schemes into which they could be grouped argued for their being a finite set of "*real* timeless structures" determined by physics rather than being mutable "Lego-like" aggregates of amino acids determined by selection.

LAWS OF FORM

Additional support for the Platonic idea that the protein folds represent a set of natural kinds is provided by the fact that their structures can be accounted for by what amounts to a *rational and generative morphology* consisting of a set of rules which govern the way that the various secondary structural motifs such as α helices and β sheets can be combined and packed into compact 3D structures (Ptitsyn & Finkelstein, 1980; Finkelstein & Ptitsyn, 1987; Chothia, 1993; Chothia *et al*., 1997; Taylor *et al*., 2001). In the words of Chothia *et al*. (1997): "In most proteins the α helices and β sheets pack together in one of a small number of ways. The connections between secondary structures obey a set of empirical topological rules in almost all cases … Subsequently it was argued that these similarities arise from the intrinsic physical and chemical properties of proteins and a great deal of work was carried out to demonstrate that this is the case."

It is impossible here not to be reminded of the constructional rules which govern the assembly of subatomic particles into atomic structures, and generate the finite set of 92 atoms which make up the periodic table of the elements, or the rules of grammar, which restrict grammatical letter strings to a tiny finite set of all possible sequences. These "laws of fold form" represent a set of pre-existing abstract prescriptions, specifying a finite set of allowable "material" forms and they therefore provide for a rational deductive derivation of *all possible fold morphologies*. They self evidently represent a set of "Laws of Biological Form" of precisely the kind sought after by Geoffroy, Carus and many other Platonic biologists of the pre-Darwinian era.

The Platonic ethos of the field is captured in the title of a recent paper by Taylor (2000) entitled "Searching for the ideal forms of proteins" in which he talks of: "unifying structural principles that can be represented as idealized proteins–protein archetypes, or their underlying Platonic forms" and goes on to cite the Platonic approach of other researchers in the field (Murzin & Finkelstein, 1988) who: "represented the all α helix class of folds, by a set of quasi-regular polyhedra."

Consideration of the various physical constraints which restrict the folded spatial arrangements of linear polymers of amino acids—the laws of fold form—suggests that the total number of permissible folds is bound to be restricted to a very small number. One recent estimate based on possible arrangements of typical structural elements gave a maximum of 4000 folds (Lingard & Bohr, 1996). Based on similar considerations, the authors of another recent paper suggested that the maximum is likely to be no more than a few thousand (Chothia *et al*., 1997). A different type of estimate based on the rate of discovery of new folds—rather than permissible spatial arrangements—suggests that the total number of folds utilized by organisms on earth might not be more than 1000 (Chothia, 1993). In many recent reports the total number of different folds is often cited to be somewhat less than 1000 (Holm & Sander, 1996; Orengo *et al*., 1997; Zhang & DeLisi, 1999; Holm & Sander, 1999).

Whatever the actual figure, the fact that the total number of folds represents a tiny stable fraction of all possible polypeptide conformations, determined by the laws of physics, reinforces further the notion that the folds like atoms, represent a finite set of allowable physical structures which would recur throughout the cosmos wherever there is carbon-based life utilizing the same 20 amino acids.

Some idea of how enormously restricting these "laws of fold form" are may be gained by consideration of the fact that the total number of theoretically possible protein structures that an individual amino acid chain 150 residues long might adopt, assuming that each peptide group

has only three conformations, is 3^{150} or 10^{68} (Brandon & Tooze, 1999). And the total number of possible 3D conformations that all possible amino acid sequences 150 residues long could theoretically adopt would be vastly greater, while, as we see, the total number of stable 3D structures allowed by physics would appear to be restricted to a tiny finite set of about 1000 unique conformations.

FUNCTIONAL ADAPTATIONS: SECONDARY MODIFICATIONS OF PRIMARY FORMS

Further evidences consistent with the Platonic conception that the protein folds represent a set of lawful immutable natural forms, ''primary givens of physics,'' are those many cases where protein functions are clearly secondary adaptations of a primary, immutable form (Gerlt & Babbitt, 2001). This is spectacularly true in the case of some of the more common folds also known as superfolds (Orengo *et al.*, 1994; Gerlt & Babbitt, 2001). In the case of one superfold the so-called triosephosphate isomerase (TIM) barrel, an eight-stranded alpha/beta bundle (see Fig. 1), essentially the same fold, has been secondarily modified for many completely unrelated enzymic functions occurring in such diverse enzymes as — triosephosphate isomerase, enolase and glycolate oxidase (Orengo *et al.*, 1994). Another example, where a basic fold has been secondarily modified for various biochemical functions, in this case closely related functions, is the various elegant functional adaptations to oxygen uptake and carriage exhibited by the globin fold in myoglobin and the various vertebrate hemoglobins. The fact that in many cases where the same fold is adapted to different functions, no trace of homology can be detected in the amino acid sequences, suggesting multiple separate discoveries of the same basic structure during the course of evolution (Orengo *et al.*, 1994; Brandon & Tooze, 1999), further reinforces the conclusion that the folds are a finite set of ahistoric physical forms.

Moreover even the extreme Platonic position of some of the pre-Darwinian biologists, such as Goethe—that form directly determines function, without any or only minimal selective modification—may also be true in certain cases. Consider those cases where a particular protein function, such as that of hemoglobin or cytochrome *c*, arises from the association of a particular fold with a particular prosthetic group, co-factor or metal ion. Because these associations are ultimately determined by the unique 3D forms of the folds, then in all such cases the biochemical function is in effect given deterministically, by the spontaneous combination of the basic fold with its prosthetic group.

PROTEIN FOLDING: MATTER DRAWN INTO A PRE-EXISTING *PLATONIC MOLD*

During folding the amino acid sequence of a protein appears to be searching conformation space for increasingly stable intermediates which lead it step wise toward the deepest energy minimum for that sequence, which corresponds to its final native conformation (Ptitsyn & Finkelstein, 1980; Finkelstein & Ptitsyn, 1987). The process is driven thermodynamically via a succession of free energy decreases (Dinner *et al.*, 2000). The process of folding is often pictured as being analogous to a ball finding its way down the sides of a complex rather irregularly shaped bowl to the bottom of the bowl, its final preordained and natural resting place, where the bottom of the bowl represents the natural free energy minimum of the fold. Extending this analogy we can think of there being a preexisting energy landscape containing 1000 or so uniquely shaped bowls or free energy minima. This picture lends itself to Platonic interpretation. Even the terms used in the literature reflect the Platonic concept of matter ''finding'' or ''filling'' a pre-existing mold. Thus the folding process is often described as a mechanism by which ''sequence selects structure.'' As a recent author commented: ''Thus the notion that sequence determines structure might be more precisely formulated with the concept that sequence chooses between the limited number of secondary structures available to the polypeptide backbone'' (Honig, 1999). In other words, it is not the sequence which specifies the mold but the mold which specifies which sequences can be accommodated. For the mold is prior to the sequence, although of course

during folding each particular sequence is prior in time to the form which it finally makes manifest. The ubiquitous text book claim that "the amino acid sequence determines the 3D form of the protein" is a mechanistic interpretation of the folding process which might be more accurately stated Platonically as "the prior laws of form determine which amino acid sequences can fold into a stable 3D form." If the sequence contains any information, it is not information to create or generate a unique artifact-like assemblage analogous to a Lego construct or a watch, but more of a guide through a pre-existing Platonic landscape to an already pre-figured end.

Aristotle expressed this Platonic conception of nature (and of the folds) in his *Parts of Animals* (Aristotle, 1937) in the famous analogy, where he envisages the pre-existing plan of a house acting as an "attractor" molding the material constituents, the bricks and stones, during building of the house into conformity with the pre-existing plan of the house: "Now the order of things in the process of formation is the reverse of their real and essential order … bricks and stone come chronologically before the house …but logically the real essence and the Form of the thing [the pre-existing plan of the house or the pre-existing Form of the fold] comes first."

However we may wish to interpret the folding of a protein, there is no question that the process is nothing like the assembly of a contingent mechanical construct like a Lego model or a watch.

ROBUSTNESS: THE MAINTENANCE OF FORM

The free energy difference between the native conformation of the fold and its denatured state is small. A consequence of this is that folds are only marginally stable. This allows for a crucial degree of structural flexibility, which is necessary for many catalytic and other functional activities, which often necessitate the fold adopting slightly different conformations. But it also means that because of the continual buffeting and bombardment arising from molecular collisions in the turbulent interior of the cell, the configuration of each fold is continually subject to conformational disturbances which may involve anything from the movement of a few atoms to the unfolding of sections of the amino acid chain (Brandon & Tooze, 1999).

However a fold is able to maintain and regain its native conformation in the face of these microchallenges because its native conformation, being a natural free energy minimum, acts as a natural attractor "continually drawing" all the parts of the fold back into its proper native conformation (the natural free energy minimum of the fold). And just as a ball in a bowl always ends up at the bottom of the bowl, a fold is also able to get back "home" or to recover its proper conformation along an *infinity of different paths*. In short, the folds are robust natural existents, whose proper forms are under the governance and supervision of natural law.

In the case of artifactual contingent assemblages of matter such as a watch or "Lego construct," there is no natural agency or natural guarantor of "proper form"—no eternally present Platonic mold or free energy minimum acting as an attractor continually drawing or guiding the assembly of its components to a pre-ordained end. Consequently artifacts are not robust and are incapable of recovering their form after rearrangements of their components. Natural forms are robust, contingent artificial forms are fragile. In the case of natural forms, the agency of natural law acts "freely" as the guarantor of form. In the case of artifacts there is no such guarantor.

These considerations highlight an interesting difference between natural and artificial forms. In the case of artifactual forms such as Lego assemblages or watches or other types of machines which are put together from the bottom up mechanically, we have an *infinity of forms*, but each is led up to or assembled by only a few or even only one unique constructional path. In the case of natural forms on the other hand—and the folds are classic examples—there is a *finite number of forms* but an *infinite* number of paths via which their actualization may be achieved.

ROBUSTNESS: RESISTANCE TO MUTATION

The 3D conformations of the folds exhibit another sort of robustness—they are

remarkably resistant to evolutionary changes in their amino acid sequences (Brandon & Tooze, 1999). As referred to above, although there are only a limited number of folds permitted by physics, many very different apparently unrelated amino acid sequences can fold into the same form (Gerlt & Babbitt, 2001). Because protein functions depend on the maintenance of a stable scaffold, the tolerance of the folds to sequential changes may well confer another crucial element of robustness making them "immune to most mutational insults" and hence evolutionarily reliable constructional units of the cell. The self-organization of the same fold from very different amino acid sequences again underscores the natural autonomy and Platonic primacy of the fold over its material constituents and highlights the fact that the folds are natural existents and not artifactual aggregates of matter like machines, which do not possess any natural autonomy over their components, and are far less tolerant of variations in their basic constituents.

We speculate that the fact that the robustness of the folds [which enables them to maintain their forms and dependent functions in the face of both mutational challenges and conformational disturbances due to the turbulence of the cell's interior] is "natural" may have deep evolutionary implications. The robustness of biological systems is generally conceived of as being analogous to that of advanced machines utilizing such devices as feedback control, parallel circuitry, error fail-safe devices, redundancy and so forth (Keller, 2000; Kitano, 2002; Csete & Doyle, 2002). But such robustness which we suggest might be termed "artifactual robustness" is inherently complex and can only be arrived at after millions of years of evolution and is necessarily a secondary and derived feature of any biological system or structure. The robustness of the folds is a natural intrinsic feature of the folds themselves and not a secondarily evolved feature. Robustness of this sort is "for free" and does not require the intervention of natural selection. Such robustness has the enormous evolutionary advantage in that it provides evolution with "ready-made" stable structures upon which to build more complex structures and functions. We speculate that an element of natural robustness may be a necessary feature of all biological forms utilized by evolution from the molecular to the organismic level.

PARTS AND WHOLES: A UNIQUE RECIPROCAL FORMATIVE RELATIONSHIP

If the folds were contingent assemblages of matter like Lego constructs, watches or other sorts of artifacts—where the parts are the primary things and pre-exist the whole—then the various parts of each fold, the constituent α helices and β sheet conformations and higher order submotifs which make up the whole should be stable structures (like Lego bricks) which should exist prior to and independent of the wholes in which they occur. And if this were the case then the structure of the whole fold should be easily predictable (as in the case of an artifactual assemblage like Lego) from the character and properties of its parts in isolation. But this is evidently not the case.

On the contrary, many of the unique secondary structural motifs which make up a mature fold are in most cases, either highly metastable outside the fold or non-existent. We can state this formally by saying that most submotifs which make up a fold are existentially dependent on being part of the native conformation of the whole fold, outside of which they have no independent existence.

Evidence that the specific confirmation adopted by particular segments of the amino acid chain is determined by the whole was gained some time ago in the classic complementation experiments of Anfinsen, which showed that the isolated S-peptide of ribonuclease, which comprises residues 1–20 of the enzyme, is a random coil in isolation, whereas most of it forms an α helix in combination with the rest of the molecule. Similarly, the fragments of myoglobin released from the molecule after cyanogen bromide treatment, the so-called peptides 1, 2 and 3, have very little residual secondary structure after their removal from the myoglobin molecule. Peptide 1 aggregates, peptide 2 is a random coil and the large central peptide, peptide 3, has only residual α helical properties (Anfinsen, 1973). Evidently the structures

adopted by the different sections of the amino acid sequence are ''context dependent.'' The case of the prion proteins illustrates again that the same sequence may fold into two alternative structures depending on context (Prusiner, 1995; Manson, 1999). Studies on the Arc repressor molecule (Cordes *et al.*, 2000) have revealed that one section of the protein can switch from a helical to a sheet conformation depending on minor environmental changes, including temperature and solvent conditions. The fact that the same or very similar sequences may adopt different secondary structures dependent on context is at the heart of the whole problem of predicting the 3D structure of a protein from its amino acid sequence. If the same sequences always adopted the same structures whatever the context—if in other words the substructures of proteins were like the building blocks of Lego, or the cogs of a watch—then the problem of prediction of ''global form'' from the form of the building blocks would be easy—but of course this is not the case and the problem is far from solved.

The linguistic analogy again springs to mind. The meaning of a word in a sentence may vary depending on the sentence in which it occurs —its context. In a spoken sentence of English for example, the sound represented by the letters ''rite'' or ''right'' may refer from anything from a medieval ritual to a movement or a moral judgement. Outside the context in which it occurs it is impossible to determine its meaning. Consequently it is impossible to determine the meaning of a ''whole'' sentence from the study of its constituent words in isolation.

Proteins, like sentences, are intensely holistic entities. All the current evidence suggests that the various parts of the fold—the various constituent α helices and β sheet conformations and higher order submotifs—exert what appears to be a mutual and *reciprocal formative influence on each other and on the whole*, which itself in its turn exerts *a reciprocal formative influence on all its constituent parts*. In this characteristic proteins are unlike any other material objects with which we are familiar.

It is interesting to recall that Kant (1790), in a well-known section of his *Critique of Judgement* argues that the formative reciprocity of parts and whole is the defining characteristic of a natural unity. In such a unity: ''the parts combine themselves into the unity of a whole by being reciprocally cause and effect of their form … [and] the whole may conversely, or reciprocally determine in its turn the form and combination of all its parts … [in other words as he continues] every part may be thought of as owing its presence to the agency of the remaining parts, and also as existing for the sake of the others, and of the whole … [and he concludes] *an organized natural product is one in which every part is reciprocally both end and means*.'' [our emphasis].

Kant also pointed out that objects whose parts exhibit such a reciprocal formative influence on each other belong to a completely different order of being to that of the contingent artifactual assemblages. In the case of a watch for example, although it is true that each part is there for the sake of the others—so that they might coherently function together for a purpose (telling the time)—the parts do not owe their presence to the agency of the others. In Kant's words: ''Hence one wheel in the watch does not produce the other, and still less does one watch produce other watches … An organized being is not therefore, a mere machine … An organized being possesses inherent formative power.''

Whatever may be the merits of Kant's views, there is no doubt that the parts to whole relationship in the case of the folds is nothing like that of any contingent artifact ever built or conceived of—as Kant insists: ''no instrument of art can answer to this description'' and the type of organization ''has nothing analogous to any causality known to us.''

Evolution by Natural Law

If the folds are indeed lawful natural forms arising out of the intrinsic physical properties of amino acid sequences, it is hard to see how selection for function can have played a significant role in their origin or evolution. The problem is somewhat like trying to provide a selectionist/functional explanation for the spherical shape of a cell, or the flat shape of the cell membrane! Selection may select a whole fold and then modify it for function and this is clearly

what has happened in many cases, but it is very hard to see how selection for biological function can put together a fold in the first place. To explain the origin and evolution of the folds we are forced, it seems, to turn to explanations in terms of natural law, as we would in the case of other types of natural forms, such as atoms, crystals or galaxies and more specifically to the twin 19th century concepts of origin by saltation or via Owen's "preordained paths."

PER SALTUM

In the case of the folds, origin *per saltum* or in Finkelstein's (1994) words "by choice from random [amino acid] sequences" is only a feasible mechanism if the protein folds are easy to find by chance and therefore common in amino sequence space. There is no doubt that individual α helices and β sheets are very common in sequence space and as the folds arise naturally from combinations of these subunits one might presume that the folds themselves would be relatively common. Whether they are or not has been a subject of considerable discussion (Finkelstein, 1994; Cordes *et al.*, 1996; Sauer, 1996; Axe, 2000). Thermodynamic considerations of the random characteristics of fold sequences support the contention that stable folds are common in sequence space (Finkelstein & Ptitsyn, 1987; Finkelstein, 1994; Finkelstein *et al.*, 1995). In Finkelstein's (1994) words "little editing of a random sequence is necessary for the formation of the protein globule itself." In libraries of random amino acid sequences, α helical proteins displaying cooperative thermal denaturation and specific oligomeric states have been recovered at frequences of 1% (Cordes *et al.*, 1996). Evidence that different stable structures may be close in sequence space is supported by the case of the prion proteins and other cases where different structures may be adopted by the same sequence, such as the Arc repressor mutant, referred to above. Discussing the implications of the conformational switch in the Arc repressor the authors (Cordes *et al.*, 2000) comment: "The intermediate can adopt either fold … Thus distinct protein folds need not be isolated islands in sequence space, but can be linked by evolutionary bridges where multiple native structures coexist." Another line of evidence which suggests that the folds must be relatively common in sequence space is the existence of overlapping genes. These appear to occur in the genomes of almost all organisms. New functional proteins could never have been discovered or evolved, embedded in existing gene sequences, if the folds were not relatively common in sequence space.

The current consensus view (Finkelstein, 1994; Plaxco *et al.*, 1998; Brandon & Tooze, 1999) is that stable folds are in fact quite common in amino acid sequence space. Brandon & Tooze (1999) have even speculated that as many as one in a hundred random amino acid sequences may fold into a stable form. If indeed folds are that common in sequence space—occurring, say, at a frequency of one in a hundred random sequences—then, this might mean that wherever random polypeptides are synthesized anywhere in the cosmos—in the laboratory, in a pre-biotic soup or in a primeval cell—all the basic folds utilized by life on earth would be bound to be generated after only a few thousand trials. And this leads us to surmise that if at some stage in cellular evolution "random polypeptides" were synthesized in great numbers, then all the folds might have been discovered quite easily by chance. This would mean that in the right environment, just as atoms are assembled in the stars and crystals form when rocks cool slowly, so the protein folds would also be formed automatically, wherever conditions permitted the synthesis of any quantity of polypeptides to occur. And because the association of many proteins with their prosthetic groups is basically spontaneous, and does not require the intervention of an enzyme, this raises the possibility that many protein functions may also have been generated deterministically in the protocell without the necessity for selection, by the direct association of small organic compounds with particular protein folds. In effect this means that merely by synthesizing a few thousand random polypeptides in a "broth" containing the basic biochemicals used by life on earth, including enzymic prosthetic groups, all the necessary proto-functional enzymes needed for cellular metabolism might be generated *per saltum* by natural law. These primitive enzymes could then

be fine tuned by selection to generate the highly efficient enzymes of modern cells. In effect the origin and evolution of the protein-based biochemistry of modern cells may be "a free lunch." Evidence that intermediary metabolism may also have been given deterministically in the protocell was obtained in a recent study (Morowitz *et al.*, 2000) which concluded that: "The chemistry at the core of the metabolic chart is necessary and deterministic and would likely characterize any aqueous carbon based life anywhere it is found in this universe."

PRE-ORDAINED PATHS

The alternative to *per saltum* models is to envisage physically determined "constructional series" or evolutionary pathways starting from say, a simple single α helix structure and leading via a series of small motifs to the final fold. One example of a simple, two-step "constructional sequence" for the evolution of the classic TIM barrel from a half barrel was reported recently (Lang *et al.*, 2000). But how feasible might such constructional pathways be in the case of many folds? In the case of some folds—the globin fold for example—no one has yet been able to provide a credible constructional sequence from simple motif to final fold to show how the fold might have come about via a series of stable intermediate forms. Some folds, like perhaps the TIM fold, may lend themselves to construction from simpler motifs but this may not be true of all folds. Of course as any set of small stable motifs and constructional sequences in pre-fold space would also be a finite set and very much given by physics, then such constructional sequences if they exist, would be no less "built-in" than the final set of stable folds to which they lead. However as there would be different routes through this pre-fold space to the 1000 folds, there would inevitably be an element of contingency in the actual routes taken. Nonetheless, the 1000 protein folds would still represent a physically determined or "built-in" bottle neck through which protein evolution had to pass and through which it would have to pass on any earth-like planet, where life uses proteins constructed out of the same 20 amino acids.

If it is possible to derive folds via evolutionary constructional sequences starting from say a simple α helix and leading via a double helical motif and so on, then selection for biological functions may have played at least some role. However as many authors have pointed out, in the context of protein evolution, selection must have a detectable proto-function to start with (Ohno, 1970, 1984; Keese & Gibbs, 1992). This means that *before* selection begins there must be at least some sort of stable scaffold on which a function can be hung. And in the case of some functions, this might necessitate a stable scaffold plus a prosthetic group! In the end, whatever role selection might have played it must obviously have been highly constrained by the various types of stable submotifs permitted by physics.

The idea that the origin and evolution of the folds occurred essentially "for free"—either *per saltum* or via physically pre-determined constructional paths, because of the essential "naturalness" of the folds and their frequency in amino acid space—is enormously attractive because it provides a plausible framework for understanding the origin of protein-based life. Such a model contrasts very favorably with the traditional selectionist scenario and the need for a long involved process of cumulative selection (Dawkins, 1986) to direct the assembly of highly improbable contingent "artifact-like" structures. If the folds were indeed "artifact-like" structures, if their complexity entailed what Davies (2001) terms "instructed complexity" it is not only difficult to see how they could possess the prerequisite robustness to function reliably as the constructional units of the cell (discussed above) but also difficult to see how they could ever have originated naturally (Davies, 2001). The re-adoption of the 19th century metaphor of the crystal and the reassigning of biological order as "natural" rather than "artificial," at least in the case of the folds, goes a long way to solving the problem of the origin of protein-based life and may provide a general naturalistic explanatory framework for understanding other aspects of the origin of life.

LIFE AS INTEGRAL TO NATURE

The discovery that the folds are natural forms, whose evolution is determined largely

by physical law and which are bound to arise spontaneously—"for free"—in any large set of random amino acid sequences, strongly supports the widely held belief among origin of life researchers (already mentioned above) that life is itself an inevitable end of chemistry—a phenomenon which is bound to arise in the correct environmental conditions, perhaps in space or perhaps on the surfaces of newly formed planets (Lehninger, 1982; De Duve, 1991; Sowerby *et al*., 2001). It also provides new support for the currently fashionable Anthropic view that the laws of nature appear to be fine tuned for life (Barrow & Tippler, 1986; Denton, 1998; Davies, 2001). For the lawful nature of the folds provides for the first time evidence that the laws of nature may not only be fine tuned to generate an environment fit for life (the stage) but may also be fine tuned to generate the organic forms (the actors) as well, in other words that the cosmos may be even more biocentric than is currently envisaged!

The lawful nature of the folds together with the intriguing fact that many of the 20 protogenic amino acids—out of which the folds are constructed—are amongst the commonest amino acids found in meteorites and the easiest amino acids to generate in pre-biotic syntheses (Miller & Orgel, 1974) is surely of considerable significance, consistent with and supporting a deterministic theory of the origin of life (or at least of proteins) and by extrapolation the whole Platonic cosmogony—raising the possibility that all organic forms and indeed the whole pattern of life may finally prove to be the determined end of physics and life a necessary feature of the fundamental order of nature. In their book *Biochemical Predestination*, the authors Kenyon & Steinman (1969) echoing the early 19th century views of Owen and Schwann, conclude with sentiments consistent with the pre-Darwinian concept of evolution by natural law and the viewpoint defended in this paper: "the ultimate development of the living cell is determined by the physiochemical properties possessed by the starting compounds from which these systems evolved. In other words the ultimate characteristics of the living cell can be traced back to the nature of the starting compounds from which it was produced. Therefore we should not look upon the appearance and development of the living cell as an improbable phenomenon but rather as one which followed a definite course governed and promoted by the properties of the simple compounds from which the process began."

Discussion

The protein folds clearly represent a finite set of about 1000 natural forms determined like atoms and crystals and other natural forms by the laws of physics. They can be classified into distinct structural types, their structures can be accounted for in terms of a rational morphology of constructional rules, they have remained invariant for billions of years, and in many cases their functional adaptations are clearly secondary modifications of what are evidently ahistoric primary forms. They are robust both in their capacity to resist long-term evolutionary mutational challenges and in their capacity to maintain and regain their proper form in the face of the destabilizing challenges posed by the buffeting and turbulence of the cell's interior. Depending on their frequency in sequence space, their evolutionary origin may have occurred either by saltation, or via physically determined intermediates, in other words by pre-ordained constructional pathways. In short, they do not conform in any way to the Darwinian conception of organic forms as contingent "Lego-like" functionally contrived assemblages of matter. On the contrary, they are wonderful exemplars of the pre-Darwinian and Platonic conception of organic forms as abstract, lawful and rational features of the eternal world order, which will occur throughout the cosmos wherever the same 20 protogenic amino acids are used to make proteins.

This is a remarkable, even historic discovery. *For the folds represent the first case in the history of biology where a set of complex organic forms can be shown to be unambiguously lawful natural forms in the classic pre-Darwinian sense*. In the realm of the proteins, Owen's metaphor of the crystal has displaced Darwin's metaphor of the watch. That what are perhaps the most important set of forms in the biological realm, the fundamental constructional units of life and

more specifically of the protein-based biochemistry of the modern cell, the most thoroughly characterized of all known organic forms to date, known and understood to the ultimate atomic level, should have turned out to be such perfect exemplars of the pre-Darwinian Platonic cosmogony, and the idea of natural law as the major determinant of organic form and evolution, is an important and intriguing discovery in its own right. But the real significance of this finding lies in the deep implications it holds for two key areas of biology, namely the origin of life and the fundamental nature of organic form. We have already seen that in the area of the origin of life this new ''naturalistic'' conception of the folds goes a long way to providing a rational explanation for the origin of protein-based life. No less significant is the radical challenge it poses to the current deeply held Darwinian presumption, that all complex organic forms (like the folds?) are the contingent artifact-like products of selection—''chance caught on the wing''—to cite Monod's (1972) famous phrase, and that natural law has played no role in directing the course of evolution or in the determination of organic form. In the context of the ''folds'' these presumptions are no longer secure.

It is more than anything else the complex hierarchic structure of the folds—their being composed of clearly defined substructures and submotifs combined together into what appear seemingly to be irregular complex hierarchic wholes, the sort of order which is so characteristic of that of a machine or artifact—which conveys the irresistible feeling that such forms *could not possibly be natural or lawful*. The fact that such structures should be lawful natural and in essence ''simple'' is entirely counterintuitive. And the question obviously arises, might there be other sets of lawful self-organizing organic forms which arise like the folds out of the intrinsic properties of their basic material constituents? A number of considerations suggest that this possibility cannot be so easily dismissed. We believe that in the case of at least two other classes of self-organizing forms—microtubular and cell forms—there is at least some preliminary and intriguing evidence for believing that these might turn out to represent like the folds, preferred arrangements of matter which are determined by various ''constructional rules'' or ''organizational laws.''

There is no doubt that the bipolar aster arises by the self-organization of microtubules and molecular motors (Kirschner & Mitchison, 1986, Kirschner *et al*., 2000) or in the words of Hyman & Karsenti (1996) from ''the intrinsic characteristics of its parts.'' And the aster is not the only microtubular form that arises in this way. This is beautifully shown by studies of the various forms which arise *in vitro* out of the interaction of molecular motors and microtubules (Nedelec *et al*., 1997). The results are very instructive. Merely by altering the concentration of the motor protein kinesin in the presence of stabilized microtubules, a variety of spatially organized structures, including the aster can be generated. Clearly these forms are not specified directly in the DNA, nor does any genetic program direct their assembly. They are to be explained as the ''lawful'' outcome of local dynamic interactions between a few basic components in the cell. They appear to be another set of natural forms no less natural than the protein folds, although the laws involved are of course very different. The authors (Nedelec *et al*., 1997) summarized the results of their work thus: ''These simplified experiments show that the basic structural ''vocabulary'' used by the cell is extremely rich. By using just two basic components and simple local rules of interaction we obtained a large variety of assembled structures. By extending our system to include other interacting components, such as nucleation centers for microtubules or different motors, it will be possible to explore the conditions needed for the formation of other structures. Another direction of study would be to introduce some elements of regulation, making it possible to search for *rules* underlying the choice of different ''words'' from this large vocabulary of self organized structures.''

These results are consistent with the notion that microtubular forms represent another set of lawful forms which arises spontaneously, like the folds, out of the intrinsic properties of their basic material constituents and where it may be possible eventually to derive them deductively from a unique set of contructional rules

(analogous to those that apply to the folds) which we might term "laws of microtubular form." It is surely not so far fetched to see the aster (like a protein fold) as a primary "given of physics" which has been secondarily co-opted and modified by selection for its crucial functional role in cell division.

Might cell forms also be lawful? The observation that some cell forms are almost as ancient and invariant as the protein folds lends some support to the notion. The cell form of *Tetrahymena* for example, has remained basically unchanged for perhaps 1000 million years among the set of sibling species termed the *Tetrahymena* swarm (Nanney, 1982). Even details of the arrangements of the cilia round the oral apparatus have remained constant over the millions of years since the various species of *Tetrahymena* diverged—almost certainly some time before the origin of the vertebrates, deep in pre-Cambrian times, hundreds of millions of years ago (Nanney, 1982). What makes the situation even more curious is the fact that the molecular constituents which generate the identical forms vary tremendously. So there is invariance of cell form with considerable variation in the building blocks. (Again the similarity here to the folds is striking, where as we have seen, the same 3D form may be specified by many different sequences, sometimes so diverse that no sequential relationship can even be detected.) Frankel (1983) has speculated that part of the explanation may be that developmental constraints impose restrictions on allowable forms. But this may not be the whole story. As Nanney (1982) has commented: "It must be concluded that few, if any, improvements are possible, at least by the usual one-step opportunistic methods of microevolutionary progression. The phenotype is prevented, by its very, perhaps *arbitrary complexity*, from any kind of substantial change." This is close to acknowledging that certain features of *Tetrahymena* cell form may be lawful afunctional structures, given by physics not selection.

Reinforcing the possibility that ciliate cell forms may be lawful is the extraordinary ability of ciliate cells like *Stentor* to recover their "proper form" after microsurgical manipulations. These recoveries are consistent with the possibility that the whole cell is behaving like a natural form, which can recover its proper form by searching a conformational space for its correct conformation, just like a folding protein, or a self-assembling aster. In the case of *Stentor* the process can take up to 40 hr (Tatar, 1961) far longer than the time taken by a protein (1 s) or an aster (several minutes) to search their respective conformational spaces.

Commenting in a recent *Cell* review Kirschner and his colleagues (Kirschner *et al.*, 2000) remark: "In the case of *Stentor* pattern reformation after surgery suggests that incorrect assembly states of the cortical cytoskeleton are less stable than correct ones, and that off rates sufficient to explore new configurations must exist." This is consistent with the view that the regeneration of cell form in Stentor involves (as is the case in the self-assembly of proteins or the self-assembly of the bipolar aster) an energy-dependent exploration of a higher order structural landscape which leads eventually as in the case of a fold to a *preferred state or free energy minimum*.

Conclusion

Whether or not there are other sets of lawful organic forms, there is no doubt that the universe of protein folds represents a Platonic universe of precisely the kind sought after by pre-Darwinian biology. There is no question that in this universe, functional adaptations are secondary and trivial modifications of what are evidently essentially invariant crystal-like "givens of physics" and evolution is by *law* not *selection for function*. It would have delighted Goethe and Owen! It is a universe where abstract rules, like the rules of grammar, define a set of unique immaterial templates which are materialized into a thousand or so natural forms—a world of rational morphology and pre-ordained evolutionary paths. We may have as yet, no evidence to support Owen's belief in the existence of vertebrates on extraterrestrial planets, but at least as far as the 1000 protein folds are concerned, we may be sure that they will be present everywhere in the cosmos where there is carbon-based life, utilizing the same 20 protogenic amino acids.

The possibility that organic forms may prove eventually to be intrinsic features of nature rather than contingent artifacts has of course immense intellectual appeal. For it holds out the prospect that the study of organic form, might eventually become as the pre-Darwinian biologists such as Goethe, Goeffroy and Owen had always hoped, a fully rational and predictive science, and that biology may in the end be unified with physics in Plato's timeless realm of the gods.

We wish to thank Dr G. Kumaramanickavel, Reader and Head, Department of Genetics and Molecular Biology, Vision Research Foundation, 18 College Road, Chennai 600 006, India, and other members of his staff, especially Dr V. L. Ramprasad for critically reading the manuscript and for editorial assistance.

REFERENCES

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181,** 223–230.

Aristotle (1937). *Parts of Animals*. London: Heinemann (A. L. Peck, English translation).

Axe, D. (2000). Extreme functional sensitivity to conservative amino acid changes on enzyme exteriors. *J. Mol. Biol.* **301,** 585–596.

Barrow, J. D. & Tipler, F. J. (1986). *The Anthropic Cosmological Principle*. New York: Oxford University Press.

Ball, P. (1999). *The Self Made Tapestry*. Oxford: Oxford University Press.

Brandon, C. & Tooze, J. (1999). *Introduction to Protein Structure* 2nd edn. New York: Garland Publishing Inc.

Chambers, R. (1969). *Vestiges of the natural History of Creation*. Leiester: Leicester University press. New York.

Chothia, C. (1993). One thousand families for the molecular biologist. *Nature* **357,** 543–544.

Chothia, C., Hubbard, T., Brenner, S., Barnes, H. & Murzin, A. (1997). Protein folds in the all α and all β classes. *Annu. Rev. Biophys. Biomol. Struct.* **26,** 597–627.

Cordes, M. H. J., Davidson, A. R. & Sauer, R. T. (1996). Sequence space, folding and protein design. *Curr. Opin. Struct. Biol.* **6,** 3–10.

Cordes, M. H. J., Burton, R. E., Walsh, N. P., Mcnight, J. & Sauer, R. T. (2000). An evolutionary bridge to a new fold. *Nat. Struct. Biol.* **7,** 1129–1132.

Csete, M. E. & Doyle, J. C. (2002). Reverse engineering of biological complexity. *Science* **295,** 1164–1169.

Davies, P. (2001). *The Fifth Miracle*. Middlesex, U.K.: Penguin Books.

Darwin, C. (1958). *The Autobiography of Charles Darwin*. London: Collins.

Dawkins, R. (1986). *The Blind Watchmaker*. London: Longman Scientific and Technical.

de Duve, C. (1991). *Blueprint for a Cell*. North Carolina: Neil Patterson Publishers.

Denton, M. J. (1998). *Nature's Destiny*. New York: Free Press.

Denton, M. J. & Marshall, C. J. (2001). Laws of form revisited. *Nature* **410,** 417.

Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M. & Karplus, M. (2000). Understanding protein folding via free energy surfaces from theory and experiment. *TIBS* **25,** 331–339.

Driesc, H. H. (1914). *The History and Theory of Vitalism*. London: Macmillan, Co. Ltd. (C. K. Ogden, English translation).

Driesc, H. H. (1929). *The Science and Philosophy of the Organism*. London: A. & C. Black.

Finkelstein, A. V. (1994). Implications of the random characteristics of protein sequences for their three-dimensional structure. *Curr. Opin. Struct. Biol.* **4,** 422–428.

Finkelstein, A. V. & Ptitsyn, O. B. (1987). Why do globular proteins fit the limited set of folding patterns? *Prog. Biophys. Mol. Biol.* **50,** 171–190.

Finkelstein, A. V., Badretdinov, A. Y. & Gutin A. M. (1995). Why do protein architectures have Boltzman-like staistics? *Proteins* **23,** 142–150.

Frankel, J. (1983). What are the developmental underpinnings of evolutionary changes in protozoan morphology? In: *Development and Evolution* (Goodwin *et al.*, eds), pp. 279–314. Cambridge: Cambridge University Press.

Gerlt, J. A. & Babbit, P. C. (2001). Divergent evolution of enzymic function. *Annu. Rev. Biochem.* **70,** 209–246.

Goodwin, B. C. (1994). *How the Leopard Changed its Spots*. London: Weidenfeld & Nicholson.

Gould, S. J. (1989). *Wonderful Life*. New York: W. W. Norton.

Gould, S. J. & Lewontin, R. C. (1979). The spandrels of San Marco and the panglossian paradigm: a critique of the adaptational program. *Proc. R. Soc. London B* **205,** 581–598.

Holm, L. & Sander, C. (1996). Mapping the protein universe. *Science* 273, 595–602.

Holm, L. & Sander, C. (1999). Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* **27,** 275–279.

Honig, B. (1999). Protein folding: from the Levinthal paradox to structure prediction. *J. Mol. Biol.* **293,** 283–293.

Hyman, A. A. & Karsenti, E. (1996). Morphogenetic properties of microtubules and mitotic spindle assembly. *Cell* **84,** 401–410.

Kant, I. (1790). *The Critique of Judgement*. Oxford: Oxford University Press (J. C. Meredith, English translation).

Kauffman, S. A. (1993). *The Origins of Order*. New York: Oxford University Press.

Keese, P. K. & Gibbs, A. (1992). Origin of genes: "big bang" or continuous creation? *Proc. Natl Acad. Sci. U.S.A.* **89,** 9489–9493.

Keller, E. F. (2000). *The Century of the Gene*. Cambridge, MA: Harvard University Press.

Kenyon, D. H. & Steinman, G. (1969). *Biochemical Predestination*. New York: McGraw-Hill Book Co.

Kirschner, M. & Mitchison, T. (1986). Beyond self-assembly: from microtubules to morphogenesis. *Cell* **45,** 329–342.

Kirschner, M., Gerhardt, J. & Mitchison, T. (2000). Molecular vitalism. *Cell* **100,** 79–88.

KITANO, H. (2002). Systems biology: a brief overview. *Science* **295,** 1662–1664.
LANG, D., THOMAS, R., HENN-SAZ, M., STERNER, R. & WIMMANNS, M. (2000). Structural evidence for evolution of the α/β barrel scaffold by gene duplication and fusion. *Science* **289,** 1546–1550.
LEHNINGER, A. L. (1982). *Principles of Biochemistry*. New York: Worth Publishers Inc.
LIMA-DE-FARIA, A. (1988). *Evolution Without Selection*. New York: Elsevier.
LINDGARD, P. & BOHR, H. (1996). How many protein fold classes are to be found? In: *Protein Folds* (Bohr, H. & Brunak, S., eds), pp. 98–102. New York: CRC Press.
MANSON, J. C. (1999). Understanding the transmission of the prion diseases. *Trends Microbiol.* **7,** 465–467.
MCDOUGAL, W. (1938). *The Riddle of Life*. London: Methuen.
MILLER, S. L. & ORGEL, L. E. (1974). *The Origins of Life on the Earth*. Englewood Cliffs, NJ: Prentice-Hall.
MONOD, J. (1972). *Chance and Necessity*. London: Collins.
MOROWITZ, H. J., KOSTELNIK, J. D., YANG, J. & CODY, G. D. (2000). The origin of intermediary metabolism. *Proc. Natl Acad. Sci. U.S.A.* **97,** 7704–7708.
MURZIN, A. G. & FINKELSTEIN, A. V. (1988). General architecture of the α helical globule. *J. Mol. Biol.* **204,** 749–769.
NANNEY, D. L. (1982). Genes and phenes in tetrahymena: the epigenetic paradox in protozoa. *Bioscience* **32,** 783–788.
NEDELEC, F. J., SURREY, T., MAGGS, A.C. & LIEBLER, S. (1997). Self-organization of microtubules and motors. *Nature* **389,** 305–308.
OHNO, S. (1970). *Evolution by Gene Duplication*. New York: Springer-Verlag.
OHNO, S. (1984). Birth of a unique enzyme from an alternative reading frame of a preexisting internally repetitious coding sequence. *Proc. Natl Acad. Sci. U.S.A.* **81,** 2421–2425.
ORENGO, C. A., JONES, D. T. & THORNTON, J. M. (1994). Protein superfamilies and domain structures. *Nature* **372,** 631–634.
ORENGO, C. A., MICHIE, A. D., JONES, D. T., SWINDELLS, M. B. & THORNTON, J. M. (1997). CATH—a hierarchic classification of protein domain structures. *Structure* **5,** 1093–1108.
OWEN, R. (1849). *On the Nature of Limbs*. London: Jan Van Voorst.
OWEN, R. (1866). *Anatomy of Vertebrates*. London: Longmans and Green.
PLAXCO, K. W., RIDDLE, D. S., GRANTCHAROVA, V. & BAKER, D. (1998). Simplified proteins: ''minimalist solutions to the protein folding problem.'' *Curr. Opin. Struct. Biol.* **8,** 80–85.
PRUSINER, S. B. (1995). The prion diseases. *Sci. Am.* **272,** 30–37.
PTITSYN, O. B. & FINKELSTEIN, A. V. (1980). Similarities of protein topologies: evolutionary divergence, functional convergence or principles of folding. *Quart. Rev. Biophys.* **13,** 339–386.
RICHARDS, R. J. (1992). *The Meaning of Evolution*. Chicago: University of Chicago Press.
RICHARDSON, J. S. (1981). The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34,** 167–339.
RUPKE, N. A. (1994). *Richard Owen*. New Haven, CT: Yale University Press.
RUSSELL, E. S. (1916). *Form and Function*. London: John Murray.
SAUER, R. T. (1996). Protein folding from a combinatorial perspective. *Fold Des.* **1,** R27–R30.
SCHWANN, T. (1847). *Microscopical Researches*. London: Sydenham Society. H. Smith, English translation; (Original published in German in 1839).
SOWERBY, S. J., HOLM, N. G., & PETERSEN, G. B. (2001). Origins of life: a route to nanotechnology. *Biosystems* **61,** 69–78.
TATAR,V. (1961). *The Biology of Stentor*. Oxford: Pergamon Press.
TAYLOR, W. R. (2000). Searching for the ideal forms of proteins. *Biochem. Soc. Trans.* **28,** 264–269.
TAYLOR, W. R., MAY, A. C. W., BROWN, N. P. & ASZODI, A. (2001). Protein structure: geometry, toplogy and classification. *Rep. Prog. Phys.* **64,** 517–590.
THOMPSON, D' AREY, W. (1942). *On Growth and Form*. Cambridge: Cambridge University Press.
WADDINGTON, C. H. (1962). *New Patterns in Genetics and Development*. New York: Columbia University Press.
WEBSTER, G. & GOODWIN, B. C. (1982). The origin of species: a structuralist approach. *J. Soc. Struct.* **5,** 15–47.
YOCKEY, H. P. (1992). *Information Theory and Molecular Biology*. Cambridge: Cambridge University Press.
ZHANG, C. & DELISI, C. (1999). Estimating the number of protein folds. *J. Mol. Biol.* **284,** 1301–1305.