

The protein folds as *complex natural forms*: Evidence that the properties of matter may be '*fine tuned*' for protein based life.

MJ Denton
Biochemistry Department
University of Otago
Dunedin

Abstract

In the *Fitness* Henderson argued that many of the basic *natural* constituents of carbon based life on earth, such as water, carbon dioxide, simple carbon compounds and basic *natural* processes such as oxidations appear to form a *uniquely fit ensemble* of constituents for “the organic mechanism which we call life.” He concluded that as it is unlikely that any other ensemble of constituents exists [none is known ninety years later] which “could possess . . . such great fitness” then carbon based life is a unique phenomenon and the universe specially ‘biocentric.’ Here I show that the one thousand protein folds represent another ensemble of *natural* constituents of life on earth (vastly more complex than the simple constituents considered by Henderson) which may also be uniquely fit for their biological role. This suggests that the properties of matter may be ‘fine tuned’ not just for generic carbon based life but for the specific type of protein based life as it exists on earth.

Introduction

In the *Fitness* Henderson commences his classic argument for the unique fitness of nature for carbon based life with a consideration of the thermal properties of water. He considers its high thermal capacity, high latent heat of evaporation, heat of fusion, expansion before freezing, and expansion on freezing. He shows how these thermal properties work together as if ‘by magic’ to preserve liquid water on the earth’s surface. By comparing the thermal properties of water with those of other fluids, he concludes that its thermal properties endow water with a unique fitness for its role in life. He then examines another key building block of life, carbon dioxide and the mild acid it forms when it combines with water—carbonic acid. The gaseous nature and absorption coefficient of carbon dioxide provide a means of efficiently distributing carbon throughout the biosphere. He shows how the properties of carbonic acid act together to maintain the acid-base balance of the hydrosphere and living tissues of unrivalled efficiency. The result is again ‘like magic.’ All the properties of carbon dioxide and carbonic acid act together to the twin ends of distributing the carbon atom throughout the hydrosphere and at the same time providing a uniquely fit buffer system to maintain its neutrality. Henderson continues his systematic examination of the basic constituents of life, including a review of the properties of organic compounds, and the process of oxidation.

He concludes with the following words: “Hydrogen, oxygen and carbon, water and carbonic acid are not to be rivaled. The fitness of water, carbonic acid, and the three elements make up a unique ensemble of fitness for the organic mechanism. No other [set of constituents] could possess such highly fit characteristics . . . to promote . . . the organic mechanism we call life.” After ninety years no evidence has come to light of any alternative ensemble of natural (or even synthetic) ingredients which could promote the “organic mechanism we call life.” When NASA looks for life in space the search is directed uniquely for the chemical signatures of carbon based life.

Henderson’s core point that the existence of carbon based life depends on a remarkably fit ‘ensemble of natural constituents’—that carbon based life as it exists on earth is integral to

nature and in a sense *preordained in the properties of matter*—can hardly be contested. However, as to whether the ensemble is uniquely or maximally fit, as Henderson (1913) claims, exhibiting “not a single disability,” as might be required to defend a teleological conclusion, remains a question for debate [would snow be fitter if it was black? would water be fitter if it was less reactive?]. And of course the possibility of alternative, non carbon based types of life, has even to day, not been finally excluded. Thus Henderson’s claim that the cosmos is “in essence biocentric” remains a very likely, but still ultimately provisional hypothesis.

Henderson’s method: The evidence for ‘biochemical fine tuning.’

The logic of Henderson’s argument is simple and empirical. He identifies a key *natural* ingredient of life say, A, and then he examines exhaustively its various chemical and physical properties and asks the question - how fit is A for the role it plays in biological systems? He then asks what other *natural* alternatives (B, C, etc) might exist which could perform the role of A. He concludes that as there are no alternatives, that A therefore, exhibits unique fitness for its biological role. Henderson’s method would appear to be a general approach or procedure for determining the ‘Fitness’ of any *natural* constituent of life for its biological role.

In the *Order of Nature* (1917) Henderson makes it clear that his argument depends on “a harmonious unity among the abstract changeless [Platonic] characteristics of the universe.” And in another place he stresses that the properties he is considering are “changeless properties of the elements” (Henderson, 1917). The logic of Henderson’s approach seems to me, as far as I can judge as a non physicist, similar to that of Paul Davies (1982). Davies is considering the unique fitness of an ensemble of ‘changeless’ constants of physics for life, Henderson (1917) is also considering the unique fitness of an ensemble of changeless natural constituents or ‘constants’ of life.

Extending Henderson’s argument

The *Fitness* raises the obvious possibility that it might not just be the basic *generic* features of carbon based life on earth that is predetermined by the properties of matter but that many of the more *specific* characteristics [such as protein based life] and perhaps even the whole course of evolution might be *predetermined* by the properties of matter. The possibility of such a grand extension is hinted at when Henderson discusses the fitness of water and carbon dioxide for “animals like man” and at the end of the *Fitness* where he turns to “properties of matter omitted in our study.” The extension is implicit also in the phrase Henderson (1917) uses in the *Order of Nature* to describe the mutual fitness of water and carbon dioxide etc., as lying “masked within” the properties of the periodic table of elements and where he speaks of the constraints on the “freedom of evolution” inherent in the properties of matter, made manifest in the outworking of the evolutionary process itself.

The protein folds

Proteins have been long invested in a mystical aura. Even the word ‘protein’ first suggested by the Swedish chemist Berzelius in 1838 and derived from the Greek meaning ‘holding first place’ conveys the mystical idea that these molecules play a unique and vital role in biology. One hundred and sixty years later it is quite clear that proteins do indeed play a central role in biology. They carry out most of the catalytic functions of the cell and the core of nearly all the cell’s supramolecular assemblies, such as microtubules and microfilaments are composed of proteins. Indeed complex cellular life would be inconceivable without what Jacques Monod (1972) called their ‘demonic’ abilities. Their importance is such, that much of the organization of the cell, including the role of DNA and RNA and the basic genetic system is devoted primarily to their manufacture. If the cosmos is indeed uniquely fit for life as it exists on earth, then given their central importance, one might predict that the folds would be another *uniquely fit ensemble of natural forms* (analogous too but far more complex than the ensemble of simple natural forms discussed by Henderson in the *Fitness*) and provide evidence that the ‘fine tuning’ goes far beyond that required merely for ‘generic carbon based life’ extending to the very specific form of protein based advanced cellular life that currently exists on earth.

Until recently the essential molecular nature of the protein folds was not understood. In the early seventies when the structure of only a handful of folds had been determined it was still possible to believe that they represented contingent assemblages of matter put together piece by piece by natural selection during the course

of evolution. However, as the structures of more proteins were determined, it became clear that the folds could be classified into various structurally distinct types. Moreover some, such as the globin, the cytochrome and the Rossman folds appeared to be essentially immutable, having been conserved over billions of years of evolution despite extensive changes to their amino acid sequences (Richardson, 1981; Brandon and Tooze, 1999). It also became apparent that the structure of the folds is hierarchical—consisting of basic submotifs combined into higher order structures (Rose, 1996). These early results suggested that the folds might be a set of natural forms determined by physical principles rather than contingent LEGO like assemblages of matter.

More recent studies have largely confirmed these early hints and shown how the chemical and physical properties of the amino acid residues in the linear sequence, and particularly the steric constraints they impose on the conformation of the chain limit the number of ‘allowable’ folds to a very finite number of about one thousand. At a lower level, these constraints include the ϕ ψ torsion angles which limit allowable conformations to a finite number of submotifs such as various helices and sheets. At a higher level, different packing constraints determine the ways these submotifs may be folded into higher-order structures (Chothia and Finkelstein, 1990; Srinivasan and Rose, 2002; Banavar and Maritan, 2003). Folding models based mainly on consideration of the steric constraints which restrict the conformation of the polypeptide backbone such as the recently developed ‘tube picture’ (Banavar and Maritan, 2003) and the hierarchic prediction program, LINUS (Srinivasan and Rose, 2002) are able to predict higher order structures with remarkable accuracy. The various physical constructional rules—‘laws of fold form’—which determine fold structure and limit the number of folds to a small finite set of natural forms, can be likened to the hierarchical constructional rules of physics and chemistry, which determine the way subatomic particles are combined into atoms and atoms into molecules.

Like all other natural forms the folds self assemble into their native forms by free energy minimization, by a process which is akin to a physical phase transition (Scheraga, 1966) between a liquid phase, the unfolded polymer, and the solid phase, the native conformation. The relationship between the unfolded polymer and its native condensed conformation is as *natural* and determined as that between water and ice, between a solution of sodium and chloride ions and a salt crystal. Like all natural forms each fold crystallizes in a unique physico-chemical environment (the interior of the cell) from a unique set of material constituents (a subset of amino acid sequences).

In sum: The thousand folds have been revealed to be a unique set of natural forms which arise deterministically from particular material constituents (different subsets of amino acid sequences) in the physico-chemical environment of the cell (Przytycka et al., 1999; Denton et al., 2002; Banavar and Maritan, 2003a). Just as 92 atoms form the periodic table of chemistry, the one thousand protein folds form another periodic table of basic physical structures (Taylor, 2002) representing what might be termed the ‘periodic table of biology.’

[Note: The discovery that the folds represent a set of natural forms has far reaching consequences because it implies that, aside from the question of the actual biological fitness of the folds, protein based life is an integral part of nature and built into the basic properties of matter. Life, or at least protein based life, is thus revealed to be a natural emergent property of matter.]

The fitness of the folds

Having established that the protein folds are an *ensemble of natural forms*, the next question that arises is are they also *uniquely fit*?

One of the most striking features of proteins, is their extraordinary structural diversity [proteins form materials as diverse as the hard substance of nails and hair, the transparent substance of the lens, the elastic substance of collagen and so on] and the seemingly limitless variety of functions they carry out in the cell compacted into one polymer class. No other single polymer class is known which possess such a rich set of diverse atomic architectures and associated range of different properties. The fact that these *one thousand diverse molecular architectures* and the very great range of structural and functional properties they give rise too, are all derived from the same *basic polymer type*—linear sequences of amino acids contributes an important element of fitness which may be critical. The logistics of assembly and synthesis of life’s key working components would be greatly, if not impossibly complicated, if the cell had to manufacture its key working components, its tool kit, from widely different sorts of materials and macromolecules. By building virtually all its supramolecular assemblies and carrying out most of its metabolic and catalytic activities using a set of basic building blocks—

one periodic table of natural forms—fabricated from only one basic type of polymer the organization of the cell system is greatly simplified.

The protein folds are far more than a just set of diverse atomic architectures. As Anfinsen first showed in the 1950s proteins are capable of *self assembly* without any external assistance from any other molecules in the cell. This ability, which is of course a property of all natural material forms, contributes another essential element of fitness. The daunting challenge the cell would face if it had to actively specify and piece together an array the two hundred amino acids of a typical fold, into a unique complex three dimensional conformation can be easily imagined. Even at the macro scale complex self assembly is still a major technological challenge (Brooks, 2002) and at the nano level where the behavior of matter is subject to all manner of chemical and physical forces remote from everyday experience, *self assembly* may be the only way that matter can be arranged into complex material forms (Lehn, 2002). Being natural forms the self assembling ability of proteins is also *robust and necessarily so*. If they are to maintain their native conformations in the face of the perpetual diverse perturbations imposed on them (by phenomenon such as Brownian motion) in the turbulent nano world of the cell they need be able to recover their native forms from all manner of minor conformational disturbances. If the folds were only able to assemble themselves along a few highly defined paths to their native conformations they would not be able to perform their biological role in the turbulent interior of the cell.

The fitness of proteins also depends on another remarkable property—their *marginal stability* (Banavar and Maritan, 2003a) which allows them to adopt slightly different conformations. The regulation of metabolism via allosteric control mechanisms which links logical control circuits with catalysis in the same molecular fabric, a phenomenally sophisticated mechanism—which Monod saw as ‘demonic’ (Monod, 1972) and the ‘second secret of life’ (Judson, 1980). For as he pointed out, without this capacity for sophisticated control, the regulation of the flow of metabolites through the cell would be chaotic and uncontrolled and probably too chaotic for the maintenance of viable complex living systems (Monod, 1972). Marginal stability is also important for another reason as it allows turnover and catabolism—components which were too stable or too rigid might be difficult if not impossible to recycle. The characteristic of *marginal stability* and the ability to adopt a variety of related conformations may also be important because it also greatly facilitates the evolution of new protein functions (James and Tawfik 2003). Stiff rigid macromolecules would not only be incapable of allosteric regulation or efficient catalysis but would also be unfit for the evolutionary exploration of variant related conformers and functions. Marginal stability, and the consequent flexibility and conformational diversity of the folds confers therefore a wonderful fitness for adaptive evolution—‘evolvability’ (James and Tawfik, 2003)—which has been exploited by life to spectacular effect over the past 3 billion years as witness the limitless variety of protein functions identified and exploited by life on earth. Their evolvability is further enhanced by the fact that many different sequences can fold into the same fold.

A characteristic of the protein folds which confers another important element of fitness is that folds contains a relatively compact hydrophobic core. The hydrophobic cores provide a ‘convenient’ reaction chamber for organic syntheses which are for the most part difficult to carry out unless water is excluded. The dense hydrophobic core may also confer on many folds the ability to stack together into various stable supramolecular assemblies, nucleosomes, cytoplasmic filaments and microtubules and so forth, which form the basic elements of the cytoskeleton.

A final aspect to the fitness of proteins which relates to their accessibility to evolution, especially in the early stages of cellular evolution, close to the origin of life is 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 prebiotic syntheses. Remarkably the folds—so fit in so many other ways—are not so many steps removed from the synthesis of the elements in the stars! A dramatic witness to the fitness of matter for the evolution of protein based life.

The folds provide the cell with one thousand self assembling complex and diverse atomic architectures all compacted into one single polymer class each exhibiting an exquisite balance of robustness and meta stability which allows allosteric control, catalysis and a whole suit of intricate properties ideally suited in so many ways to carry out a vast range of structural and functional biological roles.

Conclusion

Proteins are clearly wonderfully fit. However many of the individual properties of proteins are clearly not unique. Their ability to self fold is shared by certain ribozymes and may be shared by other polymer types. However RNA folds lack the rigid tightly packed hydrophobic cores of proteins and therefore their hydrophobic chemical reaction chambers. The lack of a tightly packed core might also render RNA folds incapable of forming stable filamentous assemblies (like microtubules and so forth). And it is certainly a significant fact that despite the advantage that would accrue to the cell by using RNA polymers for its supra molecular assemblies no cell today utilizes RNA for such structures. But even although some of the properties of folds may be shared with other polymers the evidence currently available is still consistent with the possibility that the total suit of properties making up the biological fitness of the folds is unique. Further evidence regarding the uniqueness of the folds may come from new theoretical approaches to modeling the essential properties of polypeptide chains. One such model is the 'tube model' recently developed by Banavar and Maritan (2003a). Such models may eventually provide the means of estimating what fraction of all polymers might be capable of folding into a set of complex atomic architectures exhibiting the critical mix of robustness and marginal stability of the folds. Considerations based on the 'tube model' suggest (Banavar and Maritan, 2003b) that few other polymers will exhibit this mix of properties.

Altogether the folds do appear to represent an ensemble of *natural* forms which also exhibit a unique fitness for the mechanism we call life," supporting not only Henderson's contention that the cosmos is 'fine tuned' for carbon based life but more specifically for the special protein based variety of life that exists currently on earth. Henderson would I think have been pleased by developments in this field.

References

- Banavar, J. R., Maritan, A. (2003a) Colloquium: Geometrical approach to protein folding: a tube picture. *Rev. Mod. Phys.* 75: 23-34.
- Banvar, J. R., Maritan, A. (2003b) Personal communication.
- Brandon, C., Tooze, J. (1999) *Introduction to Protein Structure*. Second edition. Garland Publishing Inc., New York.
- Brooks, R. A. (2002) *Robot*. Penguin Books, London.
- Chothia, C., Finkelstein, A. V. (1990) The Classification and Origins of Protein Folding Patterns. *Ann. Rev. Biochem.* 59,1007-39.
- Davies, P. C. W. (1982) *The Accidental Universe*. CUP, Cambridge.
- Denton, M. J., Marshall, C., Legge, M. (2002) The Protein Folds as Platonic Forms: New Support for the pre Darwinian Conception of Evolution by Natural Law. *J. Theoret. Biol.* 219, 325-342.
- Henderson L. J. (1913) *The Fitness of the Environment*. Macmillan, new York.
- Henderson L. J. (1917) *The Order of Nature*. Harvard University Press, Cambridge.
- James, L. C., Tawfik, D. S. (2003) Conformational diversity and protein evolution—a 60 year-old hypothesis revisited. *Trends in Biochemical Sciences* 28: 361-368.
- Judson. E. (1979) *The Eighth Day of Creation*. Simon Schuster, New York.
- Lehn, J. M. (2002) Towards Self-Organization and Complex Matter. *Science* 295, 2400-2403.
- Monod, J. (1972) *Chance and Necessity*. Collins, London.
- Przytycka, T., Aurora, R., Rose, G. D. (1999) A protein taxonomy based on secondary structure. *Nature Structural Biology* 6, 672- 682.
- Richardson, J. S. (1981) The Anatomy and Taxonomy of Proteins. *Adv. Protein Chem.* 34, 168-339.
- Rose, G. D. (1996) No Assembly Required. *Sciences (N.Y.)* 36, 26-31.
- Scheraga, H. A., 1963. Intramolecular Bonds in Proteins. II. Noncovalent Bonds, in *The Proteins*, Ed. Neurath, H. Vol 1. 2nd Ed. New York: Academic Press. pp 477-594
- Srinivasan R and Rose G. D. (2002) Ab Initio prediction of Protein Structure Using LINUS. *Proteins: Structure, Function and Genetics* 47, 489-495.
- Taylor, W. R. (2002) A periodic table for protein structures. *Nature* 416, 657-660.