Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Notes of a protein crystallographer: the molecular structure of evolutionary theory

Cele Abad-Zapatero

Center for Pharmaceutical Biotechnology, University of Illinois at Chicago, 900 So. Ashland St, MBRB Building, Room 3020; (M/C 870), Chicago, IL 60607-7173, USA

Correspondence e-mail: caz@uic.edu

By the nature of our profession, macromolecular crystallographers study and analyze the structures of the molecular components of life in the utmost atomic detail. They expend a considerable amount of time expressing those delicate entities in bacterial or cell systems, purifying them, crystallizing them and, of course, solving their three-dimensional structures in their endless search for understanding the connection between their structure and their function in permitting the processes necessary for life. They seek the variants that can be crystallized and frozen in time and space to be able to unravel their three-dimensional structure. We do know what evolution is – or rather, how evolution operates – as we apply the forces of human selection into our expression systems to select the constructs, molecules and mutants of our interest for structural studies.

Nonetheless, when we write or talk about the evolution of the function of those molecules we tend to take a very cavalier attitude. I am referring to the way we talk in our meetings and write in our papers about 'the theory of evolution' or simply 'evolution' for short, and how it relates to the changes that we observe and document in the macromolecular systems that we so dedicatedly study. Perhaps, in this, the 200th anniversary of the birth of Charles Darwin, we should reflect upon this fact and also try making a connecting bridge between our way of understanding evolutionary theory and the way that researchers think about evolution at the level of the organism.

Steven J. Gould (1941–2002), one of the most eloquent science writers of our century and also one of the most prominent scholars of evolutionary theory, has written masterful essays on many aspects of the theory of evolution in a monthly column for the *Natural History* magazine. He wrote 300 of them, without missing a beat, during 25 years and I have devoured many of them, giving me innumerable hours of sheer joy and insightful science reading. More importantly, just in the nick of time and within the same year of his untimely death, he published what will probably be regarded as the 21st century synthesis of evolutionary theory: *The Structure of Evolutionary Theory* (Gould, 2002) (Fig. 1). A monumental treatise of more than 1400 pages, reviewing in his inimitable prose and style the status of the concepts that Charles R. Darwin published 150 years ago. I must confess that I am a great fan of Professor Gould. When the book came out, I purchased a copy right away because even though it was not in my area of scientific expertise I knew that I could look forward to many hours of superb science writing as well as perceptive insights and perspectives into evolutionary theory. I considered writing a review for some of our professional journals as soon as it was published but soon realised that it would probably not have been of general interest; thus I put the project on the back burner. And then came 2009.

All over the world, there have been celebrations dedicated to the work and life of Charles R. Darwin (1809–1882) on the 200th anniversary of his birth, as the most visible figure in the search for understanding biological diversity and the mechanisms responsible for it. Coincidentally, this year also marks the 150th anniversary of the publication of *On the Origin of Species*, undoubtedly one of the most influential books of Western culture. From my personal perspective, on my birthday this year, I treated myself to the newly illustrated edition (Quammen, 2008) of the classic book. This anniversary edition also contains excerpts from his diaries, his biography and his account of the five-year trip (1831–1836) around the world aboard the surveying ship H.M.S. Beagle, under the command of Captain R. T. FitzRoy (1805–1865). Several threads converged in my mind to convince me that it was worth writing this essay. First, celebrating within our scientific field of enquiry Darwin's anniversary; its connection to the publication of the *Origin of Species* and the desire to pay homage to the 'old man'. Second, the need to clarify the key tenets of Darwin's ideas and how our discoveries and those of many others in related (molecular) and unrelated (cell or more traditionally biological) domains are being used

to refine, revise and extend the original concepts of Darwin. Finally, briefly reviewing the importance and subtlety of those evolutionary concepts in structural biology; this is something that I thought needed to be addressed so that we can establish a bridge with the evolutionary biologists speaking in their own terms.

Darwin was the son of a well known physician (Robert W. Darwin) and the grandson of a scholar physician 'evolutionist' (Dr Erasmus Darwin 1731–1802) who wrote about biology and evolution in poetic and flowery discourse. Charles was supposed to continue the family tradition and went to study medicine in Edinburgh. However, that was not to be. He soon found the sight of human blood disturbing and preferred taking excursions in the countryside riding horses, hunting and collecting beetles. A second attempt by his father to provide a university education for his unsettled son almost failed.

This time the idea was to have Charles study theology and focus on being a country minister. This was a quite an acceptable option for a country gentleman. He could preach on Sundays, cultivate his natural philosophy interests and devote some of his spare time to collecting specimens around the countryside to support the views of the most widespread natural philosophy of the time. This was the notion championed by the influential Sir John Herschel who praised the notion that the wonders of nature, in all their richness, should be interpreted as evidence for the omnipotence, the benevolence and the hands-on management style of God. The title of William Paley's book published in 1802 said it all (Paley, 1802). According to this view of biological diversity, undoubtedly, each species reflected a divine act of special creation.

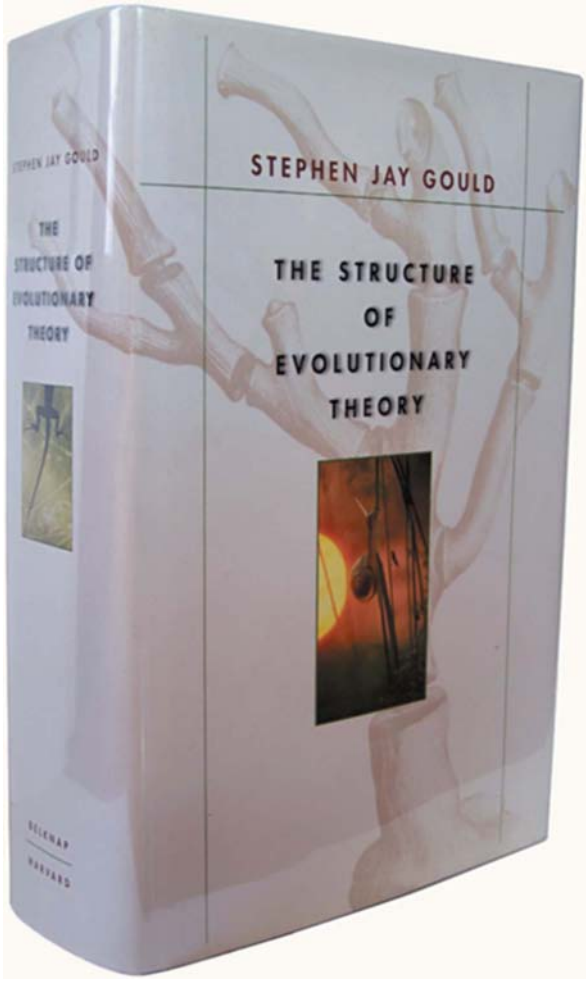

Figure 1
The book cover from *The Structure of Evolutionary Theory* by S. J. Gould (2002) which appears courtesy of Harvard University Press, Copyright © by the President and Fellows of Harvard College. The background image on the cover is a reproduction of the drawing of a certain form of coral described by the Italian naturalist and artist Agostino Scilla (1629–1700) that the author uses as a metaphor to frame the new vision of the Darwinian theory.

Darwin did receive his degree from the University of Cambridge in 1831 but the rest was not to follow. The contingency of history interjected and with the intervention of his uncle (Joshua Wedgewood, of Wedgewood pottery fame) both of them convinced Darwin's father of the unique opportunity offered by the invitation to travel around the world in the surveying ship H.M.S. Beagle. When he returned home in November 1836, after a five year journey around the globe, his life was to take a definitive course: he would be a professional naturalist. He stayed a couple of years in London and connected with the most respected naturalists of the time while he published the narrative that turned out to be *The Voyage of the Beagle* (Darwin, 1839), married his first cousin Emma Wedgewood and settled in a house in Kent for the rest of his life. Without having to work for a living, he secluded himself at the family state in Down House and devoted his life to his observations and experiments, his science, his publications and his large family. He published other monographs on different aspects of natural history, most notably coral reefs, barnacles and orchids, but the stature of his intellect was revealed in the publication in 1859 of a book that, for the first time, presented and documented a feasible and even compelling mechanism to explain the adaptation of the organisms to their environment. The book *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life* is one of the most influential books of Western thought (Darwin, 1859).

What is the structure of evolutionary theory? *On the Origin of Species* was not a book written for experts. It was written for everybody who reads, thinks and wonders about the natural world that surrounds us. A book only for experts does not sell the entire edition (1250 copies) on the first day. By the standards of the time the book was a best seller. Six editions were published between 1859 and 1872, the last one dropping 'On the' from the title to leave the classic abbreviated name for the book: *Origin of Species*. But the first edition expresses in its most pristine way Darwin's 'theory of descent with modifications through natural selection' in what he refers to in the last chapter as a 'long argument'. Only in the fifth edition (1869) did he add the controversial phrase 'survival of the fittest', excerpted from the British philosopher Herbert Spencer, providing an unsuspected link to the 'social Darwinism' that has been used historically to support unfair and even inhumane social and political policies. It is worth outlining the five

pillars of the Darwinian mechanism for the adaption of organisms to their environment and, in doing so, providing 'some light on the origin of species – that mystery of mysteries' (Darwin, 1859), in the language of the natural philosophers of the time. The five pillars are well documented in the book thus giving credence and making a compelling case for the overall mechanism:

(1) Within a given environment, all organisms produce more offspring than can survive to reproduce themselves.

(2) All organisms naturally vary, typically by small differences from one another.

(3) Offspring inherit characteristics from their parents and tend to be more like their parents than others.

(4) At least some of the variations in an organism lead to a greater number of its offspring surviving and reproducing relative to the offspring of others.

(5) This 'natural selection' means that these particular variations will become more prevalent in the population (as a consequence of the differential survival and laws of inheritance) as they are passed on to future generations. This translates to an adaptation of the organism to the environment.

It is important to emphasise that Darwin, being a child of Victorian and conservative England, could only conceive these processes acting but in a very slow and incremental way, by 'graduated steps' (*Natura non facit saltum*). In this 'gradualist view' he followed in the footsteps of one of his mentors, Professor C. Lyell author of the book *Principles of Geology* (1830–1833), which so strongly influenced Darwin. Thus, the notion of small, gradual, changes is an inherent part of classical Darwinism. Those five processes, operating gradually for long periods of time, are the fundamental tenets of Darwinism.

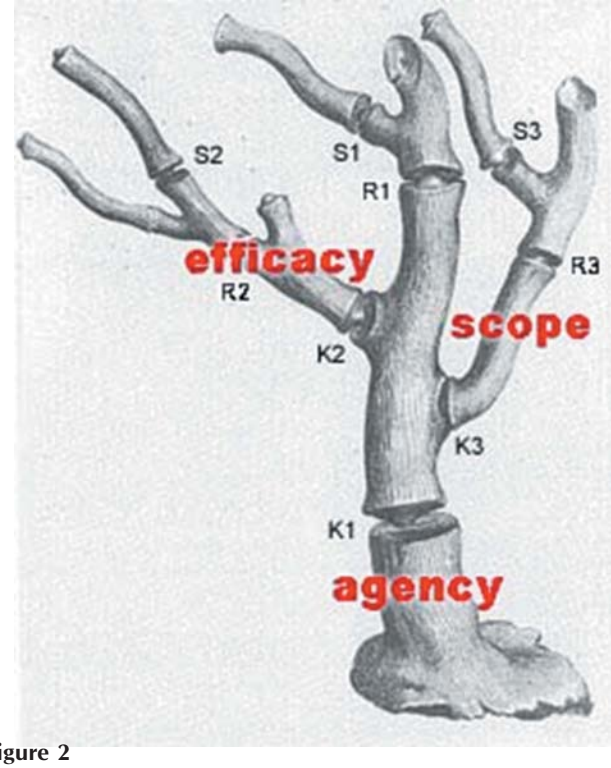

Figure 2
The fossil coral used in the background of the cover of Gould's synthesis to illustrate the current status of evolutionary theory. Image reproduced with permission from Gert Korthof (from his review at http://home.planet.nl/~gkorthof/korthof63.htm).

After a century and a half of analysis and discoveries (especially during the 20th century), followed by a maturation process, this begs the question: What is the current 'structure' of evolutionary theory? Whether you agree or not with the views expressed by the author, Gould's *magna opus* 'presents a fully articulated vision of the history and current status of evolutionary thought, written by one of the most influential biologists of the 20th century' (review by Todd Grantham, 2004). The content and the style of the book have been praised or criticised by a wide variety of authors. However, the importance of the synthesis is indicated by the presence of more than 20 published reviews, each one addressing different aspects of the work (http://www.stephenjaygould.org/reviews/).

A brief summary of the most relevant changes to the Darwinian logic presented in the *Origin of Species* will help us understand the nature of the alterations and revisions accepted or being considered today, the various and novel trends in evolutionary thought and shine some light on what macromolecular crystallography has contributed to the current edifice. There is no substitute for reading Gould's eloquent and even rhetoric prose but the reviews (written by experts in the different areas of evolutionary theory) help to put Gould's contributions and ideas in the right perspective, and leading non-experts to find a more or less straight path in the often conflicting thoughts expressed by the author. There are four that I have found useful but each one is valuable in its own way (*e.g.* Ayala, 2005).

Gould uses the metaphor of a three-branched coral fossil supported by a dominant trunk, as drawn by the Italian naturalist and artist Agostino Scilla (1629–1700), to frame the Darwinian 'logic' (not the evidence to support it) and his vision of the current status of evolutionary theory (Fig. 2). The trunk and the central branch represent the mechanism of natural selection as the central support that cannot be severed, without killing the entire theory; he refers to this trunk as the 'agency', or the 'agent' of change if you wish. The left-hand branch represents the 'efficacy' or the claim that natural selection acts as the primary creative force of novel forms. This might appear to be redundant to the central branch but it is not. Early critics of Darwin's proposal opposed his views with an idea that is still prevalent in anti-evolution circles; namely, that natural selection can weed out the ill-adapted but cannot create new forms. The right-hand branch represents the notion of the 'scope' of the theory. The micro-evolutionary processes acting *via* the two the other branches, can be extrapolated through geological time and explain the entire diversity of life. The cuts (or negation) of those concepts at the base (K-cuts) will kill the theory, while revisions (R-cuts) will maintain the basic structure but will produce a more highly branched coral. The S-cuts represent superficial changes with no fundamental changes. In a nutshell, Gould's synthesis proposes that through his work and that of many others, evolutionary theory

is undergoing substantial R-cuts and reviews in the three branches but the edifice of Darwinian logic is sound: 'The strict Darwinian form of explanation has thereby been greatly changed and enriched, but in no way defeated' (Gould, 2002). Some of the reviewers mentioned above would argue that Gould's 'long argument' is contradictory and unclear but let us use this summary as an essential core of his *magna opus*.

Of particular interest to structural biologists are the advances and revisions on the 'efficacy' branch since they are connected to the structural component of living systems and are also related to the 'scope' of the theory. The key question is: are genetic and structural changes in the molecular and genetic machinery of the living systems enough to generate new forms *de novo*, or are selective pressures needed? This is a crucial point where 'structural biology' in its broadest sense has contributed and will continue to contribute. The most recent advances in developmental genetics and structural biology have had a tremendous impact on evolutionary biology. The details are presented in chapters 10 and 11 of Gould's synthesis.

To my surprise and the best of my knowledge, books and treatises about evolutionary theory, including Gould's *Evolutionary Theory* do not discuss the sickle-cell variant of hemoglobin or the role of proteins and catalysts in life processes. But do not despair! This may change in the future; both chapters 10 and 11 mention explicitly structure and function in their title. The first in the context of the *Historical Constraints and the Evolution of Development* while chapter 11 is devoted to *The Integration of Constraint and Adaptation (Structure and Function) in Ontogeny and Phylogeny: Structural Constraints, Spandrels, and the Centrality of Exaptation in Macroevolution*. Thus, the importance of structure and function in the evolutionary process is well recognized among evolutionists; but at a different level than ours. They prefer to say 'form and function'. They talk about organs, bones, body plan (*bauplan*) in organisms, the development of the shell in snails, or fossils; whereas, we talk about atoms, amino-acid residues and domains in describing the molecular components of the cell at the atomic level. And here they (and we) have to tread delicate waters since if the physical constraints are the dominant force in the shaping of the living forms as suggested by D'Arcy Thompson in his classical book *Growth and Form* (1942), then the role of natural selection in creating new forms is probably secondary and this does imply a critical revision of evolutionary theory. Cautiously, in chapter 10, Gould expands the structure–function dualism and proposes an equilateral triangle where 'functional–historical–structural' forces act in the shaping of biological forms and discusses in detail the implications of the discoveries in the field of developmental biology. From this chapter, it is worth pointing out the tremendous importance of the recent discoveries of *Hox* genes and their associated homeobox DNA regions in directing the segmentation and segment identity in the body plan from insects to mammals, thus providing genetic explanations for the morphological blueprints suggested by D'Arcy Thompson's influential work (Gould, 2002; Coen, 1999).

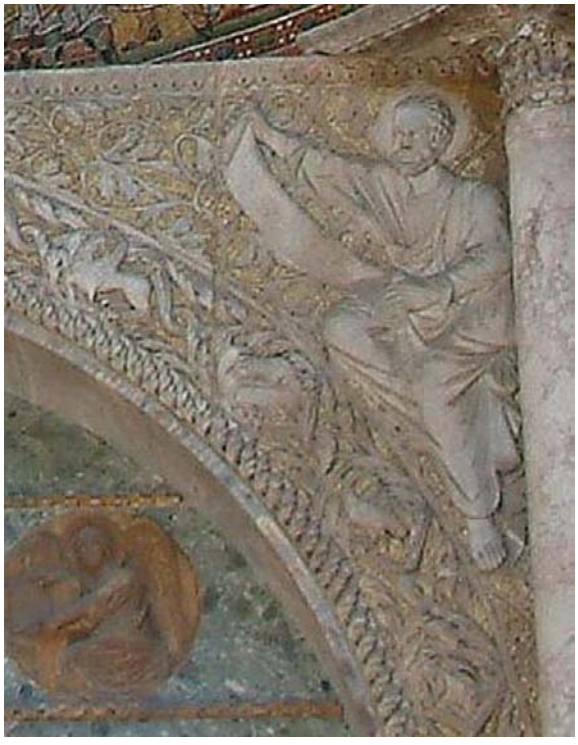

Figure 3
An example of an architectural spandrel: a two-dimensional spandrel between arches in a linear row in the Basilica di San Marco, Venice. Image by Maria Schnitzmeier and reproduced under a Creative Commons Attribution ShareAlike 3.0 License (http://creativecommons.org/licenses/by-sa/3.0/). An example of a three-dimensional spandrel can be found at http://www.bun.kyoto-u.ac.jp/~suchii/spandrel.html.

Structural biologists should also read chapter 11 of Gould's volume for two main reasons. The first reason is that we need to become better acquainted with the specific meaning of the terms 'spandrel', 'adaptation', 'exaptation' and others in evolutionary jargon. The second reason is to find ways to highlight and put our discoveries at the molecular level into the context of evolutionary theory as understood by biologists. It will be a tough read at the beginning but the intellectual insights will overcome the initial pain, especially if we are to create a durable bridge with the evolutionary biologists.

What is a spandrel? This is purely an architectural term (see Fig. 3). It refers to the additional, some might say superfluous, space that appears in an enclosed space limited by semi-circular arches as exemplified in the Cathedral of San Marco in Venice. The word and the context were carefully selected in the original article (Gould & Lewontin, 1979) to avoid controversy among biologists by choosing an architectural 'appendage' and not a biological one. Three-dimensional spandrels are the tapering triangular spaces formed by the intersection of two rounded arches at right angles that are necessary architectural byproducts of mounting a dome on rounded arches. These extra spaces were left blank for about 300 years and only later were decorated to serve 'a different function'. This concept could help us understand the significance of some of the structural features of macromolecular structures for which we desperately seek a functional significance (see below).

Adaptation as the fitness of organisms to their environment is a more common term but we should be careful in not overextending this term indiscriminately to situations where it does not apply. Not every trait or character in an organism (or physicochemical property in a macromolecular entity) can be given a definitive selective advantage. The subtle coupling between selection and adaptation among organisms can vary across a wide spectrum, as detailed by Gould & Lewontin (1979). The extension of this concept to mean that every single minute variant of a macromolecule can play a role in the survival of the organism that contains it cannot possibly be true. I fear that we do not know enough of the properties of single macromolecules to address this question yet. The physicochemical properties of the individual molecules, and certainly of the statistical ensembles of macromolecules, have certainly played a critical role at the prebiotic level but that is a completely different issue.

'Exaptation', 'co-option' and 'preadaptation' are related terms, which are now part of new concepts introduced and developed to explain better the interplay of structure and function in evolution. They refer to the change in function of a trait, an organ, a structure or physical characteristic during evolution. A classical example is the feathers of birds, which were originally developed as organs to regulate the temperature and were later co-opted and further adapted for flying. The origin of this strict biological meaning can be traced to the subtle philosophical distinction remarked by Nietzsche between 'current utility' and 'historical origin' (the connection is detailed by Gould, 2002; pp. 231–258).

Gould introduces these concepts from the historical perspective, when Darwin confronted the objections to natural selection as the creative force behind complex organs (*e.g.* the eye). The objections were published by St George Mivart in a book with a title very similar to Darwin's but with a strikingly different message: *On the Genesis of Species* (1871). In this book, Mivart raises the now familiar objection to evolution that natural selection can only have a 'secondary and subordinate' role and is, by itself alone, unable to create complex structures. The argument is based on the idea that natural selection cannot provide any selective advantage to incipient structures. This has been translated in the popular press as the value (or lack thereof) of the '5 percent of a wing principle'. These issues have been explored and exploited *ad nauseum* by the groups seeking an 'intelligent design' explanation for biological forms.

To avoid ambiguities, Gould & Vrba (1982) introduced the term exaptation (from *ex* derived from and *aptus* useful, referring to the suitability of the form) to explain 'features co-opted for a current utility following an origin for a different function (or no function at all)'. In brief, adaptations have functions derived through natural selection and exaptations have effects that might have not been intended by the original structure.

The new terminology notwithstanding, these concepts refer to things that we have seen and recognized quite often in our macromolecular studies and probably every reader can provide his/her favorite example. Gould mentions two examples from the domain of macromolecular structure that are probably the tip of iceberg. I will briefly discuss them to open the eyes of the community to many more examples that can illustrate the use of the concepts of evolutionary biology to our domain of atomic structure.

Weiner & Maizels (1999) reviewed in *Science* a series of papers related to the 'deadly double life' of the carboxy-terminal domain of the human tyrosyl-transfer RNA synthase (Tyr-tRNA-synthase). The key observation is that the carboxy terminal domain of Tyr-tRNA-synthase shows clear homology (49 percent sequence identity) with a cytokine performing a quite different function related to driving phygocytic cells to apoptotic sites. This novel role suggested that the abundance of Tyr-tRNA-synthase and its subsequent secretion would lead to the shutting down of the residual protein synthesis in the dying cell. Indeed, a deadly double job. In explaining these observations Weiner and Maizel used explicitly the term exaptation in their commentary and in doing so clarify and illustrate the use of both terms at the molecular level:

> those with an evolutionary bent sometimes use the word 'exaptation' to describe the appropriation of a molecule with one job for a completely different purpose. Exaptation contrasts with 'adaptation', a seemingly natural extension of preexisting functions,

(Weiner & Maizels, 1999).

At the end of the brief review, Weimer and Maizels refer to a second phenomenon with which we are fully familiar: the repeated appropriation of metabolic enzymes for a variety of tasks sharing (or not) a common catalytic mechanism. The example has been known for a while in structural biology but it is worth reviewing for the younger generations. Crystallins are structural proteins which constitute about 90 percent of the total soluble protein of eye lenses in most vertebrates. According to the traditional concept, the eye is an extremely specialized organ with an exquisite design, and one would have thought that crystallins represent a very limited set of highly specialized proteins designed for their unique refractive properties. The first structures were solved in the early 1990s showing very high structural similarities to various catalytic enzymes [aldehyde dehydrogenase, lactic dehydrogenase, transketolases and glutathione transferase: *e.g.* see PDB entry 1gsq (Ji *et al*, 1995), and related entries]. Structural, functional and genetic studies of the 1980s and 1990s have established that crystallins are a set of proteins 'exapted' from a variety of enzymes with originally different functions, some of which are still maintained. These findings debunk the idea that crystallins are as specialized as the eye itself and illustrate the richness and possibilities of the concept of exaptation at the molecular and structural level.

As far as I know, the evolutionary concept of spandrel defined by Gould & Lewontin (1979) has not yet entered the lexicon of molecular structural biology although there have been lots of analyses and publications on the structure and evolution of proteins (see Brändén & Tooze, 1999, for a summary). In my view, our analyses of structure–function in protein structure have put too much emphasis on functional

explanations and bypassed the subtlety of the concept introduced by Gould and Levontin. I will put forth just a few ideas for reflection, relating important aspects of structural biology.

First, the relation between the binding of the coenzymes to the corresponding apo-proteins. The concepts of apo-enzyme and holoenzyme are well known in biochemistry and structural biology. The fold of the apo-proteins (*e.g.*, lactic dehydrogenase, LDH) are well characterized as an independent entity independent of the bound NAD cofactor. It is quite conceivable (although not investigated or established) that the stable NAD-binding fold originated independently of the cofactor and that the NAD-binding pocket at the carboxy end of the Rossmann fold was an unintended result of the structural constraints imposed by the elements of secondary structure. Possibly, functional pockets in enzymes might have started as spandrels on the surface of stable protein folds.

Second, it is a familiar concept that the vast majority of active sites in enzymes or the recognition parts of the macromolecules involved in biological process are located in loops, extending from the core three-dimensional structure. For instance, antigen-binding sites in antibodies are built from the loop regions extending from the immunoglobulin fold. Other examples abound in structural biology.

Finally, stable loop structures such as Asp-box motifs are β-hairpin loops that play exclusively a structural role in β-propellers with unique loop properties. One side (residues 3, 5, 7 and 10 of the sequence: -X-X-S-X-D-X-G-X-T-W-X) is structurally conserved with a critical Asp residue making the necessary hydrogen bonds to stabilize the structure. On the other side residues 2, 6, 8 and especially 9 are only partly conserved. The Asp-box motif, or its shorter version s-Asp-box, has now been found in several protein families, and in RNase Sa (PDB entry 1rge; Sevcik *et al.*, 1996) is exposed to the surface and the residues H85-Y86 within the non-conserved side of the Asp-box bind GMP (PDB entry 1gmp; Quistgaard & Thirup, 2009). An initial 'loop-spandrel' may have been later 'decorated' for a binding interaction of biological importance.

These examples indicate to me that the doors are open for the full and detailed analysis of macromolecular structure in the explanation of the molecular mechanisms of evolution (see for example, Wagner, 2005). Most current evolutionary studies overemphasise the role of genes and replication as the dominant forces shaping evolution at the molecular level (*e.g.* 'selfish genes') but in my view underestimate the power and subtlety of the forces operating at the protein level (catalysis, regulation, chemistry and other). By being able to dissect at the atomic level the separate contributions of physicochemical constraints *versus* selection forces in the evolution of proteins and biological systems we can make our strong future mark in the field of evolutionary biology. A timely review on the structural and functional constraints in the evolution of protein families (Worth *et al.*, 2009) is a significant step in the right direction.

We should also consider what structural biology (in its wider sense) has contributed to our current understanding of evolutionary theory. The framework of the theory of evolution by natural selection was proposed by Darwin prior to the explosion in biological knowledge that took place in the 20th century heralded by the rediscovery of Mendelian genetics in 1900 by Carl Correns, Hugo De Vries and Eric von Tschermak. It has been well documented that Darwin was not aware of the mechanisms of inheritance, even though Mendel's work on the hybridization of pea plants and his novel concepts of heredity were published in 1866. Further insights into the mechanisms of heredity were to follow.

Historically, it might be argued that the three-dimensional structure of DNA is a major contribution of structural biology to evolutionary theory in that it provided the chemical and structural basis of heredity. The structure provided explanations for the structural basis of mutations and recombination, and established what was and was not possible at the DNA level. A completely new view of the mechanisms of inheritance was possible.

Darwin presented the notions of variation in the *Origin* (both under domestication in chapter 1 and under natural conditions in chapter 2) purely from the observational, macroscopic, point of view; he could not go any further. This has certainly changed. Our work and our results derived from macromolecular crystallography unveil variation at a level that Darwin could never have imagined. Just look at the catalog of structures deposited in the PDB. This represents a catalog of variation at the molecular and atomic level, a level that was inconceivable in Darwin's times. However, by delving so deep into the structures the relationship to function at the organism level is often not so clear.

Just by sheer coincidence the structures of the first proteins unveiled by macromolecular crystallographers were structurally related. This observation provided a new twist to our structural results. Georgina Ferry's biography of Max Perutz (Ferry, 2007) dramatically relates the moment when Rossmann interpreted the electron-density map of hemoglobin and related its structure to the previously determined structure of myoglobin by Kendrew and colleagues. Perutz's reaction to this finding was far from enthusiastic and Rossmann was mortified by this. Unexpectedly, this amazing fact revealed, a century after Darwin's insights, that these two proteins were structurally and functionally related, and provided the first evidence of evolutionary processes taking place also at the molecular level; the concept and field of 'molecular evolution' was born (Ferry, 2007).

This concept and a myriad of additional examples of evolutionary processes taking place at the molecular level have been a constant theme throughout the development of macromolecular crystallography, and the depositions at the PDB are a testament to it. A detailed analysis of these discoveries is needed, from viruses to the macromolecules involved in complex regulatory systems in higher organisms and passing through the critical and constant theme of the evolution of enzyme catalysis. But we have to be very cautious. Proteins themselves do not evolve in the Darwinian sense. It is the organisms within which the molecules are embedded that do. Organisms are subject to the selective pressures (natural or artificial) that Darwin's insight showed was so critical for

adaptation, and thus evolution. Natural selection acts at the level of organisms not molecules. Strictly speaking, evolution of a population of organisms, genes or molecules is characterized by the changing in frequencies of a certain variant within that population. The amazing 'inventions of nature' that we unveil in our crystals are the result of those processes of descent with modification and selection, but by themselves, proteins do not evolve inside the cells. The genetic makeup of the organisms and its frequency within a population does change with time by natural or artificial selection; different macromolecular structures are the result of that change and this is what we unveil in our crystal structures.

In my view, our major and undisputable contributions to evolutionary theory have been and continue to be: (i) showing in atomic detail the richness of matter in shape and form; how those minute spheres that we call atoms are combined through intermediate units (amino acids, nucleotides, lipids, hydrocarbons, *etc.*) to create the infinite variety of forms and machinery that make life possible; (ii) documenting the immense variation among those macromolecular components that provide the raw material for evolution at the atomic level, and the intrinsic constraints that the atomic structure imposes in what is possible; and (iii) proving unambiguously sometimes (but, let's not forget, not always) how all those macromolecular forms and variants relate to or alter the major function that they perform embedded within the matrix of the organisms.

From the standpoint of its impact on evolutionary theory the last two of our contributions will probably continue to be the most important of the three. That is to say, understanding the constraints that the atomic fabric of matter imposes on what is possible and the basic relationship between structure and function. The latter of these was revealed soon after the determination of the first protein structures and is the core concept of structural biology. We continue to understand painstakingly, day after day, the details and implications of the former; there is a long road ahead.

The relation between the structure of hemoglobin and myoglobin and their common role as oxygen transporters justified the vision of the first macromolecular crystallographers, J. D. Bernal, W. L. Bragg, D. C. Hodgkin and others, who firmly believed that determining the structure of the macromolecules involved in life processes would provide a molecular understanding of their mechanism. The seeds for understanding the implications of structural biology on understanding of human health and medicine had been planted earlier in three classic papers. First, there was the insight of Pauling, Itano, Singer & Wells (Pauling *et al.*, 1949) in showing that sickle-cell anemia was a 'molecular disease' related to changes in the electrophoretic mobility of the hemoglobin molecule of patients with the disease when compared to normal individuals. Second, Vernon Ingram (the same V. Ingram who produced the first multiple isomorphous derivatives of hemoglobin: Green *et al.*, 1954), showed later that the difference in charge was due to the replacement of a glutamic acid residue by a valine at position six in the β-chain (Ingram, 1956). Thirdly, Perutz himself followed these observations from the beginning and was able to place the culprit amino acid on the newly determined three-dimensional structure of hemoglobin. It is worth noting that these findings were before the cracking of the genetic code and the globin sequences imposed constraints on the nature of the code itself.

These initial findings rapidly established the connection not only between structure and function but also, and from our anthropocentric perspective more importantly, between health and disease. How these themes developed in the early days of macromolecular crystallography have also been beautifully narrated in Ferry's biography (Ferry, 2007). The triumphs of structural biology in explaining the structure–function nexus should not overlook the subtlety of evolutionary processes at the 'organismic' level. It is well known that a functional alteration as serious as the sickle-cell hemoglobin (so serious that from the structural point one might call it lethal) has been retained in gene populations because in heterozygous individuals it conveys some degree of protection against deadly malaria infection. The subtlety of the connection between the altered structure (no matter how drastic and detrimental it might appear at the molecular level) and its 'fitness for survival' value always has to be kept in mind. The gap between the atomic changes or alterations that we can see in our three-dimensional structures and the survival values those novel structures might have for the organisms is related to a long cascade of causality events for which we might not know all the steps.

In the future, our efforts will continue to characterize in greater detail the structural and atomic alterations ('restraints and constraints') that the atomic structure imposes on the evolutionary processes at the molecular level. Much emphasis has been placed on the evolution of genes, and the concept of the 'selfish gene' has captured the imagination of large audiences *via* the writings of R. Dawkins. I think that much more attention should be given to the 'constraints and restraints' that the physicochemical properties of matter making up the living systems play in 'presenting' or 'exposing' what is possible to the selective forces. An appropriate metaphor to counterbalance selfish genes might be the notion of 'intelligent molecules' or better still 'efficient molecules' that could capture the concept of the importance and uniqueness of the macromolecular machinery as vehicles and facilitators of the chemical processes upon which natural selection acts.

What is left to be understood? The connection between structure and function will continue to be the key factor; this is nothing new to structural biologists. However, in the strict evolutionary context it can be phrased as follows. How do the different atomic variants of the macromolecular components of life affect the function and, more importantly, the viability of the organisms of which they are a vital part? This general question needs to be answered at the level of enzymes, multienzyme complexes and eventually complex systems. But more importantly, it needs to be answered at the 'single-molecule' level.

Current studies of structure–function activities of any enzymatic system proceed by producing single or multi-amino-acid mutations of the wild-type and comparing the

macroscopic kinetic properties of the corresponding ensemble of proteins: K_m, K_{cat}, K_{cat}/K_m; against the corresponding ones for the native enzyme. As we know, quite often, the conclusions are far from clear cut. We tend to forget that the cell is not a macroscopic system (or reaction vessel) in the thermodynamic sense: systems involving numbers of molecules of the order of N_A (the Avogadro constant). The volume of a typical *E. coli* cell is approximately 1 μm^3 (or one femtolitre) and it has a mass of approximately one picogram (Phillips & Quake, 2006). Thus, the concentration of a single molecule inside the cell in that small volume is roughly 1.6 n*M* (Bustamante, 2008). This would imply that perhaps a 'new thermodynamics' (Zewail, 2008) is probably necessary to explain certain phenomena of life, rooted in the far from equilibrium conditions where the organisms operate (Abad-Zapatero, 2007; Phillips & Quake, 2008; Zewail, 2008, and authors therein).

The seeds of a new way of analyzing structure–function at the single molecule level are beginning to sprout and its practitioners are using tools that we could not have imagined when structural biology began in the early 1960s. The field is growing its first seeds with extraordinary vigor and rigor. A recent book entitled *Physical Biology, From Atoms to Medicine* (Zewail, 2008) presents the full scope of physical biology at the crossroads of the 21st century. From novel (four-dimensional) imaging techniques of tissues and organs to the single-molecule studies of enzyme catalysis, DNA-packing and molecular motors, the structure–function studies of the current century are equipped to rigorously test, at the exquisite level of a few piconewtons nanometre, the predictions of our functional hypothesis for the work of molecular machines (Zewail, 2008, and chapters therein). As a reference point it is worth mentioning that the hydrolysis of an ATP molecule liberates approximately 80 pN nm of energy. Thus, a new frontier of quantitative structural biology is dawning that has been presented as the biological frontier of physics (Phillips & Quake, 2006). One last word of caution is needed, though. In our zeal to explain *all* the functions as strictly derived from the atomic structure we need to be cautious and open minded. Especially, when we are only beginning to appreciate the subtleties of single-molecule behavior and the role that thermal forces (providing energy levels given by kT, equivalent to 4 pN nm) can play in biological processes (Phillips & Quake, 2006).

Finally, in his review of the new developments of complex systems biology in evolutionary biology, Gould is fascinated by the concept that biological systems work 'on the brink of chaos' as championed by S. Kaufman and expounded in his book *The Origins of Order* (Kauffman, 1993). Following in the tradition of D'Arcy Thompson, Kaufman reasserts the power of physical processes to generate internal structures. Given a set of conditions, he argues, there is a tendency to spontaneous order in complex systems. The effect that natural selection could have on those spontaneously formed structures is rather limited. He draws some evidence for such statements from the fields of development and complex-systems physics but his treatment of the effect that the physicochemical forces play in such a process is rather limited. I much prefer the approach presented by Andreas Wagner in his seminal book on the critical role that the physicochemical machinery plays in allowing the robustness and evolvability of living systems (Wagner, 2005). These are the two critical components of biological systems upon which our detailed structural work can provide invaluable clues. Further detailed structural work is needed to see how systems biology at the molecular level (*i.e. via* the interaction of multiple macromolecules or multiple macromolecular systems) interplays with the selective forces in shaping the organisms.

Darwin was an extremely thoughtful and cautious thinker. In spite of the importance that natural selection had in explaining the adaptation of the organisms to their environment, he did write as the last sentence in the first edition of the *Origin of Species*: 'Furthermore, I am convinced that Natural Selection has been the main *but not exclusive* means of modification'. This cautious remark was essentially ignored and the subsequent interpreters of the Darwinian ideas have proudly put the iconic image of Darwin in a Panglossian view of the world only driven by the strict forces of natural selection. This was so much so that in the last edition of the *Origin*, he introduced as the last sentence the following semi-bitter remark:

> As my conclusions have lately been much misrepresented, and it has been stated that I attribute the modification of species exclusively to natural selection, I may be permitted to remark that in the first edition of this work, and subsequently, I placed in a most conspicuous position – namely at the close of the introduction – the following words: 'I am convinced that natural selection has been the main, but not the exclusive means of modification.' This has been of no avail. Great is the power of steady misinterpretation.

I do hope that this essay has helped to understand better what the structure of the theory of evolution is and what macromolecular structure has contributed and will continue to contribute towards the understanding of the theory of evolution, as well as its implication for the understanding of living systems at the atomic and molecular level. There is still much work to be done by new generations. Gould's *magna opus* represents the conceptual synthesis and the current status of the biological, developmental and possibly genetic structure of evolutionary theory. The *Molecular Structure of Evolutionary Theory* is still to be written. The gauntlet has been cast down for the community of structural biologists at large to accept the challenge and to assess what have we contributed and, moreover, how our findings will affect the future development of the concepts laid down by Darwin 150 years ago. In the process we need to remember that 'non-adaptive' explanations of the structure–function connection do not mean 'non-intelligible' causes, keeping in mind terms like 'exaptation', 'co-option' and others for a comprehensive view of the evolutionary processes.

We can enrich and be enriched by a closer dialog with evolutionary biologists in the future but we need to understand and speak their language and the nuances of their concepts. Let this essay be a bridge for this rapprochement.

Also, let this essay be our small homage to Charles R. Darwin in the 200th anniversary of his birth and on the 150th anniversary of the publication of the *Origin*. I think that it is appropriate to close these brief reflections with the final sentence of the first, original and unaltered, edition of *The Origin of Species*,

> there is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed laws of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being evolved,

(the final sentence in the first edition of 24 November 1859).

Although the statement obviously referred to living forms, we macromolecular crystallographers should have no qualms about applying it to the precious macromolecular forms that we uncover and study daily in our molecular crystals, which, in ways that still we do not quite fully understand, provide the immensely rich raw material for the evolution of the living world of which we are an amazing part.

I wish to thank Professors Ted Baker and Michael Rossmann for their encouragement to bring this project to completion. Also appreciated are the comments, suggestions and support by Drs I. Fita, R. Recacha and F. Gago as well as from an anonymous referee of this work. Just by the contingency of life processes, I found myself completing this essay in Cambridge, UK, enjoying the walks and the scenery, following the shadow of Charles R. Darwin in Christ College, and particularly following his footsteps in the Sandwalk at Down House on this special year of his celebration. The personal hospitality of Mrs Beth Weller and her husband Alex is greatly appreciated as well as the professional support of John Overington's group at the ChEMBL group of EBI. The assistance and kindness of the staff at the library of the Sanger Institute, Genome Campus, Hinxton, UK, especially Mr Andrew King, is greatly appreciated.

References

Abad-Zapatero, C. (2007). *Acta Cryst.* D**63**, 660–664.

Ayala, F. J. (2005). *Theol. Sci.* **3**, 97–117.

Brändén, C.-I. & Tooze, J. (1999). *Introduction to Protein Structure*, 2nd ed. New York: Garland Publishing, Taylor and Francis Group.

Bustamante, C. (2008). *Physical Biology. From Atoms to Medicine*, edited by A. H. Zewail, pp. 115–135. London: Imperial College Press.

Coen, E. (1999). *The Art of Genes. How Organisms Make Themselves*. New York: Oxford University Press.

Darwin, C. R. (1839). *Journal of researches into the geology and natural history of the various countries visited by H.M.S. Beagle under the command of Cpt. FitzRoy, RN from 1832 to 1836.* London: Henry Colburn.

Darwin, C. R. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Ferry, G. (2007). *Max Perutz and the Secret of Life*, pp. 183–186. *The Origins of Molecular Evolution*. London: Chatto and Windus.

Gould, S. J. (2002). *The Structure of Evolutionary Theory*. Cambridge, MA, USA: The Belknap Press of Harvard University Press.

Gould, S. J. & Lewontin, R. C. (1979). *Proc. R. Soc. London Ser B*, **205**, 581–598.

Gould, S. J. & Vrba, S. (1982). *Paleobiology*, **8**, 4–15.

Grantham, T. (2004). *Biol. Philos.* **19**, 29–43.

Green, D. M., Ingram, V. M. & Perutz, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287–307.

Ingram, V. M. (1956). *Nature (London)*, **178**, 792–794.

Ji, X., von Rosenvinge, E. C., Johnson, W. W., Tomarev, S. I., Piatigorsky, J., Armstrong, R. N. & Gilliland, G. L. (1995). *Biochemistry*, **34**, 5317–5328.

Kauffman, S. (1993). *The Origins of Order: Self Organization and Selection in Evolution*. New York: Oxford University Press.

Lyell, C. (1830–1833). *The Principles of Geology* or *The Modern Changes of the Earth and Its Inhabitants as Illustrative of Geology*. London: Murray.

Mivart, St G. (1871). *On the Genesis of Species*. London: Macmillan.

Paley, W. (1802). *Natural Theology: or, Evidence of the Existence and Attributes of the Deity, Collected from the Appearances of Nature.* London: R. Faulder.

Pauling, L., Itano, H., Singer, S. J. & Wells, I. (1949). *Science*, **110**, 543–548.

Phillips, R. & Quake, S. R. (2006). *Physics Today*, pp. 38–43.

Quammen, D. (2008). Editor. *On the Origin of Species by Means of Natural Selection* by C. R. Darwin. Illustrated edition. New York: Sterling Publishing Co. Inc.

Quistgaard, E. M. & Thirup, S. S. (2009). *BMC Struct. Biol.* **9**, 46–64.

Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). *Acta Cryst.* D**52**, 327–344.

Thompson, D. W. (1942). *On Growth and Form*, 2nd ed. Cambridge University Press.

Wagner, A. (2005). *Robustness and Evolvability in Living Systems*. Princeton University Press.

Weiner, A. N. & Maizels, N. (1999). *Science*, **284**, 63–64.

Worth, C. L., Gong, S. & Blundell, T. L. (2009). *Nature Rev. Mol. Cell Biol.* **10**, 709–720.

Zewail, A. H. (2008). Editor. *Physical Biology. From Atoms to Medicine*. London: Imperial College Press.