The origami of life

Timothy R Lezon[1], **Jayanth R Banavar**[1] **and Amos Maritan**[2]

[1] Department of Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802, USA
[2] Dipartimento di Fisica 'G Galilei' and INFN, Università di Padova, Via Marzolo 8, 35131 Padova, Italy

E-mail: lezon@phys.psu.edu, jayanth@phys.psu.edu and maritan@pd.infn.it

Abstract
All living organisms rely upon networks of molecular interactions to carry out their vital processes. In order for a molecular system to display the properties of life, its constituent molecules must themselves be endowed with several features: stability, specificity, self-organization, functionality, sensitivity, robustness, diversity and adaptability. We argue that these are the emergent properties of a unique phase of matter, and we demonstrate that proteins, the functional molecules of terrestrial life, are perfectly suited to this phase. We explore, through an understanding of this phase of matter, the physical principles that govern the operation of living matter. Our work has implications for the design of functionally useful nanoscale devices and the ultimate development of physically based artificial life.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Cooperation is an essential part of life at all levels. At the top level, the different species within an ecosystem cooperate through the mechanism of their food chain to ensure the optimal use of the available resources. The flow of resources from species to species guards against extinctions or sudden changes in population of a single species, and benefits all. There are countless examples of cooperation between individuals within a species. Birds and fish travel in large groups and move in coordinated patterns to evade predators. Likewise, many predatory mammals hunt in packs in order to better surround and capture prey. Insects form structured colonies in which the job of each individual is both highly specific and critical to the survival of the colony as a whole. Even individual prokaryotes in a bacterial plaque are capable of sensing the presence of other members of their species and redistributing their vital resources accordingly, allowing the plaque as a whole to remain alive [1].

Within individual organisms, cooperation is also vital. The tissues of a multicellular organism, for example, must operate in concert such that none drains the organism of its energy or poisons it with waste. Cells within a tissue must cooperate with each other, not only in the timing of their operation but also in their spatial orientation. Nerve cells in the brain must be able to sense one another and to break and re-form synapses as required for the survival of an animal. Furthermore, they must cooperate through general means; that is, two neurons must not have a specific mode of communication between them that is not shared by all the neurons, but each neuron should be able to communicate without prejudice with any other neuron in a common language. It is this ability of nerve cells to spontaneously organize [2] and universally communicate that makes the brain the tremendously powerful machine that it is. Indeed, the network of neurons is far greater than the sum of its parts.

This theme of cooperation and organization between constituent parts continues to the subcellular level. The machinery of the cell operates in a coordinated way to process resources, expel waste, perform the cellular functions and react to the environment. A particularly strong example of coordination within the cell is mitosis, when the very cell itself is ripped apart. Were it not for each constituent organelle playing a specific role in this carefully orchestrated manoeuvre, the cell would surely die.

The lowest level at which we can observe similar life-like cooperation is that of the molecule. The atoms that compose living organisms are, of course, identical to those that constitute inanimate matter. The molecules that are found in living material are, however, quite different from the vast majority of those found in naturally occurring inanimate matter. The difference, most easily stated, is that the molecules of life have the intrinsic ability to operate together to perpetuate the same 3.5 billion-year-old chemical reaction from which they were formed.

Today 'supramolecular chemistry' [3, 4], the study of systems of molecules that interact cooperatively, is a field of intense research. The generally large and somewhat complex molecules that make up these systems interact with each other through specific non-covalent bonding. Recent advances in technology have permitted the design and synthesis of several species of molecules that spontaneously assemble into predetermined structures [5]. Such systems are of interest to materials scientists because they exhibit a variety of unique electrical [6] and optical [7] properties. Furthermore, there is hope that supramolecular chemistry might one day find use in the fabrication of nanoscale devices [8, 9] and in biological applications [10]; self-assembling materials are much more convenient to work with than traditional wet chemistry or photolithography. An infinitely more interesting application at the fundamental level is using supramolecular chemistry to create a system of cooperatively interacting functional molecules, the first step in the development of physically based artificial life.

According to Lehn [3], supramolecular chemical systems share three common features: molecular recognition, self-organization and an emergent phase. Recognition between molecules provides the basis for supramolecular chemistry; were it not for the highly selective nature of molecules in such systems, their associated chemistry would be of the common variety observed in much simpler molecules. Instead, the molecules in a supramolecular chemical system are tuned to interact only with a specific subset of other molecules, and only in very limited ways. Self-organization results from a collection of molecules that interact specifically with each other, resulting in a cascade of preferential attachments that leads to an organized structure. This emergent phase is the true benefit of supramolecular chemistry: from a collection of molecules spontaneously emerges an ordered macromolecular structure.

The universe is filled with examples of systems of particles that assemble into ordered states with bulk properties that differ vastly from the properties of the individual components.

A crystal serves as an example of such a self-organized system. The critical difference between a crystalline solid and a supramolecular chemical system lies in the interaction of each with its environment. An example of a crystal responding to an environmental cue is the electrical discharge produced when a piezoelectric crystal is compressed. This phenomenon is absent from the individual atoms and arises only upon their assembly into a crystal. A supramolecular system might exhibit similar reactive properties that are present only in the aggregate system, but its constituent molecules may also retain some of their individual reactive character even when assembled into a macroscopic structure. Thus, these systems can be highly sensitive to their environments, as a small external stimulus that affects one component of the system may cause drastic changes in the organization of the entire system. It should not be surprising, then, that the threshold between living and non-living matter exists within the confines of this class of complex chemical systems. It is, in fact, within these very systems that we should look for the basic science of life.

In this paper we explore the physical principles that underlie life at the molecular level. We begin by listing the essential features that a chemical system must exhibit in order to be considered alive, and we construct a set of desired criteria for the constituent molecules of a living chemical system. Having identified the properties of the functional molecules of life, we explore the method that Nature has already implemented in the synthesis of molecular life on Earth. We consider various ways in which matter is commonly organized into distinct phases, and we present a novel phase of matter that meets the requirements for functionality in a living molecular system. Finally, we demonstrate that the proteins are well-suited for this phase, and we discuss some of the implications that the phase has on our understanding of nanobiology.

2. Desirable properties of the living molecular system

What properties must a molecular system have in order to be considered alive? Several answers to this question have been postulated [11–13], and although there is still no universally accepted set of criteria that establishes whether or not a system is alive, some general conditions are commonly agreed upon. Quite obviously the system must be *stable*. Certainly it must meet the above-stated requirements of self-assembly and emergent phase: self-assembly is necessary in order for biological processes to proceed in a natural and spontaneous fashion, and the emergent phase is the living phase itself. Moreover, the system as a whole must be *functional*. Living organisms are active; they respond to and interact with their environments. Whether they are transmitting signals across synapses in the brain of a human who is performing a complicated task or suppressing motion of flagella in an *Escherichia coli* bacterium swimming toward a nutrient source, the systems of chemicals that define living things are in a constant state of activity. Finally, a living chemical system must be able to *adapt* on a range of timescales: it is not sufficient for life to exist only within a highly specialized environment, but it should be able to adjust to natural fluctuations of temperature and atmosphere.

The essence of a living chemical system is its constituent molecules. For a system to display supramolecular properties, we require that its constituent molecules are able to recognize each other chemically and are therefore highly specific. Because they are functional, they must be specific not only with respect to one another, but also with respect to the environment in which they function. As the functions of the molecules are limited only by their interactions with the environment and with each other, it is further demanded that they are able to interact diversely with one another and with their environments. These constraints, high specificity and diverse functionality, greatly limit the classes of molecules that are capable of playing this role. At the molecular level, functional specificity implies

structural specificity: if a molecule is chemically specific, then its atoms are arranged in a structured fashion that permits interactions only when precise requirements are met. It may therefore be assumed that the molecules of life will contain at least pockets of rigidity where the atoms assemble in a highly structured configuration. In addition, these molecules need to be somewhat complex in order to permit a sufficiently high degree of specificity. Structural diversity in molecules is limited by the lengths and angles of the bonds permitted by their atoms. If a collection of atoms in a molecule are to be oriented precisely with respect to each other, then the molecule must be large enough to overcome the structural constraints imposed by the discrete nature of the atomic bonds.

Functionality itself does not demand that the constituent molecules possess a high degree of order; it only requires that the molecules are active. Chemical activity in turn requires that some property of the molecule changes depending on the presence or absence of an external stimulus. The molecules of a living system must therefore be poised between two states, manifested as distinct physical forms. One phase might guarantee stability, and the other sensitivity. In terms of supramolecular chemistry, the interactions between the molecules should change depending on the functional states of the individual molecules. As these intermolecular interactions rely on the structural properties of the molecules, it follows that the activity of the molecules involves a change of structure. The molecules ought to be stable and highly specific, but once again this is achieved through large molecular size. It is possible that some areas of the molecules are structurally rigid while other areas are flexible. This kind of construction is not only reasonable, but it is ubiquitous in functional apparatus at all scales. The vertebrate skeleton, for example, is composed of rigid structures that come together at joints that are free to move in predetermined ways. Our physical movements are the result of a sophisticated combination of alternately rigid and flexible components. In a similar fashion, the constituent molecules of a living chemical system might be constructions of rigid subunits separated by joints. Diversity of function then results from a diversity of physical form, as the motions of the molecules are restricted to those allowed by their structures. Furthermore, chemical specificity enhances functional diversity. In much the same way that a rotating shaft can either drill a hole or drive a screw depending on its bit, a molecular machine with a given motion can perform a variety of tasks, depending on the specific chemical nature of its binding sites.

The requirement of system adaptability might be met on the molecular level through two paths. The first is that the network of interactions between the molecules in the system can readily change its topology in order to fit the environment. In this case, the individual molecules do not have to change in order for the system to adapt, but they must be sufficiently complex to allow multiple interactions with other molecules in the system and also to be able to determine which interactions are appropriate given the present conditions. The second way that molecules can enable a chemical system to adapt is if the chemical properties of the molecules themselves are subject to modification. Here, the individual molecules need not be complex, but they must be constructed in such a way that only small changes in their atomic configurations produce a wide range of functional variations. The primary functional difference between these two paths to chemical adaptability is the timescale to which each applies. The former method, where the molecules remain constant but their interactions change, is most effective on short timescales where a quick adjustment to an immediate and short-lived stimulus is required (acclimatization). The latter path, that of changing the molecules themselves, is a much more permanent change and is appropriate for adjustments over very long timescales (evolution). To readily permit such long-term adaptability, we desire that the molecules be constructed of only a small number of pre-fashioned building blocks. This modular form carries with it not only the advantage of easy assembly, reassembly and modification, but also that of minimization of resources. Ideally only a single class

of molecules will be required, with all of the molecules in the class constructed from the same building blocks. This way, even when the molecular interaction network undergoes a short-term topological change, the resources that were used in molecules that were active in one topology can be redirected to the molecules emphasized in the alternate topology. The molecular building blocks themselves should be chemically diverse, yet they should assemble unambiguously into the larger molecules of our living system. A living system should be capable of assembling its constituent parts from raw materials, and as the complexity of the constituents increases, so must the complexity of the associated assembly process, as well as the resources required for assembly. In fact, by requiring that all constituent molecules are formed from the same set of building blocks, we minimize the resources required for assembly. Furthermore, if there is only a single way that one building block can attach to another, then the assembly of all molecules in the system is limited to a single process that minimizes assembly error. The only candidate that meets these requirements—finite building blocks with a single method of assembly—is a linear chain.

Interestingly, the linear chain carries the additional advantage that it provides a means for encoding structural information. Associated with a chain is the sequence of its building blocks, such that each distinct arrangement of the same set of building blocks describes a unique chain. This genotypic property of chains allows a system to efficiently manage its resources by fashioning a great variety of molecules from the same raw materials. However, the immense size of the sequence space for a chain of moderate length could actually work to the disadvantage of a species if each sequence were to adopt a distinct structure. Certainly this arrangement provides structural diversity, but if each genotype corresponds to a unique phenotype, then only a vanishing fraction of genetic mutants will be physically compatible with the remainder of the system. If this were the case, the constituent molecules could not evolve individually, but would instead have to co-evolve in a coherent fashion; each time a molecule were mutated, all of the molecules with which it interacts would also have to mutate in a carefully orchestrated manner in order to retain their interactions. Because of the great improbability of such co-evolution, an abundance of structural variety restricts genetic diversity and evolution. If, on the other hand, most genetic mutations have no effect on phenotype, then they will be accepted by the system as neither a boon nor a hindrance. This process of neutral evolution [14, 15] greatly promotes genetic diversity, but it consequently restricts structural variety. Evolution then imposes a final constraint on the molecules of life, which is that they exist in only a limited number of phenotypes.

In summary, we have constructed a list of the desired properties that the molecules of life ought to have. These need to be stable molecules that are probably large and display functional diversity. Each molecule must be chemically specific. They should be easily constructed from a small number of building blocks using only a minimal blueprint, and are therefore likely to be chain molecules. Although they display structural variety, the molecules should only have a limited number of phenotypes—not so many that genetic mutations are overwhelmingly disadvantageous, nor so few that evolution is prohibited. They also ought to be quite sensitive to their environment, suggesting that they reside in the vicinity of a phase transition. In order to assess the validity of these hypotheses, however, it is useful to compare this picture with terrestrial life.

3. Terrestrial life

3.1. Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA)

Almost the entire surface of the planet is covered with living matter. Even in places that are enormously inhospitable to human life—near volcanic vents on the ocean floor, for example—

some organisms have found a way to flourish. Despite the enormous dissimilarities in their outward appearance and habitat, all of the organisms on Earth are functionally identical on the molecular level. Every organism has as its vital centre a network of DNA, RNA and proteins: DNA encodes the genetic information necessary for metabolism and reproduction, proteins perform the work in the cell and RNA acts primarily as an intermediary between the information-laden DNA and the physically functioning proteins. Although it is not impossible that some alternative molecular scheme for life may have evolved over the course of the planet's history, the triad of DNA, RNA and proteins has prevailed with authority. What is it about this scheme that makes it ideally suited for life?

To a molecular biologist, the difference between a tree and an orang-utan swinging amongst its branches is reduced to the difference between the sequences of bases contained in the organisms' DNA. This holds not just for plants and apes: the genetic difference between any two organisms on the planet can be quantitatively expressed in terms of DNA. The staggering biodiversity on Earth is a reflection of the molecular diversity of DNA, and it is mutations of DNA that are responsible for phylogeny. DNA is not usually considered to be a functional molecule, as it does not actively alter the chemistry within an organism. Instead, it is a repository for genetic data, carrying the necessary instructions to construct and activate all of the functional molecules of an organism. Because all genetic information is contained within it, DNA is the single most important molecule to any species. Mutations in DNA are capable of altering the functional molecules within a species, and if the mutations produce a selective advantage, they are capable of generating an entirely new species. All of evolution is then an inherently molecular process, and because it is at the heart of evolution, DNA must have a design that permits both molecular replication and the accurate encoding of information, yet resists decay.

DNA is a linear chain of nucleotides, each made from a sugar and one of four heterocyclic amine bases: adenine, cytosine, guanine and thymine. It has a backbone composed of sugars interspersed with phosphates, and genetic information is encoded as the sequence of bases along the DNA chain. This encoding scheme by itself already reveals a glimpse of the efficiency of DNA's construction. There exists no simpler method for encoding information than a linear sequence, and there is no simpler assembly of physical building blocks than a linear chain. Here molecular structure and encoded information are intimately tied together. Genetic information is encoded directly as the structure, and the structure at this level is based on fundamental laws of physics. Whereas a human may choose to store information in an endless variety of forms—on a scrap of paper, a CD-ROM, a magnetic tape, a photographic plate—molecular information is subject to the laws of molecules. The simple form of a linear chain provides a wonderful way to store information in a molecule. Another advantage of the chain is the ease with which it allows retrieval of encoded information. In order for a gene— an arbitrary piece of information—to be retrieved from whichever medium it is encoded on, the gene must be physically accessible and have well-defined boundaries. The constraint of physical accessibility is most easily dealt with by keeping the number of dimensions of the encoding medium as small as possible. The medium for information is embedded in three-dimensional space, and will therefore be folded upon itself for efficient storage. If the medium is linear, then its full structure will be something like a loose ball of yarn; strands of arbitrary length can be pulled to the outside of the ball without greatly disturbing the rest of the structure. If the medium is planar, then the three-dimensional compact structure will be like a crumpled ball of paper; the entire sheet must be unwrapped in order to read a paragraph of its text. Furthermore, if a gene is encoded in only a single dimension, then only two points are required to define its boundaries. This is exactly how genes are defined in DNA, with a 'start' codon signalling the gene's beginning, a 'stop' codon signalling its terminus and

everything in-between implicitly understood to be the gene. Information encoded in higher-order manifolds requires more points to specify its boundaries. Although the linear encoding of genetic information may seem like the obvious choice *a posteriori*, the benefits of this simple form should not be overlooked.

Aside from retaining all of the genetic information for an organism, DNA must be capable of indefinite self-replication. Life requires evolution, which in turn implies reproduction. The molecular structure and chemical properties of DNA must therefore permit replication. In order to copy anything it is necessary first to recognize the parts that are being duplicated, and then to accurately reproduce them in their original context. If the details of the object being copied are inaccessible—for example, if one attempts to copy a watch without knowledge of its internal workings—the duplication will not be successful. Thus the linear nature of DNA aids in its replication, because the entire molecule can conceivably be copied from one end to the other, much as an ancient scribe might copy a text by hand. At the molecular level, however, recognition is achieved through molecular specificity: a molecule can only be identified if it matches some complementary molecule. The situation is akin to trying to open a door for which the lock is unknown: one might try each key on a ring in turn until one fits the lock. When the matching key is found, it unambiguously identifies the lock. In the case of DNA, recognition is achieved through the bases. Each base has a complement with which it may form hydrogen bonds: adenine and thymine are complements, as are cytosine and guanine. This base pairing serves as another example of the clever employment of physical laws in the genetic material. Nucleotide bases readily form hydrogen bonds and self-assemble into supramolecular structures [8], and they are therefore an ideal medium for encoding molecular information. The pairing of the naturally occurring bases allows an economic use of resources during genetic transcription and replication. The information carried by DNA is encoded with a four-letter alphabet, and any four arbitrary bases would suffice to code genetic information. To permit recognition, though, each base must have a complement. Rather than requiring eight bases—four to encode the information and four to act as complements that recognize the encoded sequence—Nature uses a closed set of four bases that are capable of both encoding and recognition. The genetic code is not a sequence of four locks, each with its own matching key, but of two locks and their matching keys. This arrangement additionally helps to stabilize the molecule and provides a mechanism for replication.

In the native conformation of DNA, each base is paired with its complement such that the full structure of the DNA molecule is two anti-parallel sugar–phosphate chains joined by hydrogen-bonded base pairs in the centre. The base-pairing stabilizes the molecule, and the double-strand conformation permits semi-conservative replication. During mitosis, the two strands separate, and through base pairing each forms a new double-stranded structure identical to the original. Each daughter DNA then contains one of the chains from the original molecule. The beauty of this arrangement is that once the two chains are separated, the self-organization of the bases into complementary pairs drives replication. Again, were it not the case that the bases form complementary pairs, DNA replication would require a more complicated process. The double-helix DNA conformation is only allowed because of the precise geometry of the four bases. The distance between the sugars in an adenine–thymine pair is the same as for a cytosine–guanine pair, so two hydrogen-bonded strands of DNA are in fact parallel. Were the base pairs to force different lengths between their sugars, the paired DNA strands would zig-zag with respect to each other, and the structure would be less stable. Furthermore, the double-stranded conformation is crucial for minimizing replication errors because it permits semi-conservative replication. If the double-stranded conformation were forbidden, the simplest replication scheme would be fully conservative: the base sequence would be matched by its complementary sequence, and then the complementary sequence

matched by its complement, which is identical to the original sequence. This two-step process increases both the time required for replication and the probability of replication errors, and it physically demands more resources than a semi-conservative process. Thus we can see that while information might be encoded into a molecule in any conceivable number of ways, the application of the four bases found in DNA takes advantage of physical laws to permit extraordinary efficiency in the storage and replication of genetic information.

DNA not only encodes information and permits replication, but it allows mistakes in replication to be copied as well. Random variations must be introduced into and retained in the genetic material in order for species to evolve, and DNA is a stable molecule that reliably retains and reproduces the replication errors that it accumulates over generations. Its chemical stability prevents DNA from changing its information content spontaneously, but the replication process is prone to errors. The lock-and-key mechanism of base pairing is not perfect; hydrogen bonds can occasionally form between non-complementary bases, modifying the base sequence of one of the DNA strands. This modification is then passed to one daughter at mitosis, and is retained in future offspring. Evolution then results from the ability of DNA to indefinitely reproduce its replication errors. A similar property of DNA, recombination, has been exploited by Nature as a means of accelerating the evolutionary process. Species that reproduce sexually bestow upon their progeny a mixture of genes from two parents, such that each offspring has a unique genotype. Many species that reproduce asexually tend to transfer genes laterally between organisms. The advantage of sexual reproduction and lateral gene transfer is that they generate genetic diversity without requiring the lengthy timescale that is associated with mutation, so that populations will consistently contain enough genetic diversity to adapt easily to environmental changes. In order for a diploid parent to pass a sampling of its genes to each of its offspring, it selects one allele of each of its genes and combines them such that each gamete contains a random combination of alleles. This process of genetic recombination is due entirely to the double-stranded nature of DNA. During meiosis, one chain is broken on each of two homologous stretches of DNA. Each broken chain then separates from its complement and reattaches to its homologous complement on the other chain. The complementary chains are then broken and attached to each other, resulting in two daughter DNA molecules, each of which contains parts of both of the original chains. What is critical for recombination is that homologous stretches of DNA can attach to each other's complements. This process would be considerably more involved in a single-stranded molecule, and is incredibly complicated in a nonlinear medium. Complementary linear chains for encoding genetic information, then, are naturally suited for the promotion of diversity.

The choice of bases is also important for the translation of genetic information into functional molecules. A sequence of bases in DNA directly maps onto a sequence of amino acids in a protein, and RNA is the conduit through which DNA directs protein synthesis. RNA is a chain molecule that is quite similar to DNA in that it is composed of a sequence of four bases linked along a sugar–phosphate chain. Despite the similarity, its role in cellular metabolism differs vastly from that of DNA. The main roles of RNA are in transcription and translation of genes. Messenger RNA (mRNA) copies from DNA the sequence of bases that encodes a gene and transports it to an organelle called the ribosome, which reads the code and synthesizes the corresponding protein. The physical mechanism underlying the transcription of a gene into mRNA is the same base pairing that is seen in DNA, with the sole exception that all thymine bases in DNA are replaced by uracil in RNA. Using a stretch of the base sequence from DNA as a template, mRNA constructs the complementary sequence, which is then free to move about the cell. It is important to note that the same lock-and-key base pairing that permits semi-conservative replication also provides the mechanism for gene transcription,
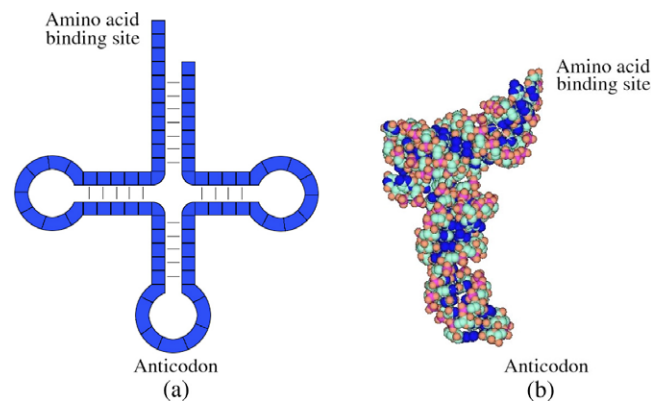
Figure 1. Structure of tRNA. (a) Diagrammatic representation of secondary structure in a typical tRNA molecule. The amino acid binding site is formed by a few bases at one end of the chain, and the anticodon is three bases on a loop. The structure is stabilized by base pairing, and each of the four branches twists into a double helix. The tertiary structure (b) takes the form of the letter 'L' or the Greek 'Γ', with the binding site and anticodon maximally separated.

once again reinforcing the physical justification for the choice of bases in the nucleic acids. The mRNA carries the genetic information to the ribosome, which itself is a complex structure formed of proteins and RNA. The ribosome translates the base sequence from mRNA to an amino acid sequence with the aid of transfer RNA (tRNA) molecules that recognize the base sequence in mRNA and match it with the appropriate amino acids. It is at the ribosome, then, that for the first time we encounter a mechanism other than base pairing that is required in order to transfer information from one molecule to another, and it is tRNA that translates information from the base sequence to the amino acid sequence.

The structure of tRNA, shown in figure 1, deserves brief discussion. Here is a truly bilingual molecule that owes its linguistic abilities to its chemistry alone. A typical tRNA molecule is a single strand of about 80 nucleotides, and the nucleotide sequence allows base pairing in specific regions of the chain. Motivated by base pairing, the tRNA molecule folds into a structure that resembles the capital letter 'L', where the stem and base of the 'L' are formed by double helices similar to those observed in DNA. At the two free ends of the structure (the top and lower right serifs of the 'L') are the anticodon for the base sequence and the binding site for the corresponding amino acid. tRNA plays a passive role in protein synthesis: it acts as an adapter that fits one type of socket at one end and its foreign counterpart at the other end, while the ribosome performs the mechanical synthesis of proteins. The two key properties that make tRNA indispensable in its role in the cell are that it folds into a well-defined shape and it binds specifically to molecules other than nucleic acids.

The accurate folding of tRNA is vital for two reasons. The first is that it specifically positions the amino acid relative to the anticodon. Because of the rigid 'L' shape of tRNA, the ribosome can easily locate the amino acid once tRNA binds to the mRNA codon. Its well-defined structure makes tRNA a useful tool on an assembly line. The second way that folding is vital to the performance of tRNA is that it permits recognition of the codon and the amino acid. The genetic code maps each triplet of nucleotide bases to a single amino acid, and in order for tRNA to translate the base sequence of mRNA to an amino acid sequence, it must be able to recognize triplets of bases. Because molecular recognition requires a lock-and-key mechanism, each tRNA has a triplet anticodon that is the complement of a triplet codon in mRNA. In order for tRNA to effectively recognize a specific codon on the mRNA

chain, its anticodon—and only its anticodon—must be accessible to the mRNA. Physically this is achieved through folding. The anticodon resides sequentially between two stretches of complementary bases that pair to form the stem of the 'L', and when tRNA is folded, the anticodon exists in the tight turn at the top of the 'L'. Thus, while most of the bases in tRNA are inaccessible because of base pairing, the exposed anticodon is physically located in a region where it is easily readable. Folding similarly places the amino acid binding site at the end of the base of the 'L'. This binding site consists of several bases at one end of the tRNA chain, and when the molecule is properly folded, these bases remain unpaired and are free to attach specifically to an amino acid.

RNA is thought to be more primitive than both DNA and proteins [16]. Its ability to both carry genetic information and act as a catalyst has prompted the theory of an 'RNA world' that existed before DNA and proteins entered the picture. It is also frequently cited as the most likely candidate for the constituent molecules of a synthetic living system [17]. Why, then, did the RNA world disappear? DNA is known to be chemically more stable than RNA, so its position as the carrier of genetic information is justified, but RNA is still effective as the primary genetic material in many viruses. The burden that accompanies the additional complexity resulting from an organism utilizing both DNA and RNA must be offset by some evolutionary advantage, and this advantage may come from the selection of proteins as the functional molecules of life. The catalytic properties of RNA are modest at best when compared with the staggering diversity displayed in the molecular capabilities of the proteins. Indeed, the chemical brawn of the proteins is thought to be responsible for pushing RNA to its present position of intermediary [18, 19]; catalytic RNAs were stabilized by bonding with amino acids, and eventually RNA enzymes lost their usefulness.

3.2. Proteins

The Nobel laureate Arthur Kornberg [20] wrote, 'What chemical feature most clearly enables the living cell and organism to function, grow and reproduce? Not the carbohydrate stored as starch in plants or glycogen in animals, nor the depots of fat. It is not the structural proteins that form muscle, elastic tissue, and the skeletal fabric. Nor is it DNA, the genetic material. Despite its glamour, DNA is simply the construction manual that directs the assembly of the cell's proteins. The DNA is itself lifeless, its language cold and austere. What gives the cell its life and personality are enzymes. They govern all body processes; malfunction of even one enzyme can be fatal. Nothing in nature is so tangible and vital to our lives as proteins, and yet so poorly understood and appreciated by all but a few scientists'. The truly functional molecules of life are the proteins, which are linear chains of amino acids. There are 20 amino acids found in natural proteins, and each differs from the other only in its side chain. Thus the primary structure of proteins—a sequence of amino acids—is quite simple, but that is where the simplicity ends. These chains of amino acids perform nearly all of the chemical functions of life. They are responsible for muscle contraction, oxygen transport, regulation of sugars and other chemicals, genetic transcription, digestion, genetic suppression, structural support, signal transmission, regulation across membranes, waste management and a host of other activities that are vital to the health of an organism. So powerful are the catalytic properties of proteins that they can increase reaction rates by several billion-fold. Indeed, without proteins, metabolism as we know it would slow to a near halt; we measure events in the convenient timescale of seconds, and it is the efficiency of proteins that allows life to proceed at such a fast pace. Proteins do not act alone, however; each protein performs the same insignificant monotonous tasks repeatedly throughout the course of its existence, and it is only through a complex network of interactions that the tiny effects of the individual proteins accumulate to produce the living state that we observe.
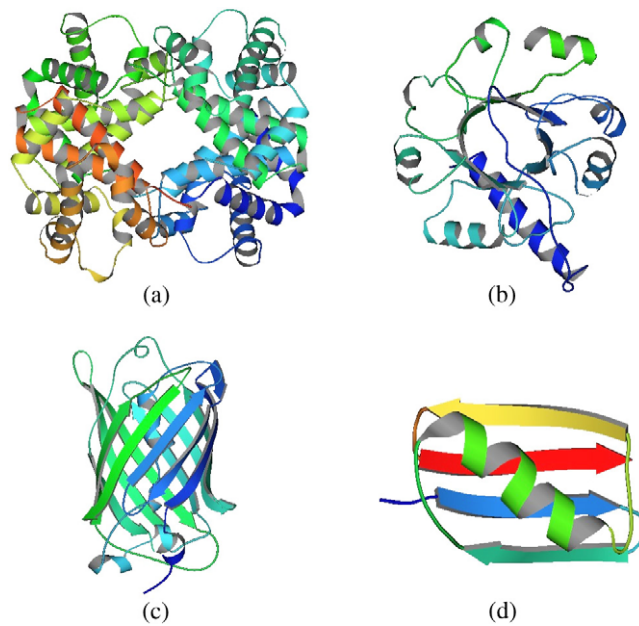
Figure 2. Ribbon views of several proteins. (a) Haemoglobin, which is widely used for oxygen transport in animals, contains four chains folded into α helical domains. (b) The TIM barrel of β strands surrounded by helices is a common structural motif. Proteins with this structure commonly metabolize sugars, peptides and nucleic acids. The structure shown here is one domain of an isomerase that is used in the biosynthesis of the amino acid tryptophan in *E. coli*. (c) The structure of green fluorescent protein is a single β sheet that is bent to form a cylinder. This protein, found in jellyfish, fluoresces green when exposed to ultraviolet light. Its high visibility gives it great experimental value as a genetic marker. (d) The single-domain immunoglobulin binding protein G. Its structure consists of a single α helix situated across a β sheet made up of four strands. This protein, found on the cell walls of pathogenic streptococcal bacteria, is thought to prevent immune response from the host.

In order to accommodate such a broad functional spectrum, the proteins fold into well-defined structures when they are in their natural physiochemical environments (figure 2). The folding of the chain into a compact structure has two major physical effects on a protein. First, it allows distant parts of the chain to come physically close together, creating local sites with chemical properties that are not present in the unfolded chain. The folding of the protein chain, in conjunction with the disparate chemical properties of the amino acid side chains, therefore permits an enormous variety of binding sites to populate the surfaces of the proteins. This fulfils the requirement of chemical specificity that we had previously placed on the functional molecules of life. Furthermore, the folding of the protein chain permits localized motions of different regions of the protein. A chain that is folded into a compact form still retains its chain nature, and therefore structural fluctuations of the folded protein are locally correlated. Although the amino acids are covalently bonded along the chain backbone, folding is governed by non-covalent forces, permitting some flexibility to the folded structure. The nature of the folded state then also allows sensitivity and functional diversity.

Folding is driven by the tendency of some amino acids to avoid contact with their aqueous solvent. Hydrophobic amino acids tend to congregate in the core of the protein, and the amino acids that are more amiable to water form a shell on the protein surface. The constraint of the

chain generally forbids every hydrophobic amino acid residue from reaching the protein's interior or every polar amino acid from attaining full exposure to the solvent, so the folded state does not necessarily represent an obvious. It is, however, a conformation that is reached with great alacrity and regularity by the chain of amino acids [21]. Proteins tend to fold on the scale of milliseconds, which is incredibly fast from an entropic point of view. If we consider as a rough example a typical protein of 200 residues, each of which can adopt five possible conformations, then all of the conformations available to a protein number 5^{200}–10^{139}. Even after discarding those conformations that are prohibited by excluded volume effects, the conformation space for a protein is unimaginably large, and the odds of randomly stumbling upon one particular fold are naught. It is then accepted that protein folding takes place pictorially along a funnel-like free energy landscape [22–26]. Non-native state conformations reside on the funnel's lip, and the native state is in its centre. As the protein folds, its free energy decreases and it is urged toward the funnel's centre. This picture rationalizes the fast folding rate of proteins and is supported by experiments in which proteins that are repeatedly chemically denatured and then reintroduced to their native environments unfold and then refold into their native state conformations [21]. It is a physiological necessity that proteins fold rapidly and reproducibly. If the folding rate were much longer than the rate of protein synthesis, then the cell would be rife with unfolded proteins, each interfering with the attempts of the others to fold correctly. The funnel-like energy landscape also demands that the native state conformation is a stable free energy minimum, preventing the protein from unfolding once it is folded. Thus the quick and definite folding of proteins allows the DNA to control all cellular functions: when a gene is expressed by the DNA, it can quickly be transcribed by RNA and translated into a protein, which immediately goes into action and remains functional until the DNA signals for its digestion. Functional molecules that fold slowly or unreliably fail to provide the sensitive response required for life.

The structures of folded proteins display a remarkable amount of regularity. All protein structures are assemblies of helices and sheets, both of which were first predicted by Pauling and Corey [27, 28] based on the geometry of hydrogen bonds between amino acids. The α helix, which holds nearly one-third of the amino acids in all known proteins [29], is characterized by 3.6 residues per turn. The β sheet that accounts for approximately one-fifth of all amino acids in known proteins is constructed of several extended strands laid parallel to each other in an almost planar fashion. Both the helix and the strand typically have normally distributed characteristic lengths of about 20 Å. Secondary structures in proteins serve three main purposes. Primarily, helices and sheets help to stabilize the structure by permitting the folded protein to hydrogen bond with itself. This hydrogen bonding contributes to the enthalpy, allowing the folded state to be favoured over the entropically more attractive random coil. Secondary structures also provide a degree of specificity to the protein structure, permitting residues to adopt precise orientations with respect to each other. Much as the anticodon on tRNA is chemically useful because its position at the end of a helix makes it physically accessible to the ribosome, so can a combination of amino acids in a protein form a binding site because secondary structure positions them in some convenient configuration. Secondary structure then takes the first step in providing the chemical specificity that is demanded of the functional molecules of life. The final benefit of the secondary structures is their efficiency in creating compact global protein structures. As we shall see later, there exists no better way to physically pack a chain molecule into a small volume than to use the same geometrical forms that are present in the α and β secondary structures. Thus, the hydrogen bonding properties of the amino acids make them useful as the building blocks of the functional molecules of life because they provide a compact form as well as chemical specificity.

Independent of Pauling's earlier work, Ramachandran, Ramakrishnan and Sasisekharan [30] showed that the sterically allowed backbone conformations of the amino acids select the α and β forms, and that a significant fraction of the remaining conformation space for each amino acid in a peptide chain is excluded due to steric interactions. This observation boosts the suitability of amino acids for forming the molecules of life, as the atomic configuration of the amino acid reinforces the secondary structures produced by hydrogen bonding. The limited conformation space also encourages fast folding of the protein because the geometry of the amino acids urge the protein chain toward conformations in which it can easily form hydrogen bonded secondary structures [31].

Unlike those that occur in the base pairing that dominates secondary structure in DNA and RNA, the hydrogen bonds that form between amino acids in proteins often do so without prejudice for which *types* of amino acids the bonds join. Whereas a folded tRNA requires specific local sequences in order to ensure proper base pairing and folding, a protein needs no such sequential specificity in order to hydrogen bond with itself. This permissiveness of sequence partially accounts for the great increase in catalytic power that the proteins demonstrate over RNA; because the number of sequences permitted for a folding, functioning protein is vastly greater than the number of sequences allowed for a similarly folding, functioning RNA molecule, the protein design is expected to support a greater variety of functions. Furthermore, amino acid side chains are placed on the outside of protein secondary structures, allowing a great deal of chemical variety between structures that have identical backbone conformations. The double helices of the nucleic acids, on the other hand, are all chemically identical to their solvent because each buries its distinguishing base sequence and exposes its sugar–phosphate backbone. It is then clear that amino acids are well-suited for constructing enzymes not only because they readily permit folding into compact structures, but also because the structures that they adopt highlight the uniqueness of the amino acid sequences.

While the rate at which protein structures are being solved continues to increase, novel protein folds are being discovered with decreasing frequency [32, 33]. All proteins are believed to be housed in only a few thousand folds [34], which might be understood as the enumeration of compact configurations formed of helices and sheets. While a short protein chain—one containing about 100 or fewer amino acids—will fit neatly into a single fold, longer protein chains fold into several *domains*, each housing part of the chain. The domains are themselves self-contained folds that fit together neatly to form a larger, composite structure. Thus, even though it is conceivable that a very long chain will adopt a similarly large and complex single fold, what is observed is that such chains instead adopt several smaller folds. Just as the secondary motifs are the structural building blocks of protein folds, the folds are the emergent building blocks of protein structure, and as such are an invaluable property of the proteins. Neutral evolution requires that the majority of genetic mutations cause no corresponding change in phenotype [14, 15]. Indeed, the mutation of a single amino acid frequently has no effect on protein structure, permitting a variety of genotypes to fit identically functioning proteins. It is also not uncommon for non-homologous amino acid sequences to share a fold. Such situations are also understood through neutral evolution: a protein subjected to many genetic mutations retains its form and function until one final mutation causes it to adopt a new fold altogether, or not fold at all and lose its functionality altogether. If this new fold has a function that carries with it a selective advantage, then it is retained; the fold has been stumbled upon by neutral evolution. Interestingly, it is primarily the protein fold, and not so much the amino acid sequence, that is responsible for determining the folding rate [35–37]. Furthermore, recent studies have shown that non-homologous proteins that share a fold also have similar transition state conformations [35, 38, 39]. The limited number

of folds, in conjunction with the knowledge that the folds determine the physical properties of the proteins, suggest that, contrary to common belief, the amino acid sequence is not the cause of protein structure, but rather that the sequence is merely the factor that determines which of many putative folds the protein will adopt.

Before we are blinded by the brilliance with which proteins fill their niche, a caveat: a host of illnesses, including Alzheimer's disease, Creutzfeldt–Jacob disease and type II diabetes result directly from the misfolding and aggregation of proteins [40–46]. These illnesses are believed to occur following the formation of fibril protein amyloids of hydrogen bonded β strands. The hydrogen bonding properties of the amino acids that permit protein folding also permit the formation of devastating amyloid plaques. A great effort is currently under way to understand the causes and prevention of amyloid plaques.

Despite the vast quantities of protein data that have been amassed over the past several decades, the protein folding problem—the task of determining how a protein folds given only its amino acid sequence—has remained unsolved. Although we know many of the generalities concerning the folding problem, the details remain elusive. To date, the most successful structural prediction methods employ knowledge-based techniques and rely heavily upon statistical analysis of the vast database of known protein structures. Even so, most prediction methods remain rather unreliable. In general, the reliability of a prediction method varies inversely with the detail of the prediction. It is easiest to predict whether a sequence of amino acids is more likely to form a protein heavy in helices rather than one that is composed mostly of sheets, but it is more difficult to pinpoint the locations of the helices and sheets along the sequence. Even if the secondary structures are correctly predicted, it is yet another task to find the three-dimensional fold that correctly assembles the substructures into the native state conformation. The problem is a bit menacing: on the one hand, we have been able to distil the key elements of protein folding as compaction driven by hydrophobicity and a tendency to form hydrogen bonded secondary structures, but on the other hand we find that the details are crucial and frustratingly difficult to deal with. An array of molecular dynamics programs have failed to solve the problem because of its enormous complexity, yet the proteins themselves fold quite rapidly and reliably in their native state, mocking the computer modeller who attempts to reproduce their dynamics. The problem has the additional complication that it is heavily dependent on history; the proteins that we see today have been selected through billions of years of evolution, during which each successful mutation could erase evidence of its predecessor. Homology matching across species can partially reconstruct ancestral proteins, but many gaps remain. History persists in complicating the problem because even if the sequences of ancestral proteins can be deduced, their relevance will remain unknown, as we have no knowledge of the specific environmental conditions that selected for any proteins in the distant past. As stated by Francis Crick [47], 'in biology, some problems are not suitable or not ripe for a theoretical attack . . ... This appears to be true of the protein-folding problem'.

It is clear that the proteins are superbly suited for their role as the functional molecules of life. Because they are chain molecules, they are easily constructed from a small number of building blocks. They are stable molecules with well-defined native state structures. Proteins display the enormous range of functional diversity necessary to sustain a living system, and they interact with each other through complex supramolecular networks, creating the emergent living phase. The rapidity with which proteins fold and their superb catalytic properties make them available for use on short timescales, and their limited phylogenetic variety promotes neutral evolution. All of these individual properties of proteins combine to make them suitable as the functional molecules of life, yet no blanket explanation has yet been provided to divine the common source of the valuable properties of proteins. The hydrophobic folding mechanism, the formation of helices and sheets, the limited number

of protein structures, the functional diversity and the occasional misfolded amyloid all stem from the nature of the protein, and the time is ripe for us to cease blindly viewing proteins as collections of weakly related properties and to instead see what the entire elephant truly is.

4. Organization of matter

4.1. From beliefs to facts

The ancient Greeks noted that matter is organized into only a limited number of forms; Empedocles hypothesized that the entirety of matter in the universe is composed of the four elements of earth, water, air and fire. This elemental concept was a novel idea, but his particular choice of the elements was obviously based on observations of everyday substances. Empedocles believed that the constituent parts of matter ought to reflect the properties observed in the bulk substance. Leucippus, and later Democretus, argued that matter cannot be infinitely subdivided, but that there exists some *atomos*, or indivisible unit of matter. Our modern understanding of the nature of matter incorporates Empedocles' elements with the *atomos* of Leucippus and Democretus: all ordinary matter is made up of atoms of various elements, and the properties of a material emerge from its elemental composition. Observations of a bulk substance will not yield an understanding of its constituent parts, and a thorough knowledge of the properties of the constituents will not necessarily provide a comprehensive understanding of their aggregate [48, 49].

The entire physical universe is governed by simple rules of symmetry that work together to bring about higher states of order, and a paradigm of modern science is that when order is observed, an underlying mechanism ought to be present. Based on the observation that their chemical properties are cyclical with mass, Mendeleyev arranged the elements into a periodic table. Even though only 66 elements were known at the time, the observed periodicity correctly predicted the properties of elements yet to be discovered. Despite the simple beauty and power of the periodic table, the observation of patterns of properties in the elements does not adequately explain the reason for or extent of the observed periodicity. Why, for example, do sodium and potassium behave similarly? The observed order fails to describe the underlying mechanism for ordering, but it indicates that such a mechanism exists. We know now that the ordering mechanism is explained through quantum mechanics; elegant mathematical laws give rise to atoms with quantized electron orbitals, which in turn produce the periodicity of the table. Before this precise mechanism came to be known, however, it was understood that there must be some physical similarities between atoms of elements that fall on the same column of the periodic table. The ordering of matter demands a physical basis, and the observed ordering of matter at any level must be an extension of some physical order at a lower level. The example of order in the chemical elements does not end at quantum mechanical electron orbitals. Atoms themselves are not truly indivisible particles, but each is a collection of smaller protons, neutrons and electrons. These particles, along with a host of more exotic particles that exist mostly in cosmic rays and particle accelerators, can be classified according to physical properties such as charge, spin, mass and strangeness. In his formulation of the quark theory, Gell-Mann employed symmetry arguments to explain that the ordering of these various properties results from an underlying physical mechanism. The Standard Model that is the cornerstone of high-energy physics is based on these same symmetry arguments, and experiments continue to validate its predictions. Indeed, arguments of symmetry have proved to be such a powerful force in explaining the nature of the physical universe that supersymmetry—the attempt to account for all matter in the universe through laws of symmetry—is a widely accepted approach to cosmology.

4.2. Some common phases of matter

The states of matter can be understood as a tower built upon a hierarchy of organization. Quarks organize to form nucleons, which organize with electrons to form atoms, which organize into the macroscopic materials that we commonly encounter. In the words of Philip Anderson [48], 'The behaviour of large and complex aggregates of elementary particles, it turns out, is not to be understood in terms of a simple extrapolation of the properties of a few particles. Instead, at each level of complexity entirely new properties appear'. Why, then, should a similar hierarchy not apply to living organisms? We began our discussion with an explanation of cooperation in life, and demonstrated how cooperation at the molecular level leads to cooperation at the level of the organism. An analogous observation might be made in the static case; just as cooperative activity at the level of the organism stems from the cooperative interactions of functional molecules within the organism, so does structural order at the level of the organism result from structural order at the molecular level. We are all physical entities created of atoms, and are therefore similar to the inanimate matter that we contact every day. We are, however, different from all inanimate matter, not because of some ethereal *vis vitalis*, but because of our material properties. Erwin Schrödinger contrasted a periodic crystal and an organic form in his statement that [50], 'The difference in structure is of the same kind as that between an ordinary wallpaper in which the same pattern is repeated again and again in regular periodicity and a masterpiece of embroidery, say a Raphael tapestry, which shows no dull repetition, but an elaborate, coherent, meaningful design traced by the great master'.

So it might be that the building blocks of human beings—our *atomos*—also cannot be easily categorized into some typical state of matter. This is clearly the case if the cell is considered to be the *atomos* of living matter; certainly a living cell is unlike any inanimate matter. But the cell can be further reduced: a cell is unlike other matter because of the symphony of chemistry that takes place within its walls. If the *atomos* of life is the protein, it is still quite unlike any non-living matter. Nowhere else do such molecules exist, each repeatedly performing the same step in what amounts to an intricate chemical dance. Life emerges from an organized collection of matter, and the nature of this organization ought to be investigated on the molecular level, where the boundary between living and non-living lies.

What organizational principles govern the ordering of common matter? Free atoms may be point-like particles that interact isotropically with each other on the average, and the dynamics of an atomic collection are often approximated by modelling the atoms as hard spheres. Even such a description, in which each atom occupies a point at the centre of a sphere and has unto itself a private volume that extends to the sphere's surface, is an emergent phenomenon [49]. Although it is somewhat crude, this model captures the physics of both the crystalline and the fluid phases of matter, and it provides both intuitive and quantitative descriptions of these phases. As Kepler first observed while pondering the stacking of cannon balls within a ship's hold, the most efficient way to fill space with a collection of hard spheres is to arrange them in a regular lattice (see figure 3). A crystal is then the emergent phase of a collection of tightly packed hard spheres, and the regularity of the crystal lattice results from the symmetry of its constituent parts. Fluids may also be modelled as collections of interacting hard spheres; the ideal gas and a fluid may be visualized as dilute collections of hard spheres that interact through elastic collisions. Unlike their highly constrained counterparts in crystals, the atoms in fluids are free to explore the full volume of the system, and this additional motility combines with the spherical geometry of the constituent particles to give rise to the macroscopic properties of the fluid.

The properties of crystals and fluids arise from the ordering of their spherical constituents, but many systems display properties that hard sphere models cannot account for. Liquid
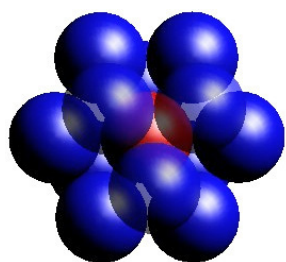
Figure 3. Spheres packed optimally into a face-centred cubic (fcc) lattice configuration. In this view, a single sphere is shown surrounded by its nearest neighbours. Each sphere in an infinite fcc lattice is in contact with exactly 12 other spheres. The resulting structure demonstrates how the translational ordering of the lattice can emerge from a collection of spheres, each of which is itself isotropic.

crystals [51, 52] are materials that exist in phases that display some of the order of crystals while maintaining a bit of the isotropy found in fluids. The emergence of novel liquid crystal phases results from breaking the spherical symmetry of the system's constituent particles. Consider, for example, the case in which the molecules of a liquid crystal are rod-like, rather than spherical. This uniaxial structure provides the liquid crystal with orientational order that is absent in a system of hard spheres, and the phase of a liquid crystal depends on both the relative positions and the orientations of its molecules. As shown in figure 4, a liquid crystal transition from the disordered isotropic phase to the highly ordered crystalline phase passes through mesophases of intermediate order. The nematic phase displays orientational order, but remains translationally isotropic. In this phase, the molecules are aligned axially in a common direction, but their translational motion remains fluidic. The smectic phases are characterized by orientational order as well as a limited degree of translational order. Molecules in the smectic phase separate into planar layers, and all of the molecules in a layer have a common orientation. Molecules can move freely within a layer, but they are restricted from moving between layers. The smectic phase then behaves like a stack of two-dimensional liquids, and the loss of translational freedom in one dimension is attributable to the uniaxial geometry of the constituent molecules. Both the nematic and the smectic phases occur within only a narrow range of environments, so liquid crystals in these phases are quite sensitive to minor environmental perturbations. Their proximity to the fluid and crystal phases provides liquid crystals with the exquisite sensitivity that makes them commercially useful. It has been pointed out [53] that the order and sensitivity of liquid crystals resemble those of living systems. Although living systems are clearly not liquid crystals, and vice versa, we shall see that the observed similarities between the two are explained through the geometries of the building blocks of their structures. Liquid crystals serve as a fine example of how certain crucial ingredients such as symmetry can significantly affect the behaviour of a model, and it is important that we do not lose sight of this as we move toward an understanding of the organization underlying the functional molecules of life.

4.3. Polymer phases

Matter is not always a system of freely interacting unconstrained particles; the particles in many systems are subject to constraints that influence the system's order. An important example of such a system is a chain molecule. Each monomer on a chain is constrained to be in contact with its nearest neighbours, so the motions of all the monomers along the chain are correlated. The most common examples of chain molecules are polymers [54–56]
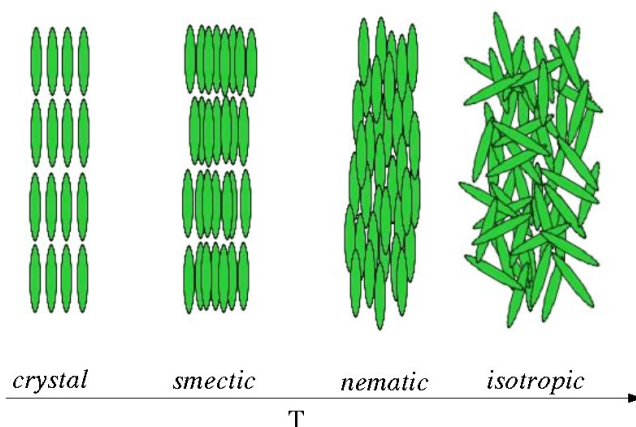
Figure 4. Liquid crystal phases. The constituent molecules of a typical liquid crystal are uniaxial and are depicted here as rods. In the high-temperature isotropic phase, shown on the far right, the molecules of the liquid crystal are translationally and orientationally disordered, and the system behaves as a classical liquid. Upon cooling to the nematic phase, the molecules of the system obtain orientational order but remain translationally isotropic. In this phase, all molecules tend to align their primary axes in the same direction. When cooled further, the liquid crystal enters the smectic phases, (second from left) which exhibit translational periodicity in a single dimension as well as orientational order. Molecules in the smectic phase separate into discrete layers, and while the molecules can move freely within their layers, they are restricted from moving between layers. The smectic phases display crystalline order in the one dimension perpendicular to the layers, and fluid order in the other two dimensions. When cooled sufficiently (far left), the liquid crystal reaches a phase of fully crystalline order.

that are typically longer than several tens of thousands of units in length, and in solution they have two distinct phases: the coil phase, in which the chain is extended and interacts very little with itself; and the globule phase, in which the polymer adopts a compact form to maximize self-interaction. Neither phase exhibits a great deal of order, and typically the properties of an entire system of polymers or polymer melt are given more consideration than the details of individual polymer structures. A solution laden with polymers in the globule phase, for example, resembles a colloidal suspension, whereas a solution with the same concentration of polymers in the extended conformation might display entirely different non-Newtonian properties that result from the interlinking of the polymer chains. Thus, the macroscopic properties of a polymer solution emerge from the microscopic phase of its constituent molecules.

The standard cartoon model for a chain molecule is a string of hard spheres. Although this picture accurately reproduces the swollen and compact phases of solvated chain molecules, it is unable to describe the phase of matter adopted by globular proteins in their native states. The surprisingly simple reason for this is that a chain of spheres intrinsically lacks the correct symmetry of a chain molecule. Thus, even on refining such a model, it fails to explain the phase behaviour and the nature of the low-energy conformations of proteins.

A sphere is an isotropic object, and the optimal close-packed conformation of a dense collection of hard spheres is a fcc lattice (figure 3). This result, conjectured several centuries ago by Kepler [57] and recently proved by Hales [58], demonstrates how crystalline ordering can emerge from a collection of spatially isotropic objects. A tethered chain of hard spheres would also have the fcc lattice as its optimal close-packed structure, provided the tethers do not forbid this arrangement. If the tethers are too short to permit the fcc arrangement, then the

chain's ground state will be some other conformation that is determined by the spatial isotropy of the spheres as well as the tether length. When this ground state becomes inaccessible due, for example, to dynamical effects, other tightly packed structures would be expected to have short-range order resembling crystalline packing at local scales. Structures comprising helices and almost planar sheets do not arise naturally in such a model.

There are additional problems with this simple description of a chain molecule. A powerful technique for understanding many scientific problems is to work in the continuum limit. For example, many of the advances in understanding fluids have been facilitated by the methods of continuum fluid mechanics, in which one treats a fluid as a continuous medium (the technique, of course, works best when the molecular aspects of fluid structure do not play an important role) and solves Newton's laws of motion, also called the Navier–Stokes equation, for the continuous fluid. Likewise, in polymer science it is occasionally useful to consider a continuous chain and use analytical techniques to obtain quantitative insights. In order to obtain the continuum limit of a chain of spheres, one needs to put in more and more spheres closer and closer along the chain. Because the self-avoidance criterion requires that the centres of no two sphere are located closer to each other than a sphere diameter, the sphere diameters must shrink as the continuum limit is approached. Indeed, in the continuum limit, one necessarily has to reduce the sphere size to zero, resulting in a continuous string of infinitesimal thickness.

The self-avoidance of such a string is typically handled by taking the limit of zero sphere diameter of the usual hard core potential that characterizes the non-overlap of a pair of spheres. In this limit, one obtains a singular pairwise δ-function potential between points along the string (see figure 5). The potential energy is then infinite when two points on the string exactly coincide (corresponding to an infinite cost of the intersection between two hard spheres), and is zero otherwise (corresponding to non-intersection of hard spheres). Interestingly, the standard treatment of a continuous polymer chain [59] uses precisely this type of singular potential and, even more curious, in order to study the collapse of a string, the typical potential employed is a pairwise *attractive* delta function and a three-body *repulsive* delta function. The ratio of the coefficients of these two terms determines whether one obtains a swollen, random coil phase or a collapsed globule phase[3].

---

[3] Conventional polymer science techniques do not address self-avoidance in everyday objects like ropes of non-zero thickness, and the necessity of dealing with singular potentials has profound and disturbing physical and mathematical consequences. Consider a rope making a closed loop with well-defined knot topology. One might wonder how certain quantities, such as its radius of gyration, scale with rope length for a fixed knot number. A δ-function type potential will not resolve such a quandary. Although the δ function provides an infinite energy cost when one part of the string precisely overlaps with another, the cost for going from one knot topology to another is finite and can be accomplished with non-zero probability. Thus, a knot treated with a δ-function potential might achieve the mathematical impossibility of changing its knot number, much like a rope in the hands of a magician. Indeed, investigation of such problems is more readily accomplished by computer studies than by analytic methods employing continuous strings.

Yet another consequence of an infinitesimally thin string arises when one considers a well-studied and well-understood problem in polymer science. How does the end-to-end distance, $R_{\text{end-to-end}}$, of a chain scale as the length of the chain L_{chain}? The well-known result is that in the swollen phase, $R_{\text{end-to-end}} \sim L_{\text{chain}}^{\nu}$ where the exponent ν is around $3/5$ in three dimensions. For the continuous string, even the statement of this result poses a problem: how does one get consistency in the dimensionality of the left- and right-hand sides of the equation? Because both $R_{\text{end-to-end}}$ and L_{chain} have dimensions of length L, there must be a proportionality constant on the right-hand side with dimension $L^{1-\nu}$; however, because the chain has no thickness, there is no other length scale in the problem. This problem is nicely resolved by using a Nobel prize-winning mathematical machinery [116] called the renormalization group method in which one introduces a cut-off length scale to 'regularize' the theory and then shows rigorously that the results are independent of this artificial length scale. The powerful renormalization group method comes to the rescue but one might wonder whether it is central to the problem being studied or is mere mathematical baggage which is peripheral to the issues of interest.
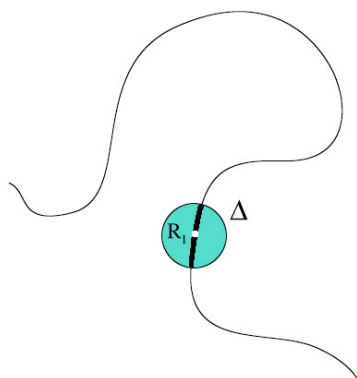
Figure 5. Cartoon depicting the necessity of δ-function interactions in the infinitesimally thin strings that are the common continuum model of chain molecules. If one assumes that pairwise interactions are sufficient to account for self-avoidance, then one provides each point on the curve with its own private volume, defined as the volume within a sphere of radius Δ centred at the point in question. Here the point R_1 has as its private volume the shaded region within the circle bounded by Δ. However, because the curve is continuous, there exist an infinite number of points along the curve—shown here as the heavy line—that will always lie closer to R_1 than Δ. Using a pairwise interaction to treat self-avoidance, then, requires special treatment for points that are near each other along the curve. The simplest solution is to shrink Δ to zero, resulting in a singular δ-function interaction.

4.4. Polymer phases revisited

The solution to the puzzle comes from noting that the symmetry of a chain is not accurately captured by representing it as a chain of spheres. As seen in the preceding subsections, symmetry plays a key role in determining the nature of ordering of the phases of matter. In a similar vein, an object that is part of a chain cannot be thought of as being isotropic. At the very least, such an object is characterized by one special local direction given by the locations of the adjoining objects along the chain, or the tangent to the chain. Thus in the simplest representation, one would need to replace the spheres along the chain with objects shaped like discs or coins (with the heads-to-tail direction representing the chain tangent).

A chain of discs lends itself rather naturally to a continuum limit, as illustrated in figure 6. As the number of discs along the chain increases and the separation between successive discs decreases, one obtains an object akin to a rope or tube of non-zero thickness. We now turn to the crucial question pertaining to how one would, in the simplest way, describe the self-avoidance of such a tube. Consider first the self-avoidance condition for a collection of hard spheres. This condition is met if for each pair of spheres, the sphere centres are no closer than the sphere diameter. Unfortunately, a pairwise interaction is not sufficient for describing the excluded volume constraint of a flexible tube in the continuum limit; in addition to the distance between two points on the tube axis, one also needs to know the context of the points (i.e. whether they are close by along the tube axis) in order to determine their interaction. Instead of a pairwise interaction, let us consider a three-body potential that characterizes the interaction between three particles on the axis of the tube. Given three points on the axis of a tube, one seeks to find a length scale, r, that is invariant under translation, rotation and permutation of these three points and that can characterize the self-avoidance, or lack thereof, of the tube. One can readily construct three independent length scales from the triangle defined by the three points $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$—the perimeter of the triangle, the area of the triangle divided by the perimeter and the product of the three sides of the triangle divided by its area. Unlike the
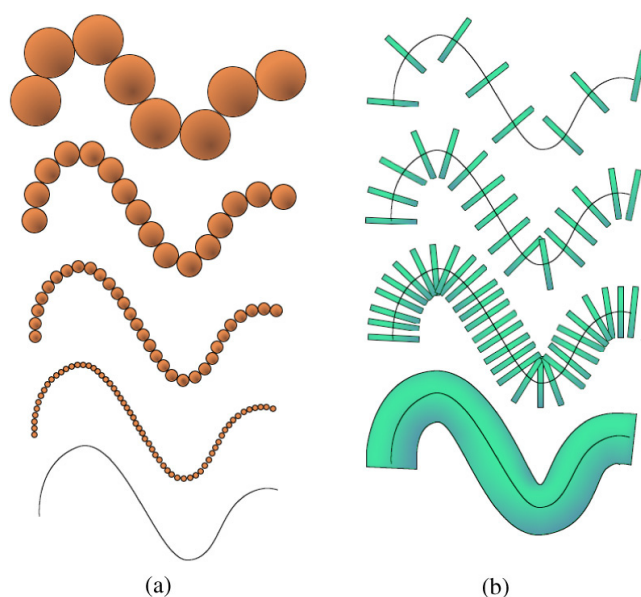
Figure 6. Continuum limits of (a) a chain of spheres and (b) a chain of coins or discs. The figure in (a) depicts a discrete chain of non-overlapping spheres taken to the continuum limit by iteratively halving the sphere radii and placing adjoining spheres in closer. Because the chain thickness is defined as the sphere diameter, the chain becomes an infinitesimally thin curve as the sphere diameter approaches zero. The figure in (b) shows the analogous progression for a chain of discs. While the separation between successive discs decreases in the continuum limit, the disc diameter can remain fixed, resulting in a continuous tube of non-zero thickness.

first two, which approach zero when the three points coalesce along the tube axis, the third length scale (which is proportional to r, the radius of a circle passing through the three points, as illustrated in figure 7) serves our purpose admirably [60, 61]. Indeed, one could use for $V(r)$ the potential commonly used in the hard sphere problem, i.e. $V(r) = \infty$ when $r < \Delta$ and $V(r) = 0$ otherwise (in the hard sphere problem r is half of the distance between a pair of sphere centres). This length scale neatly solves the contextual problem mentioned above: when two parts of a chain come together, the radius of a circle passing through two points on one side of the chain and one point on the other side turns out to be a measure of the distance of approach of the two sides of the chain, or the *non-local radius of curvature*. On the other hand, when one considers three contiguous points on the chain, the radius of the circle passing through them is simply the chain's *local radius of curvature* [60]. Indeed, when three such points form a straight line, the radius goes to infinity and the chain does not interact with itself locally. The straight line configuration is the situation of maximum self-avoidance for a tube of non-zero thickness. In the case of a chain molecule, such as a protein, a tube whose axis is a *smooth* string is clearly an approximation. One ought to introduce a *discrete* curve $\{\mathbf{r}_1, \mathbf{r}_1, \ldots, \mathbf{r}_N\}$. In correspondence with the considerations above, one may again define the thickness of a discrete curve, $\mathcal{C}$, as [60, 61]

$$\Delta(\mathcal{C}) = \min_{i,j,k} r(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) \tag{1}$$

where now i, j and k are all distinct.

The notion of a tube of non-zero thickness leads to a singularity-free description of self-avoidance in the continuum limit; indeed, the correct description of *any* chain molecule must
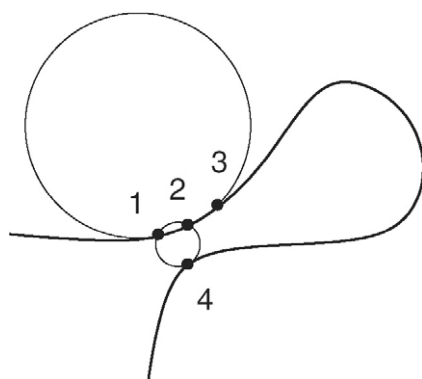
Figure 7. The definitions of local and global radii of curvature for a discrete chain. The local radius of curvature at i is defined as the radius of the circle passing through $i - 1$, i and $i + 1$. If the points 1, 2 and 3 are consecutive units on a chain, then the radius of the large circle that passes through all three points defines the local radius of curvature at point 2. The global radius of curvature at i is the radius of the smallest circle passing through the points i and any two other points on the chain. In the case shown here, the circle passing through points 1, 2 and 4 has the smallest radius of all the circles passing through point 2, and therefore defines the tube thickness at 2.

contain the inherent anisotropy that is implicit in a chain [62]. Conventional polymer phases that are well described by a chain of spheres or the continuum string of infinitesimal thickness obviously lie in the limit in which the tube thickness is small compared with other length scales in the problem. Biomolecules such as DNA or proteins, on the other hand, present a somewhat different situation because their bulky side groups confer non-zero thickness to their chains. Furthermore, for proteins, the attractive force promoting compaction occurs between the outer atoms of the adjoining side chains and is necessarily short in range because of the screening influence of the water surrounding the protein. Indeed, in that case, the range of attractive interactions is comparable to the tube diameter and has important consequences, as seen below.

Consider a discrete tube of length L, radius Δ and range of attraction R—the tube axis is made up of discrete points, each representing a monomeric unit of a chain molecule. Let us postulate a pairwise attractive interaction so that there is an energy reward when two monomers are within a distance R, and the energy is zero for all pairs separated by more than R. The self-avoidance of the tube is ensured by requiring that none of the three-body radii is smaller than Δ. Note that this model does not encode any heterogeneity and is a simple variant of the chain of spheres model with the only difference arising from the way the self-avoidance is captured. Yet, as we shall see, one obtains qualitatively new features on incorporating the inherent anisotropy of a tube.

Let us consider a short tube equivalent to a chain molecule made up of about a hundred monomers. The phase diagram of this tube is shown in figure 8. When R is much larger than Δ, the tube is in the conventional compact polymer limit and one obtains an energy landscape with significant degeneracy. In this region, there are a multitude of tube conformations that permit a large number of pairs to avail of the attraction, and the vast majority of these are structureless (not made up of any distinctive structural building blocks). At the other extreme, when Δ is sufficiently large compared to R, one again obtains many degenerate conformations. In this case, the tube is too fat for the monomers to undergo attractive interaction and one essentially obtains all self-avoiding conformations of a tube as ground
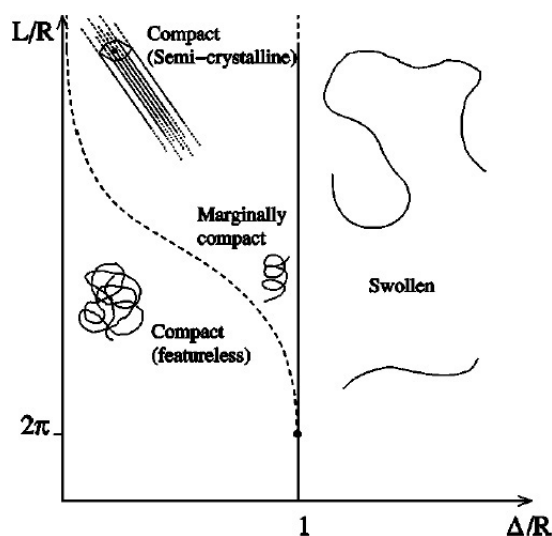
Figure 8. Sketch of the phase diagram (reproduced from [71]) of a discrete tube of length L and thickness Δ subject to a pairwise compacting potential with effective range of interaction R. Very long, thin tubes tend to pack together into bundles with hexagonal symmetry, much like a collection of tightly bound sticks or pencils. Short, thin tubes collapse into featureless, compact conformations that allow a great number of pairwise contacts. Thick tubes of all lengths reside in the swollen phase—the large tube thickness does not permit the tube to undergo the attraction promoting compaction. Short tubes for which $\Delta \sim R$ are in the marginally compact phase and display a high degree of order and are characterized by the presence of helices and paired strands.

states. As one varies the dimensionless ratio Δ/R, one obtains [62, 63] a phase transition between these two degenerate phases when $\Delta \sim R$. This phase transition is first order (akin to the melting of ice) but with a divergent persistence length. (The persistence length is a measure of how the tangent to the chain at one location is correlated with the tangent a certain distance away measured along the chain.) The phase transition from the swollen phase occurs when the attraction between the tube segments barely kicks in on lowering Δ while holding the range of attraction constant. In the vicinity of this transition, one finds *marginally compact* tubes in which the pairwise attraction competes with the three-body constraint. Because of their proximity to the swollen and compact phases, one obtains exquisite sensitivity to the right types of perturbations and confers flexibility to the structures in the marginally compact phase. This sensitivity emerges from geometrical considerations and, as illustrated in figure 9, is reminiscent of the sensitivity of the liquid crystal phases.

The nature of the energy landscape becomes very simple in the marginally compact phase [62–66]. The structural degeneracy is much lower and the structures of choice are modular and made of two kinds of building blocks—helices and strands assembled into almost planar sheets. In order to understand how this comes about, let us consider the optimal conformations of a very short tube. An optimal helix is obtained by locally bending a tube as tightly as possible (recall that the smallest local radius of curvature allowed is equal to Δ, the tube radius) and by placing successive turns of the helix right on top of each other (figure 10). Such a space-filling helix [65, 67] has both its local radius of curvature *and* its non-local radius of curvature equal to Δ and a pitch to radius ratio of $2.512\ldots$. Interestingly, the inherent anisotropy of a tube enforces parallel placement of nearby tube segments [62–66, 68], in this case the successive turns of a helix.
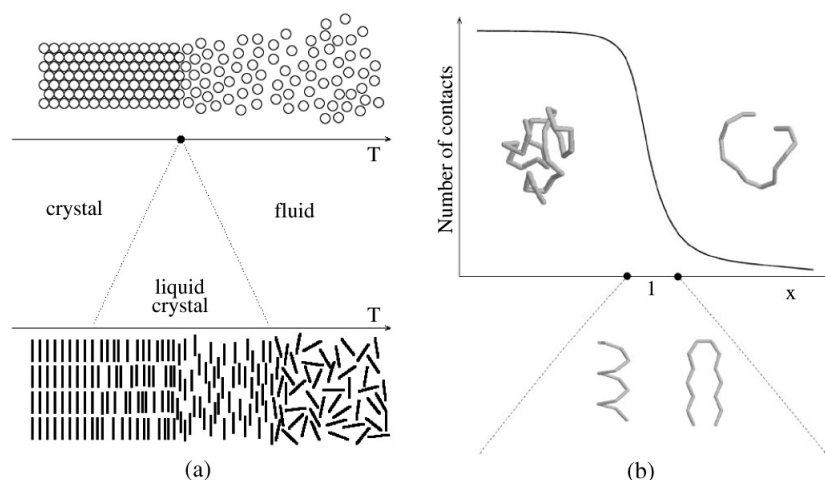
Figure 9. Schematic phase diagrams for (a) a fluid–crystal transition and liquid crystals and (b) a tube subject to a compacting potential. The cartoon in (a) shows how a collection of hard spheres (top) undergoes a phase transition from a disordered fluid state to an ordered crystal state upon cooling or densification. Similarly, a collection of uniaxial rigid rods (bottom) will make a transition from fluid to crystal, with partially ordered liquid crystal states in between. The liquid crystal states exist for only a small range of temperatures, making them exquisitely sensitive to small changes in the environment. The diagram in (b), taken from [63], shows a similar phase diagram for tubes. When X, the ratio of the effective distance of interaction to the tube thickness, is small, one finds that the tube adopts a disordered compact phase. When X is sufficiently large, one finds a swollen phase. In the region where $X \sim 1$, the tube adopts structured marginally compact phases that are characterized by helices and strands forming almost planar sheets. Its existence over a narrow range of X between the compact and swollen phases provides the marginally compact phase with liquid crystal-like sensitivity.

The formation of almost planar sheets from a series of parallel tube segments is also straightforwardly explained [66, 69]. If the tube thickness Δ is sufficiently large compared with the range of the pairwise attraction, the tube configured as an optimal helix is unable to undergo attraction but must instead adopt some other conformation in order to encourage pairwise contacts. As illustrated in figure 11, the convenient solution *for a discrete tube* is for nearby tube segments to adopt zig-zag conformations that permit non-local pairwise contacts. Unlike the helix, which is a uniaxial object, the strand has biaxial symmetry: the overall direction of the chain defines one axis and the plane of the zig-zag defines a second. The biaxial symmetry of a strand then translates into the planar geometry of a sheet when the strands aggregate. In the absence of the zig-zag pattern, the tube would be completely straight, and tube segments would stack in a hexagonal configuration like logs on a truck. The zig-zag that breaks the uniaxial symmetry of a straight tube also prevents the isotropic stacking of tube segments, promoting the formation of almost planar sheets. For a long tube or for many short tubes stacked together, the planar sheet structure can be continued indefinitely and indeed the formation of amyloid, implicated in debilitating human diseases, arises from the formation of cross-linked β structures. Thus, in the marginally compact phase, for short tubes one obtains the two key building blocks of modular structures—the space-filling helix and the zig-zag strands assembled into almost planar sheets. These building blocks themselves are emergent structures—they arise in a non-trivial manner from the constituent amino acids but in an amino acid aspecific manner. Interestingly these building blocks themselves are anisotropic, a feature reminiscent of the constituents of the sensitive liquid crystal phase of matter.
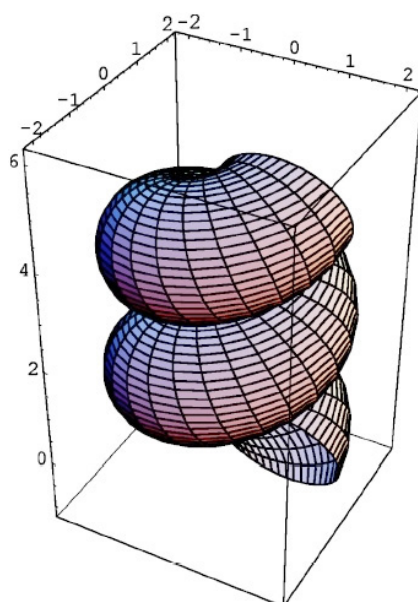
Figure 10. (a) Rendering of a tube segment curled into a space-filling optimal helix, reproduced from [65]. In such a helix, the local radius of curvature, global radius of curvature and tube thickness are all equal. The structure is said to be space-filling because there is no empty space either along the axis of the helix or between successive turns.

4.5. Facts about proteins and the unifying picture

There is a striking similarity between the structural building blocks of the marginally compact tube and protein secondary structures. The α helices that are ubiquitous in proteins display a typical pitch-to-radius ratio that is within a few per cent of the ratio for an optimal helix [65]. The β strands found in proteins form almost planar sheets through non-local contacts in much the same way that a discrete marginally compact tube forms sheets of parallel tube segments. Furthermore, the anisotropy of the tube allows for an all-or-nothing folding— nearby tube segments have to snap into place alongside each other and parallel to each other to undergo attraction in the marginally compact phase. The most remarkable feature of the marginally compact phase is that the structures of choice are determined not from the details of the chemical propensities of the amino acids and their varied side chains but rather by the overarching features of geometry and symmetry. The shadows of protein structure emerge from a phase that is described by a simple model that properly accounts for the geometry of the chain molecule.

One can consider a more refined model [70, 71] for protein structures than a humble garden hose by carefully studying the geometrical constraints imposed by backbone hydrogen bond formation and the effects of sterics (or non-overlap of atoms), again in a manner independent of the specific amino acid involved. The basic lesson that we have learned is that when dealing with a chain, it does not suffice to simply know where two objects are relative to each other, but one must also be aware of the context in which they occur. It therefore becomes necessary to additionally consider how the local coordinate systems at the two locations are oriented with respect to each other. The simplest way of defining a local coordinate system (see figure 12) is through Cartesian coordinates with the three axes defined by the tangent, normal and the binormal at a given location. A
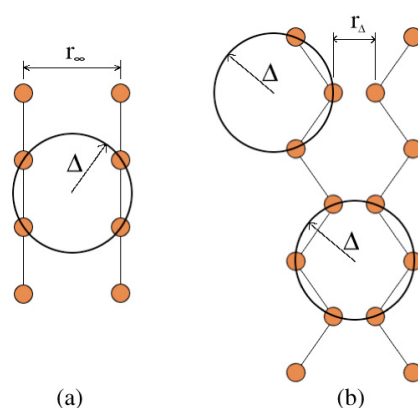
Figure 11. The formation of planar sheets from strands. For a discrete tube of thickness Δ, the self-avoidance constraint is met as long as no triplets of points lie on a circle of radius less than Δ. The pairwise compacting potential is such that any pair of points within a distance R from each other contributes favourably to the conformational energy. The marginally compact phase sets in when R takes on the minimum possible value which still permits an effective attraction between neighbouring strands. Figure (a) shows the extreme case in which two straight tube segments, each of thickness Δ, run parallel to each other. In this case, the local radius of curvature—that is, the radius of a circle passing through three consecutive points on a single tube segment—is infinite, and the non-local radius of curvature, the radius of a circle passing through two points on one tube segment and a nearby point on the other, is Δ. It can be shown that if Δ is held constant, sliding one of the tube segments parallel to the other only serves to increase r_∞, the distance of closest approach between points on opposing chains. The geometry shown then provides the minimum value of r_∞ for straight, parallel segments of tube with fixed Δ. Measured in units of the separation between consecutive beads of the chain, this is given by $r_\infty = \sqrt{4\Delta^2 - 1}$ (note that for $\Delta < \frac{1}{2}$, three-body self-avoidance is no longer a factor). If $R < r_\infty$, parallel straight chains are unable to accommodate the pairwise attraction. In order to yet undergo attraction, the uniaxial symmetry of the straight chain must be spontaneously broken by creating a zig-zag conformation in which each tube segment is locally bent by a radius of no less than Δ. Figure (b) shows two parallel zig-zag strands that are maximally bent such that both the non-local and local radii of curvature are Δ. The distance of closest approach between two points on opposing zig-zag chains with equal local and non-local radii is given by $r_\Delta = 2\Delta - \frac{1}{\Delta}$. Marginally compact tubes in which $r_\Delta < R < r_\infty$ can undergo pairwise attraction by adopting a conformation with a local radius of curvature that is greater than Δ while retaining a non-local radius equal to Δ. Such is the case in proteins, in which the local radius of curvature of a strand is about 1.07 times the separation between consecutive amino acid residues, while the non-local radius in a sheet of strands is about 0.71 times the amino acid separation. A consequence of the biaxial symmetry of the zig-zag conformation is that the chains aggregate into planar sheets. Straight chains have uniaxial anisotropy, and a collection of such chains would pack together into a hexagonal array. Breaking the uniaxial symmetry of the chain eliminates the possibility for hexagonal packing and instead promotes the planar stacking of neighbouring tube segments.

study [70, 71] of experimentally determined protein structures in the Protein Data Bank [29] reveals that there are indeed strong amino acid aspecific constraints on both the local radius of curvature and on the geometrical relationships between amino acids that form backbone hydrogen bonds with each other. These constraints turn out to be consistent with and wholly captured by the tube paradigm presented above. The incorporation of these constraints into a simple geometrical model for a tube made up of 48 amino acids leads to assembled structures that bear a striking resemblance to real protein structures, as shown in figure 13. These considerations underscore the important role played by geometry and symmetry in determining the nature of proteins, and they shed light on the
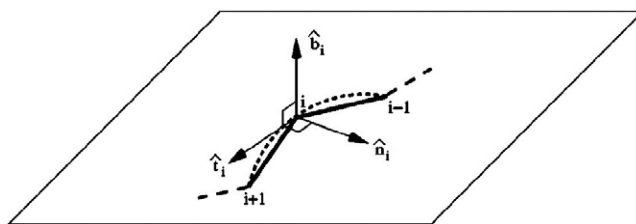
Figure 12. The natural coordinate system of a discrete chain [71]. The tangent $\hat{t}_i$ at i runs parallel to the segment joining $i - 1$ and $i + 1$. The normal $\hat{n}_i$ to the chain at i points from i to the centre of the circle formed by $i - 1$, i, and $i + 1$. The binormal $\hat{b}_i$ is defined such that $\hat{t}_i \times \hat{n}_i = \hat{b}_i$.
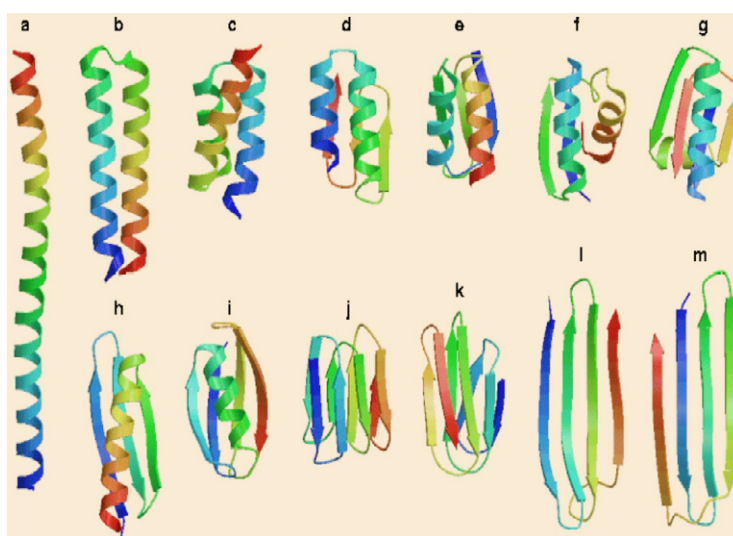
Figure 13. A sampling of conformations of a homopolymer modelled as a marginally compact tube of 48 residues. All structures are compositions of helices and strands, and most also have the low radius of gyration that is characteristic of globular proteins. The image is reproduced from [105].

numerous attendant advantages of the novel phase of matter in which Nature houses protein structures.

Let us briefly summarize several key results on proteins and assess how well the physical phase of matter of a marginally compact short tube is useful for understanding them.

Globular proteins share many common characteristics [72, 73] in spite of having very different sequences of amino acids. They fold rapidly and reproducibly into their native state structures [21]; in all cases, the geometry of the protein in its folded state controls its functionality; they share a limited number of topologically distinct folds [34]; protein structures are modular forms made up of simple building blocks—helices and almost planar sheets assembled from zig-zag strands; these structures are flexible, accounting for the ability of proteins to carry out a wide variety of tasks; proteins are able to interact with each other and with ligands in a very versatile yet robust manner; proteins are able to act as molecular targets of natural selection; and proteins have a tendency to aggregate and form amyloid. These stunning similarities between proteins can all be explained by understanding the marginally compact phase of a short tube. We discuss these properties in turn below.

Under physiological conditions, proteins fold rapidly and reproducibly into their native state structures [21]. A corollary of the rapid folding is that proteins often fold in an all-or-nothing manner without encountering dead-ends or misfolded states. The folding is driven by the aversion to water of some amino acid side chains leading to the creation of a hydrophobic core in the folded state. The marginally compact phase of a tube accounts for this cooperative folding by encouraging global geometries in which nearby segments run parallel to one another.

The geometry of a folded protein determines its functionality. Anfinsen [21] wrote in 1973, 'Biological function appears to be more a correlate of macromolecular geometry than of chemical detail'. All globular proteins have helices and sheets as their building blocks. Pauling and co-workers [27, 28] showed that helices and sheets are repetitive structures for which hydrogen bonds provide the scaffolding. A decade later, Ramachandran and co-workers [30] showed that steric effects also lead to helices and sheets as the preferred building blocks of protein structures. As shown earlier, helices and almost planar sheets occur naturally in the marginally compact phase of tubes. The tube picture provides a novel explanation of how the works of Pauling and Ramachandran, though seemingly quite different, both lead to the same helix and sheet structures. The laws of quantum chemistry and sterics conspire to independently provide a marvellous fit to the preferred structures in the marginally compact phase—an example of Nature adapting to her own laws. Indeed, contingency, the opportune application of 'historical mistakes' to select one evolutionarily course over other possible evolutionary pathways, seems to have played a role in selecting the proteins. Nucleic acids, such as DNA, also exist in the marginally compact phase [74]. Similarly, the atomic configurations and chemical properties of the amino acids permit them to form chains that are an excellent fit to this phase of matter.

There is considerable evidence, accumulated since the pioneering suggestion of Kimura [14] and King and Jukes [15], that much of evolution is neutral. Evolution can be thought of as a 'random walk' in sequence space that forms a connected network [75]—there is no similar continuous variation in structure space. Evolution and natural selection allow Nature to use variations on the same structural theme facilitated by the rich repertoire of amino acids to create enzymes that are able to catalyse a remarkable array of diverse and complex tasks in the living cell. This picture of molecular evolution is well supported by the tube model. Because the menu of possible structures is determined by geometry and symmetry, a stunningly simple picture emerges in which protein sequences and functionalities evolve within the fixed backdrop of the geometrically determined folds.

In order for protein native state structures to be targets of an evolutionary process, they must be stable, sensitive and diverse. Stability is needed because one would not want to mutate away a DNA molecule that is able to code for a useful protein; sensitivity is required in order to accomplish the myriad tasks that proteins perform; and diversity allows complex and versatile forms to evolve. With these three factors in place, selection occurs naturally—genes that code for stable proteins with useful functions thrive at the expense of genes that create unstable or useless polypeptides. The marginally compact tube provides stability, sensitivity and diversity in its low-energy conformations, creating a natural medium for evolution.

Protein interactions are at the heart of the network of life, and it is important that their interaction network does not disintegrate as protein sequences evolve. In order to maintain and especially enhance the interactions among proteins, the native state structures may evolve in one of two distinct ways: either the structures co-evolve in a coherent manner, retaining the classic lock-and-key mechanism that defines network connectivity, or there exists a menu of folds that are determined not by the sequence but by considerations common to all proteins. Such a menu provides a fixed backdrop [76] for evolution of sequences and functionalities.

Because the folds are limited in number, and because they are the physical means through which proteins interact, coherent co-evolution is not required in order for phylogenetic mutations to gain acceptance into the interaction network. Were the folds not immutable but themselves subject to Darwinian evolution, the possibility of creating so many subtle and wonderful variations on the same theme would not exist. These facts point to the picture of a pre-sculpted energy landscape [70, 71] that is shared by all proteins and has around a thousand local minima corresponding to putative native state structures—not so few that structural and functional diversity are impeded, nor so many that the landscape becomes too rugged to permit the rapid and reproducible folding of proteins.

Indeed, the total number of distinct folds is only of the order of a few thousand [34, 77, 78], and this fact is often exploited in structure prediction techniques [35, 79, 80]. Proteins are relatively short chain molecules, and longer globular proteins form domains which fold autonomously [81]. Many proteins share the same native state fold [82–85] and often the mutation of one amino acid into another does not lead to radical changes in the native state structure [38, 39, 86–100]. In addition, multiple protein functionalities can arise within the context of a single fold [101]. Recent experiments [38, 39, 86, 87, 102] have been successful in mapping out the nature of the transition state in several proteins. Interestingly, proteins that have similar native state topologies also have similar folding rates [35, 37], even if their amino acid sequences differ significantly [36, 37, 103]. Furthermore, mutational studies [35, 38, 39, 86, 87, 104] have shown that, in the simplest cases, the structures of the transition states are also similar in proteins sharing the same native state topology.

So what, then, is the role played by the amino acid sequence of a protein? The reproducibility of protein folding after chemical denaturation [21] has led to the view that folding proceeds down a funnel-like free energy landscape [24, 26]. A dominant belief in the field is that the folding funnel is created by the amino acid sequence [22, 23], and that only those sequences that produce funnel-like energy landscapes will be able to fold rapidly and reproducibly. This view is difficult to reconcile with the above-mentioned observations of the sequence-independent nature of proteins; because many sequences may share a fold, there has so far been little success in determining the precise nature of the interactions between amino acids that will lead to a folding funnel. If one considers a model of a globular protein as a chain of spheres, for example, and then attempts to determine a set of amino acid specific parameters that will universally predict the native states of proteins, one will be disheartened to find that those parameters that work best for one protein are not necessarily effective in discerning the native states of other proteins. To begin with, a hard sphere model does not readily lend itself to formation of secondary structure, forcing one to impose artificial constraints on the model. Even with such constraints properly in place (or alternatively, with a more complicated model that continues to rely on pairwise interactions), one finds that the differences between amino acids do not by themselves universally explain the folding of proteins.

We suggest that the folding landscape is *not* determined by the amino acid sequence, but that it is the pre-existing landscape corresponding to a marginally compact phase determined by the common attributes of all proteins. This phase has only a few thousand stable structures, each of which lies at the bottom of its own folding funnel and provides a stable state for a specific amino acid sequence. The role of the amino acid sequence is not to create the folding funnel, but merely to select one of around a thousand pre-existing folding funnels. Simulations have shown [105] that by using a model with only two types of amino acid, one can design protein sequences that fold reproducibly into specific structures (figure 14).

There is an additional necessity for variation among protein sequences. A useful protein is one that can interact with other proteins and cell components in a synergistic manner. There
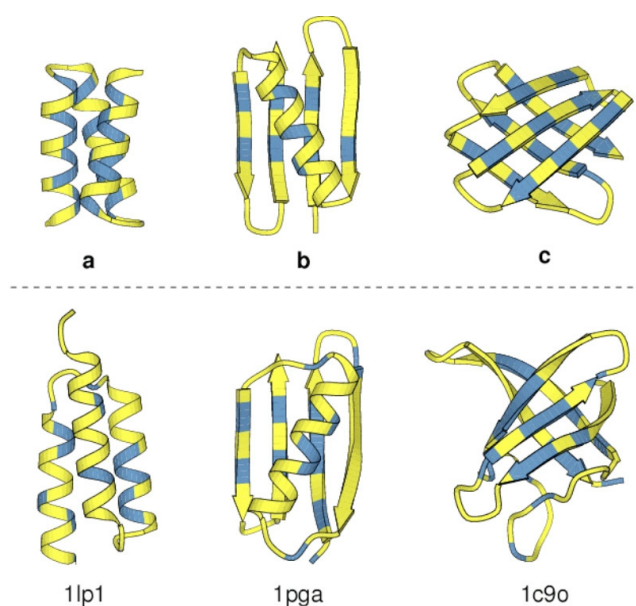
Figure 14. Ground state structures for designed heteropolymers of 48 units, taken from [105]. In the model considered, each monomer is either hydrophobic (blue, dark) or polar (yellow, light), but there is otherwise no distinction between them. Varying the hydrophobicity pattern produces different ground state conformations shown in (a), (b) and (c). Each of these is a stable conformation for a marginally compact homopolymer, but the sequence of hydrophobic and polar monomers serves to select a single conformation as a well-defined ground state.

has been much recent progress in extracting information on biological function and protein interactions [106] from the structures of proteins and the complexes that they form [107]. The existence of a pre-sculpted energy landscape with broad minima corresponding to the putative native state structures and the existence of neutral evolution demonstrates that the design of sequences that fit a given structure is relatively easy, and that many sequences can fold into a given structure [82–85]. This freedom facilitates the accomplishment of the next level task of evolution through natural selection: the design of optimal sequences, which not only fold into the desired native state structure but also fit into the environment of other proteins and the surrounding cell products.

A range of human diseases such as Alzheimer's, spongiform encephalopathies, type II diabetes and light-chain amyloidosis lead to degenerative conditions and involve the deposition of plaque-like material in tissue arising from the aggregation of proteins [40–46]. A variety of proteins not involved in these diseases also form aggregates very similar to those implicated in the diseased state [41–45]. The tendency for proteins to aggregate is a generic property of polypeptide chains with the specific sequence of amino acids again playing at best a secondary role. The vast number of experimental data on proteins suggest that both the class of cross-linked β structures and the menu of native state structures are determined from geometrical considerations (see figure 15). This picture suggests that the native state structures of proteins are determined not by the details of their sequences but by a phase of matter that exists due to the common attributes of all proteins. There are two classes of structures that exist in this phase: the thousand or so folds which sequences can choose from for their native states and the aggregated amyloid phase, in which the importance of the sequence is diminished even further.
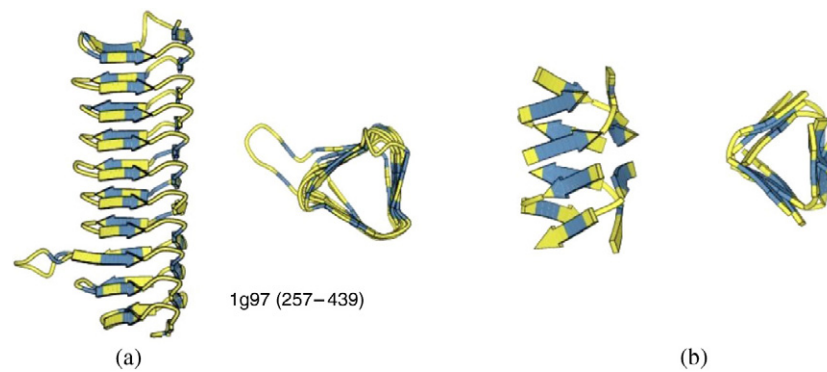
Figure 15. Sections of amyloid structures, taken from [105]. The structure in (a) is a β helix taken from the protein 1G97. The amyloid plaques that are associated with diseases like Alzheimer's and spongiform encephalopathy are formed from similar fibres, which are formed of cross-linked β strands that run perpendicular to the fibril axis. In the radial view on the left, the topmost β strands are only paired on one side, providing a series of hydrogen bonding sites for nearby peptide chains. Proteins that have misfolded into the amyloid form tend to promote the misfolding of other proteins, extending the fibrils indefinitely. The structure in (b) results from a simulation of a protein modelled as a marginally compact tube. The original structure, shown in figure 14(a), is a three-helix bundle that forms in a model with two types of amino acid: hydrophobic (blue, dark) and polar (yellow, light). When the protein is cut into six pieces of equal length, the three-helix bundle dissolves in favour of the β-linked amyloid structure.

These experimental and theoretical findings strongly suggest that the topology of the native state structures is by and large determined not by the details of the amino acid sequence but rather by some overarching principles of geometry and symmetry. This behaviour is somewhat analogous to that of crystal structures, which are determined by the requirements of periodicity and space filling and not by the material that is housed in any given crystal structure. Of course, unlike crystals, protein structures are neither infinite nor periodic. The unified picture leads to a single free energy landscape with two distinct classes of structures. The amyloid phase is dominated by β strands linked to each other in a variety of forms, whereas the structures of the native state are assemblies of α helices and β structures. Nature has exploited these native state structures in the context of the workhorse molecules of life. The selection mechanism for genetic evolution at the molecular level lies in the ability of proteins to fold comfortably into one of the predetermined folds and have useful function. Unfortunately, however, the proximity of this beautiful phase to the generic amyloid phase underscores how life can easily malfunction as soon as the aggregative tendencies of proteins come to the fore.

5. Discussion

The functional molecules of life, in order to give rise to the network of molecular interactions that animates the cell, must be stable, sensitive and chemically specific. These molecules must be able to support the wide range of functions that are necessary for metabolism, but they must also show limited structural diversity in order to accommodate evolution. Ideally they are simple constructions of only a few kinds of reusable building blocks, and are therefore likely to be chain molecules. The physical realizations of these molecules are the proteins, which are chain molecules with stable, well-defined native state structures. Chemical specificity in proteins arises from the combinations and arrangements of amino acids in their active sites.

Correlated movements in different regions of the proteins provide them with diverse functionality, and they interact with each other either directly or indirectly via other molecules in the cell. The proteins meet some of our criteria for the functional molecules of life in ways that we have not anticipated. For example, all globular protein structures are constructed from helix and strand building blocks, and these combine into only a thousand or so acceptable folds. The protein secondary structures are not specified features of functional molecules of life, but Nature employs them in order to fulfil other specified criteria. Secondary structures not only give proteins well-defined and stable conformations, but they also play a role in limiting the number of protein phenotypes, permitting evolution. There are additional features of proteins that we did not initially specify for the functional molecules of life, such as their length. Protein chains are relatively short—on the order of a few hundred amino acids—allowing them to fold rapidly upon synthesis and permitting many proteins to function within the cell without interfering with each other. The folding rate of proteins is more rapid even than is expected for chains of their length, suggesting that their folding is a directed, rather than random, process. Furthermore, proteins have the undesirable tendency to occasionally misfold into amyloid fibres. This is not one of the desiderata of the functional molecules of life, but is instead an unfortunate generic consequence of the phase of matter that they reside in.

A description that captures the physical properties of the proteins is a marginally compact tube. Unlike conventional continuum models of polymers, the tube picture correctly accounts for the inherent anisotropy of the chain. Furthermore, this description is the first of its kind that allows one to unambiguously and easily describe self-avoidance in real tube-like objects, such as thick ropes or garden hoses. In the case of proteins, the range of attraction and the effective thickness are self-tuned to be comparable to each other, and the corresponding tube is marginally compact. Such a tube provides a description of the phase of matter that exhibits all of the properties that are essential to the functional molecules of a living system, expressly: a limited number of stable conformations, sensitivity, functionality, diversity and specificity. In much the same way that the structures of crystalline solids are limited to only 230 space groups, the structures of the marginally compact tube are limited to only a few thousand geometrically determined conformations. Like the liquid crystal phases, which are poised between fully isotropic fluids and completely ordered crystals, this phase is sensitive to environmental perturbations because it resides in the vicinity of the transition between swollen and compact phases. Unlike the liquid crystal phases that are made up of many independent objects, the marginally compact tube provides a context for the identification of its parts. It is this context that allows the chain to function as a whole, rather than as a collection of loosely interacting parts. In addition, the features of this phase arise without consideration of the chain sequence; the chemical properties of the monomers are not of primary importance to the characteristics of the phase. Specificity then results from the chemical properties of only a small fraction of the monomers on the chain, and chemical diversity immediately follows as a result of permutations of the active site monomers.

What is truly remarkable is that the marginally compact tube phase that describes the functional molecules of life also reproduces details of the proteins that are peripheral to their function. This phase is characterized by helices and sheets that are almost identical to those found in protein structures. No requirement for secondary structure is made of the functional molecules of life, yet the very secondary structures that occur in proteins also emerge from strictly geometrical considerations in a tube. The few thousand stable conformations of this phase are arrangements of these structural building blocks and are geometrically similar to the protein folds. Furthermore, the tubes fold cooperatively, reflecting the observed alacrity with which proteins fold. Finally, this phase has as a competitive native state an extended sheet that is reminiscent of the amyloid plaques formed by proteins. These features are not

explicitly built into the model, but instead they emerge naturally as characteristics of the phase that carries proteins.

Of the two major consequences of the description of the molecules of life as existing in a novel phase of matter, the first is conceptual. The detailed workings of the universe are seemingly beyond human comprehension, and scientists therefore find it necessary to use approximations and model building to explain natural occurrences. Descriptions that encapsulate the gross features of natural phenomena tend to promote an understanding of the interplay between their various components. A relevant example is the popular conceptual tool of viewing the ribosome as a nanobiological machine: given a sequence of nucleobases in the form of mRNA, the ribosome constructs a protein with an associated amino acid sequence. The exact mechanism that it employs need not be understood in order to conceptualize genetic translation.

A vexing problem in molecular biology is that there currently exists no analogous conceptual model that describes how an arbitrary sequence of amino acids folds into its native state conformation and subsequently functions within the cellular environment. At present, the models that best describe various aspects of proteins are highly compartmentalized, such that each describes only the limited aspects of proteins for which it is specifically designed. Statistical methods and data mining are useful for understanding protein evolution or predicting structure, but they are not particularly useful when trying to determine or understand the method of interaction between two proteins in a cell. Likewise, quantum molecular dynamics simulations are quite helpful for understanding the behaviour of a protein binding site, but they are unable to reproduce broad motions of the protein, such as the folding pathway or low-frequency motions. For understanding these, modellers turn to coarse-grained normal mode analysis or models that presume preferential contacts. There does not currently exist, even at a very coarse level, any model that encapsulates the chemical, structural and evolutionary properties of the proteins. Issues such as the universal presence of helices and strands, the limited number of protein folds, the formation of amyloid plaques, the exact sequence–structure relationship and the folding pathway are not fully explained by any of the conventional protein models. Furthermore, proteins are so complex that experiments can only probe a tiny corner of their universe. Experimental studies are frequently limited to investigating only a single protein or family of proteins, leaving the universal features of proteins unexplained. The proposed phase of matter recovers many of the universal features of proteins in a conceptually accessible manner. By visualizing them as folded states in the vicinity of a transition to the swollen phase one is more easily able to intuitively grasp the multifarious nature of proteins.

The second significant impact of this description is in its potential to provide a foundation for the physical basis of life. As all organisms are material, biology is necessarily compatible with physical laws. However, biological systems often appear to obey rules that are far removed from those of classical physics. Newton's third law, for example, cannot explain the reaction that a schoolboy will warrant upon being unexpectedly struck by his classmate. Even though schoolboys are physical systems, they are so enormously complex that their behaviour is quite unpredictable. Indeed, human beings are arguably the most complex systems in the known universe, but even simpler organisms like bacteria are far too complex to behave predictably. The complex behaviour that we associate with life emerges from the network of interactions between proteins and other cell products; while a cell might be said to be alive, an isolated protein within the cell is simply a molecule and is not alive. The details of how a collection of lifeless molecules combine to animate an organism have yet to be distilled. Surely the structure of their network is vital, but more fundamentally it is the physical properties of the molecules that enable them to form this network. Thus it is at the

level of the proteins that the physical laws that apply to all inanimate matter first yield to the rules governing life. In a natural way that is consistent with known physics, the living molecular phase provides an explanation for how the functional molecules of life can emerge from common matter.

The mere suggestion that biological phenomena fall under the blanket of physics may be taken as rather contentious. Many biologists vehemently defend their discipline as an autonomous science that is not subject to the laws of physics. The great evolutionary biologist Ernst Mayr wrote [108], 'To the best of my knowledge, none of the great discoveries made by physics in the twentieth century has contributed anything to an understanding of the living world'. In truth, physics has contributed significantly to our understanding of the living world, although perhaps not in the areas where Mayr focused his attention. For example, quantum mechanics explains the hydrogen bonds that stabilize protein structures, and it is at the heart of nuclear magnetic resonance that is commonly used to establish the structures of biomolecules. On the other hand, the complexity and emergent properties of biological systems prevent them from being described, as are complex physical systems, by a concise set of equations. Mayr's assertion [108] that 'none of the autonomous features of biology can ever be unified with physics' highlights the rift between these scientific disciplines, and while the autonomy of biology is not in question here, the claim that biology and physics are incompatible demands address. Mayr provides four concepts—essentialism (typology), reductionism, universal natural law and determinism—that he claims are central to physical sciences but are absent from biology. We shall address these in turn and consider the implications that a living molecular phase of matter has for each.

Essentialism, the idea that all natural objects or actions can be categorized into exhaustive and non-overlapping types, permeates the physical sciences. The elementary particles and chemical elements fall neatly into discrete classes; there exists no intermediate element between nitrogen and oxygen, for example. This is clearly not the case with organisms, which display an enormous amount of variation even within a single species. While it is impossible to hybridize chemical elements, it is both possible and common to form hybrid organisms. The scope of this hybridization, however, is limited by physical properties at the molecular scale. Although there are seemingly limitless variations to the sequence of bases on a strand of DNA, the bases themselves are the subjects of essentialism. There are only four types of bases that can be present in DNA, and the chemistry of hybrid bases prevents them from successful substitution in functional genes. This immutability of form extends beyond the properties of the nucleobases to the structures of the proteins. The number of protein folds is limited to a few thousand, and the explanation for this limitation is that the folds correspond to the low-energy conformations of molecules in a unique phase. That the structures of the functional molecules of life are determined from completely physical considerations does not detract from the autonomous nature of biology; instead, it reinforces the true nature of life as a phenomenon that emerges naturally in ordinary matter. Typology is not absent in biology, but its presence at the molecular level may go unnoticed when studying life at the organismal and phylogenetic levels. The continuity of form that one observes in organisms does not extend to the proteins. Indeed, it is the immutability of the protein folds that provides protein interaction networks with stability in the face of genetic mutations. Were the proteins able to adopt a continuous range of folds, arbitrary mutations would have the power to destroy protein networks, preventing evolution altogether.

In direct opposition to the reductionism that dominated the early years of physical sciences, a prevailing paradigm in modern physics is the study of emergence. An increasing variety of physical phenomena, from superconductivity to the recently discovered supersolid [109, 110] and superatomic [111] states, can be understood only as characteristic

properties of an aggregate, and are not present in its individual constituent particles. Living organisms are the quintessential example of emergence: organismal behaviour emerges from a collection of tissues, which in turn behave according to the emergent features of a collection of cells, each of which emerges from a collection of molecular networks. Although cellular activity is too complicated to understand through straightforward application of chemical laws, the molecules within the cell must adhere directly to physical law. By considering the proteins as belonging to a novel phase of matter that is uniquely suited for life, we remove the mystery surrounding the emergence of life from a collection of inanimate particles. Instead of thinking of proteins as large and complicated molecules whose behaviour is determined by a complex network of chemical interactions, it is useful to think of them as entities in a state of matter that has its own emergent properties that are intractable in terms of the chemistry of the individual amino acids. Such a view allows one to understand that the properties of proteins arise from physical law, and it places them in the ranks of numerous physical phenomena that are understood only as emerging from a collective.

The scarcity of universal laws in biology is a testament to the complexity of biological systems. Mayr wrote [108] that 'Most theories in biology are based not on laws but on concepts. Examples of such concepts are, for instance, selection, speciation, phylogeny, competition, population, imprinting, adaptedness, biodiversity, development, ecosystem, and function'. It is absurd to anticipate the existence of a set of equations that can determine the life expectancy of any organism to arbitrary precision, or that will accurately predict the features that are evolutionarily selected within a population in some microclimate. The living world is far too complicated to be governed by a simple set of rules, and therefore the investigative techniques that are employed in many fields of biology are dramatically different from those typical of physics research. It is important to note, however, that the relevant length and timescales in both physics and biology span several orders of magnitude, and that in both fields the techniques that are useful at one scale are often inapplicable at other scales. The disparity between microbiology and ecology, for example, is akin to that between quantum mechanics and geophysics. The living molecular phase extends the reach of physical laws in biological systems beyond atoms to macromolecules, explaining the *concepts* that biologists hold about proteins in terms of physical *laws*. Because the catalytic properties of the proteins and the immutability of their folds have already been established, the physical explanation for these characteristics as features of a material phase need not startle biologists. There is no reason why a system, just because it is labelled as 'biological', should not be described by physical laws if such laws exist. Indeed, the discovery of physical laws that influence biological systems will only deepen our insight into the nature of life.

The physical sciences rely heavily upon mathematical descriptions of natural phenomena, and because of the precise nature of its underlying mathematics, physics is frequently believed to be a science that demands casual determinism. Since the development of quantum theory in the 1920s, the role of determinism in physics has been the topic of much discussion. Although there exist interpretations of quantum mechanics as a deterministic field, most physicists accept that the quantum universe is inherently rife with random events. Regardless of how one interprets them, the equations of quantum mechanics are fixed, and quantum determinism is essentially a matter of philosophy.

There exist many macroscopic systems that have nonlinear equations of motion and are subject to chaotic behaviour. Such systems are deterministic, but because their dynamics depend heavily upon the details of their initial conditions, their behaviour has the appearance of randomness. The nonlinearity of the weather is commonly explained through the well-known (though entirely fictitious) example of a butterfly in Paris causing a tornado in Texas: the seemingly insignificant effect of a butterfly's wings on atmospheric conditions is amplified

through the nonlinear dynamics of the atmosphere and may result in catastrophic weather patterns.

The intrinsic unpredictability of quantum mechanics demonstrates that interesting physics is not always deterministic, and the apparent randomness of chaos shows us that fully deterministic processes can be completely unpredictable. Biological systems are subject to both quantum noise and nonlinear effects—the former introduces randomness into the system behaviour and the latter amplifies the effects of random variations from both quantum and external noise. It is the complexity of life that makes biological systems unpredictable. In what other system can a single molecular event alter the face of a planet? Yet this is exactly what happens on Earth, as the random walk of evolution that occurs on the molecular level manifests itself in macroscopic phylogenetic variation and impacts upon the planet as a whole. Indeed, there exists no more powerful amplifier in the universe than the living world, and the echoes of genetic mutations that occurred billions of years ago still ring in the DNA of modern organisms. By recognizing the protein folds as stable conformations of matter in a particular phase, we can begin to understand the relationship between the biological world and the physical universe in which it is embedded. Bernal [112] wrote, in 1939, 'The problem of the protein structure is now a definite and not unattainable goal, but for success it requires a degree of collaboration between research workers which has not yet been reached. Most of the work on proteins at present is uncoordinated; different workers examine different proteins by different techniques, whereas a concentrated and planned attack would probably save much effort which is now wasted, and lead to an immediate clarifying of the problem'. Undoubtedly the clarification of the protein problem will result only from an accurate physical description that accounts for the key features common to all proteins, such as is provided by the paradigm of the tube model. A deeper understanding of the nature of life might then result from studies of life's higher ordering, such as the laws governing the networks of interactions among proteins and other biological components. Whether or not our future understanding of biological systems finds them to be largely deterministic, it is important that we keep in mind that these remain physical systems that cannot violate physical law, even if their behaviour is too complicated to easily understand.

In recent years the global physics community has grown increasingly interested in biological phenomena. This biological renaissance stems from advances in biological research that permit living systems to be described in terms of mathematics that is familiar to physicists. While some biologists may insist that modern physics is useless to biology, few physicists would claim that modern biology has had no impact on physics. The living world is rife with fascinating structures and events that are absent from the realm of inanimate matter. Proteins are no exception, as Flory [113] noted: 'Synthetic analogs of globular proteins are unknown. The capability of adopting a dense globular configuration stabilized by self-interactions and of transforming reversibly to the random coil are peculiar to the chain molecules of globular proteins alone'. It is quite exciting to physicists to uncover new physics in any form, but when—like the phase of matter that describes the proteins— it is found in a biological system, its implications extend across the borders of scientific disciplines. However, there are fundamental differences between the scientific approaches that physicists and biologists take toward a topic of research. In the eyes of a physicist, both the atmosphere of a planet and a single bacterium are complex systems that abide by physical laws. A biologist recognizes that only one of these systems is *alive*. Our research suggests that deeper biological insight can indeed be gleaned from a comprehension of the underlying physics of life, and that the *vis vitalis* has a physical origin.

Despite mounting evidence to the contrary, ideas persist that living systems display an irreducible complexity that implies that they have been crafted by an intelligent designer.

We suggest that the elegant forms of the protein folds are no more mysterious than the crystals of ice on a windowpane: both proteins and crystals have structures that characterize a phase of matter. The protein folds do not require the intercession of an intelligent designer; helices and sheets emerge spontaneously from the phase of matter occupied by the proteins. Furthermore, this model subverts the alleged 'irreducibly complex' nature of the protein interaction networks from which life emerges. Because there are only a limited number of physically allowable stable conformations for matter in their phase, all functional proteins must be housed in only a small number of folds. Because their allowed conformations are local energy minima of homopolymers that encapsulate the common characteristics of all proteins (i.e. they are not strongly dependent on amino acid sequence), a single fold can house many amino acid sequences. Evolution proceeds naturally from here. The robustness of the folds permits a random exploration of sequence space without structural changes to most proteins or significant disturbance of their network. Occasionally a mutation causes a change in phenotype, and if an advantageous effect accompanies the change, the mutation is retained. Most importantly, the limited number of folds provides novel proteins with a fair chance of being chemically active with an existing network. Thus, even though a protein system *in its present state* may exist as a closed set of indispensable components, the universal structural properties of proteins allow protein networks to evolve. Our work suggests that neither is the blueprint of life mysterious nor is there any reason to doubt that life as we know it could have arisen merely from chance and natural selection.

Henderson [114] pondered whether the nature of our physical world is biocentric: is there a need for fine-tuning in biochemistry to provide for the fitness of life in the cosmos or even for life here on Earth? It is remarkable that the lengths of the covalent and hydrogen bonds and the rules of quantum chemistry conspire to provide a perfect fit to the basic structures in the novel phase of matter studied here. One cannot but be amazed at how the evolutionary forces of Nature have shaped the molecules of life ranging from DNA, which carries the genetic code and is efficiently copied, to proteins, the workhorses of life, whose functionality follows from their form, which, in turn, is a novel phase of matter. Protein folds seem to be immutable—they are not subject to Darwinian evolution and are determined from geometrical considerations, as espoused by Plato. It is as if evolution acts in the theatre of life to shape sequences and functionalities, but does so within the fixed backdrop of these Platonic folds.

A great deal of the technology of the last century resulted from the exploitation of the emergent properties of material phases. Solid-state electronics are perhaps the most obvious example of technology emerging from a phase of matter. Our modern lives are filled with fascinating devices that all owe their functional existence to the electronic properties of the solid state. It is interesting to note that much of the technology that we have derived from the common states of matter has some natural analogue in the living state. The photoelectric effect, in which light incident on a crystalline metal produces an electric current, has as a natural analogue in the chlorophyll molecules that have been converting radiation to chemical energy for billions of years. Synthetic fabrics such as nylon have natural analogues in the silks spun by arthropods. The lac extracted from the scale insect is the natural precursor to synthetic polymer solids such as vinyl. Bioluminescent proteins like luciferase have been producing light of specific colours for millions of years before the invention of the LED. The list goes on, but the theme remains: through the exploitation of the living phase and with billions of years of trials, Nature has spontaneously produced substances that our technology is only beginning to mimic. As Darwin [115] stated, 'Slow though the process of selection may be, if feeble man can do much by his powers of artificial selection, I can see no limit to the amount of change, to the beauty and infinite complexity of the coadaptations between all organic beings, one with another and with their physical conditions of life, which may

be effected in the long course of time by nature's power of selection'. Indeed, 'feeble man' has only managed to produce enormously simplified versions of the novel materials found in biological systems. A synthetic polymer thread may not share the vibrant elastic properties and great tensile strength of silk, but it is far easier to produce in a laboratory. Simplification is the rule for human selection of technology, and human-made materials almost always take the most direct path to their objective. Our association of synthetic materials with phases of matter other than the living molecular phase is understandable because the molecular phase of living matter is not easily mastered.

Not only do natural phenomena precede much of technology, but they also inspire a great deal of it. It is hard to imagine that, without the birds to envy, humans could have possibly conceived of either the idea or the means of conquering the skies in flight. Nature, on the other hand, produced an array of flying animals through selection. Early aircraft designers looked to the avian world for guidance, and 100 years later modern jumbo jets still retain the birdlike qualities of a central body, two wings and a tail. This example of human kind's emulation of nature demonstrates that even though the details of a natural phenomenon may be impossible to reproduce, central themes in the biological world can guide innovation. Constructing a machine that propels itself with flapping wings is a feat even today, and constructing the muscle tissues and feathers that aid the flight of a bird is currently impossible. An excellent approach to biologically inspired technological advancement is using the natural form as a guide and improvising where Nature has employed too sophisticated a solution. Thus, technology replaces muscle with engines and motors as the source of mechanical energy. As our knowledge and capabilities advance, however, the gap between biological templates and our most sophisticated imitations of them continues to narrow. Drug development, for example, requires the synthesis of chemicals with properties that are identical to those of the organic molecules that they mimic. Here there is no room for improvisation; if the drug does not exactly match its target, it will assuredly fail to produce the desired effect. Meanwhile, our ability to manipulate the molecules of life has advanced to a state where a molecular biologist can create such oddities as fluorescent mice. Our understanding of the molecular mechanisms of life is converging with our ability to synthesize pseudo-organic chemicals, and the near future may see the synthesis of artificial living systems based on the framework for life that Nature has provided.

Before this scenario becomes a reality, a multitude of details needs to be sorted out, not the least of which is pinpointing the chemical magic that must occur in order to transform inanimate matter into a living system. The functional basis for life is a network of interacting proteins, and while the functional basis for artificial life should likewise be a network of interacting molecules, there is no reason to assume that those molecules need to be proteins. The origin of proteins in terrestrial life is likely due to the contingent availability of amino acids on pre-biotic Earth, and the persistence of protein-based life is due to selective preference. The presence of proteins has never been proved to be a necessity for life. Although the amino acids are well-suited for their role in proteins, it is the *phase*, and not the amino acids, that empowers proteins to behave as the functional molecules of life. Thus, just as an aspiring aviator need not create a bird in order to fly, a scientist need not create proteins in order to have a synthetic living system. The critical element to flight is not the bird, but the pressure differential caused by air flow over the wing; the critical element of the functional molecule of life is not its exact chemistry, but its existence in a living molecular phase.

This thought ought to be kept in mind by those who endeavour to gain insight into nanobiological materials. Research progresses steadily toward the development of molecular machines and mechanical nanodevices, and our ideas concerning operational devices at this scale need neither descend from our knowledge of the macroscopic world nor ascend from

our understanding of quantum mechanics. In the development of nanoscale machinery, we must be acutely aware of the template that Nature has provided and of the theme that she has triumphantly employed in the functional molecules of life. Proteins fold and function because they are in a marginally compact phase, and they are in this phase because they are chain molecules with effective distances of self-attraction and self-avoidance that are roughly the same. The design of functionally useful nanoscale machines should be accompanied by verification that the machines themselves exist in this phase of matter. Such devices ought to be self-tuning to fit the marginally compact phase and designed to self-assemble unambiguously into a single stable conformation. Attention to these details will set research on the path to the creation of nanoscale machines that follow the same design principles as those used in the universe's most elegant chemical systems.

Our work provides hints to the answers to deep and fundamental questions that have been pondered by our ancients. Was life on Earth inevitable? For about a billion years into its existence, the Earth, while impressive, was bleak and made up primarily of inorganic matter with the largest molecules having less than a hundred atoms. And then there was life. Once life began, the random walk of evolution took its course and the forces of natural selection shaped life into what it is today. The key question of course is what the essential difference is between inanimate and living matter. Both kinds of matter are governed by physical law. While we have a reasonable understanding of the gross behaviour of inanimate matter, a similar simple understanding, even in principle, has been missing for living matter. Specifically, what is it about proteins that allow them to carry out a dizzying array of functionalities with aplomb and at the same time serve as the molecular targets of natural selection?

The answer lies in the fact that the novel phase of physical matter populated by protein native state structures has all the attendant advantages needed to accomplish this. Even for a homopolymer, one obtains a simple energy landscape with around a thousand minima, not so many that rapid or reproducible folding cannot take place and not so few that a lack of diversity thwarts the development of complexity. These minima correspond to geometric structures for which a simple lock and key mechanism can be operational making possible catalytic mechanisms speeding up reactions by factors of tens of billions. The phase of matter lies in the vicinity of a phase transition providing exquisite sensitivity of the structures to the right types of perturbations. The inherent anisotropy of a chain molecule provide a simple mechanism for rapid cooperative folding of proteins—nearby tube segments need to snap into place parallel to and alongside each other to avail of the attraction promoting compaction. Finally, there is a remarkable self-tuning of two length scales—the effective thickness of the tube controlled by the sizes of the side chains of the amino acids and the range of attraction which is also controlled by the locations of the outer atoms of the same side chains. To our knowledge, this marginally compact phase of matter of short tubes is unique in its ability to provide the key attributes of life. Thus it does not seem surprising that, given the right environment and resources, Nature would have stumbled into this phase with her chain molecules and set out on the road to living matter.

We close with the thoughts of Erwin Schrödinger, who is perhaps the most visible of the twentieth-century physicists to seriously contemplate the existence of life based on modern physical law. As one of the many fathers of quantum theory, Schrödinger was well aware that an entirely new branch of physics had to be formulated in order to explain the world of the very small. Perhaps it was with this in mind that he wrote [50], 'We must therefore not be discouraged by the difficulty of interpreting life by the ordinary laws of physics. For that is just what is to be expected from the knowledge we have gained of the structure of living matter. We must be prepared to find a new type of physical law prevailing in it'.

Acknowledgments

References

[1] Miller M B and Bassler B L 2001 *Annu. Rev. Microbiol.* **55** 165
[2] McAllister A K, Katz L C and Lo D C 1999 *Annu. Rev. Neurosci.* **22** 295
[3] Lehn J-M 2004 *Rep. Prog. Phys.* **67** 249
[4] Lehn J-M 1992 *Nobel Lectures, Chemistry 1981–1990* ed T Frängsmyr and B G Malström (Singapore: World Scientific)
[5] Weiss P S 2001 *Nature* **413** 585
[6] Ballardini R, Balzani V, Clemente-Leon M, Credi A, Gandolfi M T, Ishow E, Perkins J, Stoddart J F, Tseng H R and Wenger S 2002 *J. Am. Chem. Soc.* **124** 12786
[7] Goto H and Yashima E 2002 *J. Am. Chem. Soc.* **124** 7943
[8] Whitesides G M, Mathias J P and Seto C T 1991 *Science* **254** 1312
[9] Gates B D, Xu Q, Stewart M, Ryan D, Willson C G and Whitesides G M. 2005 *Chem. Rev.* **105** 1171
[10] Whitesides G M 2003 *Nature Biotech.* **21** 1161
[11] Cleland C E and Chyba C F 2002 *Orig. Life Evol. Biosphere* **32** 387
[12] Ruiz-Mirazo K, Peretó J and Moreno A 2004 *Orig. Life Evol. Biosphere* **34** 323
[13] Eschenmoser A and Kisakürek M V 1996 *Helv. Chim. Acta* **79** 1249
[14] Kimura M 1968 *Nature* **217** 624
[15] King J L and Jukes T H 1969 *Science* **164** 788
[16] Mojzsis S J, Krishnamurthy R and Arrhenius G 1999 *The RNA World* 2nd edn, ed R F Gesteland, T R Cech and J F Atkins (New York: Cold Spring Harbor Laboratory Press)
[17] Szostak J W, Bartel D P and Luisi P L 2001 *Nature* **409** 387
[18] de Duve C 2005 *Nature* **433** 581
[19] Dworkin J P, Lazcano A and Miller S L 2003 *J. Theor. Biol.* **222** 127
[20] Kornberg A 1989 *For the Love of Enzymes* (Cambridge, MA: Harvard University Press)
[21] Anfinsen C B 1973 *Science* **181** 223
[22] Gō N 1983 *Annu. Rev. Biophys. Bioeng.* **12** 183
[23] Bryngelson J D and Wolynes P G 1987 *Proc. Natl Acad. Sci. USA* **84** 7524
[24] Wolynes P G, Onuchic J N and Thirumalai D 1995 *Science* **267** 1619
[25] Bryngelson J D, Onuchic J N, Socci N D and Wolynes P G 1995 *Proteins* **21** 167
[26] Dill K A and Chan H S 1997 *Nat. Struct. Biol.* **4** 10
[27] Pauling L, Corey R B and Branson H R 1951 *Proc. Natl Acad. Sci. USA* **37** 205
[28] Pauling L and Corey R B 1951 *Proc. Natl Acad. Sci. USA* **37** 729
[29] Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I and Bourne P 2000 *Nucleic Acids Res.* **28** 235
[30] Ramachandran G N and Sasisekharan V 1968 *Adv. Protein Chem.* **23** 283
[31] Srinivasan R and Rose G D 1995 *Proteins* **22** 81
[32] Hou J, Jun S-R, Zhang C and Kim S-H 2005 *Proc. Natl Acad. Sci. USA* **102** 3651
[33] Vendruscolo M and Dobson C M 2005 *Proc. Natl Acad. Sci. USA* **102** 5641
[34] Chothia C 1992 *Nature* **357** 543
[35] Baker D 2000 *Nature* **405** 39
[36] Kim D E, Gu H and Baker D 1998 *Proc. Natl Acad. Sci. USA* **95** 4982
[37] Perl D, Welker C, Schindler T, Schroder K and Marahiel M A 1998 *Nat. Struct. Biol.* **5** 229
[38] Chiti F, Taddei N, White P M, Bucciantini M, Magherini F, Stefani M and Dobson C M 1999 *Nat. Struct. Biol.* **6** 1005
[39] Martinez J C and Serrano L 1999 *Nat. Struct. Biol.* **6** 1010
[40] Kelly J W 1998 *Curr. Opin. Struct. Biol.* **8** 101
[41] Dobson C M 2003 *Nat. Rev. Drug Discov.* **2** 154
[42] Radford S E and Dobson C M 1999 *Cell* **97** 291
[43] Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J S, Taddei N, Ramponi G, Dobson C M and Stefani M 2002 *Nature* **416** 507

[44] Dumoulin M *et al* 2003 *Nature* **424** 783
[45] Chiti F, Stefani M, Taddei N, Ramponi G and Dobson C M 2003 *Nature* **424** 805
[46] Prusiner S B 1998 *Proc. Natl Acad. Sci. USA* **95** 13363
[47] Crick F 1988 *What Mad Pursuit* (New York: Basic Books)
[48] Anderson P W 1972 *Science* **177** 393
[49] Laughlin R B 2005 *A Different Universe: Reinventing Physics from the Bottom Down* (New York: Basic Books)
[50] Schrödinger E 1962 *What is Life?* (Cambridge: Cambridge University Press)
[51] Chandrasekhar S 1977 *Liquid Crystals* (Cambridge: Cambridge University Press)
[52] de Gennes P G and Prost J 1995 *The Physics of Liquid Crystals* (Oxford: Oxford University Press)
[53] Needham J 1936 *Order and Life* (Cambridge, MA: MIT Press)
[54] Yamakawa H 1971 *Modern Theory of Polymer Solutions* (New York: Harper and Row)
[55] de Gennes P G 1979 *Scaling Concepts in Polymer Physics* (Ithaca, NY: Cornell University Press)
[56] de Cloiseaux G and Jannink J F 1990 *Polymers in Solution: Their Modeling and Structure* (Oxford: Clarendon)
[57] Szpiro G G 2003 *Kepler's Conjecture* (New York: Wiley)
[58] Sloane N J A 1998 *Nature* **395** 435
[59] Doi M and Edwards S F 1993 *The Theory of Polymer Dynamics* (New York: Clarendon)
[60] Gonzalez O and Maddocks J H 1999 *Proc. Natl Acad. Sci. USA* **96** 4769
[61] Banavar J R, Gonzalez O, Maddocks J H and Maritan A 2003 *J. Stat. Phys.* **110** 35
[62] Marenduzzo D, Flammini A, Trovato A, Banavar J R and Maritan A 2005 *J. Polym. Sci.* **43** 650
[63] Banavar J R and Maritan A 2003 *Rev. Mod. Phys.* **75** 23
[64] Banavar J R, Maritan A, Micheletti C and Trovato A 2002 *Proteins* **47** 315
[65] Maritan A, Micheletti C, Trovato A and Banavar J R 2000 *Nature* **406** 287
[66] Banavar J R, Flammini A, Marenduzzo D, Maritan A and Trovato A 2003 *Complexus* **1** 4
[67] Snir Y and Kamien R D 2005 *Science* **307** 1067
[68] Banavar J R, Cieplak M and Maritan A 2004 *Phys. Rev. Lett.* **93** 238101
[69] Banavar J R, Maritan A and Seno F 2002 *Proteins* **49** 246
[70] Hoang T X, Trovato A, Seno F, Banavar J R and Maritan A 2004 *Proc. Natl Acad. Sci. USA* **101** 7960
[71] Banavar J R, Hoang T X, Maritan A, Seno F and Trovato A 2004 *Phys. Rev.* E **70** 041905
[72] Baldwin R L and Rose G D 1999 *Trends Biochem. Sci.* **24** 26
[73] Baldwin R L and Rose G D 1999 *Trends Biochem. Sci.* **24** 77
[74] Stasiak A and Maddocks J H 2000 *Nature* **406** 251
[75] Smith J M 1970 *Nature* **225** 563
[76] Denton M and Marshall C 2001 *Nature* **410** 417
[77] Chothia C and Finkelstein A V 1990 *Annu. Rev. Biochem.* **59** 1007
[78] Chothia C, Gough J, Vogel C and Teichmann S A 2003 *Science* **300** 1701
[79] 2003 *Proteins* **23** (Suppl 6)
[80] Jones D T, Taylort W R and Thornton J M 1992 *Nature* **358** 86
[81] Privalov P L 1982 *Adv. Protein Chem.* **35** 1
[82] Bowie J U, Reidhaar-Olson J F, Lim W A and Sauer R T 1990 *Science* **247** 1306
[83] Lim W A and Sauer R T 1991 *J. Mol. Biol.* **219** 359
[84] Heinz D W, Baase W A and Matthews B W 1992 *Proc. Natl Acad. Sci. USA* **89** 3751
[85] Matthews B W 1993 *Annu. Rev. Biochem.* **62** 139
[86] Villegas V, Martinez J C, Aviles F X and Serrano L 1998 *J. Mol. Biol.* **283** 1027
[87] Riddle D S, Grantcharova V P, Santiago J V, Alm E, Ruczinski I and Baker D 1999 *Nat. Struct. Biol.* **6** 1016
[88] Richardson J S and Richardson D C 1989 *Trends Biochem. Sci.* **14** 304
[89] DeGrado W F, Wasserman Z R and Lear J D 1989 *Science* **243** 622
[90] Hecht M H, Richardson J S, Richardson D C and Ogden R C 1990 *Science* **249** 884
[91] Hill C P, Anderson D H, Wesson L, DeGrado W F and Eisenberg D 1990 *Science* **249** 543
[92] Sander C and Schneider R 1991 *Proteins* **9** 56
[93] Kamtekar S, Schiffer J M, Xiong H Y, Babik J M and Hecht M H 1993 *Science* **262** 1680
[94] Brunet A P, Huang E S, Huffine M E, Loeb J E, Weltman R J and Hecht M H 1993 *Nature* **364** 355
[95] Davidson A R and Sauer R T 1994 *Proc. Natl Acad. Sci. USA* **91** 2146
[96] West M W, Wang W X, Patterson J, Mancias J D, Beasley J R and Hecht M H 1999 *Proc. Natl Acad. Sci. USA* **96** 11211
[97] Wei Y, Kim S, Fela D, Baum J and Hecht M H 2003 *Proc. Natl Acad. Sci. USA* **100** 13270
[98] Shakhnovich E, Abkevich V and Ptitsyn O 1996 *Nature* **379** 96
[99] Mirny L A, Abkevich V I and Shakhnovich E I 1998 *Proc. Natl Acad. Sci. USA* **95** 4976

[100] Vendruscolo M, Dobson C M, Paci E and Karplus M 2001 *Nature* **409** 641
[101] Holm L and Sander C 1997 *Proteins* **28** 72
[102] Fersht A 1999 *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*
      (New York: Freeman)
[103] Riddle D S, Santiago J V, BrayHall S T, Doshi N, Grantcharova V P, Yi Q and Baker D 1997 *Nat. Struct.
      Biol.* **4** 805
[104] Ding F, Dokholyan N V, Buldyrev S V, Stanley H E and Shakhnovich E I 2002 *Biophys. J.* **83** 3525
[105] Banavar J R, Hoang T X, Maritan A, Seno F and Trovato A, unpublished
[106] von Mering C, Krause R, Snel B, Cornell M, Oliver S G, Fields S and Bork P 2002 *Nature* **417** 399
[107] Thornton J M, Todd A E, Milburn D, Borkakoti N and Orengo C A 2000 *Nat. Struct. Biol.* **7** 991
[108] Mayr E 2004 *What Makes Biology Unique?* (Cambridge: Cambridge University Press)
[109] Kim E and Chan M H W 2004 *Nature* **427** 225
[110] Kim E and Chan M H W 2004 *Science* **305** 1941
[111] Bergeron D E, Roach P J, Castleman J A W, Jones N O and Khanna S N 2005 *Science* **307** 231
[112] Bernal J D 1939 *Nature* **143** 663
[113] Flory P J 1969 *Statistical Mechanics of Chain Molecules* (New York: Wiley)
[114] Henderson L J 1913 *Fitness of the Environment: An Inquiry into the Biological Significance of the Properties
      of Matter* (Basingstoke Hampshire: Macmillan)
[115] Darwin C 1869 *On the Origin of Species by Means of Natural Selection* (Cambridge: Harvard University
      Press)
[116] Wilson K G 1983 *Rev. Mod. Phys.* **55** 583