Chapter 4
Universality-Diversity Paradigm: Music, Materiomics, and Category Theory

Abstract The transition from the material structure to function, or from nanoscale components to the macroscale system, is a challenging proposition. Recognizing how Nature accomplishes such a feat—through universal structural elements, relatively weak building blocks, and self-assembly—is only part of the solution. The complexity bestowed by hierarchical multi-scale structures is not only found in biological materials and systems—it arises naturally within other fields such as music or language, with starkly different functions. If we wish to exploit understanding of the structure of music as it relates to materials, we need to define the relevant properties and functional relations in an abstract sense. One approach may lie in category theory, presented here in the form of ontology logs (ologs), that can transcend the traditional definitions of materials, music, or language, in a consistent and mathematically robust manner.

> *A sentence should contain no unnecessary words, a paragraph no unnecessary sentences, for the same reason that a drawing should have no unnecessary lines and a machine no unnecessary parts...*
>
> *William Strunk, Jr. and E.B. White, The Elements of Style (1919)*

4.1 Introduction

To solve society's most pressing problems, including medical, energy, and environmental challenges, we will need *transformative*, rather than *evolutionary*, approaches. Many of these depend on finding materials with properties that are substantially improved over existing candidates, and we are increasingly turning to complex materials. As discussed in the previous chapter, biological materials and systems present many challenges before we can "unlock" the secrets of Nature. The bottom-up approach of genetic determinism has proven insufficient to encompass the complexity of biological materials [1, 2]. At the level of coded information in DNA—of replication, inheritance, and decoding of DNA messages—the theory of the gene is elegant in its simplicity, accurately captured by:

$$\mathsf{DNA} \Rightarrow \mathsf{RNA} \Rightarrow \mathsf{Protein(s)}$$

S.W. Cranford, M.J. Buehler, *Biomateriomics*, Springer Series in Materials Science 165, DOI 10.1007/978-94-007-1611-7_4,

At this level the theory is complete or nearly so. Indeed, from this paradigm, we have successfully been able to "recreate" biological materials from proteins (such as recombinant silks [3, 4], for example), but have yet to replicate their performance or properties. We do not fully understand the following process:

$$\text{DNA} \Rightarrow \text{RNA} \Rightarrow \text{Protein(s)} \Rightarrow \text{Function(s)}$$

This "break in the chain"—the inability to predict function from DNA—has partially motivated the paradigm of *systems biology*. Systems biology can be thought of as a field that focuses on complex interactions in biological systems, through a perspective of holism (*e.g.*, system based) rather than reductionism, the modeling and discovery of emergent properties.[1] A reductionist approach can successfully identify basic components and interactions but, unfortunately, offers no convincing concepts or methods to understand the mechanisms by which system properties emerge. We encounter complex hierarchies, multiscale mechanisms, folded and entangled structures, entropic behaviors, and environmental sensitivity, all coupled with a living, growing organism. Yet, these material perform much better than engineered analogues, exhibiting adaptability, toughness, flaw tolerance, self-healing, and energy efficient production. By what means can we decode such complex materials? In other words, how can we explicitly *define the materiome*?

While biological materials are composed of complex hierarchical structures, the molecular basis is comprised of a restricted set of amino acids—simple nanoscale building blocks. In this chapter, we turn to *music* as a "model system", to teach us how complex and intricate compositions can be created (*e.g.*, composed) from simple building blocks (*i.e.*, sound). Like material characterized by mechanical, thermal, electrical, and optical properties across scale, music can be described by categories such as notes and scales, timbre and pitch, tones and instruments, tempo and style (see Fig. 4.1). We demonstrate how analogies between seemingly different fields enable the understanding of general principles by which functional properties in hierarchical systems emerge, similar to an analogy learning process. The mathematical field of *category theory* [5] provides a necessary abstraction to integrate our understanding of complex materials. Biological science has already introduced the concepts of phenotypes (*i.e.*, observed properties) and genotypes (*i.e.*, hereditary information), similar to the morphisms and structures central to category theory. The extension to materials, mechanical properties, and most importantly functionality, is a consequent extension.

Specifically, natural hierarchical materials (such as spider silk) exhibit properties comparable to other constructs (such as classical music) in terms of their hierarchical structure and function. In this chapter, we demonstrate category theoretic analysis as an abstract means to describe emergence of functionality from first principles, *e.g.*, on the basis of fundamental interactions between universal building blocks. As a comparative tool, we discuss hierarchical ontology logs (also referred to as

---

[1] Systems biology can be considered an umbrella term, encompassing many of the *-omic* fields previously discussed such as genomics, phenomics, proteomics, interactomics, etc. Indeed, "materiomics" itself is concerned with the material *system*.

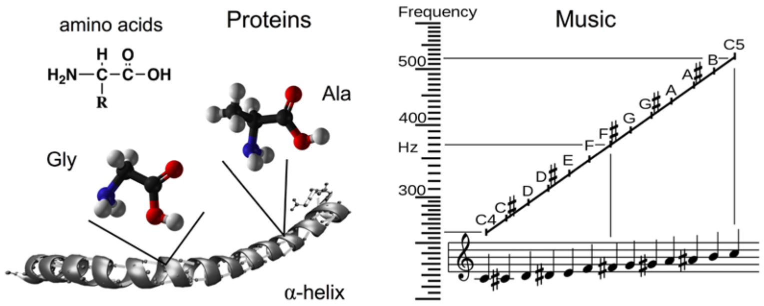

Fig. 4.1 How are proteins and music similar? Proteins are composed of a sequence of amino acids—general molecules containing an amine group (–NH_2), a carboxylic acid group (–COOH) and a side-chain (here, designated as R). Variation in the side-chain results in specific amino acids such as glycine (Gly) or alanine (Ala), which form a sequence that dictate the structure and function of larger protein biomolecules (*e.g.*, an α-helix). Music is also composed of a sequence—a sequence of notes, rather than amino acids. Each note is physically related to a frequency (*e.g.*, sound). Variation in the frequency results in specific notes such as *D* or *A*♯. A sequence of notes create a riff or melody. Can we use music to understand complex biomaterials, to design them, and can we create musical expressions of material mechanisms to gain insight into how materials work?

ologs) [6, 7] that follow a rigorous mathematical formulation based on category theory [5] to provide an insightful system representation by expressing knowledge in a conceptual map. We explain the process of analogy creation, draw connections at several levels of hierarchy and identify similar patterns that govern the structure of the hierarchical systems and discuss the impact of the derived analogy for nanotechnology. The construction of simple analogies and their accomplishment to collocate a broad picture of materials structure and function can be achieved. Through such representations, similar patterns between systems which trace back to their hierarchical build-up are identified. The broad aim of such constructions is the institution of category theoretic tools as a rigorous and comprehensive means to systematically depict and communicate hierarchical structure-function relationships for applications across fields in engineering, science and the arts—from *music* to *materiomics*.

4.2 Universality-Diversity Paradigm

One of the most striking aspects of physics is the simplicity of its laws. Maxwell's equations, Schrödinger's equation, and Hamiltonian mechanics can each be expressed in a few lines [8, 9]. Such simplicity motivates the continuing search for the proverbial "Universe on a T-Shirt", wherein the fundamental equations of Nature can be expressed on a single swath of poly/cotton blend [10]. Everything is simple and neat, except, of course, for biology. Biology is "messy" because it has many different working parts, yet derived from a few fundamental components [11]. Can we use these "limited building blocks" to formulate simple and neat laws? A vision shared by most researchers in complex systems—whether biological, economical, or sociological—is that certain intrinsic, perhaps even universal, features capture fundamental aspects of complexity in a manner that transcends specific domains [12]. The challenge, of course, is choosing the appropriate perspective. For example, most undergraduate students are introduced the concept of polar coordinates (*e.g.*, r, θ) as a means to simplify the analysis of certain problems (such as the equation of motion of a pendulum); using Cartesian coordinates (*e.g.*, x, y) complicates the analysis. How we *choose* to formulate the problem affects the *solution*.

As mentioned in Chap. 3: The Challenges of Biological Materials, the evolution of protein materials through genetic selection and structural alterations has resulted in a specific set of successful (but not necessarily optimized) protein building blocks that define their structure. While protein materials exist in an abundant variety, the need exists to formulate a widely applicable model to systematically *categorize* all such materials, to reduce the seemingly complexity of biological materials [13]. The approach of utilizing universal building blocks to create diverse multifunctional hierarchical structures has been successfully applied in current macroscale engineering paradigms. Advanced technologies and biology have extremely different physical implementations, but they are far more alike in systems-level organization than is widely appreciated [13].

The *universality-diversity paradigm* (UPD) incorporates the recognition and analysis of biological materials based on the universality and diversity of its fundamental structural elements and functional mechanisms [14, 15]. For example, proteins constitute the elementary building blocks of a vast variety of biological materials such as cells, spider silk or bone, where they create extremely robust, multi-functional materials by self-organization of structures over many length- and time scales, from nano to macro. Examples of such universal building blocks include α-helices, β-sheets or tropocollagen (triple-helical) molecules. In contrast, other features are highly specific to tissue types, such as particular filament assemblies, β-sheet nanocrystals in spider silk or tendon fascicles. Similarly, cellulosic materials [16], such as wood, grasses, and other green plants, exhibit a wide array of macro-scale mechanical properties dependent on the fiber morphology and structure [17, 18], yet are composed of similar molecular building blocks (various polysaccharides in both crystalline and amorphous phases [19]). Some of the structural features of both protein-based and cellulosic materials are commonly found in a many different systems, that is, they are highly conserved and constructed in similar structural hierarchies. It is apparent that using only a limited number of components, Nature has produced a broad range of materials with diverse properties and biological functions, and created multifunctionality (*diversity*) by changing structural arrangements of few (*universal*) constituents rather than inventing new building blocks. This paradigm is a paradox: *How can a structure be universal and diverse at the same time*? In protein materials, the coexistence of universality and diversity is enabled by utilizing hierarchies of structure, which serve as an expansion of the design space.

Table 4.1 Intermediate filaments are remarkable due to their diverse appearance in organisms, where they fulfill multiple functions at different hierarchical levels

IF	Found in	Protein level	Filament level	System level
Vimentin [20, 21]	Cell's cytoskeleton	cell signaling mechanisms, associated protein organization	responsible for location, shape and stability of cell organelles, protein targeting processes	"security belt" of the cell
Keratin [22, 23]	Hair, nails, hoofs	protein synthesis, cell signaling mechanisms, associated protein organization	cell pigmentation, organization of cell organelles	cell growth, wound healing, locomotion, prey procurement
Lamin [24, 25]	Nuclear envelope	signaling mechanisms, mechanotransduction, chromatin positioning	gene regulation and transcription, chromatin positioning	protection of the chromatin, involved in cell mitosis

Universality-Diversity Paradigm (UDP): The analysis of materials systems based on the recognition of the *universality* of structural elements (building blocks) and potential *diversity* of fundamental functional mechanisms and material behavior. Through the UDP it is possible to improve our understanding of how complex systems are capable of robustly unifying seemingly incompatible features at different hierarchical scales.

The UDP can be exemplified by a particular class of proteins called intermediate filaments (IFs). Intermediate filaments form protein networks in the cytoskeleton of eukaryotic cell, stabilize the nuclear envelope and provide the basis for extra-cellular tissues such as hair or nails. Interestingly, the elementary building block of all kind of intermediate filaments is identical—a universal α-helical coiled-coil motif. The three intermediate filaments of interest are described in Table 4.1.

The lowest level of hierarchy encodes the structure of these proteins in the sequence of amino acids. This is reflected by the fact that each intermediate filament type has a distinct amino acid sequence (diversity). Intriguingly, the differences at the lowest hierarchy do not influence the immediately following hierarchical level. This can be verified since all intermediate filaments feature the α-helical motif, despite differences at the amino acid sequence level and/or differences at larger scales (universality). However, moderate effects can be observed at the dimer level (two biomolecules). In the dimer structure subtle differences emerge. For example, amino acid inserts in the periodic heptad repeat lead to a local uncoiling of the super helix (creating a "stutter"—a distinct region where the amino acid sequence differs from other parts of the α-helical domains), which effects the assembly process as well as the unfolding mechanics [26–28]. Another example is the occurrence of mutations in desmin intermediate filament coiled-coils. It was shown that disease related mutations do not destroy the α-helix structure but build additional stutters or stammers in

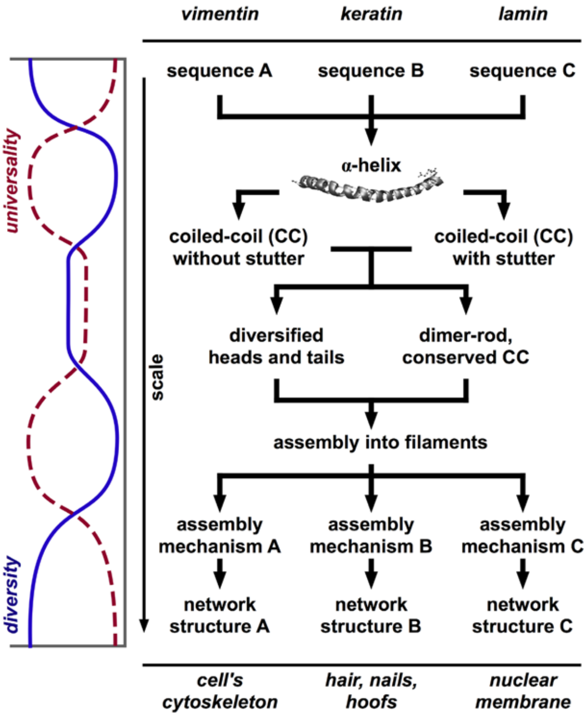

Fig. 4.2 Hierarchical biological materials, here exemplified for the example of three types of intermediate filaments, are governed through interplay of universal and diverse patterns, which, combined with silencing and activation are unified over multiple hierarchical scales. This enables to forward information that is completely coded at the lowest scale (amino acid sequence), safely by means of silencing through intermediate scales (alpha helix, coiled-coil) up to higher scales, where they are activated in order to fulfill specific requirements. There is a balancing between *universal* features (*e.g.*, α-helical structures, filament assembly) and *diverse* functionality. The *plot on the left* shows the balancing exchange between universality and diversity across scales. Adapted from [14]

the coiled-coil [29, 30]. Even though all types of intermediate filaments commonly show an assembly into filaments, lower scale differences (that is, for instance the amino acid sequence and presence of the stutter) affect the pattern and process of assembly, such as the number of proteins per filament cross-sectional area, or the way dimers associate. The differences on the filament level are of utmost importance, as they influence the properties at the network and the super-structural level, which are dominated by (but not limited to) mechanical functions. The cross-scale universality-diversity of intermediate filaments is depicted in Fig. 4.2.

As shown in this example, nanoscopic modifications do not always influence the properties at the next hierarchical layer, but potentially those of one or more hierarchical layers above. It appears as if specific functional requirements at several higher scales are "forwarded" to lower scales, where modifications are implemented. Through this mechanism biological materials are not only multi-functional but are further continuously adapted to the required scale-specific processes, with the goal to fit the diverse required functions in the best possible way.

This coexistence of universality and diversity as described in the universality-diversity paradigm, is an overarching feature in biological materials and a crucial component of materiomics. Nature has a unique capacity of creating toughness from weak components, capable of balancing multiple, seemingly incompatible properties such as strength, robustness, and adaptability. Materials like bone, being a nano-composite of strong but brittle and soft but ductile materials, illustrate this unification of components with disparate properties within a hierarchical structure. Primarily, the utilization of structural hierarchies enables the coexistence of universality and diversity. Through the UDP it is possible to improve our understanding of how materials are capable of robustly unifying seemingly incompatible features at different hierarchical scales, a fundamental keystone to materiomics. Indeed, material characteristics such as strength and robustness are contradicting properties that cannot be easily combined within a single scale of traditional materials science, requiring a materiomics approach to reconcile such disparate concepts. Further, such a perspective clearly indicates that in structures and processes are amalgamated and cannot be considered alone. Although wide and varying, biology commonly exhibits unlikely harmony within material structures and physiologic functionality.

4.3 Tu(r)ning Weakness into Strength

The importance of hierarchies is elusive for many material systems. For the sake of argument, justification of a continued focus on such material hierarchies is prudent. Indeed, one may ask why such complex and redundant hierarchies are even necessary? The simple answer is that hierarchical structures are not required under ideal circumstances (*e.g.*, unlimited resources, both material and energetic), but serve to *extend* the physical design space while limited to a restricted set of constituent building blocks. In other terms, it provides material scientists and engineers more design parameters to manipulate within the same set of building blocks (*i.e.*, structural elements) to realize multiple functions, which together create a highly functional system that is much more than the sum of its parts. Nature implements this trick frequently. A major trait of biological materials is the occurrence of not only hierarchies, but also hierarchies that: (a) contain weak interactions (such as H-bonds in spider silk); (b) exploit abundantly available materials (*e.g.*, silica in some sea creatures), or; (c) implement structurally inferior materials (*e.g.*, extremely brittle crystals in mineralized tissues like nacre or bone). Yet such materials are commonly more robust and tough than those currently designed and engineered. The paradox of a “strong” material being composed of “weak” components is difficult to theoretically reconcile. The entire materiome across multiple scales must be introduced, where universal and diverse patterns are unified hierarchically, and the integrated view of it results in a quantitative understanding of how highly functional materials are created based on inferior building blocks.

Despite this apparent complexity of functional properties, the structural designs of biological materials have evolved under extreme evolutionary pressures to ensure a species' survival, often in adverse environments. As a result, materials are created with the lowest possible energy consumption, under simple processing conditions, and are exquisite as they often form from abundant material constituents. Notably, these abundant material constituents typically represent functionally inferior material building blocks that are exceptionally brittle, such as silica or other minerals, or extremely weak, such as H-bonding or Van der Waals forces. The integration of weak building blocks is not a design *choice*, but rather a natural evolutionary constraint. Moreover, comparative studies show that most biological materials are made up from only a few select universal elements (*e.g.*, C, H, O, N, S…), despite their functional diversity [31]. This is exemplified at different levels, such as in the occurrence of only twenty natural amino acids (which provides the basis for all known protein materials), and a few universal secondary structures such as α-helices, β-sheets or random coils, or more fundamentally, the four DNA or RNA nucleotides that are universal to encode the structure of all protein materials at the genetic level. The definition of these "fundamental building blocks" is also a matter of perspective. In terms biological systems, the primary building blocks of proteins are commonly considered the amino acids (which constitute a unique sequence). However, these acids are also composed of side-chains (*e.g.*, amino acid = amine + carboxylic acid + side-chain), which can also be deconstructed to a few elements and atoms (illustrated by Fig. 4.3). Even the atoms themselves can be considered "hierarchies" of protons, neutrons, and electrons. Down the rabbit hole we go! Practical application requires a "truncation of decomposition", which is dictated by a system under analysis. For certain systems, the behavior of neutrinos and other elementary subatomic particles may be required.

The realization of how biology perpetually uses a limited number of elements to create highly diverse systems poses a fundamental question about the design of biological materials, and specifically, how such a great diversity of functional properties is achieved, despite the apparent inferiority and universality of the building blocks [15, 32]. Recent research suggests that the basis of understanding the remarkable properties of these materials lies within the biological design paradigm where multifunctionality is created not through the use of high quality, or use of a large number of, distinct building blocks. Rather, functionality is created by compiling simple and often inferior elements into assemblies where structures are created at multiple length-scales, resulting in hierarchical material architectures.

It is also within this self-assembly process that the inherent "weakness" of biological materials is a "strength". The flexibility of proteins facilitate folds, coiling, geometric manipulation and compatibility with other proteins. Weak interactions such as H-bonding, ionic cross-linking, and polarization enables the formation of structural units (*e.g.*, dimers and larger quaternary protein structures) with minimal energetic and material expense. One cannot adequately bond carbon nanotubes to a copper substrate without significant chemical and structural modifications, but it is possible to link two proteins to exploit functional groups and assembly mechanisms [33–35] as depicted in Fig. 4.4. Such approaches are further discussed in Chap. 10: Synthesis and Design.

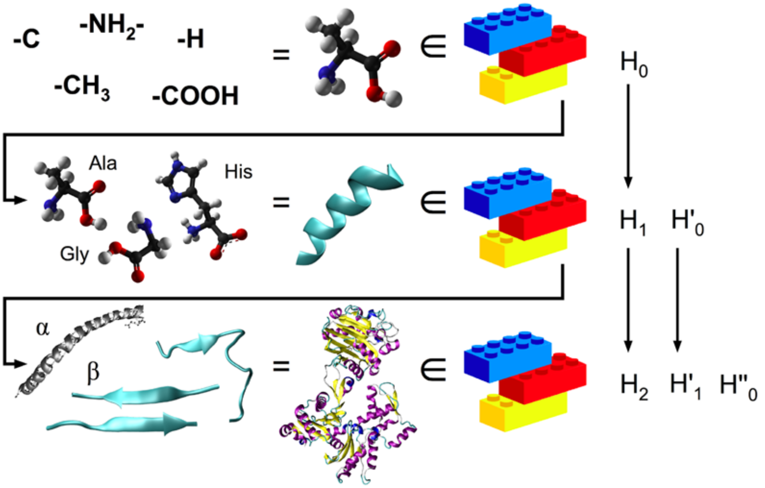

Fig. 4.3 Building blocks built by building blocks? Hierarchical systems composed of structural units are subject to functional perspective. Here, we assume the molecular groups (amine, carboxylic acid, and a side chain, such as the a methyl group depicted) are the fundamental irreducible building blocks of amino acids (such as alanine), and the set of amino acids provide the building blocks at the primary hierarchy (H_0), defined as the first structural unit. The set of amino acids (glycine, alanine, histidine, etc.) can form sequences with characteristic structures (such as α-helix), which compose the next hierarchy (H_1). Finally, many secondary structures (α-helices, β-sheets, etc.) can form a folded protein (here, actin) at the highest hierarchy considered (H_3). At higher scales, proteins themselves form the "set" of building blocks (actin is one of the three major components of the cytoskeleton). Depending on application, amino acids can be considered irreducible building blocks, and thus the secondary structures provide the primary hierarchy (H_0') followed by the protein structures (H_1'). Similarly, from a cellular sense, the secondary structures may be considered irreducible, and thus the proteins provide the first structural unit (H_0''). Definition of hierarchies and building blocks is dependent on application and the function of each block. Note too that the characterization of secondary structures is a choice—as the amino acid sequence can be considered the building blocks of entire proteins, without the intermediate definition of secondary structures

To fully exploit the "pliability" of proteins as structural building blocks, we must fully understand their potential interactions. Normal biological function requires that a systems's biochemical network be highly interconnected and interactive [36]. The reductionist approach employed by recent biologists is effective in defining the molecular structure and organization of DNA and proteins, but less effective at delineating function, because most biological molecules do not operate independently [37]—they must be considered in the context of all of the other proteins in the cell and their organization [38]. To reconcile this network complexity, the concept of "modular cell biology" was proposed by Hartwell *et al.* [37], grouping interacting molecules, including DNA, RNA, proteins, and smaller molecules, into cellular modules that represent discrete functional entities. Depending on the scale of interest, these modules can be considered the irreducible functional "building blocks", or defined by their finer scale constituents.

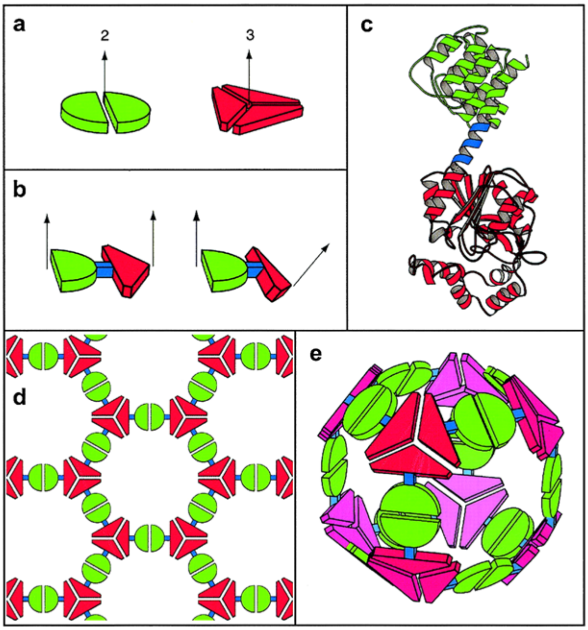

Fig. 4.4 "Weak" materials can self-assemble: a general strategy for designing proteins that form predictable nanostructures. (**a**) The semicircles represent a natural dimeric protein (*i.e.*, a protein that associates with one other copy of itself), whereas the triangle shape represents a trimeric protein. The symmetry axes of the natural oligomers are shown. (**b**) The two natural proteins are combined by genetic methods into a single protein. Each of the original natural proteins serves as an "oligomerization domain" in the designed fusion protein, with known interaction properties (*i.e.*, function). Two different hypothetical fusion proteins are shown to illustrate that the oligomerization domains can be joined rigidly in different geometries. (**c**) A ribbon diagram of a fusion protein showing one method for joining two oligomerization domains in a relatively rigid fashion. The two are linked by a short stretch of amino acids and the two oligomerization domains are joined physically in a predictable orientation. (**d**) A designed fusion protein self assembles into a particular kind of nanostructure that depends on the geometry of the symmetry axes belonging to its component oligomerization domains. A molecular layer arises from an arrangement like that in (**b**) (*left*). (**e**) A cubic cage arises from an arrangement like the one in (**b**) (*right*). From [33], used with permission, copyright © 2000 National Academy of Sciences

Certain properties of biological modules are recognizable by engineers, such as positive and negative feedback loops, and parallel (redundant) circuits [37]. To exploit a modular building block understanding of cellular function it will be necessary for biologists to understand design principles more familiar to engineers and computer scientists, such as amplifiers, oscillators, and logic circuits [39]. Instead of having to include many of the molecular details of protein-protein or cellular interactions, the functional behavior each module (or system "motif") can be described using mathematical models that have a certain degree of universality. In many cases the models do not require plunging into the details of how every protein works—they only need to include information on whether X activates or inhibits Y, and at what concentration (and perhaps a few additional parameters). Such models seem to capture the essential dynamics of protein "circuits", while being, in a sense, insulated from most of the complexity of the proteins themselves [11].

Thereby, delineated levels of resolution (*e.g.*, network module or material building block) provides access to specific *functional* properties, which can be achieved by defining a particular material structure at each hierarchical scale. This paradigm, the formation of distinct structures at multiple length scales, also enables biological materials to overcome the intrinsic weaknesses of the building blocks. This can be illustrated in the use of nanoconfinement that often results in enhanced strength and ductility despite the intrinsic brittleness of the same material in bulk form [40]. In an example relevant for sea creatures such as diatom algae, while silicon and silica is extremely brittle in bulk, the formation of nanostructures results in great ductility and extensibility, where the specific geometry used allows for a continuum of mechanical signatures. Thus, the realization of distinct structural designs provides a means to tune the material to achieve a great diversity of functional properties despite the use of the same building blocks. More generally, if it is possible to alter the material's structure at specific hierarchy levels, independently and during operation of a material, there exists the potential to realize varied material properties depending on functional needs. This is exemplified in echinoderms (*e.g.*, sea urchins and sea cucumbers) that can change their body's modulus manifold through alterations of the cross-linking of constituting collagen molecules. Mutability also occurs in plants that track the direction of sunlight, through a mechanism that involves a change of the plant wall's stiffness exposed to light and resulting in bending towards the softer part (*via* a hypothesized breaking and reforming of H-bonds).

A powerful example to demonstrate the biological material design paradigm is the case of spider silk (discussed in detail in Chap. 8: Unlocking Nature: Case Studies), a remarkable material that-in order to fulfill its biological tasks-must provide extreme levels of strength (1–2 GPa), toughness, and great deformability of up to 50 % tensile strain [41]. Yet, the structural basis of spider silk is extremely simple, and merely consists of a few distinct few amino acids, arranged in long polypeptide chains and that interact only by weak interactions. Moreover, spiders must be able to produce silk quickly, out of a limited stock of solvated protein [42]. The physiologic processing conditions in solvent, at room temperature, and at very short time-scales rules out enzymatic processes and suggests that self-assembly is the primary mechanism to form spider silk under these conditions. This necessitates the use of weak bonding in the fundamental interactions of the silk protein strands. Indeed, spider silk is known to be dominated by H-bonds, one of the weakest chemical bonds known, and also present in liquid water (for example, H-bonds have an energy of $\sim$5 kcal/mol whereas the covalent bond of a H_2 molecule has an energy of $\sim$100 kcal/mol). But how is it possible to generate such a mechanically superior material out of clearly inferior constituents and constraints?

The answer is that weak elements in the material, here H-bonds, are arranged geometrically in order to provide maximum strength and toughness (see Fig. 4.5). It has been found that the intrinsic weakness of H-bonds vanishes when grouped into clusters of $\sim$4 H-bonds, which allows them to work cooperatively and thus reach

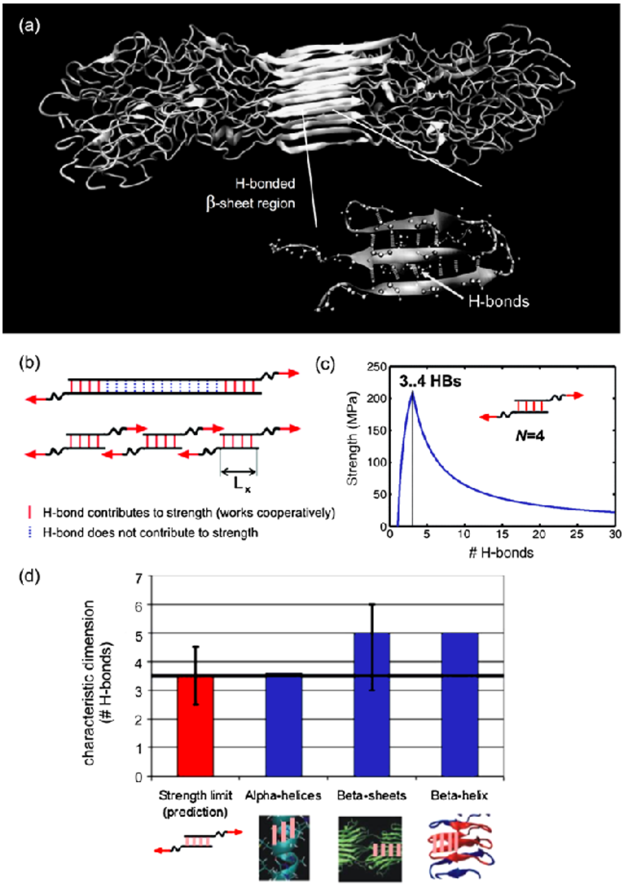

Fig. 4.5 Clustering of weak H-bonds enhances cooperativity. (**a**) The molecular structure of dragline silk is characterized by a crystalline region of H-bonded β-sheets [43]—a result of a regular repetition of alanine (poly-Ala) within the silk protein sequence. For natural silk, these poly-Ala regions are typified by six residues resulting in a finite cluster of $\approx$5 H-bonds. *Why not more*? The rupture strength of H-bonds is governed by their cooperativity *via* geometric confinement. (**b**) Schematic illustrating the difference of the strength of a single, long beta strand *vs* a combination of multiple small strands. In the former, only H-bonds at the boundary participate in the rupture process and provide resistance. In the latter, all H-bonds throughout the entire structure contribute to the strength, making the overall structure three times stronger. (**c**) Size effects of the shear strength of β-sheets. The *plot* depicts the shear strength of β-sheets as a function of N (*e.g.*, number of H-bonds). The highest shear resistance is found at a characteristic length scale of $\approx$3 to 4 H-bonds. Beyond this length scale, the shear strength drops rapidly (defined as the strength of the β-strand divided by the sheared area). (**d**) Characteristic dimensions of the size of H-bond clusters in common protein structures: α-helices, β-sheets and β-helices. Since the theoretical derivation of H-bond cooperativity considers uniform deformation of hydrogen bonds with no particular specificity to geometry, it may also apply to other protein structures where nature utilizes geometric confinement to achieve higher mechanical stability. The fact that 3 to 4 H-bonds per convolution exists on α-helices and β-sheets on the sides of helices occur in clusters of approximately four may be indicative of such a universal biological concept that may be based on the evolutionary driving force to provide maximum strength. Adapted from [44–46]

maximum strength [44, 47]. This induced cooperativity is similar to the contact-splitting mechanism that enhances the adhesion of a gecko's foot—dividing the load among many small interfaces maximizes strength, as opposed to a single, inefficient transfer [48–50]. This phenomenon can be understood by applying basic thermodynamics concepts from fracture mechanics to the protein unfolding problem [45], and explain experimental findings such as DNA unzipping studies or multidirectional loading of complex β-sheet structures.

Notably, the ability of H-bonds to work cooperatively is also critical to ensure enhanced robustness, where the loss of a single bond does not result in the breakdown of an entire system. Cooperativity is actually facilitated by the weakness of H-bonds, which implies a softness of bonding that endows them with the freedom (entropy) to explore a great variety of structural states such that they can most effectively resist deformation [44]. The assembly of H-bond clusters into geometrically confined β-sheet nanocrystals, consisting of a pancake-like stack, results in the structural basis for effective cross-linking of multiple polypeptide chains in silk [47]. Because H-bonds can be reformed easily upon breaking, β-sheet nanocrystals have another highly useful property, toughness, enabled by the ability of H-bonds to self-heal and thereby effectively preventing catastrophic brittle failure as often observed in materials with stronger bonding. Yet, due to the nature of the available building blocks it is not possible to achieve all desired properties at a single material scale. Thus, in order to achieve another functional property of silk, extensibility (while maintaining the great strength facilitated by β-sheet nanocrystals), the structural design is extended to higher structural scales, and specifically, at the next level through the formation of a nanocomposite achieved by using the same basic material building block (polypeptide) but arranged in a different geometry. Here strong and tough β-sheet nanocrystals are combined with an additional protein secondary structure that consists predominantly of so-called 3_1-helices that realize a more disordered phase. Through the provision of extreme amounts of hidden length, this constituent provides the capacity to generate large levels of deformability before the β-sheet nanocrystal cross-links are deformed and eventually broken [43]. By tuning the relative ratio of the two phases in silk, or the geometric makeup of the β-sheet nanocrystals, it is possible to achieve a diversity of mechanical signatures (*e.g.*, stiff, soft, extensible, etc.) without a need to define new constituents, a mechanism spiders use to generate different types of silks [41, 51].

The discussion of some of these most salient design features of spider silk strikingly provides an insight into a much broader design paradigm in biology at the nanoscale, pertaining to biology's use of a universal "cement"—H-bonds—in the creation of mechanically stable materials. The formation of confined clusters of H-bonds is indeed observed widely in biology and not only in silk, pointing to a universal design paradigm that enables biological systems to overcome the intrinsic weakness of H-bonds and to form mechanically strong and tough materials. A comparison of the geometric size of H-bond clusters in a diversity of protein found in the Protein Data Bank (http://www.rcsb.org) confirms that H-bonds typically organize in clusters between 4 to 6 in β-sheets and separated by disordered or differently structured elements, or in clusters of 3 to 4 in the turns of α-helices

[44] (see Fig. 4.5(d)). These H-bond clusters represent a geometric feature found across species and highly conserved in biology. Such structures provides simple, yet strikingly effective protocols to achieve mechanical strength out of weak chemical bonding.

There are important insights from the universality of geometrically confined hierarchical structures in protein materials [31]—effectively turning (or tuning) weakness into strength. The application of this concept can impact technological use of abundant, however mechanically inferior (*e.g.*, extremely brittle, or weak) materials such as H-bonded proteins or silica in the design of strong, tough and lightweight materials. The design of hierarchical structures could be the key to overcome their intrinsic weakness or brittleness, properties that currently prevent their widespread technological application. Broadly, this may result in the development of highly functional materials made of abundant, cheap and environmentally friendly constituents, and even materials that are currently considered waste. It suggests that there may exist an enormous technological opportunity that is based on the concept that functional material properties can be design based on almost arbitrary constituents, provided that multiple structural levels are created. An observation made based on the analysis of different materials suggests that universality tends to dominate at smaller scales (*e.g.*, Ångstrom and nanometer), whereas diversity is found predominantly at larger, functional scales.

4.4 Music and Nature: Complexity from Common Elements

It is clear that the synergistic interaction of structures and mechanisms at multiple scales, using multiple levels of structure, provides the basis for enhanced functionality of biological materials despite the reliance on few distinct building blocks. The biological paradigm to create diversity of function out of simple, universal elements is not unique to Nature, and can be explained by drawing an analogy to rather far and different fields, from sociology to economics. Here, the importance of multiscale interactions, hierarchical structuring, and multifunctionality is illustrated by using an analogy of *music* [7, 32, 52]. Music, akin to protein materials, is founded by a common basis that can be explained by simple physics: *sound*. Like the combination of elemental building blocks of carbon, hydrogen, and oxygen that constitute protein materials, the phenomenon of music is fundamentally the combination of traveling sound waves with different frequencies (middle C, for example, has a frequency of 261.626 Hz). Such sound waves, however, fail to encapsulate Beethoven's *Symphony No. 9*, just as a listing of atoms fails to convey the function of a protein.

Musical composition is built upon universal elements at the microscale such as basic wave forms, and gathers a small variety of available instruments into hierarchical assemblies to create macroscale functionality, such as a particular orchestral sound (*e.g.*, a symphony). In music, when one considers the synthesis of orchestral music based on universal wave forms, structures at multiple scales are similarly used to arrive at a functional system, which is the resultant assembly of multiple scales—for example in a symphony. The concept is schematically illustrated in Fig. 4.6,

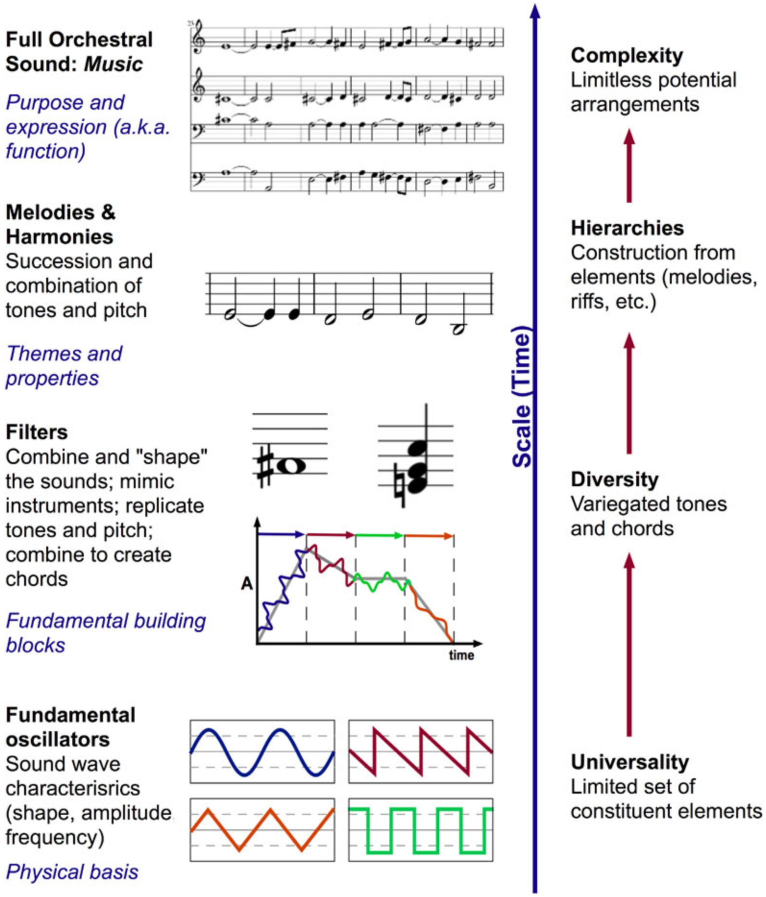

Fig. 4.6 Music as a complex functional hierarchy. Illustration of multiscale (or cross-scale) interactions in the case of music, here exemplified for the process of synthesizing orchestra-level music based on four fundamental oscillators (a simplistic model). At a fundamental level, four basic oscillators create unique sound waves characterized by physical properties (*i.e.*, shape, amplitude, and frequency). At the next level, these basic sounds are modulated using envelope generators or filters, which shape and mimic the sound of an instrument and construct various tones and pitches (the depicted filter is a schematic representation of an attack-decay-sustain-release envelope). At this level, the fundamental building blocks of music are developed beyond the simple sound waves from which they are composed. The assembly of tones with different duration and pitch over time creates melodies (sometimes referred to as theme or riffs); where all tones come from a universal, limited set of harmonics (assembled in octaves). The theme or type of music is typically dependent on the selection and construction of themes and riffs (*i.e.*, the difference between contemporary jazz and classical baroque, for example), which can be thought of as the musical "properties". Through the combination of multiple instruments, each of which plays characteristic melodies or riffs, a complex orchestral sound is produced at the largest level, fulfilling the intended purpose, or musical "function" that emerges as the hierarchical structures interact synergically across the scales. The example also illustrates how the interplay of diversity and universality provides the fundamental paradigm behind music, resulting in near limitless arrangements from the hierarchical construction of musical elements (figure adapted from [32])

where at a fundamental level, four basic oscillators (chosen here as a fundamental set of constituents) create sine, square, and other wave forms considered raw, unalterable input [53]. *Universality tends to dominate*. At the next level these basic sounds are modulated using envelope generators or filters that change the volume, pitch and duration of the waves over time, employing on-the-fly processing (through Fourier analysis, or band-pass filters, for example), which shape the sound of an electronic instrument (note that similar hierarchies can be formulated for more traditional instruments such as woodwinds, guitars, or pianos).

An assembly of these modulated tones with different duration and pitch, or combinations of several of them into chords, creates melodies or riffs, where all pitches used come from a universal and limited set of harmonics, organized into octaves. Through the combination of multiple instruments, each of which plays characteristic melodies or riffs, a complex orchestral sound is produced at the highest structural level, the *de facto* "functional" scale. Indeed, while four sound waves differ only minimally in terms of physical properties (*e.g.*, frequency, shape, amplitude), the potential for a great diversity of arrangements is vast and manifest at the ultimate functional levels. By analogy, sound waves can represent elements, tones can reflect amino acids, protein sequences embody the melody, and their combination can provide the "music" of protein based structures. Nature has indeed proven to be an adept composer.

The "function" of music, be it the aural aesthetic or emotional expression, is a result of multiscale phenomenon of resonance and dissonance, the creation of chords and harmonies, the choice of classical piano or electric guitar; combined in one way the result is Mozart, while in another the result is *The Rolling Stones*. Variation (mutability) can be achieved by changing any of the levels-leading to variations in rhythm, tones, or melody, which in turn provide a different overall musical piece, or "function". The structure of music and protein material (including how it is changed), and the particular observer (*e.g.*, the audience, or the sensing in a particular physiological environment) are inextricably linked. This is a critical issue in both music and protein materials that is due to the way a particular observer processes and interprets functional properties and how this information is used in feedback loops that can alter the structural makeup at various scales. In biology this may happen through changes in gene expression (at relatively short time-scales) and in evolution (at relatively long time-scales). In music, this may happen through alteration of music during composing (which could involve continuous revisions of a piece), or through changes in the way a particular musical piece is played. In a jazz or rock jam session, music is continuously revised during performance by the interaction between the performers or with an audience (as anyone who has attended a *Phish* concert will surely attest). Similarly, the process of composing music can be regarded as an analogy to the evolutionary process.

While the synthesis of complex sounds from the level of basic oscillators is now possible with modern synthesizers-resembling a bottom-up "nanoscale" paradigm in creating music, composers in ancient days were limited by the availability of certain instruments, such as flutes or harps created from bone. Classical composers (*e.g.*, Bach, Mozart, Beethoven, and others) subsequently used more advanced instruments such as the violin or the piano, whose design was enabled by the materials and technology that became available at the time. Despite the limited set of available instruments (the basic building blocks), composers were able to create music that is considered some of most ingenious of all time (perhaps because it required composers to utilize many levels of structure to achieve a great functionality, which is apparently found appealing by the human brain). On the other hand, the impact of novel approaches to synthesize music from the bottom-up, with the use of complex sounds realized in *de novo* instruments, remains unclear, and music with a classical structure remains preferred by many. Subjective and cultural aspects may likely play a central role in the development and experience of music, an aspect that is evident from distinct types that emerged from different geographical regions and cultures. The construction of music exemplifies how the interplay of diversity and universality provides a powerful design paradigm, which relates directly to that found in biological materials and to what kinds of materials can be designed based on synthetic approaches.

The analogy between protein material design and music illustrates how biology is indeed a proficient composer, and that information can be derived to facilitate the development of more sustainable materials that rely not merely on the strength of building blocks but rather, on the use of structures at multiple levels, to provide heightened functionality. Conversely, humans have found ways to deconstruct and express hierarchical structures in various art forms, such as music, poetry or paintings. The similarity between the design of music and that of materials in biology is intriguing and merits further investigation, perhaps with the potential to unravel more general principles that govern mechanisms by which biological systems achieve functionality. Indeed, it was pointed out that functionality in certain animals (bees, ants, birds, etc.) does not rely on the strength of individual elements, but that functionality is achieved based on collective interactions of elements in clusters (*via* swarming, schooling, flocking, etc.) [54]. This concept is akin to the concept discussed here for the cooperative behavior of H-bonds that turns their weakness into strength. A key lesson reiterated is that in order to create a diversity properties, it is not necessary to rely on strong or numerous building blocks, but rather, the design space can be expanded *via* the formation of hierarchical structures, realized in biology through the merger of the concepts of structure and material and in music through the creation of complex compositions inherent in symphonic pieces.

Although an illustrative analogy, a cursory comparison of music with natural materials (and complex synthetic materials) may appear trite at first glance. Indeed, one could pull many examples where "*the whole is greater than the sum of its parts*" from everyday usage—including recipes for cooking or abstract painting to name a few. Undeniably, a dish of *beef bourgogne* is more than just the summation of beef, wine, and spices; Jackson Pollack's *No. 5, 1948* is more than just drizzles of yellow and brown paint. Being said, music, cooking, and artistic expression do not have obvious objective functions or properties like biological materials. Whether one prefers *The Beatles* or *The Rolling Stones*, chicken or beef, a Picasso or a Renoir, is merely a subjective difference of opinion. This need not be the absolute case.

For example, the book, "*A generative theory of tonal music*" proposes a detailed theory of musical hierarchies [55]. The theory develops a grammar of tonal music based in part on the goals, though not the content, of generative linguistics. The grammar takes the form of explicit rules that assign heard structures from musical surfaces. The term "musical surface" is meant, broadly, as the physical signal (sound waves) of a piece when it is played (the components or building blocks), while the "heard structure" refers to all the structure a listener unconsciously infers when he listens to and understands a piece, above and beyond the data of the physical signal (the function or meaning) [56]. Listeners hear certain structures rather than others. How can these structures be characterized and by what principles does the listener arrive at them? One would ultimately hope to specify those cognitive principles, or "universals" that underlie all musical listening, regardless of musical style or acculturation [57, 58]. Such musical theory attempts to produce formal descriptions in a scientific sense. That is, the goal is not just the description of formal relations, but the "functional form" of music. Thus the theory is *predictive* [59, 60]. In addition to criteria of internal coherence and parsimony, its principles can be verified or falsified by comparing the analysis it generates with one's intuitions about particular pieces of music. In addition, many of its principles can be investigated through laboratory experiment. Even the subjective aesthetic of music can be quantified in a self-consistent manner.

4.5 Comprehension by Analogies: Functional Similes and Abstraction

Why even turn to music when our focus is biological materials? Comprehension by analogies is a widely applied concept in science and education [61–64]. When a concept is complex, we tend to simplify our perspectives. Einstein was once quoted as saying "*You do not really understand something unless you can explain it to your grandmother*"—try explaining molecular hierarchies and folded protein structures to Grandma! You may ultimately find yourself describing flexible and folded proteins as cooked spaghetti.

Successful pedagogical strategies comprise features such as "constructivist learning environments" that challenge the view that scientific and mathematical knowledge is static, independent from our minds, and represents a universal truth [65]. In fact, this knowledge serves as a mediator resulting from human inquiry. Children in school become introduced to sciences such as mathematics, physics, chemistry and biology *via* the link to structures and concepts they are more likely to be familiar with. For instance, atom and electron interactions are represented by a model that resembles galaxy structures—the Bohr model—or animal cells are represented by factories [66, 67]. While most people may agree that it does not represent reality even closely, the analogy provides a sufficiently thorough understanding of general mechanisms that take place on the scales of Ångstroms and nanometers. Typically, an abstract comparison can be easily made if the *function* is transferable—white blood cells of your immune system and law enforcement, for example. We can say "white blood cells protect your body like the police" in a kind of *functional simile*, and even a child understands the role of white blood cells (it doesn't matter that our leukocytes are incapable of detective or forensic work).

By systematic abstraction and the deduction of analogical steps, the process of building an analogy itself helps to boost discretion about the important properties and parameters of the inquired system—or at least to ask essential questions [68, 69]. A rigorous methodology to formulate and categorize these analogies can be provided by *ontology logs* (*i.e.*, ologs) based on *category theory* [5, 6]. Category theory is a relatively new branch of mathematics (invented 200 years after the introduction of partial differential equations), designed to connect disparate fields within the larger discipline [5]. It is both a language that captures the essential features of a given subject, and a toolbox of theorems that can be applied quite generally. Category theory originates from a mathematical concept in topology [5] and has recently been used in broader contexts to identify patterns in other fields in a mathematical framework [70–73]. A good overview for non-specialists can be found in Lawvere [74] and Awodey [75].

Category Theory: An area of study in mathematics that examines in an abstract way the properties of particular mathematical concepts, by formalizing them as collections of *objects* and *morphisms* (or *functors*), where these collections satisfy certain basic conditions related to the composition of *morphisms*. The most accessible example of a category is the category of sets, where the objects are sets and the morphisms are functions from one set to another. However it is important to note that the objects of a category need not be sets—any way of formalizing a mathematical concept such that it meets the basic conditions on the behavior is a valid category, and all the results of category theory will apply to it.

Quickly after its inception, category theorists realized that its basic ideas were applicable well beyond the borders of mathematics. Category theory has by now been successfully applied in computer science, linguistics, and physics [71]. Whereas the theory of differential equations can be applied throughout science to create *quantitative* models, category theory can be applied throughout science to create *qualitative* models. And once such a qualitative model is formed as a category, its basic structure can be meaningfully compared (again *via* functors) with that of any other category, be it mathematical, linguistic, or other [76]. The use of category theory allows many intricate and subtle mathematical results or functional relations to be stated, and proved, in a much simpler way than without the use of categories. Like a biological system, the basic building blocks of a category are simple, but the networks that can be formed out of them are as complex as mathematics itself. These building blocks are called *objects*, *arrows*, and *composition*: arrows between objects form paths which can be composed into new arrows. It is a wonder that such a simple system can account for the wide variety of forms found in the mathematical universe, but perhaps this is less of a surprise to a biologist who notices the same phenomenon in his or her field.

Categorical algebras consist of objects and arrows which are closed under composition and satisfy conditions typical of the composition of functions [75]. Indeed, category theory provides a bag of concepts (and theorems about those concepts) that form an abstraction of many concrete concepts in diverse branches of mathematics. In a linguistic version, category theory and ologs in particular describe the essential features of a given subject and represent a powerful method to store and share data, knowledge, and insights in structure and functionality.

Equivalent to the learning and understanding process applied in school physics and mathematics, the methodology and hence the above described advantages of analogical thinking can be adapted to the field of materiomics. Recent studies linked for example mechanical properties of protein networks to communication networks [77], active centers in proteins to top predators and top managers [78], and death of living organisms to the strength of solids [79]. Furthermore, the connection between grammar and protein structure has been elaborately studied [80], even in the context of category theory [81].

4.6 Category Theoretic Analysis: Linking Hierarchical Structure and Meaning

Our understanding of the synthesis of elements into multi-functional structures remains in its infancy, and is currently limited to specific protein networks or protein materials. For example, mechanistic theories are typically developed for specific systems (including bone [82, 83], cells [84], and spider silk [47]) rather than providing a unified model that is applicable to a variety of distinct materials. A major frontier in the field is thus the extraction of generic principles of how functional properties are derived in functionally diverse systems despite the presence of the same (universal) building blocks, solely by using structure as a design paradigm. The systematic characterization of this knowledge is a keystone to materiomics.

Is it possible to show universality across seemingly disparate biological material systems? That is, by the application of the mathematical field of category theory to protein materials, can the extreme diversity of protein functional properties be described in a unified model that contains only a limited number of universal elements and interaction rules? The successful application of category theory to carry out a qualitative analysis in fields such as linguistics (grammar, syntax, semantics, etc., a toolbox of key concepts that enable the understanding of language [80, 85–88]) and computer science (again modeling syntax and semantics of denotation and operation in programming languages [89–92]), provide promising results. Through the development of such an approach, it may be possible to show that the hierarchical combination of universal elements into multi-level structures enables protein materials to achieve context specific functionalities in an abstract "complexity space"—linking hierarchical structure and function (or meaning) in a logical and self-consistent manner. (See also Fig. 9.15, which shows the hierarchical structure of amyloid materials and an analogy to the structure of language.)

Such an analysis would reveal that in order to create highly functional materials it is not essential to rely on a multitude and a certain quality of building blocks (*e.g.*, with superior qualities, great material volume, strong interactions, etc.). Rather, it is sufficient to use simple interaction rules and simple building blocks—each of which does not need to possess superior qualities—but assembled into hierarchical systems, where the overall structure provides enhanced functionality. Moreover, the universal relations composing the olog would be transferable to a multitude of equivalent systems. For example, the fundamental functional relationships for β-sheet interactions is not restricted to spider silk. This insight has implications for our understanding of how nanomaterials could be utilized to create macroscopically functional materials, and suggests a paradigm that departs greatly from the current belief in engineering science that material building blocks with superior qualities at the small scale (*e.g.*, carbon nanotubes, graphene, etc.) are critical to reach high performance materials. The implication is that superior functionality can be reached with *any* fundamental building block (or *set* of building blocks), provided that the design space is expanded to incorporate hierarchical structures. Eventually, an understanding of how diverse functional properties can arise out of inferior building blocks could make a profound impact towards the development of environmentally benign and friendly materials, as it would allow manufacturers to use local, abundant, and simple building blocks with overall negative CO_2 balance (*e.g.*, wood, plants, silica, water, soy beans) to create highly functional materials and structures. But how can we find a proper mathematical description of these hierarchical mechanisms that generate functional properties? Through category theoretic analysis.

As previously discussed, rather than being selected for a specific application, biological materials evolved to perform specific biological functions [1, 31, 93–96]. The components and connections within a given biological material are analogous to a circuit diagram [38, 39]. But just as it is extremely difficult to deduce the circuitry of a device by experimenting with its inputs and outputs, it is similarly inadequate to describe the higher-level structure of a biological material using only the physical interactions between proteins and some information about gene expression. Instead, we need to take into account additional types of structural information given by the fundamental principles that govern the interactions of the building blocks that define the system and its emerging functionality as these building blocks are connected together, from the molecular to the macro scale. The above considerations are important in any synthetic science; in order to duplicate the functionality of a natural system, we do not need to understand everything about it, only the principles out of which the desired functions arise. As discussed in Chap. 1: Introduction, knowledge is more powerful (and practical) as the representative systems shifts from total mimicry (“stealing” from Nature) to an understanding of the process at its basic level (abstraction). Biological systems contain any number of copies of thousands of different components, each with very specific interactions, and each representing a microscopic device in and of itself—that is, one that works away from local equilibrium. As a result, the microscopic description of a biological system (and materials therein) is intractably complex, unless one moves to a higher level of abstraction in the analysis. In the same manner, one only needs a basic knowledge of lift and drag to understand how a Boeing 777 attains flight, but the a Boeing 777 is fully “fly-by-wire”, with 150,000 different subsystem modules, organized *via* elaborate protocols into complex control systems and networks [13]. A 777 is essentially a complex computer network that just happens to fly [13]. If the Orville and Wilber Wright had set out to build and fly a 777, they certainly would have failed.

It is exactly in the face of this complexity that ologs are so appealing. The olog presents us with an opportunity to identify patterns that describe systems and their components, to elucidate possible connections among these components, and to construct isolated functional “modules” by comparing information from many different materials or organisms. That is, by determining fundamental design principles that are simple yet functional, we can not only produce a powerful conceptual model of our system, we also create the possibility of comparing vastly different systems. For the example of a Boeing 777, the functional module that describes the lift of the wings can be equally applied to the wings of an eagle. Indeed, while there is almost nothing physical in common between a protein and a social network, or music and spider silk, we can construct a scenario in which the design principles are well-matched, and thus the systems may be compared. Such a comparison facilitates the transfer of results from other fields (social science, music, linguistics, etc.) to guide us in our study of biological materials of the same structure, and *vice versa*.

If a given study within mathematics (or materiomics) is formalized as a category, it can be connected with other categories that are seemingly far afield, as long as these structures align in the required “functorial” way. Theorems within one branch, like abstract equational algebra, can be applied to a totally different area, like geometric topology. Category theory may not only serve as an alternate foundation to mathematics [97], it unites the various distinct areas within advanced mathematics, formally proving similarities that are not apparent on the surface [98].

4.7 Language to Ontology Logs (*ologs*)

Category theoretic analysis and transformations of syntactic structures have been introduced by Chomsky in 1957 [85]. For formal language theory a well-known transformation is for example the left part transformation from non-left-recursive context-free grammars to context-free Greibach normal form where the syntactic structure is preserved during the transformation [99]. Simply put, the grammatical transformation maintains the syntactic meaning, which we can consider (for our purposes), a coherent function or property. These structure preserving transformations are morphisms (or functors) between objects and arrows among categories and constitute the essential operator to form analogies. The linguistic categorical objects are sets, with unique functions between the objects.

In this chapter we use a linguistic version of category theory in which the objects are drawn as text boxes describing some type of thing, like a protein or a genetic code, and where the arrows also have labels describing some functional relationship, as every protein has a genetic code. This notion of category theory has been

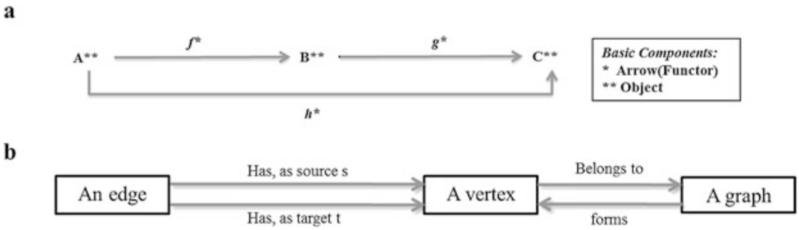

Fig. 4.7 Basic components of ologs. (**a**) Conceptual diagram of a category theoretic system consisting of objects and arrows. Objects are linked *via arrows* and these paths can again be composed into new paths. Objects in ologs are sets and the *arrows* are unique functions. For example, here, the functor f relates **A** to **B**, or $f : \mathbf{A} \to \mathbf{B}$. Moreover, $f : \mathbf{A} \to \mathbf{B}$ followed by $g : \mathbf{B} \to \mathbf{C}$ is equivalent to $h : \mathbf{A} \to \mathbf{C}$. (**b**) Olog of a graph, where subgraphs can form a new vertex for a graph on a "higher hierarchical level". Each box contains a set of instances which are subject to a functional relationship represented by the *arrow*. From [7]

introduced by Spivak *et al.* [6] and recently applied in analogical categorizations of proteins materials [7, 77], discussed in depth for the remainder of this chapter.

Chains of arrows can be composed, providing a description of how a number of small-scale relationships come together to constitute a single, conceptually simpler, larger-scale relation (like a person's father's sister's daughter is a simply their cousin; an example for "functionality" in the space of linguistics). These linguistic categories are called "ologs," short for "ontology logs" [76] (see Fig. 4.7). Ontology is the study of how or what something is, and ologs are a systematic framework in which to record the results of such as study. Simply put, an olog is a well-defined category presented as a visual abstraction. The term "log" (like a scientist's log book) alludes to the fact that such a study is never really complete, and that a study is only as valuable as it is connected into the network of human understanding. This brings us to the heart of the matter: in order to build a sufficient understanding of hierarchical materials, scientists must integrate their findings more precisely with those of other scientists.

The fact that an olog is fundamentally a category means that such connections can be formulated between ologs with mathematical rigor (see Fig. 4.8), and meaning preserved [76], to facilitate the communication with other fields of science. It is believed that this will ultimately enable the kinds of breakthroughs needed to further our understanding of how functional diversity is achieved despite severe limitations of building blocks. The generation of ologs also allows us to observe the formation of patterns that define certain functionality, and draw connections between disparate fields. A key insight used here is that although patterns of functionality generation can be quite different in the space specific to applications (*e.g.*, proteins, language, music), they are remarkably similar in the space of categories. In other words, we hypothesize it is possible to observe universal patterns of how functionality is created in diverse fields; and that it is possible to generate fundamental laws (similar to PDEs in conventional physics) that describe the emergence of functionality from first principles.

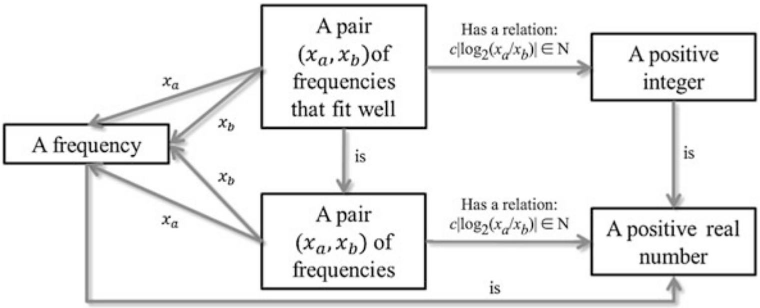

Fig. 4.8 Special paths in ologs (*i.e.*, fiber products). In case the resulting instance of two distinct paths is equal, two paths commute. These structures define new terms in the olog in terms of already existing structures. For example, consider two arbitrary sound frequencies, x_a and x_b, which define "A pair of frequencies". By construction, "A pair of frequencies that fit well" can be defined by the relation $c|\log_2(x_a/x_b)| \in N$ with $c = 12$ [100, 101] being a positive integer, rather than just a positive real number, where positive integers are itself a subset of positive real numbers. All pairs of frequencies have can be formulated by $c|\log_2(x_a/x_b)|$, but not all pairs "fit well". Note that these kind of definitions represent a kind of "worldview" of the author (*e.g.*, extraneous information). From [7]

The actual benefit of an olog is—due to a rigorous mathematical background—its unambiguous way to store and share data, knowledge, and insights in structure and functionality within a single research group and also among many disparate research groups and different fields in science and engineering. As such, ologs can be embedded into a database framework and thus are easily implemented in object based computer languages. Ologs offer means to reveal the origin of the described system property and to connect them to previous results or other topics and fields. For biological materials it is crucial to elucidate the principles from which the superior macroscopic functionality arises in order to define the hierarchical structure-function relationships or even synthesize them. These insights can be gained on the one hand from the category theoretic analysis of protein materials by describing the emergence of functionality from first principles, *e.g.*, on the basis of fundamental interactions between building blocks. On the other hand, the use of systematic analogies with the help of functorial relations supports the researcher in formulating these structure-function relationships in an abstract way, ensuring the ability to connect disparate ologs. Furthermore, by the use of ologs for knowledge creation by shared conceptual models an educational application is feasible [102].

Ontology Logs (olog): The analysis of materials systems based on the recognition of the *universality* of structural elements (building blocks and their interactions), defined with respect to one another in the context space of the olog, and potential *diversity* of fundamental functional mechanisms and material behavior.

The concept of an olog can also be easily extended across hierarchies (not limited to spatial hierarchies, but also encompassing complex functional relations and multi-phenomena coupling). Hierarchical ologs yield similar mathematical features as conventional ologs but are designed to improve the ability to overview the build-up of a hierarchical system by compiling the subunit sets together with sets of superior structural units [7] (see Fig. 4.9). A category theory expert can understand our definition of a hierarchical olog as a category C equipped with a subcategory H with the same set of objects, such that H has the structure of a forest (collection of trees). A morphism $F : (C, H) \to (C', H')$ of hierarchical categories, which we here abuse notation in calling a *functor*, consists of a functor $F : C \to C'$ such that $F(H) \subseteq H'$. A state of a hierarchical olog is just a set-valued functor on C; we can denote sequences, graphs, etc. by a simple modification of the olog, which again we abuse notation in eliding. In other words, a hierarchical olog can provide a coarse or fine categorical perspective, depending on the needs of the olog.

The "extended olog" of Fig. 4.9(a) can also be considered hierarchical, as there is no explicit definition or functional relation for amino acids (*e.g.*, molecules containing an amine group, a carboxylic acid group and a side-chain). Such information is not necessary, as it is superfluous to the function of the protein, the amino acid is considered the fundamental constituent. The olog could easily be extended to accommodate the functional build-up of amino acids, but can be encompassed by a single box.

As an example of the typical features found in hierarchical ologs we can also analyze a distinct feature of linguistics, specifically the structure of a sentence which is formed of words (Fig. 4.10). Words consist of phonemes, the smallest pronounceable segments comprising one or more letters [103]. Hence, these phonemes form the "building blocks" for spoken language and are categorized by distinctive articulatory features, *i.e.*, the description of how the sound is mechanically formed in the body [104]. Whether or not a feature is active for a certain phoneme can be indicated by a binary number. The voice laryngeal feature, for instance, determines whether a sound is formed including the vocal folds (1) or not (0) [105, 106]. The "*r*" in "theo*r*y" contains as voice laryngeal feature a (1), while the "e" does not (0).

Each path is constrained to represent a unique function between the instances of the sets. Consequently, each building block can only be uniquely assigned to its higher units by constructing pairs of building blocks and their higher structural units respectively, see Fig. 4.11. Set **A** associates units from set **B** and set **C** together with their positions, an elegant way to maintain functional relationships within the category. The checkmarks indicate commutative paths (*e.g.*, a kind of isomorphism) in the olog where, starting from the same instance of a set, two distinct paths point to the same instance of another set.

We omit further precise mathematical definition of categories and hence ologs, as we will focus on the application of this concept—the discussion will be limited to a general description of ologs, how they are constructed, and focused on the application to describe materials. Hence, we will proceed to describe ologs by example; for a more mathematically precise account of ologs, see Ref. [6]. For further details on isomorphisms and other features such as limits and colimits in ologs see [6, 77].

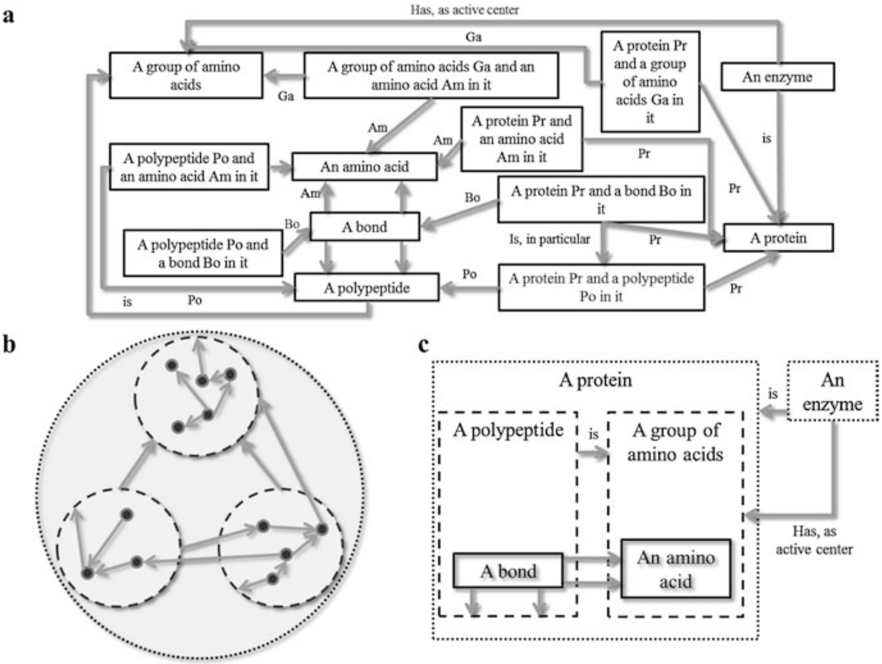

Fig. 4.9 Development of hierarchical ologs. (**a**) Expanded olog describing the structure of proteins out of their building blocks, amino acids. Each bond has as start point and end point an amino acid/a polypeptide. In order to get a hierarchical level higher a higher structure, *e.g.* the protein, is combined with one of its ingredients of a lower level, for example amino acids or bonds. This assures the unique allocation of lower level elements to superior structures and can be seen as a state or "snapshot" of the system. Apart from the hierarchical levels, a subset of proteins, the enzymes, are included to exemplify the way to include "hierarchical interaction". A function relates a certain property of a higher hierarchical structure, here active centers of enzymes, to an element of a lower level structure, a group of amino acids. A polypeptide is a linear chain of two or more amino acids connected by a bond. (**b**) General graph description of a three level hierarchical structure. Note that each cycle represents a state of its lower level elements and each level may interact (in a functional way) with higher or lower level elements. (**c**) Hierarchical olog depicting the same information included in as in panel (**a**) whereas an overview over the hierarchies is now possible. Analogous to panel (**b**) each box represents a state of its lower level elements. From [7]

The way the olog is represented in Fig. 4.11(a) correlates one-to-one to a computer implementation.

Beneficial to the possibility to overview and clearly identify the underlying hierarchical structure of the system, we introduce hierarchical ologs, Fig. 4.11(b). They yield absolutely identical information but a well-arranged notation. In this context, an increased emphasis on the discovery of the structural makeup of systems and how it relates to the emergence of specific properties can provide an important educational feature. Set **A** of Fig. 4.11(a) is now inherently included and the dashed box represents a state of the system, *i.e.* all words in the former set **B** with their current phonetics and binary information (comparable to a "snapshot"). Often, for

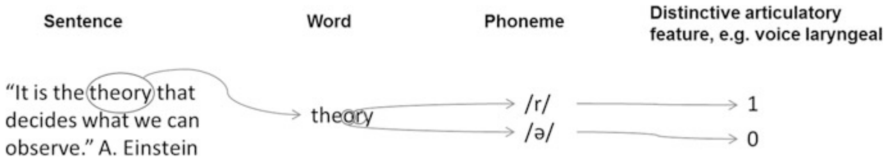

Fig. 4.10 Hierarchies of linguistic structure. The build-up of linguistics can be seen when analyzing the structure of a sentence. It is formed by words which itself are formed by phonemes, the smallest pronounceable segments of a word. These phonemes are categorized by distinctive articulatory features, *i.e.* the description of how the sound is mechanically formed in the body. Whether or not a feature is active for a certain phoneme is indicated by a binary number. From [7]

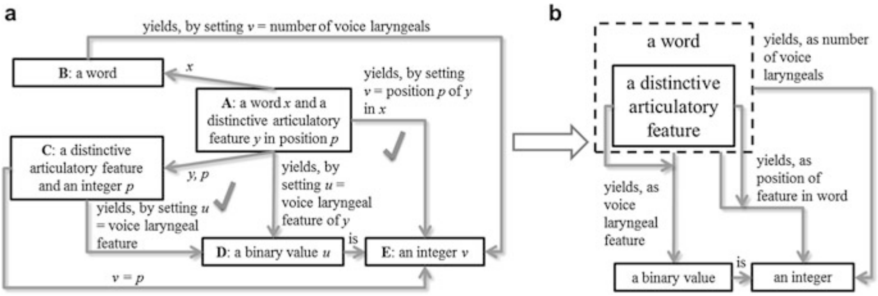

Fig. 4.11 Linguistic structure to category theory *via* an olog. (**a**) This olog describes the situation where every word consists of phonemes, hence they form the "building blocks" for spoken language. An appropriate way to assign the building blocks to its higher units is a representation of sets of words with associated phonemes (set **A**). The checkmarks indicate commutative paths in the olog where starting from the same instance of a set two distinct paths point to the same instance of another set. The way the olog is represented here correlates one-to-one to a computer implementation. Note, that the paths **A** → **D** and **A** → **C** → **D** as well as **A** → **E** and **A** → **C** → **E** commute as indicated by the checkmarks. No other paths commute. (**b**) A hierarchical olog yields identical information but a better overview of the underlying structure of the problem. From [7]

instance in the case of phonemes in words, a higher structural unit (*e.g.*, the word) is formed by a sequence of subunits (*e.g.*, phonemes). This information is conveyed by a simple dashed "hierarchy-box": the word box surrounds the phoneme box, indicating a hierarchical construction. In other cases, the structure may be arbitrary, but always describable by graph-theoretic tools. This holds true in the case of proteins, which are arrangements of (*i.e.*, hierarchically constructed from) amino acids. Paths combining "inner" and "outer" information, *e.g.*, new "double arrows" and former commutative paths, are automatically commuting and no additional checkmarks are needed.

4.8 Proteins and Communication Networks

Let us demonstrate the potential powerful application of category theory. Again, in biology, the same 20 amino acids can have different functions depending on how they are arranged in a sequence as defined by the genes. In other words, the same library of fundamental building blocks can produce different functionality depending on the precise sequence. Just so, an olog serves as a code or formula for a complex structure, but the context in which it is interpreted can lead to different results. Here we ask:

> *Can the same olog represent the structural and functional relationship within a protein filament* (*such as an* α*-helix or an amyloid fibril*) *and the relationships between a type of social network*?

In the case of a protein the building blocks are polypeptide fragments with H-bond clusters as glue, whereas in a social network the building blocks are people and communication methods. It is the interplay between *form* and *function* of few universal building blocks that ties biological structuralism and category theory, and which may produce potentially novel approaches to designing engineered systems.

The main ideas will be developed for two examples of protein materials that display a distinct mechanical behavior once exposed to mechanical force [77]. We begin the discussion with a presentation of the proteins and their functional properties, here their mechanical properties under axial extension (realized *via* the application of an axial force applied to the protein filament). The structure, mechanisms and resulting functional properties have been developed in a series of earlier studies based on computational approaches to molecular nanomechanics (for α-helices [26] and for amyloids or β-sheet crystals [47, 107], and we refer the reader to these original papers for further detail into the mechanistic analysis of the nanomechanics).

Figure 4.12 shows the visualization of the two protein materials considered here based on an abstraction of how their mechanical properties can be understood based on the interplay of a set of “building blocks” (Fig. 4.12(a)). Both protein materials considered here resemble a linear arrangement of three available elements: “bricks”, “glue”, and “lifeline”. As a (known) design rule, brick and glue need to alternate in order to achieve a stable structure. Two brick or glue elements immediately next to each other would not stick together. There is a fundamental chemical reason for this constraint as bricks represent the protein’s polypeptide backbone and glue represents H-bonding which can only occur between a cluster of amino acid residues in the backbone. The “lifeline” is a third element that is introduced here, reflecting the situation in which there is still a physical connection of bricks even after large force causes the glue to break. Chemically, this resembles the existence of a “hidden” polypeptide length such that there exists a “covalent” link between two brick elements even after the H-bond glue has broken (as present in β-helices and α-helices). This hidden length is not observed as a relevant structural property until the glue breaks, at which point the lifeline comes into play and provides an increasing resistance against further deformation. Thus, although both glue and lifeline can connect neighboring brick elements, they are differentiated in that the lifeline is much stronger than the glue and that its resting extension is roughly the failure extension of the glue.

Although this description of protein filaments is a simplification of how their mechanical properties can be described it enables us to understand the key functional properties based on the interplay of building blocks. Figure 4.12(b) depicts a model

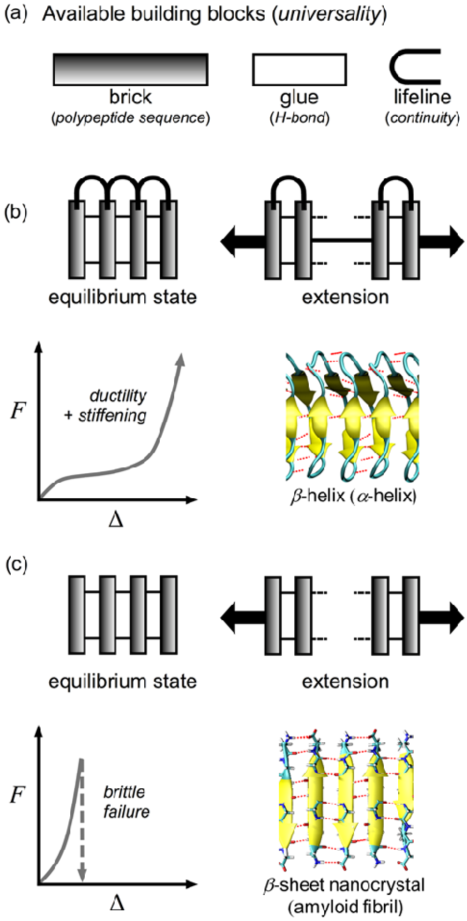

Fig. 4.12 Visualization of protein materials considered here, and abstraction of how key functional properties (here: mechanical properties under axial extension) can be understood based on the interplay of a set of "building blocks". (**a**) Fundamental building blocks of our protein materials. The protein materials considered here are composed of a linear arrangement of three potential elements, "bricks", "glue", and "lifelines". As a "design rule", brick and glue need to alternate in order to achieve a stable structure. That is, two brick or glue elements immediately next to each other would not stick together—the chemical reason is that bricks represent the protein's polypeptide backbone and glue represents H-bonding which can only occur between residues in the backbone. The "lifeline" is a third element introduced here, reflecting the situation when there is still a physical connection between bricks even after the glue breaks (continuity). (**b**) Model of an β-helical protein (or, equivalently in terms of structure and behavior, a α-helix) under axial loading [26]. This resembles a system with a lifeline, as after breaking of the cluster of H-bonds (glue) that are formed between groups of amino acids (brick) there still exists a physical connection due to the polypeptide backbone (lifeline). The existence of a lifeline has major implications on the functional properties of the overall system, resulting in a stiffening F-Δ behavior, and enhanced ductility (large extension at failure). (**c**) Model of a β-sheet crystal protein filament (or similarly, an amyloid fibril) under axial loading [107], resulting in brittle failure. This resembles a system without a lifeline since after breaking of the H-bond cluster (glue) between the layers formed by clusters of polypeptide (brick) no physical connection exists. From [77]

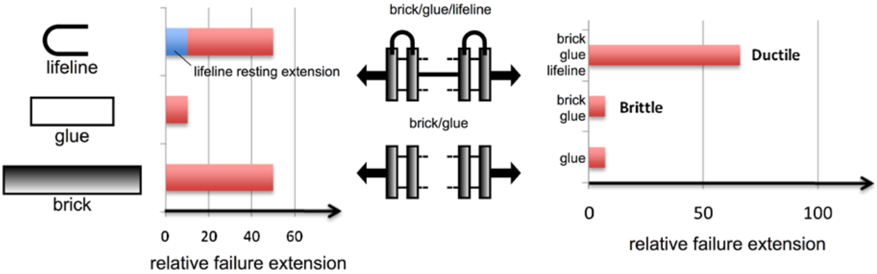

Fig. 4.13 Mechanical behavior of each of the building blocks characterized by a description of the failure extension. The hidden length of lifelines is reflected in the fact that the resting extension of the lifeline is roughly equal to the failure extension of the glue. Both the brick and the lifeline have large failure extensions relative to the glue. A system with a lifeline shows a *ductile* response, where a connection can be sustained at large extension as compared to the glue alone. A system without a lifeline shows a *brittle* response, where only a small extension can be sustained until the material breaks (which equals roughly the failure extension of the glue). From [77]

of an β-helix protein (or similarly, a α-helix protein) under axial loading, assembled based on an alternating sequence of bricks (amino acid cluster), glue (cluster of H-bonds) and a lifeline element. The lifeline element is formed by the protein backbone that still exists even after the cluster of H-bonds break after unfolding of one α-helical turn [26]; providing a physical connection that allows additional glue elements to break after more axial extension is applied. In fact, at large extensions all glue elements will have broken such that the system's overall failure extension is much larger than the failure extension of the glue, marking a "ductile" behavior. Figure 4.12(c) depicts a model of a β-sheet nanocrystal (or similarly, an amyloid fibril). The structure is realized by the assembly of on an alternating sequence of bricks (amino acid cluster) and glue (H-bond cluster). Upon the increase of the extension one of the glue elements breaks. Since there is no more physical connection between the two brick elements that were previously connected by the glue element the entire system has failed, and at an extension that is roughly equal to the failure extension of the glue. We define this behavior as "brittle".

The comparison of the distinct mechanical behavior of β-helices or α-helices and β-sheet nanocrystals or amyloid fibrils was achieved by mapping the key mechanisms that generate their specific properties into the abstract space of interactions between a set of building blocks (such as failure extension; see Fig. 4.13). What was described in words in the preceding paragraphs can be rigorously achieved using ologs, which describe the interactions between building blocks. Through the development of ologs for each system we aim to answer a series of questions:

- What are the components of the system, and how do they interact?
- How do these interactions produce the functionality we observe of the overall system?
- When does functionality break down? *E.g.*, failure of building blocks as the system is pushed to extreme conditions, or the presence of defects.

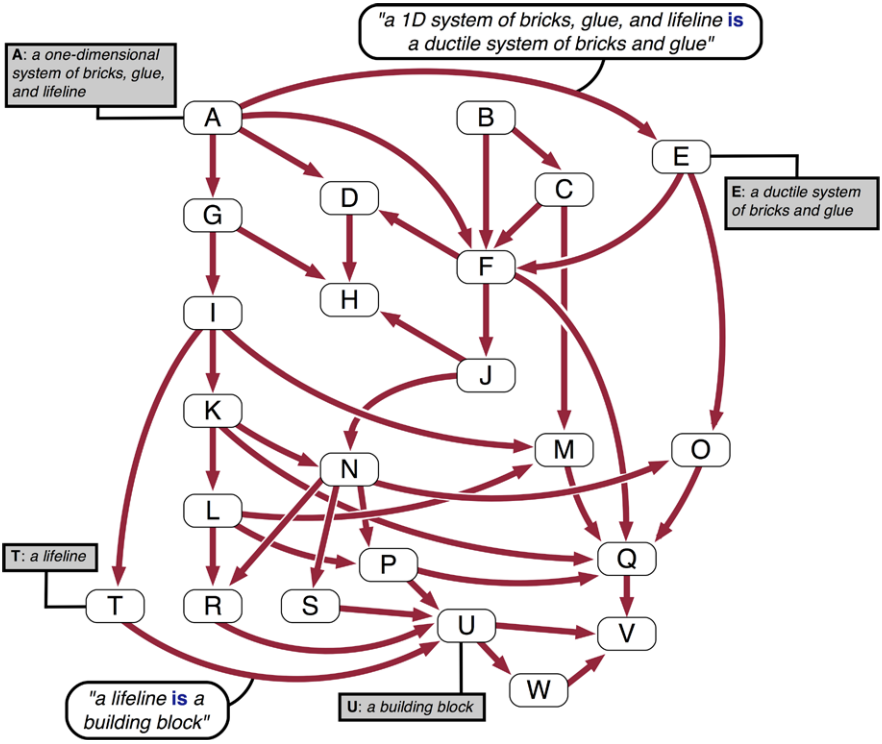

Fig. 4.14 Olog description of both proteins and a social network in abstract terms of bricks, glue, and lifelines. Each box represents an abstract type, and each arrow represents an aspect (or observable) of that type. Each type refers to a set of intended instances, which we think of as being contained in the box. For example box **E** contains ductile sequences of bricks and glue (like an α-helix), whereas box **V** contains real numbers (*e.g.*, 9.228). Each arrow from a source box to a target box refers to an observation one may make on things in the source box, for which the observed result is a thing in the target box. For example, arrow **E** → **O** indicates that one can observe of any ductile material (S) a pair of numbers (R, r) where R is much greater than r. The meaning of these numbers R and r is enforced by a "commutative diagram", illustrated in Table 4.2. Here, the number R must refer to the failure extension of the system S and the number r must refer to the failure extension of its glue. This says that a ductile system fails at a much greater extension than its glue elements do. Each box is meant to contain an intended set of instances and each arrow is meant to functionally relate two such sets. The rest of the olog is recorded in Tables 4.2, 4.3 and 4.4. Some are commutative diagrams which declare two paths through the olog to be equivalent and some are fiber products which define new types in terms of others. From [77]

- A further reaching question may be, by what process did the system come to be constructed, and what selective pressures at the macroscale induce observable changes in the system and at different levels in the structural makeup?

To eventually get us to this point, we will now discuss the components of our brick-and-glue system of proteins, and their integration within the olog presented in Fig. 4.14 which describes both the brittle and ductile protein filaments outlined above. Three universal elements, which we have been calling bricks (b), glue (g), and lifelines (L) are the abstract building blocks composing our systems, and they are defined in relation to one another. Both glue and lifeline are materials that can connect two brick elements. There are two distinctions between them: (i) the failure extension of glue is much less than that of brick, whereas the failure extension of lifeline is roughly equal to that of brick, and (ii) the resting extension of lifeline is roughly equal to the failure extension of glue. These two properties ensure that the lifeline is not detected under axial loading until a glue element breaks and that all the glue elements break long before a lifeline or brick element breaks.

This distinction between one number being roughly equal to another and one number being much greater than another is simple, yet universal in the sciences, and thus we can expect these types (**M** and **O** in the olog) to be quite common in scientific ologs. In fact, we reuse this concept within the olog when we distinguish a ductile system from a brittle one. That is, we characterize a ductile system to be one whose failure extension is much greater than that of its glue element, whereas we characterize a brittle system to be one whose failure extension is roughly equal to that of its glue element. Other common (*i.e.*, universal) patterns that we may find in biological materials is a certain shape (*e.g.*, fibers, helices, spheres), bonds of a certain form (*e.g.*, H-bonds versus covalent bonds, backbone versus side-chain), dimensionality (*e.g.*, 1D, 2D, 3D), and so on. Our olog concentrates on materials whose shape is one-dimensional, a feature we define by the use of mathematical graphs.

Once the fundamental structure of our protein materials and the definition of ductility and brittleness have been defined in the olog, we describe our hypotheses by two arrows, $\mathbf{A} \to \mathbf{E}$ and $\mathbf{B} \to \mathbf{C}$, the first of which hypothesizes that systems with lifelines are *ductile*, and the second of which hypothesizes that systems without lifelines are *brittle*. This hypothesis has now been examined in the paragraphs above, but can be even more carefully explicated using a category theoretic formulation, where each component type and aspect is laid bare. In fact, we have no hope of proving an analogy between this protein setup and the upcoming social network setup without such a formulation. In Fig. 4.14 we show the entire setup as a diagram of boxes and arrows, the precursor to an olog. However, this diagram is not sufficient in the sense that there are mathematical truths present in our system that are not present in the diagram. We include the rest of this information in Tables 4.2, 4.3 and 4.4, which we will describe shortly.

In order to explain what is missing from Fig. 4.14, we should more clearly explain what is there. Each box represents a set. For example box **H**, defined "a graph", represents the set of graphs, whereas box **J**, defined "a system consisting of bricks connected by glue, structured as in graph **G**", represents the set of such systems. Each arrow represents a function from one set to another, and its meaning is clear by reading the definition of the source box, the definition of the arrow, and then the definition of the target box. For example, we read arrow $\mathbf{J} \to \mathbf{H}$ as "a system consisting of bricks connected by glue, structured as in graph **G** is structured as a graph". Thus any element of the set **J** is functionally assigned its structure graph, an element of **H**, by an arrow. Just as the structure graph of a system is an observable of that system, any function from one set to another can be considered an observable of the former.

Table 4.2 Selection of commutative diagrams in the olog. Each sequence of consecutive arrows through the olog (Fig. 4.14) is called a path, which represents a functional relationship between its starting point and its ending point. Two such paths **A** → **B** may result in the same function. The idea is that given an instance of **A**, each of these paths returns the same instance of type **B**. By having this additional data, we confine the meaning of the label on each box and arrow—they cannot stray far from our intended meaning without "breaking" these path equalities. Thus this table serves as an additional check on our labels. Adapted from [77]

Start	End	Path 1	Path 2	Result
A: a 1D system (S) of bricks (b), glue (g), and lifelines (L)	**F**: a 1D system (S) of only bricks (b) and glue (g)	$\mathbf{A} \to \mathbf{E} \to \mathbf{F}$	$\mathbf{A} \to \mathbf{F}$	Each path simply omits the lifelines; Path 1 provides the additional data of *ductility*
C: a brittle system (S) of bricks (b) and glue (g)	**Q**: a pair (x, y) of real numbers	$\mathbf{C} \to \mathbf{F} \to \mathbf{Q}$	$\mathbf{C} \to \mathbf{M} \to \mathbf{Q}$	Each path sets the failure extension of the system (x) and glue (y); Path 2 provides the additional data that $x \approx y$.
N: a pair (b, g) of building blocks, serving as bricks (b) and glue (g)	**Q**: a pair (x, y) of real numbers	$\mathbf{N} \to \mathbf{P} \to \mathbf{Q}$	$\mathbf{N} \to \mathbf{O} \to \mathbf{Q}$	Each path sets the failure extension of the system (x) and glue (y); Path 2 provides the additional data that $x \gg y$.
K: a threesome (b, g, S) of building blocks, serving as bricks (b), glue (g), and strong-glue (S)	**R**: a brick	$\mathbf{K} \to \mathbf{N} \to \mathbf{R}$	$\mathbf{K} \to \mathbf{L} \to \mathbf{R}$	Each of these paths from **K** to **R** yields the same brick element (b); Path 1 defines a (b, g) pair; Path 2 defines a (b, S) pair.

A function may be thought of as a "black box" which takes input of one type and returns output of another type. If the output of one function is fed as input to another function and the whole system is imbedded in a black box, it is called the *composition* of functions. Finally, two functions are equal (regardless of the inner workings of their "black boxes") if, upon giving the same input they always return the same output. The first kind of mathematical truth alluded to above that is missing from Fig. 4.14 is a declaration of which compositions of functions in our system are equal. Such equalities of compositions of functions are called commutative diagrams in category theory literature. A selection of such declarations are presented in Table 4.2. These equalities can be considered as checks on our understanding of all the sets and functions in the arrows—declaring them is at the very least "good science".

Table 4.3 describes a certain class of commutative diagrams, called fiber product diagrams. In a fiber product diagram, one set and two observables of it are declared as a kind of "universal solution" to a problem posed by another diagram. In these terms, we consider the diagram $\mathbf{D} \to \mathbf{H} \leftarrow \mathbf{J}$ as posing a problem, to which

Table 4.3 Fiber product diagrams in the olog. Some boxes in the olog (Fig. 4.14) are defined in terms of others by use of so-called fiber products. For example, object **A** is defined in terms of three others in relationship, $\mathbf{D} \rightarrow \mathbf{H} \leftarrow \mathbf{G}$: given a system of bricks, glue, and lifeline (**D**), we observe its structure graph (**H**) and set it equal to a "chain" graph (**G**)—in so doing we define "one-dimensionality" for a system. The notion of one-dimensionality is not up for interpretation, but directly dependent on the other notions in this olog. Thus this table serves to anchor the interpretation(s) of the olog more firmly. Adapted from [77]

Object	Fiber product object name	Defining attributes	Equated terms	"Idea"
A	a one-dimensional system of bricks, glue, and lifeline	$\mathbf{D} \leftarrow \mathbf{A} \rightarrow \mathbf{G}$	$\mathbf{D} \rightarrow \mathbf{H} \leftarrow \mathbf{G}$	A system of bricks, glue, and lifeline is defined as "1D" if its structure graphs (brick/glue) and (brick/lifeline) are both chains
C	a brittle system of bricks (b) and glue (g)	$\mathbf{F} \leftarrow \mathbf{C} \rightarrow \mathbf{M}$	$\mathbf{F} \rightarrow \mathbf{Q} \leftarrow \mathbf{M}$	A system is defined as "brittle" if its failure extension is roughly equal to the failure extension of its glue
E	a ductile system of bricks (b) and glue (g)	$\mathbf{F} \leftarrow \mathbf{E} \rightarrow \mathbf{O}$	$\mathbf{F} \rightarrow \mathbf{Q} \leftarrow \mathbf{O}$	A system is defined as "ductile" if its failure extension is much greater than the failure extension of its glue
F	a one-dimensional sequence (S) of bricks (b) and glue (g)	$\mathbf{D} \leftarrow \mathbf{F} \rightarrow \mathbf{J}$	$\mathbf{D} \rightarrow \mathbf{H} \leftarrow \mathbf{J}$	A system of bricks and glue is defined as "1D" if its structure graph is a chain
I	a threesome (b, g, L) of building blocks, serving as bricks, glue, and lifeline	$\mathbf{M} \leftarrow \mathbf{I} \rightarrow \mathbf{K}$	$\mathbf{M} \rightarrow \mathbf{Q} \leftarrow \mathbf{K}$	A strong-glue element is defined as "lifeline" if its resting extension is roughly equal to the failure extension of a glue element
K	a threesome (b, g, S) of building blocks, serving as bricks, glue, and strong-glue	$\mathbf{N} \leftarrow \mathbf{K} \rightarrow \mathbf{L}$	$\mathbf{N} \rightarrow \mathbf{R} \leftarrow \mathbf{L}$	A "brick/glue/strong-glue threesome" is defined to be a brick/glue pair and a brick/lifeline pair where the bricks are the same in both instances
L	a pair (b, S) of building blocks, serving as bricks and strong-glue	$\mathbf{M} \leftarrow \mathbf{L} \rightarrow \mathbf{P}$	$\mathbf{M} \rightarrow \mathbf{Q} \leftarrow \mathbf{P}$	Two building blocks, one of which can connect together two instances of the other, are defined as "bricks and strong-glue" if their failure extensions are roughly equal
N	a pair (b, g) of building blocks, serving as bricks and glue	$\mathbf{O} \leftarrow \mathbf{N} \rightarrow \mathbf{P}$	$\mathbf{O} \rightarrow \mathbf{Q} \leftarrow \mathbf{P}$	Two building blocks, one of which can connect together two instances of the other, are defined as "bricks and glue" if the failure extension of the connector is much less than the failure extension of the connectee

$\mathbf{D} \leftarrow \mathbf{F} \rightarrow \mathbf{J}$ is a solution, as we now explain. The diagram $\mathbf{D} \rightarrow \mathbf{H} \leftarrow \mathbf{J}$ poses the problem "what should we call a system consisting of bricks connected by glue, structured as in graph **G**, where graph **G** is a 'chain' graph?". The declared solution is **F**, "a one-dimensional system (S) of bricks (b) and glue (g)", together with its two observables $\mathbf{F} \rightarrow \mathbf{D}$ and $\mathbf{F} \rightarrow \mathbf{J}$. Thus the second kind of mathematical truth alluded to above that is missing from Fig. 4.14 is that some boxes and attributes have fixed meaning in terms of the others. A list of these is given in Table 4.3, where we see terms such as "one-dimensional", "brittle", "ductile", and "lifeline" defined solely in terms of more basic concepts.

Thus, while it is convenient to think of the olog for our protein systems as the diagram in Fig. 4.14, in fact it is the totality of Fig. 4.14, Table 4.2, Table 4.3 and Table 4.4, which really constitute the olog. Just as in biological materials, the parts of the olog (its boxes and arrows) are not sufficient for the system to act as a whole—the less-obvious inter-relationships between these parts give the system its functionality. It is important to note that ologs can be constructed based on modeling and simulation, experimental studies, or theoretical considerations that essentially result in the understanding necessary to formulate the olog. This has been done for the proteins considered here based on the results from earlier work which provided sufficient information to arrive at the formulation of the problem as shown in Fig. 4.12.

Subsequently, we may construct a simple social network that may appear to some as vastly different than a protein filament, and to others as quite similar. The reason for the discrepancy is that semantically and physically the situations have almost nothing in common, but structurally and functionally they do. In fact, we will prove category-theoretically that they are structurally and functionally isomorphic in the sense that their ologs are identical. We now describe the setting for our simple social network as depicted in Fig. 4.15. Imagine a building with sound-proof rooms labeled 1 through 100, equipped with a controlled wireless communication system connecting each pair of consecutive rooms. In each room a human participant sits on a chair with a simple wireless transceiver that can transmit and receive text messages from the participant to the left (his or her predecessor) or the person to the right (his or her successor). We assume that participants in odd numbered rooms are women and people in even numbered rooms are men, just for pronoun clarity. The goal is to faithfully pass messages (sentences of under ten words, say) from room 1 to room 100 and back the other way as quickly as possible. The woman in room 1 (respectively the man in room 100) receives a message from the experimenter. She then inputs it into her transceiver and sends it to her neighbor (2), who passes it along to his neighbor (3), and on down the line until it is received by the man in room 100, who submits it to the experimenter there. Thus the network has a task of faithfully sending messages from one experimenter to the other; if they fail to successfully transmit at least one message per hour we say that the system has failed.

An obstacle can be added by allowing that the transmission of messages between participants is not always error-free. That is, the experimenters can adjust the amount of "noise" in the system, resulting in messages that could be anywhere from error-free to completely unintelligible. For example, the message "*the party was fun and exciting*" may arrive in the next room as "*tha partu was fon and escitin*". Upon receiving a garbled message, a participant may take the time to "fix

Table 4.4 Component analogies between protein and social network. Because the olog (Fig. 4.14) is specifically designed to abstract away the particulars of either the protein or the social network, and thus represent *both* cases, this table defines the particulars in each case. Some types, such as "a real number", stand on their own and we merely give examples. Others, such as "a one-dimensional system of bricks, glue, and lifeline" require a bit more description. This table provides the necessary description to connect the concrete formulations in the case of our protein and social network to the abstract formulation given by the olog. Adapted from [77]

Type	Label	Protein	Social network
A	a 1D system (S) of bricks (b), glue (g), and lifelines (L)	α-helix, β-helix, etc.	wireless and physical communication
B	a 1D system (S) of bricks (b) and glue (g); no lifelines	amyloid, β-sheet nanocrystal, etc.	wireless communication
C	a *brittle* system (S) of bricks (b) and glue (g)	brittle protein filament	brittle social network
D	a "chain" graph $* \to * \to * \cdots \to *$	chain shape for protein	one-to-one communication
E	a *ductile* system (S) of bricks (b) and glue (g)	ductile protein filament	ductile social network
F	a 1D system (S) of bricks (b) and glue (g)	α-helix or β-helix, amyloid or β-sheet nanocrystal	social network
G	a system consisting of bricks connected by glue (g) and lifelines (L), both structured as a graph, G	lifeline protein of specified shape	lifeline social network of specified shape
H	a graph, G	shape of protein	shape of network
I	a threesome (b, g, L) of building blocks, serving as bricks, glue, and lifelines	amino acid cluster, H-bonds, backbone	transceiver, wireless network, possible physical passage/contact
J	a system consisting of bricks (b) connected by glue (g), structured as a graph, G	protein of specified shape	social network of specified shape
K	a threesome (b, g, S) of building blocks, serving as bricks, glue, and strong glue	amino acid cluster, H-bonds, backbone	transceiver, wireless network, possible physical passage/contact
L	a pair (b, S) of building blocks, serving as bricks and strong glue	amino acid cluster, backbone	transceiver, possible physical passage/contact
M	a pair (R, r) of real numbers such that $R \approx r$	*e.g.*, $R = 19.6$, $r = 21.4$	*e.g.*, $R = 19.6$, $r = 21.4$
N	a pair (b, g) of building blocks, serving as bricks and glue	amino acid cluster, H-bonds	transceiver, wireless network
O	a pair (R, r) of real numbers such that $R \gg r$	*e.g.*, $R = 89.6$, $r = 2.3$	*e.g.*, $R = 89.6$, $r = 2.3$
P	a pair (B1, B2) of building blocks, such that B2 can connect two instances of B1	amino acid, backbone	wireless network, transceiver
Q	a pair (x, y) of real numbers	*e.g.*, $x = 17.4$, $y = 38.7$	*e.g.*, $x = 17.4$, $y = 38.7$

Table 4.4 (continued)

Type	Label	Protein	Social network
R	a brick (b)	amino acid cluster	transceiver
S	a glue (g)	H-bond cluster	wireless connection
T	a lifeline (L)	backbone	possible physical passage/contact
U	a building block	basic unit of material	basic unit of communication
V	a real number	*e.g.*, 42.6	*e.g.*, 42.6
W	a resting extension	*e.g.*, 12 Ångstrom	*e.g.*, 1 % error per bit

it up" before sending it along, thereby helping to ensure that the message can be correctly submitted at the end of the line. We define the "extension" of the system to be the amount of noise, measured as the probability that a transmission error occurs for an arbitrary letter in a message. Given sufficient noise, it may happen that no messages can be transferred successfully through the network. Thus we define the "failure extension" of the network to be the amount of noise present when this occurs. Similarly, the failure extension of a glue element is the amount of noise at which a wireless transmission cannot be successfully sent from one room to the next.

Finally, we can add lifelines to this picture by adding physical passageways between consecutive rooms. Now, in case the noise gets too high, individuals may walk or run through these "lifeline passageways" and transmit a message by voice. During low levels of noise, the doorways will typically not be used to relay information because the text messaging is much faster, and hence the existence of the lifelines will be "hidden". However, once the transmission noise is severe enough to prevent good wireless communication (that is, the glue breaks), these passageways come into effect and save the network from breaking altogether. The three basic building blocks of this social network are shown in Fig. 4.15. For a rigorous analysis we also define a failure extension for bricks and lifeline, and resting extension for lifeline (qualitatively identical to those depicted in Fig. 4.13). We can also define the failure extension of bricks and lifeline to be infinite (because messages existing on a given transceiver or passed *via* voice are unaffected by the noise level), and the functional relation is the same. We also define the resting extension of our lifeline passageways to be the amount of noise at which participants begin to use the passageways (*i.e.*, restricts the wireless communication).

We now analyze the performance of the two types of networks constructed here, without and with a lifeline (see Fig. 4.15). In the system without a lifeline as soon as the noise level is high enough to cause breakdown of one of the glue elements the system fails since no more messages can be transmitted. In the system with lifelines, even though glue elements may break there is still the possibility for signals to travel through the passageway such that a much greater noise level (or extension) can be sustained. A brittle network is one in which the failure extension is roughly the

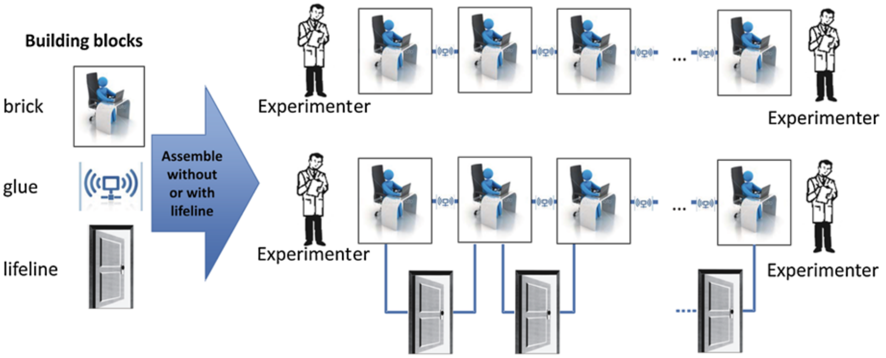

Fig. 4.15 Visual representation of the idealized social network. The social networks considered here are composed of a linear arrangement of three elements, "bricks", "glue" and in some cases, "lifeline" (*e.g.*, fundamental building blocks). Thereby as a (known) *design rule*, brick and glue need to alternate in order to achieve a stable structure. That is, two brick or glue elements immediately next to each other would not stick together; where the reason is that bricks represent participants with transceivers and glue represents wireless communication that, in our case, can only occur between neighboring participants. The "lifeline" is a third element that is introduced here, reflecting the situation when there is still a physical connection of bricks even after the glue breaks. This reflects the existence of a "hidden" connection in that there exists a physical passageway between two brick elements even after the communication over the wireless connection is no longer feasible. The hidden connection is not "visible" before the glue is actually broken because, for reasons of efficiency, participants will choose to communicate the simple messages wirelessly rather than verbally, as the latter requires much more effort. The hidden length of lifelines is reflected in the fact that the "resting extension" of the lifeline is roughly equal to the failure extension of the glue. In other words, lifeline passageways are used only when wireless communication is no longer feasible. Both the brick and the lifeline have large failure extensions relative to the glue because participants and their verbal communication function perfectly well in the presence of noise on the wireless channels. *Top*: The social network restricts face-to-face interaction (*e.g.*, no lifeline). If noise on the wireless line reaches a critical point, messages can no longer be correctly conveyed. *Bottom*: The social network allows for face-to-face interaction (*e.g.*, possible physical passage/contact), resembling a system with a lifeline. If messages can no longer be conveyed wirelessly, communication can still take place, due to the physical passageways as shown. The existence of a lifeline has major implications on the functional properties of the system. A system with a lifeline shows a *ductile* response, where a connection can be sustained at large displacements as compared to the glue alone. In contrast a system without a lifeline shows a *brittle* response, where only a small displacement can be sustained until the material breaks (roughly the failure extension of the glue). From [77]

same as the failure extension for each glue element. A ductile network is one in which the failure extension is much greater than the failure extension of each glue element. We thus hypothesize that social networks with lifeline passageways will be ductile and that those without lifeline passageways will be brittle. While the above communication network is fairly degenerate as compared with, for example, Facebook,[2] the basic idea is similar. People are connected with a set of "friends"

---

[2] http://www.facebook.com.

and the basis of this friendship is communication. Communication can be muddled by various kinds of noise, but the use of additional forms of interaction (*e.g.*, talking in face-to-face meetings in addition to using online text messages) increases the probability that the parties understand each other.

We have constructed a system so that the olog describing it is precisely the same as that defining the protein system previously described. The basic layout is in Fig. 4.14, and the tables add "rigidifying information". For example, the participants with their transceivers are the bricks, the wireless communication between neighboring rooms is the glue, the passageways are the lifelines (see Table 4.4 for all definitions). We define brittleness and ductility exactly as we did in the protein case and as described in the previous paragraph; in fact this is forced on us because boxes C and E are fiber products. The fact that the same olog describes our protein materials and our social network should be considered as a rigorous analogy or isomorphism between these two domains.

This is the analogy between the protein strands (*e.g.*, α-helix and amyloid, β-helix and β-sheet crystal, or any mechanically equivalent protein structures) and the social network experiment: In both cases a network (protein/social) consisting of bricks (amino acid clusters/human participants) connected together by glue (H-bond cluster/wireless communication) is subject to pulling (axial extension/error-producing noise) and eventually reach a breaking point (when the maximum extension is reached/transmission rate drops to zero). Lifelines (additional physical connections *via* covalent links/passageways) serve to increase the ductility (failure extension of network divided by failure extension of individual glue elements) of the network. The two situations can be modeled by precisely the same olog. Thus the olog sets out a space of possible systems that includes everything from proteins to social networks (and potentially many other realizations), any two instances of which must be analogous, at least to the level of detail found in Fig. 4.14 and the associated tables. If one desires additional detail, for example to add a precise meaning for resting extensions, or even real numbers, one would simply expand the olog to capture these ideas. The interpretation of what b, g and L mean in different systems can be distinct (*e.g.*, proteins, polymers, music, etc. can have different physical realizations of these concepts). Yet, their fundamental properties and how they relate to others—other elements, different scales in hierarchies, etc.—are defined properly in the olog, and mathematically expressed not only as a fundamental property but in addition as functors to other elements in the system. For physical systems a key aspect of understanding the interplay of building blocks can for instance be expressed in scaling laws that define properties as a function of ratios of length-scales or energy levels, which fundamentally define how elements behave and interact with others. The general presentation of such relationships in networks is what is missing in current theories, and is where ologs present a powerful paradigm for *de novo* design of biologically inspired systems that span multiple hierarchical levels. This is because ologs can achieve a rigorous description of the synergistic interactions of structures and mechanisms at multiple scales, which provides the basis for enhanced functionality despite the reliance on few distinct building blocks.

Of course, our social network was contrived to fit the olog of the protein, in order to show that two very different domains could have identical conceptual descriptions at a very high level of detail. More detail could show differences between these two domains. For example, an observation we purposely did not include is that the bricks in our social network have need to breathe and eat! It is impossible (and perhaps not desirable in some cases) to include every detail of each system—our goal was to emphasize the essential parameters, and to provide a level of abstraction that emphasizes the key elements that define functional properties. Furthermore, whereas it may be rare for two different scientific studies to result in identical ologs, finding reusable parts should be quite common. In our olog, the notion of bricks being connected together by glue to form the structure of a graph is surely reusable not only within materials science but throughout engineering. The category-theoretic notion of "functors," which formally connect one olog to another, will allow scientists in vastly different fields to share their work by rigorously connecting together their ologs. This opens enormous opportunities for design of novel functional properties by drawing from the understanding gained in diverse fields.

A unique aspect of the equivalence that we describe between two seemingly disparate systems is the rigorous analysis of the conceptual interaction rules in protein materials and the establishment of a direct link to those of a social network *via* the use of category theory. We now turn to another pairing of systems—silk and music—to illustrate the use of hierarchical ologs.

4.9 Spider Silk and Music

Returning to a functional simile with music [7], we construct an olog that reflects the hierarchical structure found in protein materials such as spider silk, Fig. 4.16. In order to form the analogy to music we must determine a way to dissect the structure to its basic constituents. A generally advisable approach is the definition of building blocks of the systems first. Depending on the level of abstraction, these building blocks can be of real nature, *e.g.* phonemes, or of abstract nature, *e.g.* the predefined "lifeline" in α-helices. Here, proteins assemble out of their building blocks, amino acids, whereas we define the building blocks of music as sound waves (sine, triangular, sawtooth, etc. [108, 109]) that are assumed to assemble *via* stacking, *i.e.* without any additional information about amplitude, frequency or pitch.

In a second step, we define the superior structural units and indicate how they are related to their basic constituents. Bonds affiliate amino acids into groups and thereby a polypeptide is a linear chain of two or more amino acids connected by a (peptide) bond. Each bond within the polypeptide has as starting point and end point an amino acid and hence they represent a subset of amino acid groups which are, in contrast to polypeptides, not necessary a linear chain of amino acids but can assemble in more complex structures (possibly describable *via* a 3D graph). In an analogous way we define the creation of musical structural assemblies where stacked groups of sound waves are called a *tone*.

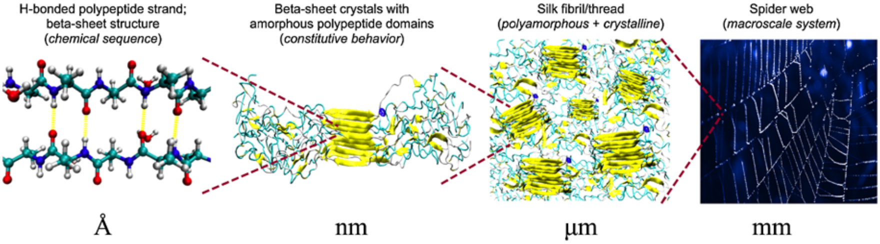

Fig. 4.16 Hierarchical spider silk structure that ranges from nano (Ångstrom) to macro (millimeters). The image displays key structural features of silk, including the chemical structure found at the level of polypeptide H-bonded β-strands, the secondary structure β-sheet nanocrystals embedded in a softer semi-amorphous phase, bulk assembly of poly-crystalline components which assemble into macroscopic silk fibres, and finally the web-structure itself. Further discussion in Chap. 8: Unlocking Nature: Case Studies

So far, the relations only concern structure terminology and the question how functionality can be addressed remains open. Proteins, *i.e.* groups of one or more polypeptides, fold into secondary structures which are crucial to their properties and functionality [31]. Hence, a precursor to the assignment of sequence-structure-function relationships is the sequence-structure identification by experiment and computational studies. Such knowledge-based assignments have already been part of preceding inquiries [110, 111]. The information gathered from these studies, for example the sequence and environmental conditions that lead to distinct structural assemblies, then become data in the olog (Fig. 4.17).

Nanocomposites consist of proteins positioned in certain secondary structures (*e.g.*, α-helix, β-sheet or amorphous phase in spider silk [43, 112]; in any case defined by a graph structure) of a specific size that assemble into higher level networks. The shear strength of these secondary structures, information stored in “a shear strength”, is directly related to properties such as size and arrangement [44, 45, 47]. Similarly, the variation of frequency and amplitude of the stacked waves leads to the formation of the functional unit “a note” defined by its property “a pitch” [113]. The pitch corresponds to the audibility which then determines together with duration, loudness and timbre the functional properties of a chord (a grouping of intervals into categories such as thirds, fourths, etc.) which assemble into harmonically stable riff structures [114–118]. Here we identify a major potential of hierarchical ologs. As all chords in the riff are assembled in a weighted graph structure, the information that riffs are made of rhythmic arrangements of chords is inherently included. This designates a novel way of writing music sheets, where chords represent graph nodes connected to their nearest neighbors by edges where the edge length (or weight) directly correlates to the length of the chord.

Apart from the hierarchical levels, a subset of proteins, the enzymes, are included to exemplify the procedure to include “hierarchical interaction”. Subsets of higher hierarchical levels contain distinguished members whose functionality is based on lower level architecture. Thus, a function relates a certain property of a higher hierarchical structure, here active centers of enzymes, to an element of a

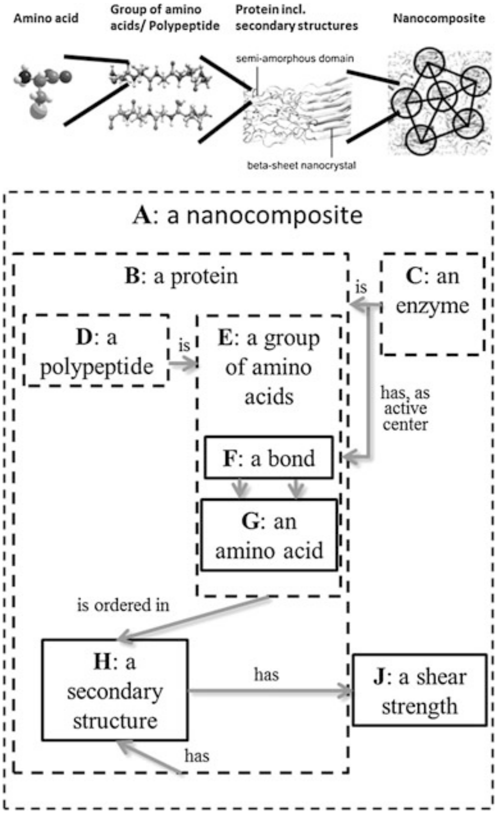

Fig. 4.17 Olog of silk, describing the hierarchical build-up of proteins out of their building blocks, amino acids. Each bond has as starting point and end point an amino acid. A polypeptide is a linear chain of two or more amino acids connected by bonds. Nanocomposites consist of proteins in certain secondary structures (*e.g.* β-sheets, amorphous phases) of a specific size that assemble into higher level networks. The geometry of these of the secondary structure directly relates to macroscopic functional properties such as shear strength and extensibility. Each hierarchy-box represents a state or "snapshot" of its inner constituents that are connected in a graph structure, as shown in Fig. 4.11. This assures the unique allocation of lower level elements to superior structures. Apart from the hierarchical levels, a subset of proteins, the enzymes, are included to exemplify the way to include "hierarchical interaction". A function relates a certain property of a higher hierarchical structure, here active centers of enzymes, to an element of a lower level structure, a group of amino acids

lower level structure, a group of amino acids. Similar to the enzyme-protein relation subsets of chords also include group members with distinguished functional meaning [116, 119]. Here, the major chords, as interval special functional class of chords, has as distinguished member the root, the base on which a triadic chord is built. This kind of relation is typical for all kinds of hierarchical organizations, *e.g.* in primate groups [120]. After assembling and relating these insights, a challenge that can be overcome by multiscale studies including graph theoretic tools,

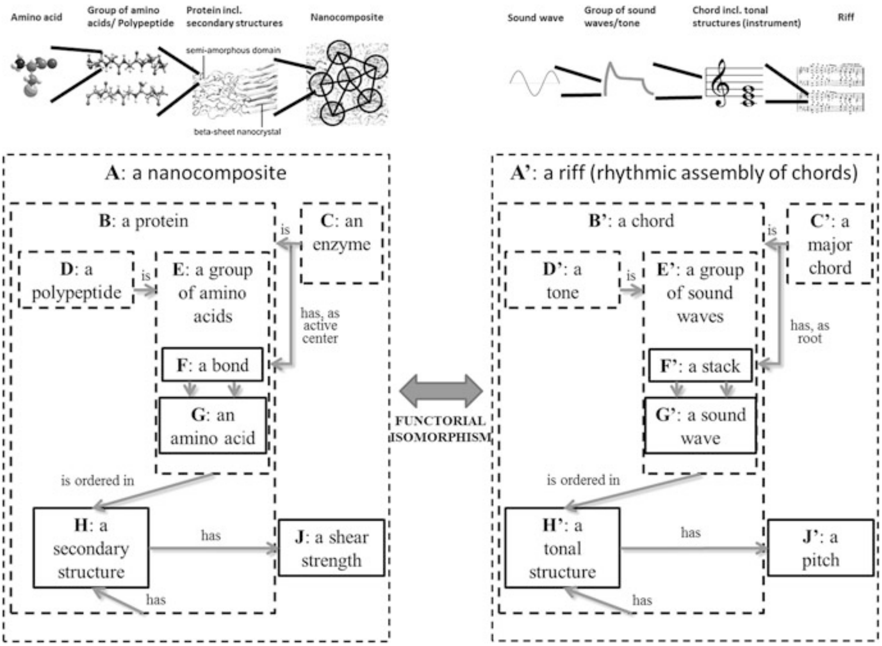

Fig. 4.18 Functorial analogy of protein materials and musical structure by *hierarchical* ologs I: structure. (*Left*) Olog describing spider silk, identical to Fig. 4.17. (*Right*) A functorial isomorphism relates objects and arrows from the protein network to a music network which shows an equivalent hierarchical buildup. The building blocks in music networks are basic sound waves (sine, triangular, sawtooth, etc.) that assemble *via* stacking. The variation of frequency and amplitude of the stacked waves leads to the formation of the functional unit "a note" defined by its property "a pitch". Similar to the enzyme-protein relation certain subsets of chords, here the major chords, include group members with distinguished functional meaning. From [7]

a deduction to fields which show an equivalent hierarchical build-up by functors is possible (Fig. 4.9). Thereby, the relations and thus the functionality within the category are maintained and the two seemingly disparate fields display their intrinsic connection. In this example the functorial transformation is an isomorphism meaning that the positions of boxes and arrows are the same in both systems; thus it requires no further clarification (Fig. 4.18).

Apart from the simple description of structural details (graph theory can provide potential means), ologs also reveal system properties in a category-theoretic framework. Such a property is for example the H-bond clustering found in protein structures like spider silk [44, 45]. Geometric confinement of protein materials at the nanoscale leads to the rupture of clusters of 3 to 4 H-bonds in the β-sheet structures and thus to an optimized shear strength. This is shown in Fig. 4.19 where a functional property of the cluster, the shear strength, is related to a structural condition, the geometric confinement. This olog is based on the insights gained from multiscale computational studies [44, 45, 47].

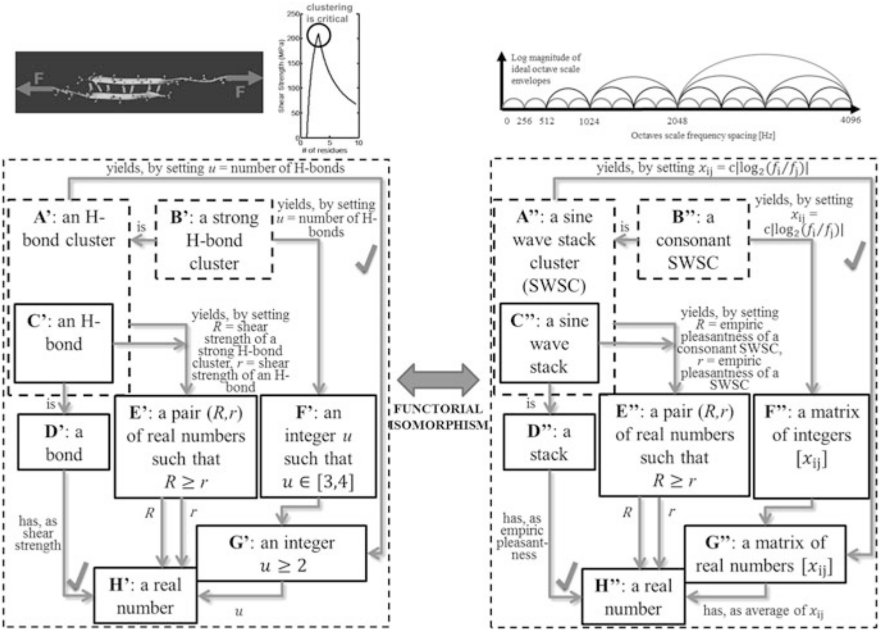

Fig. 4.19 Functorial analogy of protein materials and musical function by *hierarchical* ologs II: cooperativity. (*Left*) Olog describing the dependence of clustering to obtain functionality. For protein materials such as spider silk H-bonds cluster into groups of 3 to 4 residues. Compared to the strength of a single bond, the shear strength of such a bond cluster becomes significantly higher. In the olog this is typically modeled by relations of real numbers. Therefore, the property must be quantitatively ascertainable. (*Right*) Analogous to this, chords in music form by stacking sound waves on an equal tempered scale, *i.e.* with frequencies that "fit well". Unlike shear strength the benefit of sound wave clustering—in a so-called consonant cluster—is not easily quantifiable but *is* subject to empirical observations, the measured pleasantness. The condition for consonance is given by the condition that all entrances of the frequency matrix are integers. Note, that the paths $\mathbf{C} \rightarrow \mathbf{D} \rightarrow \mathbf{H}$ and $\mathbf{C} \rightarrow \mathbf{E} \rightarrow \mathbf{H}$ as well as the paths $\mathbf{B} \rightarrow \mathbf{A} \rightarrow \mathbf{G}$ and $\mathbf{B} \rightarrow \mathbf{F} \rightarrow \mathbf{G}$ commute. Both ologs can be adjoined to the ologs shown in Fig. 4.18 simply by connecting it to the set "a bond"/"a stack". From [7]

Similar cluster strategies can be found in music. Chords consisting of "consonant" frequencies are considered to be innately pleasant to humans and even some animals, *i.e.* these frequencies belong to an (equally tempered) linear pitch space where pairs of frequencies f_i, f_j follow the approximate relation $c|\log_2(f_i/f_j)| \in N$ with $c = 12$ [100, 101]. This can be thought of more simply as the specific collection of notes that construct a guitar chord (*e.g.*, notes played on adjacent or separate strings), for example. We check this condition for consonance by determining the matrix $[x_{ij}]$ and checking whether the entries have indeed integer values. This physically quantifiable ratio results in the empirically quantifiable sensation revealing emotions, data that can be incorporated in our olog [121, 122]. The major difference between materials sciences and social or artistic sciences is disclosed here: material properties are usually unambiguous (*i.e.*, well-defined) and objective whereas the evaluation of artistic properties is frequently subjective. The present functorial isomorphism allows the correlation of these data and subsequent statistical analysis may reveal additional insights that then lead back to the original system—an advantageous approach to recycle knowledge of well-studied systems such as music for novel applications.

Concerns could relate to the fact that the secondary structure of proteins is often not deterministic, *i.e.* the same protein can fold into more than one structure which endangers the unambiguity of the functional relationships. For our silk, analogous to frequencies that "fit well", we can define two amino acid sequences (in one-letter amino acid codes) such as:

...SQGAGRGGLGGQGAG**AAAAAA**GGAGQGGYGGLGSQG...

and

...GGYGPGQQGPSGPGS**AAAAAAAA**GPGQQGPGGYGPGQQ...

as sequences that "fit will" as the bold poly-Ala sequence constitutes the β-sheets of the silk [43, 112], but this may not be universally true. The same holds true for music, where for example pitch and timbre are sometimes ambiguous [123]. This has to be addressed by determining the environmental conditions that uniquely specify the protein's secondary structure, thus defining a protein grammar [80] or similar, again by experiment and computational modeling.

Both ologs, Figs. 4.18 and 4.19, are part of a bigger olog which would describe the material system spider silk (or classical music respectively) in total. Addressing the challenge to complete the olog immediately, it is straightforward to start with subunits as presented here. The assembly of a bigger system would work for example with the set "a bond" or "a stack", which is shared in both ologs and hence serves as an attachment point to combine them. Table 4.5 summarizes key structures and functions where connatural hierarchical ologs could be designed to uncover more analogies and then be attached to the existing parts. For instance, higher order structural assemblies such as nanocomposites or musical riffs gain functional importance by pattern building. In proteins this may relate to the repetition of secondary structural units and their overall confinement which ensures macroscale functionality, for example semi-amorphous phase and β-sheet domains in spider silk that provide a superior toughness by confinement to a fibril size of around 50 nm [124]. The corresponding pattern in music is the formation of chord sequences into riffs and phrases *via* syntactic structures that provide musical tension, an important functional focus in music [125, 126]. Yet another functional commonality of the two systems silk and music is related to the damage tolerance behavior. Localized defects in spider webs do not effect overall mechanical functionality [127] while the deletion of certain chords in a chord sequences do not affect the tonal coherence and hence the functionality [126]. Note that in hierarchical systems functionality is generally obtained by structural arrangement, *e.g.* clustering or stacking, and hence it is mostly impossible to separate structure and function.

Table 4.5 Overview over similarities between spider silk and music. Beyond the patterns shown in Fig. 4.18 and Fig. 4.19 other structural and functional similarities between spider silk and music can be identified. Adapted from [7]

General property	Spider silk	Music
Assembly of building blocks	Amino acids assemble into polypeptides *via* polypeptide bonds	Sound waves are stacked and interfere
Assembly of single units	Polypeptides assemble *via* covalent and weak bonds and form secondary structures	Sound waves with different frequency, amplitude and pitch form notes (instrument)
Assembly of functional units	Silk protein is formed in a stable structure dependent on solvent condition and ionization state	Sound wave of consonant frequency form chords on the equally tempered scale
Assembly of functional structure	Alanine rich repeat units form beta-sheets with high strength whereas glycine rich repeat units form extensible semi-amorphous phases; repetition of functional units creates nanocomposites	Harmonic sequences consist of the three main functions (tonic, sub-dominant and dominant); sequence/repetition of chords creates a melody riff
Upscaling of functionality	Nanoconfinement of composite structure ensures functionality (high strength, extensibility and toughness) on the macroscale	Phrases and climaxes within the music ensure musical tension, functional dependency of chord sequence
Damage tolerance (robustness)	Localization of deformation upon loading provides spider webs with robustness, damage mitigation, and superior resistance by nonlinear material behavior	According to the dependency structure single chords may be removed from or entered into the sequence without affecting the harmonic function

Does such a overlap of silk and music provide *practical* engineering value? Due to the repetitive nature of the amino acid sequence of silk and the length of the natural protein, genetic engineering of silk analogs is limited: creating exact replicas of the protein sequence is difficult. Synthetic silks of defined sequence (that act as minimal building blocks) that incorporate structural motifs that represent native sequence elements are therefore highly desirable. In order to explore the idea of creating structures which mimic the function of native materials, one could use music theory to compose music that 'sounds' like silk or natural silks, and variants thereof. This is, in a sense, using Nature to inform music. More importantly, once we establish this equivalency baseline, we can then ask the opposite question—can we design a material according to a musical composition? For example, one could use functional harmony to create a "chord progression", likening a material to a function in musical syntax (*i.e.*, a I-IV-V-I chord progression). Taking this further, one could choose an existing musical composition that is successful and try to follow its path of "sounding like silk". This can then be fed back into our development of the framework of using music to inform materials design through common language and vocabulary of patterning, texture, tone, and expressivity.

This qualitative account will allow us to draw direct analogies to existing models of complex hierarchical structures such as those from social networks, and potentially linguistic theory where similar problems have been studied, and enables the utilization of insights and design paradigms across disparate fields of the science of hierarchical systems. The review of the key concepts presented here provides a generic framework that has the potential to unify existing understanding derived from the myriad existing studies of individual protein materials such as bone, silk, or cells and many others, where a major limitation was that no unifying framework that applies generally to all such materials has yet been proposed. This paradigm and associated design rules, which are applicable to other complex systems such as music, engineered technology and materials, or even food recipes, could emerge as an exciting new field of study and make critical contribution to the field of materiomics for which it serves as a central tool to describe structure-function relationships for hierarchical systems.

4.10 Motivating the Abstraction

At first glance, the construction of a complex olog may seem counterproductive. After all, if one requires a thorough understanding of a materiomic system to construct the olog (and all of its relevant connections and relations) what new insight does it bring? Where is the predictive or analytical power? The motivation behind such an abstraction is threefold:

Primarily, it combines a material system's function, structure, and properties as a logical, cohesive (and unique) whole. The complexity of hierarchies and cross-scale interactions can be integrated within the olog to any desired level of precision (consider the "coarse" hierarchical olog of Fig. 4.18 or 4.19 versus the "fine" olog of Fig. 4.14). Definition of the olog embodies the holistic materiomic approach, and supplied a universal means of representation, irrespective of biological material, architectural structure, musical composition, or social network.

Secondly, the ologs provide a common platform of comparison across variegated material systems. They can be considered a kind of complex "Venn diagram" where systems can share similar characteristic and properties unobserved by cursory comparisons. This was demonstrated in the simple example of the social network, wherein wireless communication paths provide the same "transfer function" of H-bonds in protein structures. Discovering the commonalities between complex systems requires both the abstraction and reduction (fundamental building blocks, function, connection, etc.) the ologs represent. The logical reduction from:

$$\text{amino acid} \Rightarrow \text{sequence} \Rightarrow \text{structure} \Rightarrow \alpha\text{-helix} \Rightarrow \text{H-bond} \Rightarrow \text{transfer}$$

is not apparent directly from knowledge of the polypeptide sequence. Delineating the function role of H-bonds, for example, can (a) motivate the implementation in designed, synthetic systems and (b) provide a focus for similar roles in other biological molecular structures. Such investigations are currently being undertaken, such as elucidating the role of H-bonding in stacked graphene oxide systems [128] or the similarities between weak H-bonding and relatively stronger disulfide bonds in human keratin [129, 130], as examples.[3] Is the olog for collagen similar to that of an intermediate filament? Is that of keratin similar to an amyloid fibril? Compilation of this information (*e.g.*, *categorization*) in respective ologs can subsequently reveal subtle similarities and differences between diverse biological materials and systems.

Finally, the abstraction to a reduced form allows an entirely new design space, as yet unexplored (and unexploited). One of the pitfalls of analyzing the material behavior of biological systems is that we resort to common metrics of materials science and engineering. Such tools, successful in describing single scale phenomena are ill-suited in describing a holistic materiome. Undoubtedly, there are differences in stiffness of bone, the stiffness of collagen fibrils, and the stiffness of tropocollagen molecules, yet, they are all within the same complex biological system. Each scale requires a difference method to determine mechanical properties. Integrated computational-experimental approaches must be implemented and further developed to explore the potential and limits of assembling building blocks in hierarchical material design (discussed further in Chap. 5: Experimental Approaches and Chap. 6: Computational Approaches and Simulation). Moreover, even at the same scale, the material can express different mechanical properties (*e.g.*, the stiffness of bone varies depending on the location in the body [132]). While we label "bone" as a material in the traditional sense, with an associated (mechanical) stiffness, we are unintentionally neglecting the hierarchical system composing that material. We overlook the significance of the *materiome*. Would a structural engineering, for example, ever label the "compressive stiffness" of the Eiffel Tower as a single value? Such a property could be formulated (based on the deformation of the tower under load), but would obviously have little value. While such an extreme example is trivial, the determination of the stiffness of bone does have value, such as the case of diagnosing and treatment of genetic diseases such as *osteogenesis imperfecta* (see Chap. 9: Pathological Materiomics). The determined (or assigned) "property" of a system has practical value relative to its function. Again, we see the need for a holistic view, even when we probe benign properties such as "stiffness".

We can also consider the practical aspects of an olog in the *transferability* of system structures and components. Ologs provide a powerful platform to understand the connections between form and function, complex behaviors and simple building blocks. *Solving* a particular olog—that is, determining the governing functional relationships—provides a template to reproduce similar complex systems (see Fig. 4.20). Indeed, through the interaction of bricks, glue, and lifelines, one can design a multitude of synthetic systems to mimic either a protein (such as an α-helix)

---

[3] In graphene oxide, for instance, it has been shown both computationally and experimentally that there exists an optimal water content for a fully formed H-bond network between functionalized graphene layers [131]. Additional water and associated H-bonds serve to weaken the system—similar to the cooperativity of H-bonds in β-sheets.

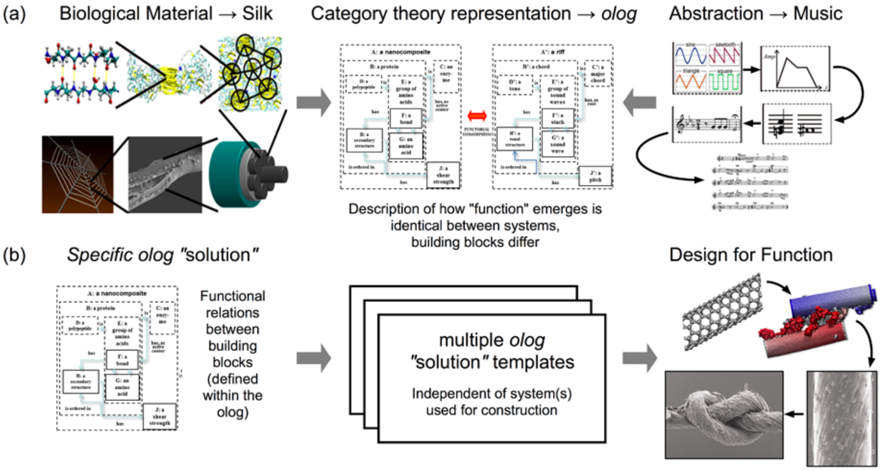

Fig. 4.20 "Solving" the *olog*—a route to function by design. (**a**) Construction of an olog through the analysis of a complex system. Here, both silk and music can be analyzed to determine the functional relationships between elements, as previously described and depicted in Figs. 4.18 and 4.19. (**b**) Once a solution is obtained (the systems and/or iterations thereof are sufficiently captured by the olog), the olog (and the information contained therein) provides a template for any system that can be defined with the identical abstract elements. The olog is self-contained in the sense that it no longer matters what system was analyzed to initially construct it—be it spider silk or music, the functional relations are the same. A collection of such ologs can be "mixed and matched" for relations that commute—introducing new ways to combine building blocks. The end result is materials with designed functionality. Depicted is possible carbon nanotube "yarn", similar to the threads of spider silk fibers (insets of CNT yarn from [133], used with permission, copyright © 2011 American Association for the Advancement of Science)

or a message-passing communication network. Such "solutions" can be coupled to analysis exploiting analogies between patterns found in the arts with those in materials design, facilitated by the application of category theory as a means to enable a rigorous translation. The realization of category theory as an engineering tool can provide a powerful means to represent and optimize such patterns.

Beginning with proteins and materials we know (such as DNA) we can begin to construct a vast array of materiomic data (*e.g.*, the materiome). Detailed design control allowed *via* genetic templates of protein sequence and length can be used to synthesize and explore the physical design space and performance of functional biomaterials. This *tailorability* and control of chemistry, sequence and size for proteins, all lead to predictable functional material features through the realizing of stable folded configurations. Such relationships in proteins are needed to draw useful analogies to music (and other fields) in a meaningful way in terms of structure and function. For example, we have proposed an olog that links silk to music (see Figs. 4.18 and 4.19). Can we then use that olog as an abstract template to construct new materials? While we construct an olog to delineate the complexity of a known material—spider silk, for example—the principles underlying the exemplary performance of silk

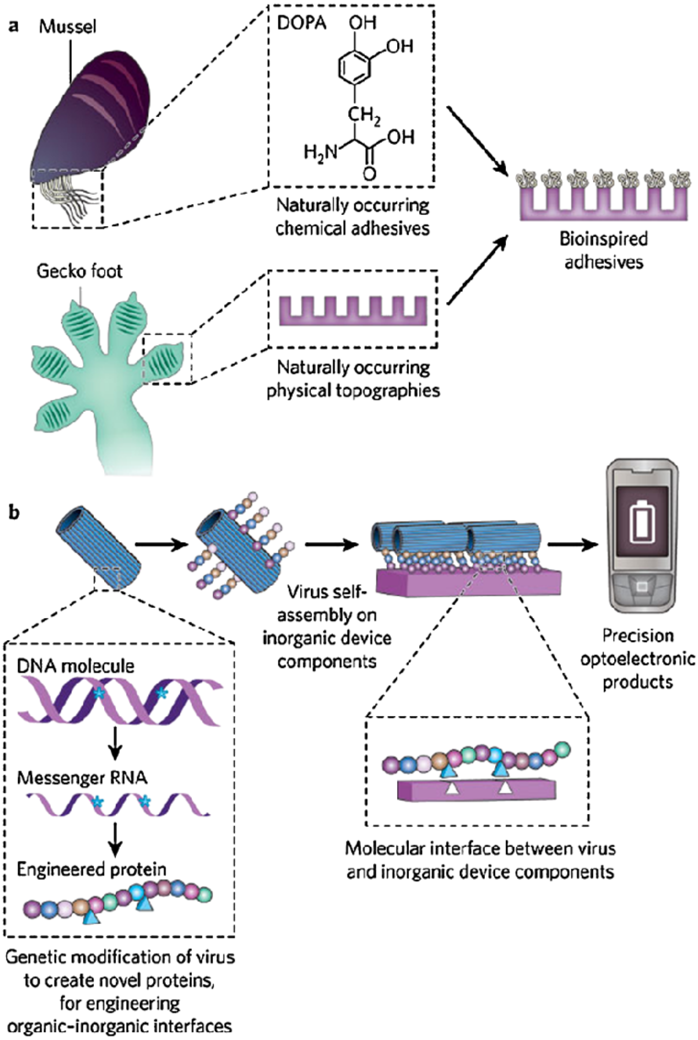

Fig. 4.21 Using *functional* components of biological organisms and materials in novel applications—the building blocks of engineered materials. (**a**) The chemical and physical properties of materials used by organisms to facilitate surface adhesion can be mimicked, allowing the generation of synthetic coatings that modify surface chemistry or prevent biofouling. For example, 3,4$(OH)_2$-phenylalanine (DOPA), a naturally occurring chemical adhesive used to facilitate the adhesion of mussels to surfaces in wet environments, has been combined with the physically patterned nanopillar topography found in the toes of geckos, which facilitates strong adhesion in dry environments, to produce novel adhesives that work in both wet and dry environments. (**b**) The molecular templating of whole viruses allows high-precision, multiscale patterning of electronic devices. Genetic modification of the organism (*left*) is used to engender biomolecular organic–inorganic interactions that lead to the coating of viruses with desired inorganic materials and their macromolecular assembly (*center*). Low-cost, high-precision energy-storage systems (*right*) are one potential application of this concept [134]. From [135], used with permission, copyright © 2009 Macmillan Publishers Limited

can be transferred to other materials (*e.g.*, carbon and other high-performance fibers) and thus help create a synthetic "super-silk" (Fig. 4.20). Exploiting the functional components of biological systems within synthetic devices and novel applications is currently being explored (see Fig. 4.21). Adoption of category theoretic analysis and materiomic descriptions can inform such biomimetic attempts.

Why is it that the fundamental self-assembly processes for materials manufacturing can draw inspiration from musicians? Universal commonalities can be found between materials other fields, from music to economics to sociology. Once discovered, the knowledge base can be shared reciprocally. The construction of a complex olog, for example, helps the creator to think about the problem from a new, abstract perspective, and discovery the "missing links" that may exist. The ideas of motifs, structure, form and function, aimed at arriving at enhanced understanding of how functional diversity can be achieved despite a limited set of universal building blocks, can be applied to the design of new mechanism-based materials.

Biological diversity that emerges from combining simple elements to generate structures that span multiple length-scales (from nano to macro) has many parallels to music, dance, language, and other expressions of art. For example, as we have seen in the case of a musical composition, individual structural motifs are assembled to form larger musical structures that emerge from wave forms, leading to instruments, and then to a simple melody and finally to large-scale symphonic works. However, the directed use of multiscale hierarchical assembly in engineering material function remains largely unexplored, and presents an exciting opportunity to realize novel material functions. From one perspective, such disparate fields can be thought of as merely providing new approaches to complex problems, using the successful insights of one field to understand another (yet another instance of scientific *convergence*). From another, bolder, perspective, the analysis of such disparate fields can be thought of as the same, much grander problem.

4.11 A New Merger of Science and Art?

Through analogy and abstraction and the use of category theory, we recognize a stunning similarity in the structure of music, language and literature, and materials. Perhaps, all expression of arts are a mere representation of humans' own inner workings to the outside world; without us (or the artists) knowing it. We make use of hierarchical structures (with a selection of forms) to express varied feelings and emotions—such as happiness, sadness, or phenomena such as life and death. While we may not have been able to interpret the arts in such a way in preceding times, we now (since the 20th century or so) have a detailed view into the physical resemblance and mechanisms of our own construction (through computational, experimental, theoretical, efforts as described in this book) so that we can identify the analogy between the structure and function of our own tissues and the structure and function of our expression of art.

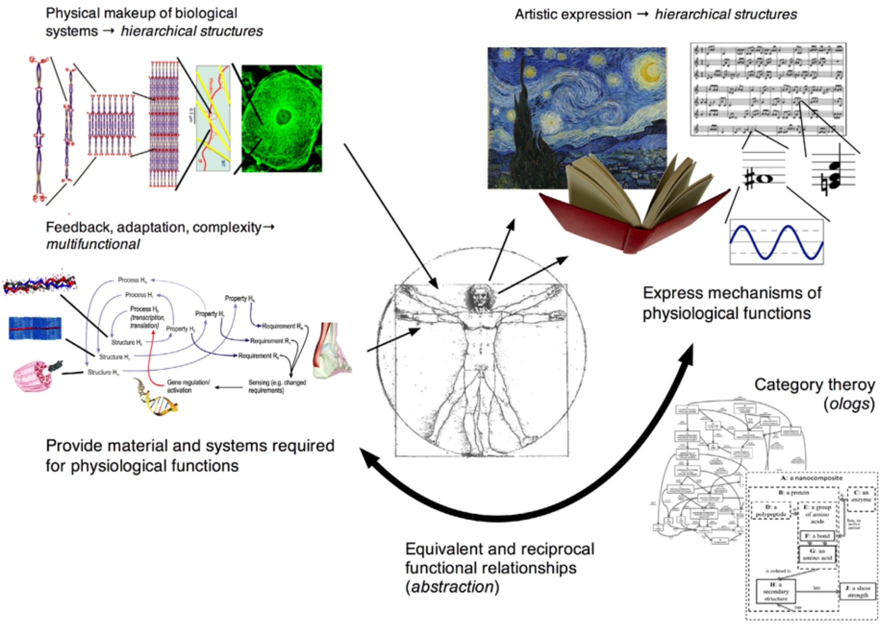

Fig. 4.22 A new merger of *science* and *art*: linking the materiome to artistic expression through category theory. The biomaterial systems that constitute our living bodies (represented here by Leonardo Da Vinci's *Vitruvian Man*) are composed by hierarchical structures, which results in the necessary feedback and adaptation (and intrinsic complexity) for physiological function (*i.e.*, life), and the emergence of a broad range of material properties including mechanical, electrical, optical, chemical, etc. We also expressive ourselves artistically through mediums that are inherently hierarchical in structure (be it physical or aesthetic), from the composition of a symphony, a painting, a sculpture, or a sonnet. What emerges is a broad range of functions, encompassing emotions, life events, messages, stories, etc. Investigation of biological systems and art reveal universalities—equivalent and reciprocal functional relationships—perhaps as they evolved from the same source (ourselves). Such commonalities can be expressed in category theoretic analysis, and represented by an *olog*, reflecting our understanding of a complex system. Such a theoretical template can be applied to materiomics, or in general to other complex systems, as it reflects our own intrinsic perception

Indeed, the description of happiness, sadness, or life and death are focus areas that reach throughout all forms of arts; and they express who we are as humans and how we function and "fail". It appears as if humans have an innate driving force to express themselves to the outside world, to enable communication and to form social interactions. In this context the way art is created may tell us much more about ourselves than we anticipated. From this angle, the study of the arts may provide a powerful window to learn about the function and failure of biological materials, and biological systems more generally (see Fig. 4.22).

Some of the earlier artists such as Leonardo Da Vinci were both scientists and artists and illustrate the close connection between these two fields. While this connection has been lost over many centuries we may now be able to reinstate and exploit it, and the concepts put forth in a materiomics framework, and with the tools of category theory, can be a powerful means to achieve this for synergistic interactions in both directions. For example, we can create materials that resemble some of the mechanisms seen in the construction of music, as another "microscope" into the inner workings of ourselves; which we can exploit for the design of synthetic materials that share some of the unique characteristics of natural ones.

There is an intrinsic disconnect between description and existence, characterization and phenomena. A material has certain behaviors, of course, subject to the observer. This is always open for interpretation—a song may be "enjoyable", a material may be "stiff". The realization remains that certain materials exist, certain materials behave, and there is some measure of performance. How one attempts to categorize, quantify, or rank such behaviors is ultimately subjective. Categorical analysis (through *ologs* for example) is one means of characterizing a material system in terms of functional relationships (indeed a powerful methodology in terms of self-consistency), but by no means is it the only method. Full understanding (*e.g.*, description, categorization, characterization, representation, modeling, templating, etc.) can take many forms—*ologs* being only one. The materiome can be explicated by an olog, but also any holistic collection of multiscale structure, interactions, and performance (in an extended structure-property-process sense) is adequate. A materiomic perspective—encapsulated by an *olog*—is not dependent on any metric of characterization, but rather, the intrinsic property of "function subject to description". But that is the point. Spider silk will still capture prey, regardless of our knowledge of amino acid sequence, and music can still be enjoyable regardless of tuning (*e.g.*, musical scale). Our knowledge—the description of functional phenomena—is subjective and requires a sort of theoretical mapping. The function (*e.g.*, behavior) and interactions within the materiome—especially in the case of natural materials—exists independent of such mapping.

The use of ologs are a powerful approach to find the level of abstractions to prove these concepts mathematically. While we have used category theory here to bridge the seemingly disparate fields of proteins and social networks, silk and music, we envision that this approach will be generalizable to other systems, thus broadening the impact of this research beyond protein materials. Understanding the inner workings of biological systems, in turn, also allows us to create technologies that are more sustainable as they allow for the continued function of a natural system while providing the kinds of technologies and innovations that improve our lives.

4.12 Summary

Here we introduced the concept of hierarchical ologs to describe typical natural hierarchical systems such as language, biological materials and music and draw intrinsic connections between the underlying structures. We showed that in protein materials such as spider silk hierarchical structures identical to classical music can be identified and properly documented by means of hierarchical ologs. Similar to an analogy learning process this method may on the one hand serve as a guide to construct ologs for data and knowledge sharing in research groups and on the other hand as a method to utilize analogies to teach structure and functionality of hierarchical material systems. Conceptually organized like Wikipedia (http://www.wikipedia.org/), but with a substantiated mathematical background, ologs may provide a powerful academic and scientific tool to categorize, organize, relate and share insights gained during research. Specifically, analogy building as a momentous instrument for human understanding and education may be formalized by the use of category theory based ontologies. The ultimate strength of this tool relates to the enforcement of rigor during the analogy building process. Each term and concept defined in one system needs to be precisely related to their analogy counterpart to obtain a structure preserving transformation *via* functors. This characteristic feature of ologs sets the fundament for their superiority over common ontological or heuristic approaches [6].

In general, what ologs can identify for each system is how function emerges (for materials, for example, function can relate to mechanical properties, such as strength, toughness, extension to failure, etc.)—and specifically, how we *understand* these functional properties based on the interplay of the basic building blocks such as molecules or musical notes. Many such systems follow similar—or *universal*—patterns by which diverse function is realized—*e.g.*, through hierarchical structures, size effects, formation of composite structures with certain patterns (*e.g.*, maximum strength for clusters of H-bonds around 3 to 4; complementary frequencies that form chords), etc. While we observe these universalities between diverse biological systems such as, say, silk, diatoms, and bone, there is no current objective mathematical approach to define the similarities between these materials. To resolve this issue we propose the use of category theory—a branch of mathematics, which serves as a powerful tool for us to capture these relationships in an abstract space. When an abstract model is constructed using category theory, the result is a so-called olog, which resembles a graph or network, and encompasses all functional relationships within and between system building blocks.

Ologs can successfully break down—with mathematical rigor—how function emerges based on the interplay of building blocks. All building blocks are defined and self-contained within an olog, *i.e.*, we are free to describe what properties the building blocks have with respect to one another, how they are connected to one another, and how these structures lead to certain material functions and mechanisms. It is important to note that mathematics cannot tell us what these mechanisms are, so we need simulation, theory or experiment to discover these (discussed within Part II: Methods and Tools). This is analogous to the fact that we needed experimentation to eventually confirm the validity of Schrödinger's equation, which serves as the math model to describe the world of quantum mechanics.

The observation of similarities across systems suggest the existence of universal mechanisms by which functional properties emerge, *i.e.*, based on the interplay of the basic building blocks that are arranged in hierarchical structures across vast time-/length-scales, and that hold for a stunning range of natural systems. If such patterns govern much of the natural world, and can be sufficiently defined, the potential for application in engineering the built environment (materials, structures, infrastructure, etc.) are immense. There are several impacts for engineering:

1. Category theory is a tool to compile information about how materials function in a mathematically rigorous way, which allows for comparative analysis between different materials. It also allows us to extract powerful mechanisms from Nature for the use in engineering. One particularly intriguing aspect of natural materials is that function emerges because of hierarchical structures; in spite of or because of the weakness or simplicity of building blocks. These are concepts that we now just begin to understand, and they still need to be translated to engineering.
2. Related to the previous point, we can also ask a question such as: What if we want to replace the building block for a material? *E.g.*, in silk the building block is a β-sheet protein (possible due to restrictions in available material, which may not be the case in engineering). For a new material, however, we may want to use a carbon nanotubes instead, or a clay mineral, or other naturally occurring structures that may be abundantly available. A successful "solving of the olog" allows us to identify what we need to do be able to use the other building blocks, or how we need to change the structure of these, to create the same function as found in the natural material like silk. What this practically means is that we can learn how to make a material that behaves like silk (*e.g.*, extreme strength, toughness, deformability, etc.) from synthetic building blocks like carbon nanotubes.
3. We can learn about (and from) Nature, and how to design technologies that are more in sync with natural processes—specifically, how function emerges in different systems is of extreme value for engineers. For example, we can learn from patterns in music or language how to make better materials.

Future directions, open research questions, and the impact of an increased understanding of hierarchical protein materials can be discussed at three levels with increasing generality: (i) impact on protein material synthesis (design, engineering and manufacturing, or novel biomaterials); (ii) impact on bioinspired nanoscale material design and assembly (*e.g.*, hierarchical materials such as fibers, yarns or armors); and (iii) impact on macro-scale systems design and engineering (*e.g.*, design of cars, airplanes etc. where the merger of the concepts of structure and material across all the scales provides opportunities for more efficient systems). Immediate future work could be directed towards applying the concept of ologs to specific hierarchical biological materials, such as to silk or bone that show a greater complexity than the simple problems reviewed here. While the resulting ologs are more complex, the basic approach is identical and the main insights discussed here should still hold. Eventually the ologs reviewed in this chapter (Figs. 4.14, 4.18, and 4.19) could be implemented in a computational model, which will open the possibility for design optimization using numerical algorithms or make it easier to reuse existing ologs for the design of new ones.

4.13 Key Citations

- D.I. Spivak, R.E. Kent, Ologs: a categorical framework for knowledge representation. PLoS ONE **7**(1), e24274 (2011)

- D.I. Spivak, T. Giesa, E. Wood, M.J. Buehler, Category theory analysis of hierarchical protein materials and social networks. PLoS ONE **6**(9), e23911 (2011)
- T. Giesa, D.I. Spivak, M.J. Buehler, Reoccurring patterns in hierarchical protein materials and music: the power of analogies. BioNanoScience **1**(4), 153–161 (2011)
- T. Giesa, D. Spivak, M.J. Buehler, Category theory based solution for the building block replacement problem in materials design. Adv. Eng. Mater. (2012). doi:10.1002/adem.201200109
- J. Schmalfeldt, *In the Process of Becoming: Analytic and Philosophical Perspectives on Form in Early Nineteenth-Century Music*. Oxford Studies in Music Theory (Oxford University Press, New York, 2011)
- D. Tymoczko, *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. Oxford Studies in Music Theory (Oxford University Press, New York, 2011)
- S. Awodey, *Category Theory* (Oxford University Press, London, 2010)
- M.J. Buehler, Tu(r)ning weakness to strength. Nano Today **5**(5), 379 (2010)
- M.J. Buehler, Y.C. Yung, Deformation and failure of protein materials in physiologically extreme conditions and disease. Nature Materials **8**(3), 175 (2009)
- U. Alon, Simplicity in biology. Nature **446**(7135), 497 (2007)
- R. Kamien, *Music: An Appreciation* (McGraw–Hill Humanities/Social Sciences/Languages, 2007)
- G. Sica, *What is Category Theory*? Advanced Studies in Mathematics and Logic (Polimetrica, Monza, 2006)
- R. Oppenheimer, Analogy in science. American Psychologist **11**, 127–135 (1956).

References

1. R.C. Strohman, The coming kuhnian revolution in biology—commentary. Nat. Biotechnol. **15**(3), 194 (1997)
2. S. Huang, Complexity—the practical problems of post-genomic biology. Nat. Biotechnol. **18**(5), 471–472 (2000)
3. A. Lazaris, S. Arcidiacono, Y. Huang, J.F. Zhou, F. Duguay, N. Chretien, E.A. Welsh, J.W. Soares, C.N. Karatzas, Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. Science **295**(5554), 472–476 (2002)
4. X.X. Xia, Z.G. Qian, C.S. Ki, Y.H. Park, D.L. Kaplan, S.Y. Lee, Native-sized recombinant spider silk protein produced in metabolically engineered escherichia coli results in a strong fiber. Proc. Natl. Acad. Sci. USA **107**(32), 14059–14063 (2010)
5. S. Eilenberg, S. Maclane, General theory of natural equivalences. Trans. Am. Math. Soc. **58**(Sep), 231–294 (1945)
6. D.I. Spivak, R.E. Kent, Ologs: a categorical framework for knowledge representation. PLoS ONE **7**(1), e24274 (2001)
7. T. Giesa, D.I. Spivak, M.J. Buehler, Reoccurring patterns in hierarchical protein materials and music: the power of analogies. BioNanoScience **1**(4), 153–161 (2011)
8. N. Goldenfeld, L.P. Kadanoff, Simple lessons from complexity. Science **284**(5411), 87–89 (1999)

9. R.B. Laughlin, D. Pines, The theory of everything. Proc. Natl. Acad. Sci. USA **97**(1), 28–31 (2000)
10. D. Falk, *Universe on a T-shirt: The Quest for the Theory of Everything*, 1st edn. (Arcade Publishing, New York, 2004)
11. U. Alon, Simplicity in biology. Nature **446**(7135), 497 (2007)
12. J.M. Carlson, J. Doyle, Complexity and robustness. Proc. Natl. Acad. Sci. USA **99**, 2538–2545 (2002)
13. M.E. Csete, J.C. Doyle, Reverse engineering of biological complexity. Science **295**(5560), 1664–1669 (2002)
14. T. Ackbarow, M.J. Buehler, Hierarchical coexistence of universality and diversity controls robustness and multi-functionality in protein materials. J. Comput. Theor. Nanosci. **5**(7), 1193–1204 (2008)
15. M.J. Buehler, Nanomaterials: strength in numbers. Nat. Nanotechnol. **5**, 172–174 (2010)
16. S.J. Eichhorn, C.A. Baille, N. Zafeiropoulos, L.Y. Mwaikambo, M.P. Ansell, A. Dufresne, K.M. Entwistle, P.J. Herrera-Franco, G.C. Escamilla, L. Groom, M. Hughes, C. Hill, T.G. Rials, P.M. Wild, Current international research into cellulosic fibres and composites. J. Mater. Sci. **36**, 2107–2131 (2001)
17. W. Hamad, *Cellulosic Materials: Fibers, Networks and Composites* (Kluwer Academic, Boston, 2002)
18. K. Abe, H. Yano, Comparison of the characteristics of cellulose microfibril aggregates of wood, rice straw and potato tuber. Cellulose **16**(6) (2009)
19. Y. Nishiyama, Structure and properties of the cellulose microfibril. J. Wood Sci. **55**, 241–249 (2009)
20. H. Herrmann, U. Aebi, Intermediate filaments: molecular structure, assembly mechanism, and integration into functionally distinct intracellular scaffolds. Annu. Rev. Biochem. **73**, 749–789 (2004)
21. N. Mucke, L. Kreplak, R. Kirmse, T. Wedig, H. Herrmann, U. Aebi, J. Langowski, Assessing the flexibility of intermediate filaments by atomic force microscopy. J. Mol. Biol. **335**(5), 1241–1250 (2004)
22. S. Kim, P. Wong, P.A. Coulombe, A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth. Nature **441**(7091), 362–365 (2006)
23. L.H. Gu, P.A. Coulombe, Keratin function in skin epithelia: a broadening palette with surprising shades. Curr. Opin. Cell Biol. **19**(1), 13–23 (2007)
24. K.N. Dahl, D.E. Discher, K.L. Wilson, The nuclear envelope lamina network has elasticity and a compressibility limit suggestive of a “molecular shock absorber”. Mol. Biol. Cell **15**, 119–120 (2004)
25. A.C. Rowat, J. Lammerding, J.H. Ipsen, Mechanical properties of the cell nucleus and the effect of emerin deficiency. Biophys. J. **91**(12), 4649–4664 (2006)
26. T. Ackbarow, X. Chen, S. Keten, M.J. Buehler, Hierarchies, multiple energy barriers and robustness govern the fracture mechanics of alpha-helical and beta-sheet protein domains. Proc. Natl. Acad. Sci. USA **104**, 16410–16415 (2007)
27. S.V. Strelkov, J. Schumacher, P. Burkhard, U. Aebi, H. Herrmann, Crystal structure of the human lamin a coil 2b dimer: implications for the head-to-tail association of nuclear lamins. J. Mol. Biol. **343**(4), 1067–1080 (2004)
28. M. Arslan, Z. Qin, M.J. Buehler, Coiled-coil intermediate filament stutter instability and molecular unfolding. Comput. Methods Biomech. Biomed. Eng. **14**(5), 483–489 (2011)
29. A. Kaminska, S.V. Strelkov, B. Goudeau, M. Olive, A. Dagvadorj, A. Fidzianska, M. Simon-Casteras, A. Shatunov, M.C. Dalakas, I. Ferrer, H. Kwiecinski, P. Vicart, L.G. Goldfarb, Small deletions disturb desmin architecture leading to breakdown of muscle cells and development of skeletal or cardioskeletal myopathy. Human Genet. **114**(3), 306–313 (2004)
30. B. Kiss, A. Karsai, M.S.Z. Kellermayer, Nanomechanical properties of desmin intermediate filaments. J. Struct. Biol. **155**(2), 327–339 (2006)
31. M.J. Buehler, Y.C. Yung, Deformation and failure of protein materials in physiologically extreme conditions and disease. Nat. Mater. **8**(3), 175–188 (2009)

32. M.J. Buehler, Tu(r)ning weakness to strength. Nano Today **5**(5), 379–383 (2010)
33. J.E. Padilla, C. Colovos, T.O. Yeates, Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proc. Natl. Acad. Sci. USA **98**(5), 2217–2221 (2001)
34. M.M.C. Bastings, T.F.A. de Greef, J.L.J. van Dongen, M. Merkx, E.W. Meijer, Macrocyclization of enzyme-based supramolecular polymers. Chem. Sci. **1**(1), 79–88 (2010)
35. Y. Mori, K. Minamihata, H. Abe, M. Goto, N. Kamiya, Protein assemblies by site-specific avidin-biotin interactions. Organic & Biomolecular Chemistry **9**(16), 5641–5644 (2011)
36. N. Barkai, S. Leibler, Robustness in simple biochemical networks. Nature **387**(6636), 913–917 (1997)
37. L.H. Hartwell, J.J. Hopfield, S. Leibler, A.W. Murray, From molecular to modular cell biology. Nature **402**(6761), C47–C52 (1999)
38. L. Hartwell, Theoretical biology—a robust view of biochemical pathways. Nature **387**(6636), 855 (1997)
39. K.M. Dipple, J.K. Phelan, E.R.B. McCabe, Consequences of complexity within biological networks: robustness and health, or vulnerability and disease. Mol. Genet. Metab. **74**(1–2), 45–50 (2001)
40. H. Gao, B. Ji, I.L. Jager, E. Arzt, P. Fratzl, Materials become insensitive to flaws at nanoscale: lessons from nature. Proc. Natl. Acad. Sci. USA **100**(10), 5597–5600 (2003)
41. F. Vollrath, D. Porter, Spider silk as archetypal protein elastomer. Soft Matter **2**(5), 377–385 (2006)
42. S. Rammensee, U. Slotta, T. Scheibel, A.R. Bausch, Assembly mechanism of recombinant spider silk proteins. Proc. Natl. Acad. Sci. USA **105**(18), 6590–6595 (2008)
43. S. Keten, M.J. Buehler, Nanostructure and molecular mechanics of spider dragline silk protein assemblies. J. R. Soc. Interface **7**(53), 1709–1721 (2010)
44. S. Keten, M.J. Buehler, Geometric confinement governs the rupture strength of h-bond assemblies at a critical length scale. Nano Lett. **8**(2), 743–748 (2008)
45. S. Keten, M.J. Buehler, Asymptotic strength limit of hydrogen-bond assemblies in proteins at vanishing pulling rates. Phys. Rev. Lett. **100**(19), 198301 (2008)
46. M.J. Buehler, S. Keten, Colloquium: failure of molecules, bones, and the earth itself. Rev. Mod. Phys. **82**(2), 1459–1487 (2010)
47. S. Keten, Z. Xu, B. Ihle, M.J. Buehler, Nanoconfinement controls stiffness, strength and mechanical toughness of beta-sheet crystals in silk. Nat. Mater. **9**(4), 359–367 (2010)
48. K. Autumn, Y.A. Liang, S.T. Hsieh, W. Zesch, W.P. Chan, T.W. Kenny, R. Fearing, R.J. Full, Adhesive force of a single gecko foot-hair. Nature **405**(6787), 681–685 (2000)
49. E. Arzt, S. Gorb, R. Spolenak, From micro to nano contacts in biological attachment devices. Proc. Natl. Acad. Sci. USA **100**(19), 10603–10606 (2003)
50. M. Varenberg, N.M. Pugno, S.N. Gorb, Spatulate structures in biological fibrillar adhesion. Soft Matter **6**(14), 3269–3272 (2010)
51. A. Nova, S. Keten, N.M. Pugno, A. Redaelli, M.J. Buehler, Molecular and nanostructural mechanisms of deformation, strength and toughness of spider silk fibrils. Nano Lett. **10**(7), 2626–2634 (2010)
52. R. Kamien, *Music: An Appreciation* (McGraw–Hill Humanities/Social Sciences/Languages, New York, 2007)
53. C. Dodge, T.A. Jerse, *Computer Music: Synthesis, Composition, and Performance* (Cengage Learning, New York, 1997)
54. P. Miller, *The Smart Swarm: How Understanding Flocks, Schools, and Colonies Can Make Us Better at Communicating, Decision Making, and Getting Things Done* (Avery, New York, 2010)
55. F. Lerdahl, R. Jackendoff, *An Overview of Hierarchical Structures in Music* (MIT Press, Cambridge, 1983)
56. F. Lerdahl, R. Jackendoff, An overview of hierarchical structure in music. Music Percept. **1**(2), 229–252 (1983)

57. D.R. Hofstadter, *Metamagical Themas: Questing for the Essence of Mind and Pattern* (Basic Books, New York, 1985)
58. J.S. Bamberger, A. Hernndez, *Developing Musical Intuitions: A Project-Based Introduction to Making and Understanding Music* (Oxford University Press, New York, 2000)
59. D. Lewin, *Generalized Musical Intervals and Transformations* (Oxford University Press, New York, 2007)
60. J. Schmalfeldt, *In the Process of Becoming: Analytic and Philosophical Perspectives on Form in Early Nineteenth-Century Music*. Oxford Studies in Music Theory (Oxford University Press, New York, 2011)
61. R. Oppenheimer, Analogy in science. Am. Psychol. **11**(3), 127–135 (1956)
62. S. Vosniadou, A. Ortony, *Similarity and Analogical Reasoning* (Cambridge University Press, Cambridge, 1989)
63. J. Bransford, *How People Learn: Brain, Mind, Experience, and School*, expanded edn. (National Academy Press, Washington, 2000)
64. D. Gentner, K.J. Holyoak, B.N. Kokinov, *The Analogical Mind: Perspectives from Cognitive Science* (MIT Press, Cambridge, 2001)
65. P.C. Taylor, B.J. Fraser, D.L. Fisher, Monitoring constructivist classroom learning environments. Int. J. Educ. Res. **27**(4), 293–302 (1997)
66. R. Stavy, Using analogy to overcome misconceptions about conservation of matter. J. Res. Sci. Teach. **28**(4), 305–313 (1991)
67. C.-C. Tsai, Overcoming junior high school students' misconceptions about microscopic views of phase change: a study of an analogy activity. J. Sci. Educ. Technol. **8**(1), 83–91 (1999)
68. L.R. Novick, K.J. Holyoak, Mathematical problem-solving by analogy. J. Exp. Psychol. Learn. Mem. Cogn. **17**(3), 398–415 (1991)
69. S. Kaniel, Y. Harpaz-Itay, E. Ben-Amram, Analogy construction versus analogy solution, and their influence on transfer. Learn. Instr. **16**(6), 583–591 (2006)
70. W. Croft, *Typology and Universals*, 2nd edn. Cambridge Textbooks in Linguistics (Cambridge University Press, Cambridge, 2003)
71. G. Sica, *What is Category Theory?* Advanced Studies in Mathematics and Logic (Polimetrica, Monza, 2006)
72. N.C. Ellis, D. Larsen-Freeman, R.C. in language learning (Ann Arbor, Mich.), in *Language as a Complex Adaptive System*. The Best of Language Learning Series (Wiley–Blackwell, Chichester, 2009)
73. W. Croft, Pragmatic functions, semantic classes, and lexical categories. Linguistics **48**(3), 787–796 (2010)
74. F.W. Lawvere, S.H. Schanuel, *Conceptual Mathematics. A First Introduction to Categories* (Cambridge University Press, Cambridge, 2009)
75. S. Awodey, *Category Theory*, 2nd edn. Oxford Logic Guides, vol. 52 (Oxford University Press, London, 2010)
76. R. Brown, T. Porter, Category theory: an abstract setting for analogy and comparison, in *What is Category Theory?* (Polimetrica, Monza, 2006)
77. D.I. Spivak, T. Giesa, E. Wood, M.J. Buehler, Category theoretic analysis of hierarchical protein materials and social networks. PLoS ONE **6**(9), e23911 (2011)
78. P. Csermely, Creative elements: network-based predictions of active centres in proteins and cellular and social networks. Trends Biochem. Sci. **33**(12), 569–576 (2008)
79. N.M. Pugno, A statistical analogy between collapse of solids and death of living organisms: proposal for a 'law of life'. Med. Hypotheses **69**(2), 441–447 (2007)
80. M. Gimona, Protein linguistics—a grammar for modular protein assembly? Nat. Rev. Mol. Cell Biol. **7**(1), 68–73 (2006)
81. S.C. Ji, Isomorphism between cell and human languages: molecular biological, bioinformatic and linguistic implications. Biosystems **44**(1), 17–39 (1997)
82. H. Peterlik, P. Roschger, K. Klaushofer, P. Fratzl, From brittle to ductile fracture of bone. Nat. Mater. **5**(1), 52–55 (2006)

83. D. Taylor, J.G. Hazenberg, T.C. Lee, Living with cracks: damage and repair in human bone. Nat. Mater. **6**(4), 263–268 (2007)
84. D.A. Fletcher, R.D. Mullins, Cell mechanics and the cytoskeleton. Nature **463**(7280), 485–492 (2010)
85. N. Chomsky, *Syntactic Structures*, 2nd edn. (de Gruyter, Berlin, 2002)
86. M. Moortgat, *Categorial Investigations: Logical and Linguistic Aspects of the Lambek Calculus* (Foris Publications, Providence, 1988)
87. M.D. Hauser, N. Chomsky, W.T. Fitch, The faculty of language: what is it, who has it, and how did it evolve? Science **298**(5598), 1569–1579 (2002)
88. D.B. Searls, The language of genes. Nature **420**(6912), 211–217 (2002)
89. E. Moggi, *A Category-Theoretic Account of Program Modules* (Springer, London, 1989)
90. P. Wadler, Monads for functional programming. Math. Struct. Comput. Sci. **2**, 461–493 (1992)
91. M. Barr, C. Wells, *Category Theory for Computing Science* (Prentice Hall, New York, 1995)
92. B.C. Pierce, *Basic Category Theory for Computer Scientists* (MIT Press, Cambridge, 1996)
93. R. Lakes, Materials with structural hierarchy. Nature **361**(6412), 511–515 (1993)
94. E.M. Marcotte, M. Pellegrini, M.J. Thompson, T.O. Yeates, D. Eisenberg, A combined algorithm for genome-wide prediction of protein function. Nature **402**, 83–86 (1999)
95. E.M. Marcotte, Detecting protein function and protein-protein interactions from genome sequences. Science **285**, 751–753 (1999)
96. D. Eisenberg, E.M. Marcotte, I. Xenarios, T.O. Yeates, Protein function in the post-genomic era. Nature **405**(6788), 823–826 (2000)
97. F.W. Lawvere, The category of categories as a foundation for mathematics, in *Proceedings of the Conference on Categorical Algebra* (Springer, New York, 1965)
98. S.M. Lane, *Categories for the Working Mathematician*, 2nd edn. (Springer, New York, 1998)
99. A. Nijholt, From left-regular to Greibach normal form grammars. Inf. Process. Lett. **9**(1), 51–55 (1979)
100. N.C. Maddage, C. Xu, M.S. Kankanhalli, X. Shao, Content-based music structure analysis with applications to music semantics understanding, in *Proceedings of the 12th Annual ACM International Conference on Multimedia* (2004), pp. 112–119
101. D. Tymoczko, *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. Oxford Studies in Music Theory (Oxford University Press, New York, 2011)
102. G. Deline, F. Lin, D. Wen, D. Gagevic, A. Kinshuk, Ontology-driven development of intelligent educational systems, in *2007 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, vols. 1 and 2 (2007), pp. 34–37
103. M. Halle, *From Memory to Speech and Back: Papers on Phonetics and Phonology, 1954–2002*. Phonology and Phonetics (de Gruyter, Berlin, 2002)
104. G.N. Clements, The geometry of phonological features. Phonology **2**, 225–252 (1985)
105. A.S. Abramson, Laryngeal timing in consonant distinctions. Phonetica **34**(4), 295–303 (1977)
106. M. Jessen, C. Ringen, Laryngeal features in German. Phonology **19**, 189–218 (2002)
107. R. Paparcone, M.J. Buehler, Failure of a beta(1–40) amyloid fibrils under tensile loading. Biomaterials **32**(13), 3367–3374 (2011)
108. J.E. Cutting, B.S. Rosner, Categories and boundaries in speech and music. Percept. Psychophys. **16**(3), 564–570 (1974)
109. J.A. Moorer, Signal-processing aspects of computer music—survey. Proc. IEEE **65**(8), 1108–1137 (1977)
110. D. Frishman, P. Argos, Knowledge-based protein secondary structure assignment. Protein. Struct. Funct. Genet. **23**(4), 566–579 (1995)
111. Z.R. Sun, S.J. Hua, A novel method of protein secondary structure prediction with high segment overlap measure: support vector machine approach. J. Mol. Biol. **308**(2), 397–407 (2001)
112. S. Keten, M.J. Buehler, Atomistic model of the spider silk nanostructure. Appl. Phys. Lett. **96**(15) (2010)

113. R. Erickson, *Sound Structure in Music* (University of California Press, Berkeley, 1975)
114. D. Deutsch, Music recognition. Psychol. Rev. **76**(3), 300 (1969)
115. J. Bharucha, C.L. Krumhansl, The representation of harmonic structure in music—hierarchies of stability as a function of context. Cognition **13**(1), 63–102 (1983)
116. B. Pardo, W.P. Birmingham, Algorithms for chordal analysis. Comput. Music J. **26**(2), 27–49 (2002)
117. D.M. Randel, *The Harvard Dictionary of Music*, 4th edn. (Belknap Press of Harvard University Press, Cambridge, 2003)
118. K. Jensen, Multiple scale music segmentation using rhythm, timbre, and harmony. EURASIP J. Adv. Signal Process. (2007)
119. C.L. Krumhansl, R.N. Shepard, Quantification of the hierarchy of tonal functions within a diatonic context. J. Exp. Psychol. Hum. Percept. Perform. **5**(4), 579–594 (1979)
120. P. Izar, R.G. Ferreira, T. Sato, Describing the organization of dominance relationships by dominance-directed tree method. Am. J. Primatol. **68**(2), 189–207 (2006)
121. J.A. Sloboda, Music structure and emotional response: some empirical findings. Psychol. Music **19**, 110–120 (1991)
122. T. Schafer, P. Sedlmeier, From the functions of music to music preference. Psychol. Music **37**(3), 279–300 (2009)
123. W.M. Hartmann, *Signals, Sound, and Sensation*. Modern Acoustics and Signal Processing (American Institute of Physics, Woodbury, 1997)
124. T. Giesa, M. Arslan, N. Pugno, M.J. Buehler, Nanoconfinement of spider silk fibrils begets superior strength, extensibility and toughness. Nano Lett. **11**(11), 5038–5046 (2011)
125. E. Bigand, R. Parncutt, F. Lerdahl, Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. Percept. Psychophys. **58**(1), 125–141 (1996)
126. M. Rohrmeier, Towards a generative syntax of tonal harmony. J. Math. Music **5**(1), 35–53 (2011)
127. S.W. Cranford, A. Tarakanova, N. Pugno, M.J. Buehler, Nonlinear material behaviour of spider silk yields robust webs. Nature (2012)
128. N.V. Medhekar, A. Ramasubramaniam, R.S. Ruoff, V.B. Shenoy, Hydrogen bond networks in graphene oxide composite paper: structure and mechanical properties. ACS Nano **4**(4), 2300–2306 (2010)
129. J. Alegre-Cebollada, P. Kosuri, J.A. Rivas-Pardo, J.M. Fernndez, Direct observation of disulfide isomerization in a single protein. Nat. Chem. **3**(11), 882–887 (2011)
130. S. Keten, C.-C. Chou, A.C.T. van Duin, M.J. Buehler, Tunable nanomechanics of protein disulfide bonds in redox microenvironments. J. Mech. Behav. Biomed. Mat. **5**(1), 32–40 (2012)
131. O. Compton, S. Cranford, Z. An, K. Putz, C. Brinson, M.J. Buehler, S.B. Ngyen, Tuning the mechanical properties of graphene oxide paper and its associated polymer nanocomposites by controlling cooperative intersheet hydrogen bonding. ACS Nano **6**(3), 2008–2019 (2011)
132. P.K. Zysset, X.E. Guo, C.E. Hoffler, K.E. Moore, S.A. Goldstein, Elastic modulus and hardness of cortical and trabecular bone lamellae measured by nanoindentation in the human femur. J. Biomech. **32**(10), 1005–1012 (1999)
133. M.D. Lima, S.L. Fang, X. Lepro, C. Lewis, R. Ovalle-Robles, J. Carretero-Gonzalez, E. Castillo-Martinez, M.E. Kozlov, J.Y. Oh, N. Rawat, C.S. Haines, M.H. Haque, V. Aare, S. Stoughton, A.A. Zakhidov, R.H. Baughman, Biscrolling nanotube sheets and functional guests into yarns. Science **331**(6013), 51–55 (2011)
134. K.T. Nam, D.W. Kim, P.J. Yoo, C.Y. Chiang, N. Meethong, P.T. Hammond, Y.M. Chiang, A.M. Belcher, Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes. Science **312**(5775), 885–888 (2006)
135. N. Huebsch, D.J. Mooney, Inspiration and application in the evolution of biomaterials. Nature **462**(7272), 426–432 (2009)