Frontiers in Bioscience Landmark

Front. Biosci. (Landmark Ed) **2022**; 27(4): 128
https://doi.org/10.31083/j.fbl2704128

Review

Recruitment: A Problem of Entangled Temporal Parts

Gustavo Caetano-Anollés[1,*], M. Fayez Aziz[1], Fizza Mughal[1], Ibrahim Koç[1], Kelsey Caetano-Anollés[2], Derek Caetano-Anollés[3]

[1]Department of Crop Sciences and Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana, IL 61801, USA
[2]Callout Biotech, Albuquerque, NM 87112, USA
[3]Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
*Correspondence: gca@illinois.edu (Gustavo Caetano-Anollés)
Academic Editor: Graham Pawelec
Submitted: 16 January 2022 Revised: 4 April 2022 Accepted: 7 April 2022 Published: 18 April 2022

Abstract

Recruitment is a pervasive activity of life that is at the center of novelty generation and persistence. Without recruitment, novelties cannot spread and biological systems cannot maintain identity through time. Here we explore the problem of identity and change unfolding in space and time. We illustrate recruitment operating at different timescales with metabolic networks, protein domain makeup, the functionome, and the rise of viral 'variants of concern' during the coronavirus disease 2019 (COVID-19) pandemic. We define persistence within a framework of fluxes of matter-energy and information and signal processing in response to internal and external challenges. A 'triangle of persistence' describing reuse, innovation and stasis defines a useful polytope in a phase space of trade-offs between economy, flexibility and robustness. We illustrate how the concept of temporal parts embraced by the perdurantist school provides a processual 4-dimensional 'worm' view of biology that is historical and atemporal. This view is made explicit with chronologies and evolving networks inferred with phylogenomic methodologies. Exploring the origin and evolution of the ribosome reveals recruitment of helical segments and/or large fragments of interacting rRNA molecules in a unification process of accretion that is counteracted by diversification. A biphasic (bow-tie) theory of module generation models this frustrated dynamics. Finally, we further elaborate on a theory of entanglement that takes advantage of the dimensionality reduction offered by holographic principles to propose that short and long-distance interactions are responsible for the increasingly granular and tangled structure of biological systems.

Keywords: Evolution; gene ontology; hierarchical modularity; horizontal exchange; endurantism; persistence; metabolic networks; molecular evolution; molecular functions; origin; perdurantism; proteome; ribosome

1. Introduction

Advances in molecular, structural, genomic and evolutionary biology have made clear that the molecular makeup of cells and viruses is a historical patchwork, a fact that complicates the definition of biological systems. The most widespread and deepest evolutionary transformations involve the process of *recruitment*, the reuse of gradually accumulating molecular innovations to perform functions in different molecular and cellular contexts [1,2]. Evolutionary genomics has also shown that molecular reuse occurs in different temporal contexts [3]. This patchwork-generating process of co-option extends to all levels of biological complexity, from loop motifs and domain structures in proteins to the structure of populations and the diversity of complex biological behaviors. The origin and omnipresence of recruitment however has not been explained in evolutionary biology, nor has its many philosophical difficulties been analyzed. In fact, the problem of recruitment interfaces with the central problem of maintaining identity through space and time, an issue already known to Presocratic scholars such as Parmenides, Heraclitus and Empedocles more than two millennia ago as exemplified by the poems of the Strasbourg papyrus [4]. We have addressed the problem of recruitment within a framework of a 4-dimensional space-time 'worm' theory of entangled temporal parts [5]. Here we extend our initial verbal elaborations. First, we describe how pervasive is recruitment in biology with a number of genomic biology examples. Second, we show recruitment is just one (but central) strategy of several operating in a *triangle of persistence*. Third, we discuss how the concept of temporal parts can help understand persistence, recruitment and evolutionary accretion. Finally, we introduce a theory of entanglement, which can be dissected and tested with networks.

2. Recruitment

Phylogenomic analysis of molecular structures and functions, both of which are highly conserved in evolution, provide decisive evidence of widespread recruitment across biological and temporal scales. Construction of evolutionary chronologies and evolving networks have been used for example to test the age of component parts of biological systems and their interrelationships and dissect recruitment patterns in evolution [3]. Chronologies arrange parts or interactions in the order of their temporal or irreversible occurrence. Networks generally model interactions between components of a system with graphs, in

which vertices (nodes) describe parts and lines (links) describe pairwise interactions between them. Value functions are often mapped onto the nodes and links of the networks. Chronologies and networks reveal evolutionary patchworks in the makeup of the ribosome [6,7], the proteome [8–10], or the functionome [11,12]. We here illustrate the central evolutionary role of recruitment with metabolic networks, protein domain makeup, the functionome, and the mutational landscape of viruses.

2.1 Recruitment in Metabolism

The existence of recruitment in metabolic evolution was already intimated by the early observation that enzymes with homologous β/α fold structures catalyzed similar metabolic reactions across metabolic pathways [13,14]. These types of structural assignments were soon extended to the entire set of the small molecule metabolic pathways of *Escherichia coli* [15,16]. The studies revealed widespread recruitment of structural domains in single domain and mutidomain enzymes present in the proteome of the bacterial model organism. Evolutionary genomic research later confirmed the validity of the patchwork recruitment model of metabolic evolution [17–19]. The availability of a larger number of proteomes enabled to trace the time of origin (age) of enzyme domains on the networks of metabolic pathways, visualized as color diagrams in the Metabolic Ancestry Network (MANET) database [17]. This was made possible by the reconstruction of phylogenomic trees of protein domains [20] and the ability to build evolutionary chronologies (reviewed in [3]). In a later database update and study, evolutionary patterns of domain recruitment were sorted out with an algorithm that derives the most plausible ancestry of an enzyme from structural and evolutionary annotations [18]. The analysis indicated that recruitment of ancient domain structures in modern enzymes was widespread. More recently, MANET 3.0 was used to systematically trace the age of enzyme domains in metabolic networks [19]. Domains were defined at lower 'fold family' level of structural granularity and their times of origin (ages) were reconstructed using structural information present in 8127 proteomes from organisms and viruses. Fig. 1 shows a collage of the 148 metabolic subnetwork diagrams of MANET 3.0 with domain age colored onto enzymatic functions defined by Enzyme Commission (EC) classification. A patchwork is evident in most subnetworks, showing there is little repetition of domain structures or ages of enzymes in consecutive enzymatic steps. A full blown 'purine metabolism' subnetwork makes that evident. The analysis revealed that other types of patterns in metabolic pathways were rare, including retro-evolving and forward-evolving pathways that would support alternative evolutionary scenarios of metabolic growth [21]. Multiple assignments of ages to enzymatic activities also revealed that enzymes were patchworks of structural domains of different origins.

2.2 Recruitment in Proteins and Proteomes

Domains are the structural, functional and evolutionary units of proteins [22–24], the collective of which make up proteomes. For that reason, the evolution of the proteome of an organism can be studied through the domain makeup of its protein constituents. Reconstructing a phylogenomic tree of proteomes and tracing changes of occurrence or abundance of protein domains along its branches revealed a wide spectrum of domain gains and losses occurring throughout the entire tree of life [25]. Consistently, domain gains overshadowed domain losses in all superkingdoms of life supporting a central and pervasive trend of growth in protein evolution. This gain and loss 'yin-yang' resulted in proteomes being highly dynamic evolutionary patchworks of component parts harboring different histories. Significant contributors to this patchwork are processes of domain rearrangement that are responsible for the large diversity of multidomain proteins, including recombination and duplication of genes. Since a significant number of proteins have more than one domain in their structural makeup [26], domains appear in different molecular contexts, individually or combined with other domains. This 'domain organization' often enhances the molecular functionality of proteins and proteomes. Domain organization, in itself, reveals the central evolutionary role of domain recruitment in protein evolution. Studying how domain organization unfolds in proteins and proteomes along a timeline makes recruitment evident. Fig. 2 for example illustrates how a time series of networks that link domains and supradomains to multidomain nodes when proteins share domain makeup make evident a chronology of domain organization [10]. Evolving networks formalize various episodes of domain recruitment by establishing links between nodes describing the domain makeup of protein architectures. Network chronologies reveal biphasic patterns of network recruitment and growth in which younger architectures coopt older counterparts in a 'combinatorial 'big bang' that occurred two-thirds of the way in protein evolution.

2.3 Recruitment of Molecular Functions

Biological functions are the actions of biological 'agents', cellular entities such as those encoded in genes. These actions are observables that are actively occurring ('occurrents'). They represent 'activities', which have a beginning and an end. Their life however extends when activities are defined at high levels of abstraction. For example, modern metabolism as a whole is a universal activity that originated with proteins and is still ongoing on our planet. The Gene Ontology (GO) database has standardized the functional annotation of genes [27]. This annotation involves developing a controlled and species-neutral vocabulary of ontological terms describing the molecular functions (*mf*), biological processes (*bp*) and cellular components (*cc*) of cellular life. GO *mf* terms describe activ-

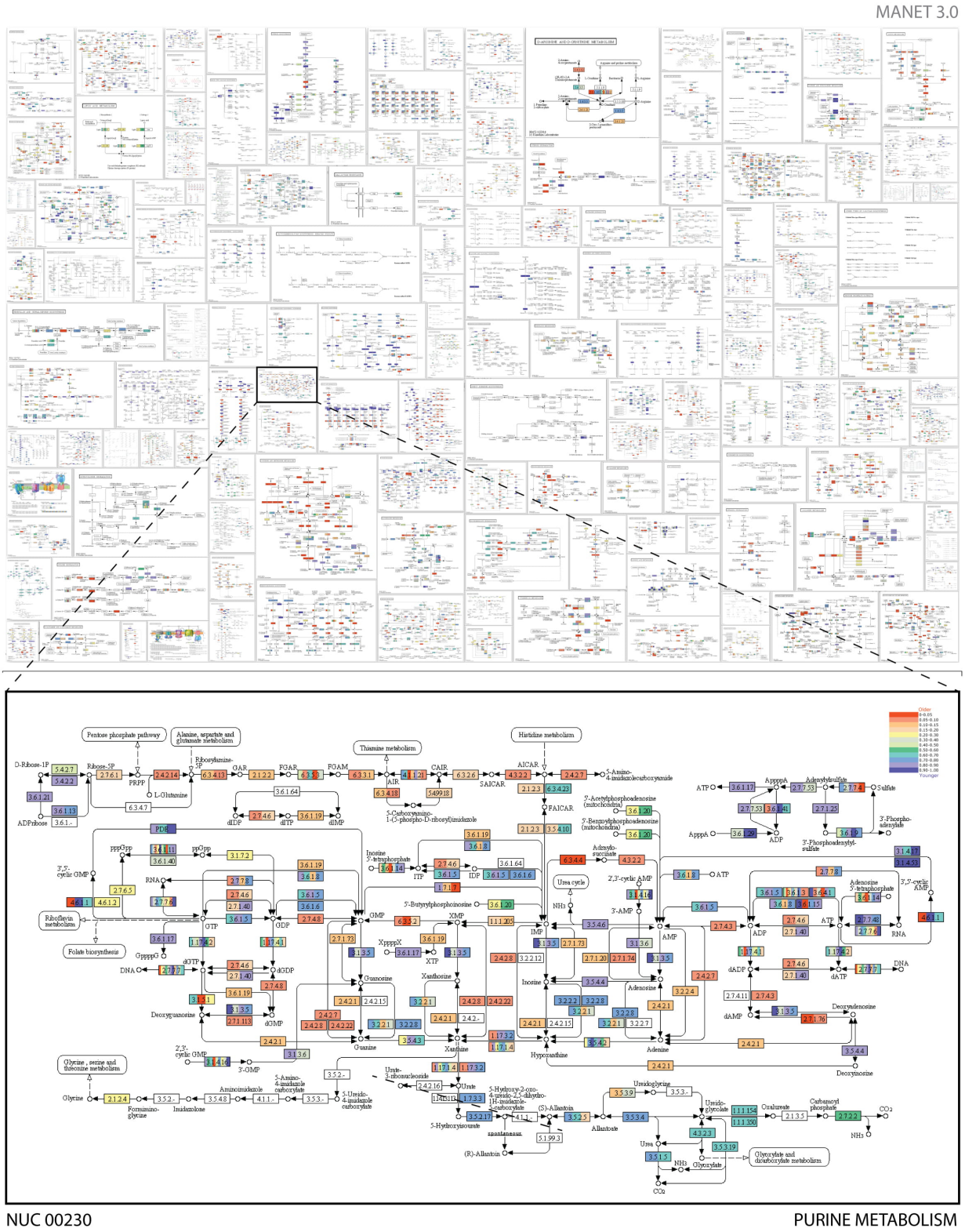

Fig. 1. Visualizing recruitment in metabolic networks. The Molecular Ancestry Network (MANET) database (https://manet.illinois.edu) 'paints' times of origin (age) of structural domains onto enzymes of metabolic pathways (Mughal and Caetano-Anollés 2019). A collage of the 148 subnetwork diagrams of MANET 3.0 (in the top) and the oldest subnetwork, NUC00230 purine metabolism (in the bottom), highlight how pathways are enzymatic patchworks of different evolutionary ages. The database makes use of domain, enzyme and pathway information from the SCOP, PDBsum and KEGG databases, respectively, and evolutionary information in the forms of a color scale of ages derived from structural phylogenomic analyses. The domains of the oldest enzymes are colored in bright reds and the most recent in dark blues.

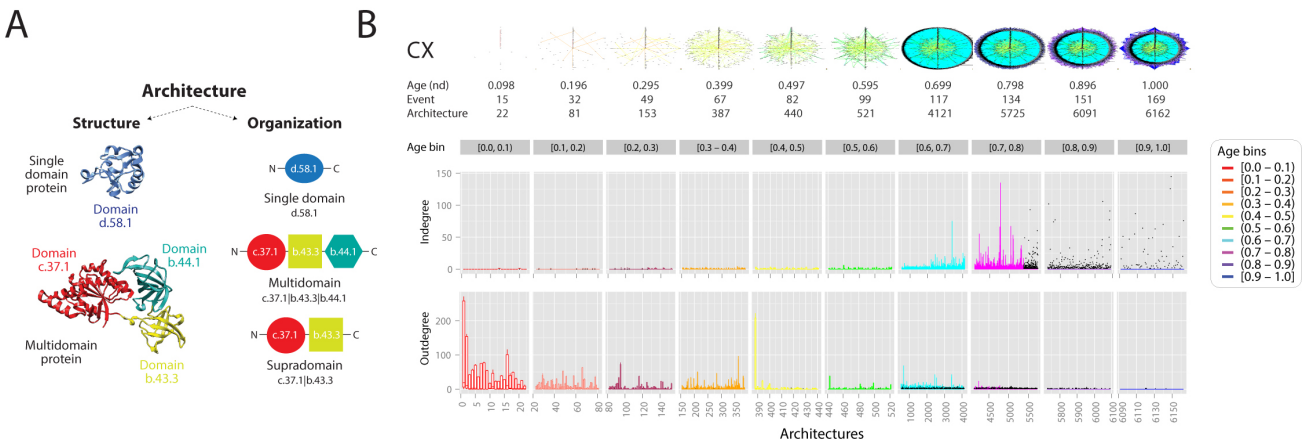

Fig. 2. Evolution of domain organization in proteins visualized with a time series of composition networks (CX) in radial layout. (A) Proteins have 'architectures' defined by their 'structure' (the folding of constituent domains in 3-dimensional space) and their 'organization' (the combinatorial ordering of domains along the polypeptide chain). Here we use the term supradomain as a sub-combination of domains that is sometimes repeatedly present in the architectural census and its modular role can be explored with networks of domain organization. (B) CX networks link domains and supradomains to multidomain nodes when proteins share domain makeup. A total of 6162 nodes appear in evolving networks. They represent protein architectures defined at SCOP fold superfamily level that were surveyed in a phylogenomic census of 749 proteomes drawn from the three superkingdoms of life. Only 10 snapshots of CX network growth are shown out of 169 networks corresponding to individual time events of the evolutionary chronology. Networks are indexed with evolutionary age (nd) of architectures, event number, and number of architectures present at each time event. Ages range from 0 (origin of architectures) to 1 (the present) and evolving networks grow from left to right as time progresses and architectures accumulate in evolution. Tuckey boxplots below networks describe network connectivity measured as node outdegree or indegree along the 169 time events. These plots provide a view of chronological accumulation of links between accumulating architectures, which portray protein recruitment events.

ities that occur at the molecular level (e.g., 'catalytic activity' [GO:0003824], 'transporter activity' [GO:0005488]) performed by individual genes products or by assembled molecular complexes (see **Supplementary Table 1** for example GO term definitions). Note that these activities are not the molecular entities themselves and do not specify where, when, or in what context the action takes place. GO *bp* terms describe series of events arising from the activity of one or more molecular functions. These biological processes are collectives of distinct steps enabled by a repertoire of GO *mf* terms. For example, the 'pyrimidine nucleotide metabolic process' [GO:0006220] is defined as chemical reactions and pathways involving a pyrimidine nucleotide. Similarly, 'signal transduction' [GO:0007165] is a cellular process in which a signal is conveyed to trigger a change in the activity or state of a cell. These example *bp* terms require the action of an ensemble of *mf* terms. Finally, GO *cc* terms describe components that are part of larger objects, such as groups of gene products associated with common functions of the cell (e.g., 'ribosome' [GO:0005840]) or an anatomical cellular structure (e.g. nucleus [GO:0005634]).

Given different levels of GO abstraction, the relationship of GO terms is not only hierarchical but also complex. It can be organized into an integrated object-oriented architecture that can improve the usability of GO terms and the depth of information they contain [28]. GO terms induce a tree-like network structure, in which each of the root terms for *bp*, *mf* and *cc* networks unfolds into an independent directed acyclic graph (DAG) where more specialized lower-level 'child' terms (e.g., 'ATP binding' [GO:0005524]) can be connected with multiple higher-level 'parents' representing broader functional categories (e.g., 'binding' [GO:0005488]) [27]. We note however that annotations begin at the lowest level by assigning GO 'terminal' terms (the lowest in the hierarchy) to genes and then building the DAG structure by associating these terms to higher level GO terms. In this regard, a series of levels of GO classification (level 1, level 2, level 3, …) can be indexed with AmiGO (http://amigo.geneontology.org/amigo), currently an Apache Solr open-source document store of GO information. In a series of classification rounds, level 1 terms are linked directly to the root GO term, level 2 terms are defined as child nodes of level 1 terms, and so on until reaching the terminal GO level. Using this strategy, a census of GO terms at these levels in proteomes can be used to reconstruct phylogenomic trees describing the evolution of GO terms and functionomes [11,12]. These analyses produced a natural history of molecular functions that could be inferred directly from GO data. Unfolding the DAG while mapping level GO *mf* terms along evolutionary chronologies for each level revealed a host of interesting recruitment

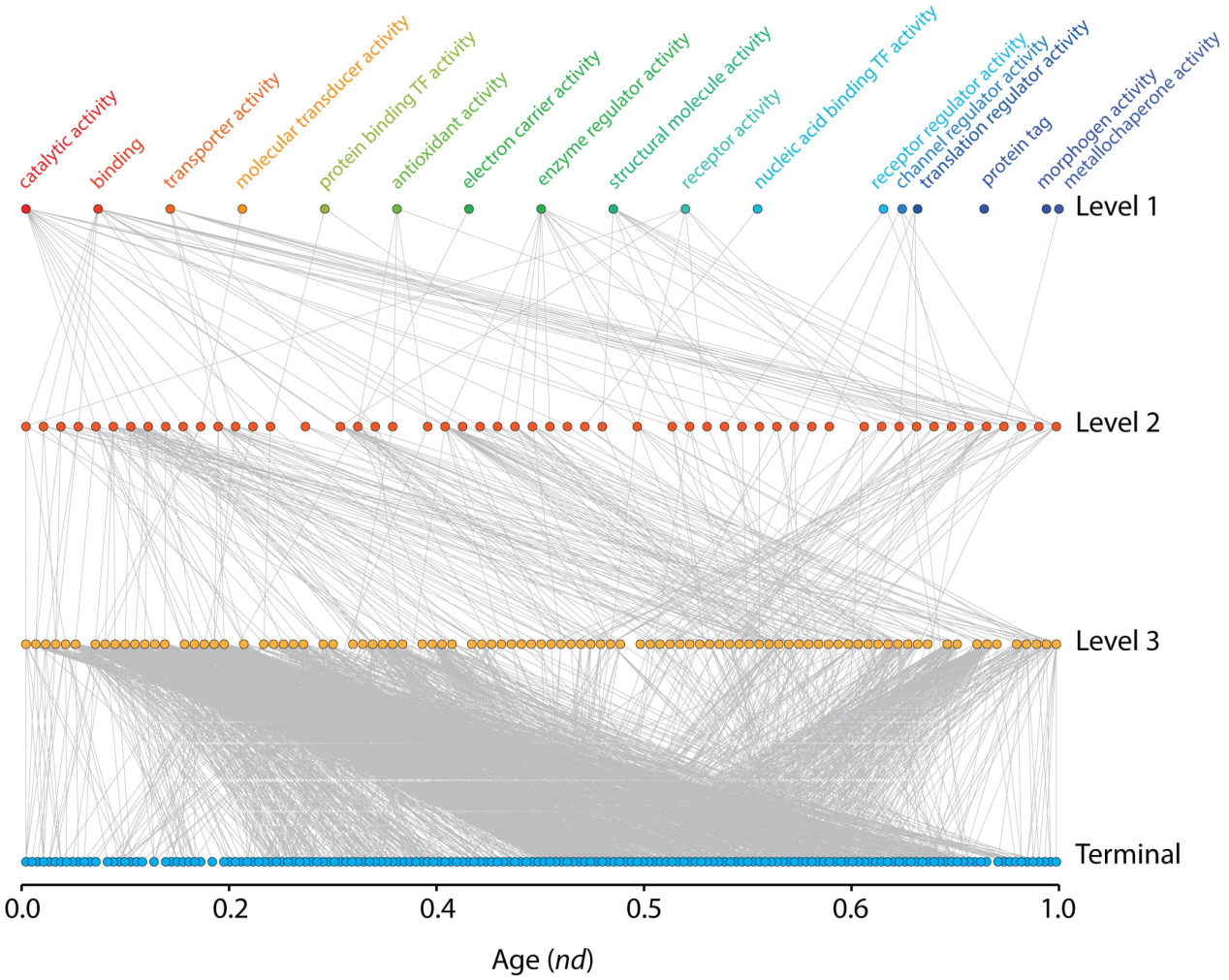

Fig. 3. A gene ontology (GO) network connects different levels of hierarchical organization of 'molecular functions' in a tree-like network structure known as the direct acyclic graph (DAG). Nodes represent GO terms and connections represent associations between lower level (child terms) and higher-level terms (parents). The lowest level involves 'terminal terms'. In the network, GO terms at each level are sorted by time of origin (*nd*), with time (age) flowing from *nd* = 0 to *nd* = 1. The crisscross patterns define a biphasic structure, which reveals the recruitment of ancient and modern functions throughout the timeline.

patterns (Fig. 3). Level 1, 2 and 3 GO terms accumulated linearly in the timeline at similar levels (R^2 = 0.94–0.98) but there was a frustrated interplay between occurrence and abundance of GO terms [12]. For example, the occurrence of individual terms 'boomed' quite late in the chronology but the newly appearing terms showed lower levels of occurrence, probably the result of not having enough time to accumulate in increasingly restricted organismal groups along branches of the Tree of Life. This frustrated interplay resulted in hourglass patterns of recruitment producing crisscross patterns in the DAG that resulted for example from older level 1 terms connecting to level 2 terms appearing throughout the timeline, and younger terms coopting ancient functions (Fig. 3).

2.4 Recruitment of Mutations in the Proteomes of Viruses

One useful strategy to study the emergence of evolutionary novelties and their recruitment in real-time is to take advantage of viral pandemics, especially those caused by RNA viruses [29]. RNA viruses are considered the most common etiological agents of human disease, likely because of their exceptionally short generation times, high infection rates, large populations, high replication error rates, and high levels of mutation and recombination [30]. These properties make them ideal to study viral evolution. Each newly replicated viral genome will carry an average of 1–2 mutations when aligned with the parental sequence. This genetic diversity manifests in the expanding virus population as a 'viral quasispecies', a dynamical collective of viral variants that show genetic linkage through mutation, have shared functions, and collectively contribute to the characteristics of the entire population [31].

We are currently studying genomic change in SARS-CoV-2, the noxious agent of the ongoing coronavirus disease 2019 (COVID-19) pandemic. Millions of genomic sequences carefully curated by the GISAID repository [32] are being sampled worldwide to study mutation pathways and seasonal behavior of the virus [29,33–35]. The appearance and accumulation of amino acid variants in the ~30 proteins that make up the viral proteome are then surveyed to identify mutations that are becoming fixed and those that are lost in the expanding virus population. One particular focus has been mutations that are part of haplotypes (mutation sets that are inherited together) and are linked to seasonal behavior. An initial analysis during the first wave of the pandemic (April 2020) revealed that a first haplotype of four mutations was overtaking the viral population worldwide [33]. This haplotype included the D614G mutation of the spike protein (S-protein) and the P323L mutation of the NSP12 polymerase, which are currently present worldwide in most viral isolates. The haplotype is believed to have increased infectivity by enhancing the flexibility of the spike protein. In fact, D614G breaks a D614-T859 side chain hydrogen bond between the neighboring S1 and S2 subunits of pairs of the three protomers of the spike enhancing flexibility and subunit interactions [36]. It also interacts with residues K854 and Y837 of the fusion peptide (FP) region contributing to linkage and/or allostery between the subunits [37]. We also identified pathways of mutational change associated with regions of protein flexibility and intrinsic disorder [32]. Some of these regions involved amino acid variants now fixed in the 'variant of concern' (VOC) Omicron that has overtaken the viral population worldwide, including an haplotype of two mutations associated with the intrinsically disordered domain linker of the nucleocapsid protein [35]. VOCs are variants of the virus exhibiting a characteristic constellation of mutations associated with statistically significant and experimentally verified increases in clinical or epidemiological criteria of significance (e.g., virus transmissibility). VOCs appeared for the first time at the end of 2020 and have been displacing each other since their inception while retaining selected amino acid substitutions and haplotypes. Fig. 4A (Ref. [35]) for example shows a network of VOCs unified by mutations in the S-protein that are shared between them. Mutation D614G is shared by all 10 VOCs, E484K is shared by 7, and N501Y and P681R by 5. The network reveals that the current VOC Omicron, which first appeared in South Africa in late 2021, harbors the most complex mutation constellation known so far, including all of the widely shared mutations mentioned above. Remarkably, the network also reveals that VOC Mu, which appeared in South America early in 2021 and vanished by the end of the year, established cryptic but common sharing routes between all VOCs. These patterns can only be explained by pervasive recruitment of mutations benefiting viral spread and infection. The Venn diagram of Fig. 4B describes how VOC Omicron has now collected 13 haplotypes in Australia alone, 4 of which were directly recruited from the once highly abundant VOCs Alpha and Delta [35]. These 4 haplotype recruitments involved 15 markers, most of which affected regions of flexibility and intrinsic disorder, including an haplotype of 3 deletions (H69del, V70del and Y144del) affecting exclusively the N-terminal domain (NTD) of the S-protein. We found that the NTD harbors a galectin-like fold structure responsible for a hemisphere-dependent seasonal pattern that was driven by mutational bursts and was consistent with a marked seasonal behavior of COVID-19 [29]. Remarkably, our exploration of differences in mutation accumulation of haplotypes and free-standing markers in regions of Australia that span different latitudes reveal a rationale behind the emergence of VOCs [35]. The noisy rise of haplotypes by recruitment and molecular optimization involved gradual coalescence into monolithic constellations that were only decoupled by virus seasonality. Thus, recruitments appear tailored by the seasonal periodicities of the planet that arise from Earth's tilted axis relative to the plane of its orbit. Our expectation is that COVID-19 outbreaks will soon follow those of other 'winter' viruses, which move across the Earth every year along a sinuous curve parallel to the 'midsummer' curve of solar radiation.

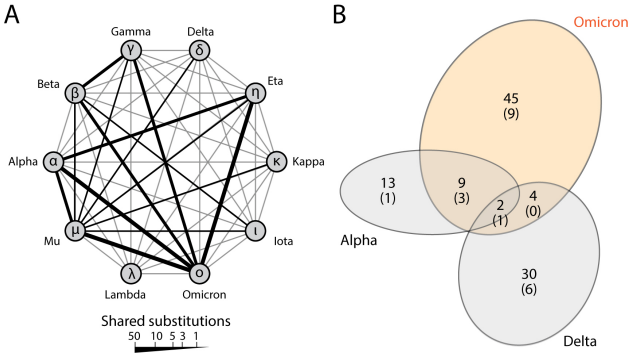

Fig. 4. Recruitment of mutations in the proteomes of Variants of Concern (VOCs) appearing in the evolving SARS-CoV-2 viral population. (A) A network of VOCs unified by the number of amino acid substitutions in the spike protein shared by the viral groups. Data was acquired from the CoVariants site of GISAID. Note that the worldwide spread of VOC Omicron (lineage BA.1) was already massive at the time of the studies of both panels (January 15, 2022). (B) Venn diagram describing how mutations and haplotypes have been recruited by VOC Omicron from the mutant constellations of VOCs Alpha and Delta in Australia. The area of ovals is proportional to the number of mutant markers. Markers and haplotypes in each Venn group are numbered without or within parentheses, respectively. Data from ref. [35].

3. Persistence

Recruitment constitutes a pervasive activity of structured entities (things) that persist (persistents). Biological systems are structured entities [5]. The interaction, behavior and goals of subsets of parts of these systems may be different from the rest. Biological systems are also persistent. They maintain identity through time. Persistence embodies for example the temporal continuance of a structural feature or lineage. Structural features can be varied (e.g., molecular, functional, physiological, behavioral) and often involve *'modules'* (e.g., protein structural domains, RNA junctions, cells), sets of integrated (coordinated) parts of the system that cooperate to perform a task and interact more extensively with each other than with other parts and modules of the system. Cells for example are modules of organisms such as bacteria or multicellular fungi, plants or animals. Because modules are structured entities that are highly conserved in evolution, they are easily recruited to perform related but different tasks. Neuronal cells for example have been classified into distinct cell types on the basis of structural, physiological and genetic characteristics, and in recent efforts on the basis of their transcriptomes (known as cellular t-types) [38]. These classifications reflect their performance of different functional tasks. In contrast, structural lineages involve sequences of structural features reflecting temporal change or unified by an evolutionary theme (e.g., those that descend from ancestors). Lineages include sequences or structures of related nucleic acids or proteins, evolving networks and cellular structure, or evolving cellular organisms or viruses. Structural features and lineages within the GO framework for example correspond to molecular functions and biological processes, respectively.

The driving forces of biological processes are in trade-off relationships and can be better described within a framework of a 'triangle of persistence' [39]. This triangle describes the impact of the environment on the persistence of a biological system and its modular parts. Note that the environment represents an influence that is external to the system and takes the form of a 'suprasystem', a system that embeds the open system under study and acts as a reference. Trade-offs exist when one trait cannot increase with a decrease in another [40]. These trade-offs are often caused by restrictions in matter-energy and information flows that unfold in spacetime and constrain the functioning and evolution of the biological system over the system's initial and boundary conditions. Constraints are here defined as historical imprintings that result from increasing molecular, cellular and higher-level structure in biological systems as well as emergent layers of biological organization. Tacking advantage of von Uexhull's organism-centric view of the environment [41] and Miller's general theory of living systems [42], the 'triangle of persistence' framework maps a budget of matter-energy costs of a system to an 'economy' axis and the system's perception of the environment to both an 'umwelt' axis of 'flexibility' and a 'gap' axis of 'robustness' [39]. Here, the umwelt ('the world around us') corresponds to the totality of signals that undergo sensory processing. In turn, the gap reflects signals that are not perceived and cannot be processed but that indirectly impact robustness through redundancy and reliability of internal constituents. The sum of umwelt and gap make up the totality of signals, the 'scope' of a system, which reflects unique experiences (genetic, epigenetic, physiological, behavioral) triggered by unique environmental (suprasystem) changes impinging on that system. Signals (also known as signs) are here defined as semiotic entities that communicate a meaning and are causally related to objects (e.g., biological entities) through interpretation of some sort.

Fig. 5 summarizes the theoretical approach. In the *economy-flexibility-robustness* trade-off (phase) space of performance of the figure, the economy vertex reflects the budget of matter-energy costs of a system, the flexibility vertex describes structural and functional mechanisms requiring processing of information needed to respond and adapt to internal and external challenges, and the robustness vertex embodies mechanisms that use information to maintain structure and function in the face of environment-induced damage and change. The spheres in the triangle diagrams depict a cloud of points in the phase space that locate in Pareto fronts, boundaries in multidimensional performance spaces that provide best fitness solutions. Performance spaces in the triangle of persistence are world of traits resulting from processes that associate with the strategies of economy, flexibility and robustness. The geometries of fitness solutions have been mathematically elaborated [43]: line segments, planes or polygons arising from trade-offs in performance spaces of 2, 3 or more dimensions, respectively.

Biological systems operate by using concrete entities, which cost matter and energy to produce and maintain. These entities must arise as biological novelties at some point in evolution and must be made persistent by incurring in information costs. For example, a multidomain protein at some point emerged as a functional unit by coopting the structure and functions of its individual domains, resulting in matter-energy costs associated with the synthesis of the longer multi-domain protein and the maintenance of its structure and function by mechanisms of information dissipation (e.g., enhancing error-correction mechanisms of proofreading). Within the context of recruitment, persistence manifests as trade-offs between levels of reuse of novelties (*reuse*), which are constrained by matter-energy, and the system's ability to introduce novelties (*innovation*) and tolerate redundancy (*stasis*), both of which are constrained by information flow. These general strategies of persistence, *reuse-innovation-stasis*, faithfully map to the *economy-flexibility-robustness* mechanisms of the triangle of persistence (Fig. 5B). Establishing levels of reuse implies modulating matter-energy costs related to the reuse of indi-

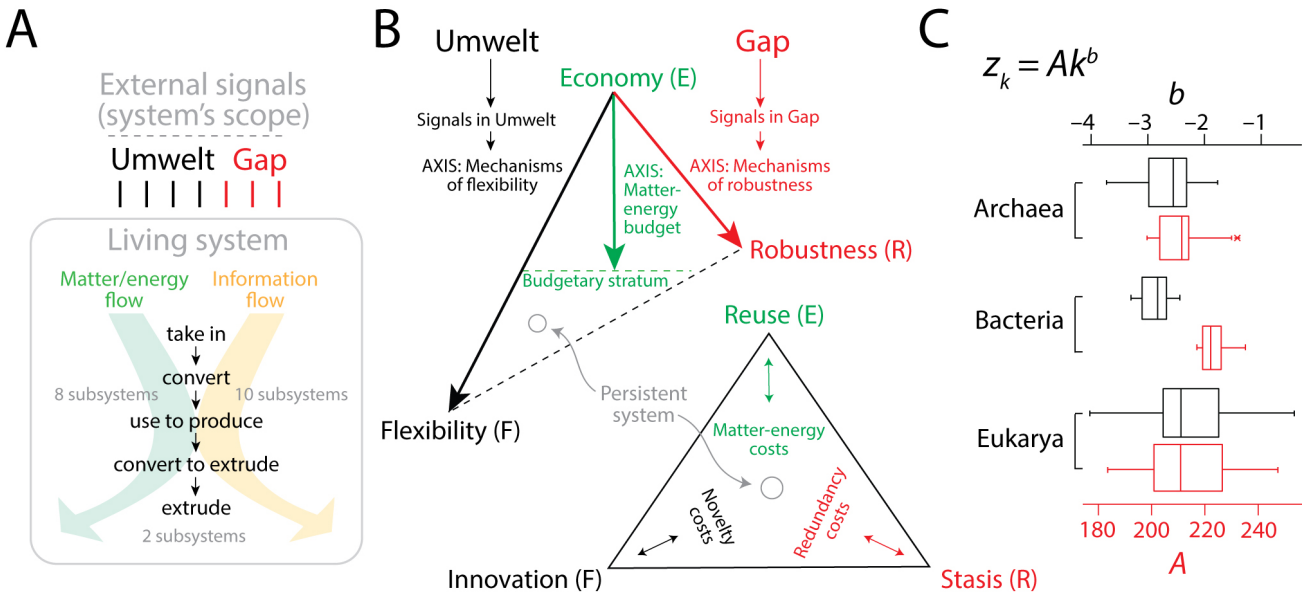

Fig. 5. Building the framework of the 'triangle of persistence'. (A) A living system is exposed to environmental signals (scope), that the system either perceive (umwelt) or not (gap). These signals result in vital matter-energy and information flows, which are processed by 20 critical subsystems. Vital flows enable biological processes and functions of 'action' and 'communication' to occur. (B) The triangle of persistence delimits a phase space of evolution-driven trade-off solutions among economy, flexibility and robustness by mapping a matter-energy budget to and 'economy' axis, umwelt signals to a 'flexibility' axis, and gap signals to a 'robustness' axis (triangle in the top). The flexibility axis measures structural and functional mechanisms that process information needed to respond to internal and external challenges. The robustness axis measures mechanisms that use information to maintain structure and function in the face of environment-induced damage and change. Within the context of recruitment, the triangle of persistence unfolds as a trade-off of *reuse*, *innovation* and *stasis* when the system balances matter-energy, novelty and redundancy costs (triangle in the bottom). (C) A Menzerath-Altmann (MA) law of language exists in the structural domains of proteins, which imposes size-dependent patterns of decreasing returns. The two-parameter law formulation describes how the average length of a domain in a protein (z_k) decreases with the number of structural domains it contains (k) as a straight line in log-log plots (R^2 ranging from 0.85 to 1). Boxplots describe fitting decay parameters b and intercepts A (when k = 1) for the proteomes of the three superkingdoms (data from [45]). Steep slopes b depict increased patterns of decreasing returns and intercept A values reflect lengths of single-domain proteins and upper bounds (economic strata) for the size-dependent minimization principle.

vidual novelties. The reuse of a fold structure to perform related functions for example introduces a matter-energy cost of maintaining structure and function of the reused fold novelty in light of stochastic effects from mutation and/or genomic rearrangements. Similarly, establishing levels of innovation implies modulating the evolutionary costs of generating novelties through exploration of spaces by diffusion (e.g., searching for new structures and functions in the search space of protein sequence) [44]. Here, dissipation of information becomes a tendency of states to diversify by random walks in a space of possibilities. Novel structures enhance the flexibility of the system at a cost of their generation. Finally, establishing levels of stasis implies modulating the evolutionary costs of redundancy, the maintenance of identity in a sea of stochastic change. Redundancy is a desirable characteristic of systems because it enhances persistence. It usually manifests as the repeated occurrence of a system's component in quantities greater that necessary for its activity and is usually defined with respect to activity performance. For example, the repeated use of novelties ensures they will not be lost by environment-induced damage and change. The downside is the cost of maintaining redundancy as redundant components are pervasively degraded by mutational diversification. The existence of hierarchical structure and levels of organization ensures evolutionary constraints will counteract the degrading force, turning redundant components into modules [44]. Thus, matter-energy, novelty generation, and redundancy costs engage in trade-off relationships necessary for recruitment to unfold in evolution.

This landscape of trade-offs was recently made evident in an analysis of 60 proteomes covering all major kingdoms of life and exploring the organization of structural domains in proteins [45]. The analysis revealed that the lengths of domains decreased linearly with increasing domain number, supporting the existence of a Menzerath-Altmann (MA) law of size-dependent decreasing returns (which follow the motto '*the greater the whole, the smaller*

its constituents' [46]). Fitting parameters revealed the broadest patterns of decreasing returns present in Eukarya and a higher push towards economy in Archaea at a lower economic stratum (Fig. 5C). To understand drivers, a persistence function P was formulated with two terms, one reflecting the matter-energy cost of adding domains and extending their length (P_{ME}) and the other reflecting how information present in domain length and number influence the flexibility and robustness of the molecules (P_{FR}) [45]. P was able to define a triangle of persistence and distinguish distinct strategies used by the proteomes of superkingdoms. For example, both archaeal microbes and fungi (and to a lesser degree plants) showed the largest push towards molecular economy, each at its own budgetary stratum. A complicated language-like behavior exists in protein structure that is constrained by universal laws and engineering principles.

4. Temporal Parts

It has become clear to philosophers and bioinformaticians alike that space, time, identity and change are interlinked and that understanding these links is necessary to address the problem of persistence. Persistence is one manifestation of existence that has been the subject of deep philosophical analysis for decades [47]. The perdurantist philosophical school proposes that systems extend through space by having 'spatial parts' in different places, and through time, by having 'temporal parts' at different times [48]. In other words, perdurantists consider things have temporal parts and persist through both time and space by 'perduring'. These things are 4-dimensional persisting entities (*occurrents*) that collectively resemble spatio-temporal worms stretching through time and space. In contrast, endurantists explain systems as being wholly present at every moment of their existence. Things persist by 'enduring' and are 3-dimensional persisting entities (*continuants*) that only exist in the present. While both perdurantists and endurantists agree that there are perduring things (*persistents*), they disagree on how things persist. Perdurantist see the world in 4-dimensions, always thinking historically or predictively. Endurantists see the world in 3-dimensions, always focusing on the mechanics of the present rather than on the history of things. Perdurantists consider all points in time are equally real. For them, existence is tenseless (eternal), with ocurrents existing in the past, present and the future. Endurantists consider that only the present is real, and that time is a transformative force. For them, time is just an ordered series of transformation events. An alternative yet radical flavor of temporal parts theory, 'slice theory' (also known as 'stage theory'), tries reconciliation by proposing that things are not 4-dimensional spacetime worms but are instead momentary slices of those worms [49]. While perdurantists consider things have temporal part slices, each existing at a different time (as long as things are worms composed of related slices, i.e., identity of things can be addressed at particular [synchronic] or multiple [diachronic] times), supporters of slice theory consider things are not the worms but rather the slices of those worms themselves. We note however that semantic 'common sense' accommodation of slice theory has been shown to make it inconsistent and has challenged its validity in favor of the classical perdurantist view [50].

Perdurantism appears in line with modern physics [51] and is more powerful in its ability to lift many philosophical objections than other philosophical explanations of existence and complexity phenomena in biology [5]. It is also compatible with the 'eternalist' B-theory of time that considers that all points in time are equally real events that can be temporally ordered and enjoy the same ontological status [52]. This tenseless view has been recently supported by modal arguments that link metaphysical, human and language conceptualization of real time [53]. While the perdurantism-endurantism debate has translated into addressing problems important for systematic biology (e.g., the species concept) or philosophy of science (e.g., ideographic versus nomothetic thinking), understanding the persistence paradigm is central to the problem of recruitment. In Caetano-Anolleś *et al*. [5], we described some difficulties of the classical perdurantist and endurantist doctrines (sometimes using the Ship of Theseus paradox [54]) and extended their impact to our understanding of causation and evolution. Here we address how the concept of temporal parts can help dissect the evolving *tela vitae* (web of life) that unfolds in living systems, especially during novelty emergence.

4.1 Temporal Parts, Growth and Change

We start by revisiting the relation between temporal parts and change and its link to evolution. When things change their existence changes accordingly, highlighting what is known as the problem of change. Because the 'problem of change' demands an 'explanation of change' [47], there is a need to explain why things change in space and/or time to begin with. For example, a protein changes with the passage of time without losing or gaining any of its physical parts when its conformation changes at nanosecond timescales. Similarly, a growing protein lodged in a ribosomal channel changes with time when ribosomal actions (mechanical movement, molecular recognition, chemical reactions) add a specific amino acid residue to its growing polypeptide chain at timescales of seconds. Both examples are similar but different because they involve spatial and temporal changes of different type. While different protein conformations involve a same set of physical parts unfolding as different slices of time, the 'elongating' protein example appears to show temporal slices composed of different physical parts. To clarify the difference, let us explore change in time and space. While a temporal part can be considered simply a part of the history of a system [55], temporal parts are also proper parts of a whole spatial en-

tity. They are spatial (physical) things with a temporal dimension. Mereologically, proper parts are parts that when combined together make up the whole, i.e., they account for the whole entity. Temporal parts are indeed proper parts of a succession of temporal parts. Conversely, spatial parts are proper parts of a given temporal whole. In addition, spatial parts are also proper parts of a temporal part. This 4-dimensional view over history is simply a consequence of parthood and its partial ordering relationships (reflexivity, transitivity and antisymmetry). The view also supports the mereological 'weaker supplementation principle' that nothing has precisely one proper part and the 'unrestricted composition principle' that for any non-empty set of things there is a mereological fusion (a composite) of all those things. The problem is that while things can be in more than one place at a time (e.g., proteins can have atoms that occupy different positions) they can also take up both space and time (e.g., atomic positions can change in a different protein conformation) and in doing so can gain and loose properties (e.g., parthood relationships, shape) that characterize things 'in and of themselves', not in relation to anything else. This is referred to as the problem of 'temporary intrinsics'. In addition, the identity of temporal parts also changes in time and space but in a relative sense because temporal parts are part of an unfolding spatio-temporal worm regardless of their morphing properties (e.g., conformations are different but yet the protein is the same). In other words, while the perdurantist view of persistence remains atemporal, the temporal parts of an entity can be used to describe change in time and space. This is particularly relevant for cases of network-generating fusions or fissions, such as our previous elongating protein example, and for growth, such as the unification of parts to form more complex biological wholes. The addition of an amino acid to the growing polypeptide lodged in the ribosomal tunnel adds an additional and previously non-existent physical part (an amino acid residue) to the physical ensemble. The initial state (elongating protein without extra residue) is a temporal part of the whole (protein holding the extra residue), even if one of its physical parts (the extra residue) was absent. For perdurantists this is unproblematic because the nonexistent part gains parthood when the whole is created (addressing Plutarch's ontological issue of 'things that grow'). In fact, it is actually the relational abstraction of how temporal parts make up the whole what is important under the tenseless paradigm of temporal parts in the past, present and future. This makes the problem of change irrelevant since explanations are confined to parthoods and properties of temporal parts and not to transformation events or mechanistic explanations of time. The identities of worms are atemporal and the historical perdurantist view allows to focus on the processes operating behind change, a property that enables a processual view of biology (e.g., [56]). In contrast, endurantists reject the proposition that a prior form of an elongating protein existed prior to the addition of the extra amino acid residue. They claim that the 'doctrine of arbitrary undetached parts' (DAUP), which supports the existence of objects being part of others, is false [57], making it illegitimate to add or subtract parts because in doing so the objects go out of existence. Of course, these claims of 'eliminativism' bring a number of additional difficulties, which we will not discuss. In particular, both the definition of objects in abstract sense [58] and a careful analysis of Chrysippus' paradox [59] of Dion and Theon [60] have been shown to nullify the falsification of DAUP and support the idea that things survive the addition or deletion of parts.

4.2 An Example with the Origin and Evolution of the Ribosome

Occurrents are collections of temporal parts. These collectives help unravel processual views. Fig. 6 for example shows a set of temporal parts making up the Eiffel Tower and a similar set making up the ribosome. Each example provides a valid description of a process, one describing the construction of an engineered object, the other describing the evolutionary construction of life's central molecular machine. Note how physical parts are added to evolving occurrents as both the Eiffel Tower and the ribosome molecular core materialize with the passage of time during their 'construction' phases (Phases 1). During their stable existence (Phases 2), the engineered and natural objects may change in ways our cartoon depictions may not show. For example, the material makeup of the tower may degrade while the ribosomal molecular ensemble may diversify. Under the temporal parts paradigm, these changes do not detract from the existence of an Eiffel Tower or of a ribosome. Note that occurrents described in Fig. 6 involve temporal parts of the past and the present. They do not involve temporal parts appearing in the future. We consider that engineering elements of the Eiffel Tower will continue to unfold and the history of temporal parts of the tower will eventually materialize in unanticipated ways. Similarly, levels of ribosomal organization will continue to unfold in the future with recruitments and cooptions. The chronological scheme we have presented to describe temporal parts of a developing 3-dimensional worm is general. For example, a similar collection of occurrents can be used to describe the development of a human individual, starting from the fertilization of the egg and ending with the death and decomposition of the human body. In these chronologies, the evolving systems can be studied by focusing on different properties. In our example, these properties are physical component parts (steel beams, RNA scaffolds, proteins). However, our focus could shift to function or fitness characteristics.

Perdurantists view systems as movies while endurantist focus on individual still frames. The historical view of temporal parts seeks to provide ideographic (retrodictive) insight into processes. The mechanistic view of endurantists renders nomothetic (universal) insight into the present state of a system (a continuant). While both seek predic-

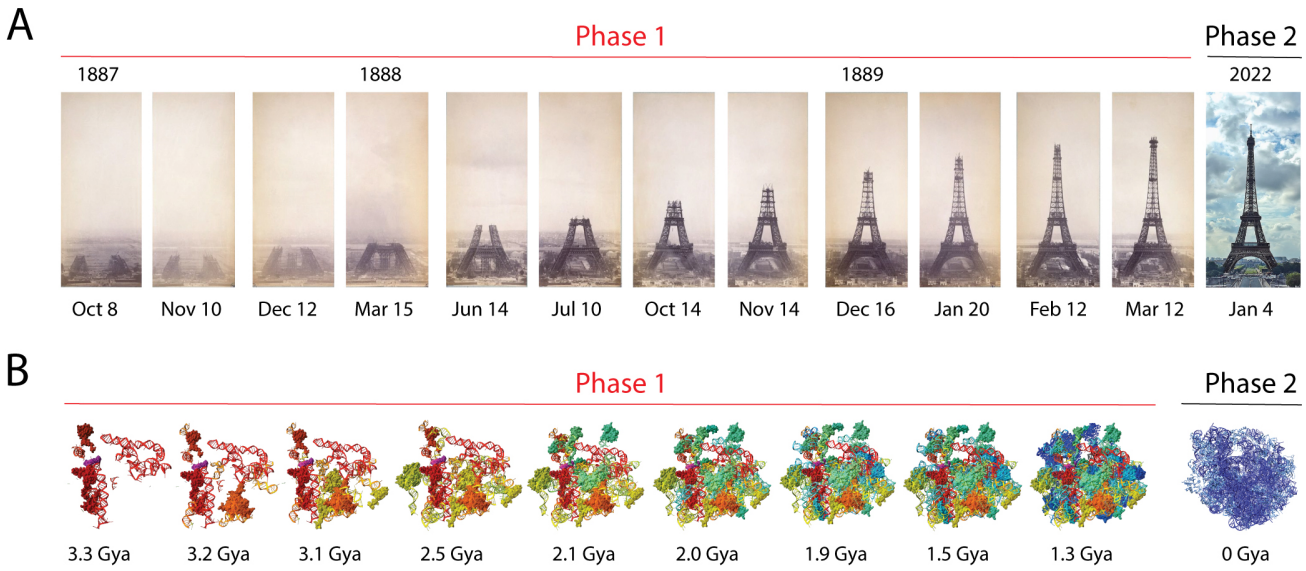

Fig. 6. Temporal parts of both an engineered and a natural object. (A) Eiffel Tower while being constructed (Phase 1) and at an arbitrary time during its present existence (Phase 2). (B) The ribosome of *Thermus thermophilus* during the origination of its universal core (Phase 1) and at present time (Phase 2), with time depicted in billions of years (Gy). Pictures of the Eiffel Tower were retrieved from public record. Crystallographic models of the evolving ribosome were renditions portraying likely temporal parts of the central molecular machine of the cell.

tion from their historical or mechanistic elaborations, their approach to exploration is different. To help unravel processual views of biology, let us focus on the origin and evolution of the ribosome. Ribosomes are the most central molecular components of the cell. They are responsible for most of protein synthesis and for facilitating the folding of emergent proteins. Ribosomes are ribonucleoprotein complexes typically composed of two subunits, a small subunit (SSU) with one ribosomal RNA (16S/18S rRNA) molecule holding ~50 universal helical structures that fold independently into 3 major domains, and a large subunit (LSU) typically containing 2–3 rRNA molecules (23S/28S and 5S/5.8S rRNA) with ~100 universal helices that fold into 6 domains and 5S rRNA. Both subunits act as scaffolds of numerous ribosomal proteins. The LSU structure holds the peptidyl transferase center (PTC) responsible for protein biosynthesis and substructures specialized in ribosomal mechanics and energetics such as the L1 and L7/12 stalks, the GTPase center of the central protuberance (CP), and the α-sarcin-ricin loop (SRL). The SSU structure holds the central ratcheting and genetic decoding mechanisms. Our understanding of the origin and evolution of the ribosome has substantially increased in this past decade, especially because of the use of phylogenomic methods that can build evolutionary chronologies (e.g., Fig. 6B; reviewed in [7]). Analyses of ribosomal history uncovered the early appearance of structures supporting mRNA decoding and tRNA translocation, the coevolution of ribosomal proteins and RNA, and a first evolutionary transition that brings ribosomal subunits together into a processive protein biosynthetic complex (e.g., [6]). Structural studies of tertiary interactions and putative ancient insertions in rRNA, together with statistical analysis of in-silico designed RNA rings, complemented and enhanced these phylogenomic findings (e.g., [61–63]).

As with the Eiffel Tower, the evolutionary origins of the multiple components that make up the ribosome are expected to have occurred gradually in a process of accretion. In this process of unification, individual parts are added piecemeal to a growing whole. Construction of phylogenomic trees describing the appearance of rRNA helices and structural domains of ribosomal proteins provide decisive evidence in support of this piecemeal model of growth [6]. Fig. 7A shows crystallographic models of subunit rRNAs with helices colored according to their evolutionary age of origin, showing a patchwork of colors indicative of their gradual and complex origin. While phylogenomic data supported the rather stochastic evolutionary addition of component parts to the growing domains of the ribosomal subunits, phylogenies could not clearly dissect alternative models of ribosomal growth in which molecular parts are added separately to growing cores to produce modern ribosomal subunits, or to separate growing substructures, which then accreted into forming those cores. These two alternative models (Fig. 7B) describe either a steady accretion process or a more hierarchical alternative in which modules first form and then combine to form more complex structure. Both recruitment alternatives are of significance.

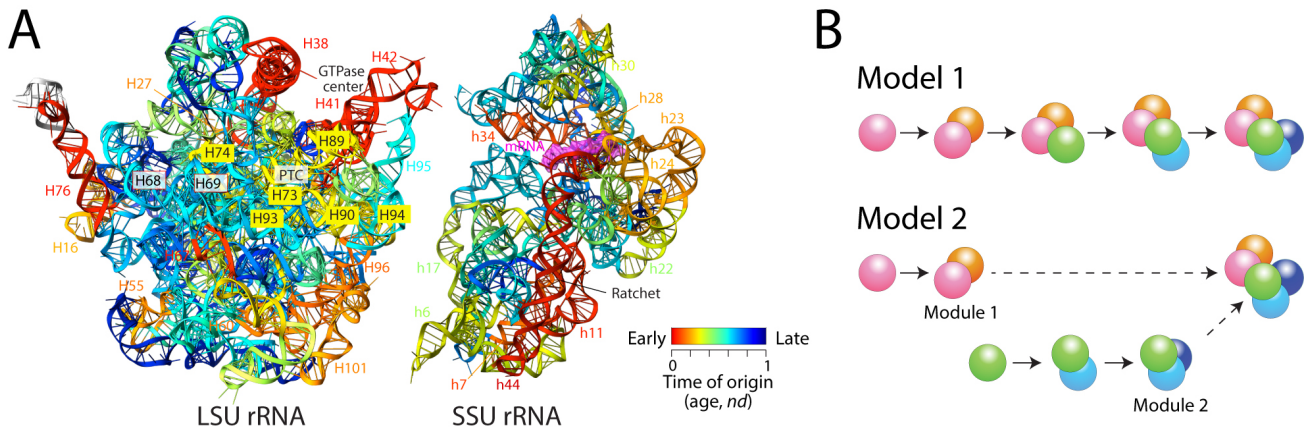

Fig. 7. A patchwork of temporal parts explains ribosomal RNA (rRNA) helical structures of different evolutionary ages. (A) Evolutionary heat maps of small (SSU) and large (LSU) subunit rRNA of *Thermus thermophilus* with helical structures mapped onto their crystal structures (PDB entries 2WDK and 2WDL) colored according to their age (*nd*). PTC, peptidyl transferase center. Data from [6]. (B) Alternative models of ribosomal growth showing how helical substructures depicted as spheres accrete to form molecular wholes. Model 1 shows a continuous scheme of accretion. Model 2 depicts clumping of substructures into modules, which are then grafted together. We note that in both models substructures can also be lost when they fail to provide structural or functional advantages to the evolving molecular complexes. Substructral loss should be included in the models if it can be measured or retrodicted.

In recent explorations, we used structural evidence to propose that the ancient ribosome was initially composed of separate interacting fragments, some of which grafted together to form larger structures [7,64]. We used patterns of coaxial helical stacking present in ribosomal junctions to identify roadblocks to outward growth of the rRNA molecules that were suggestive of ancestral fragmentation [64]. Fig. 8 describes 19 putative fragments reflecting separate structural origins in SSU and LSU rRNA molecules. These fragments likely contributed specific functions to the emerging ribosomal ensemble. Sequence and structure similarities between reconstructed ancestral ribosomal segments and *in vitro* evolved ligase and RNA polymerase ribozymes revealed ancient functional centers could have endowed ligase and polymerase-like enzymatic activities to a processive core that had common replication and translation functions [6]. Phylogenomic data suggests this core led to the formation of a functional ribosome during a first major ribosomal transition that occurred 3.1 billion years (Gy) ago [6,7]. Remarkably, many of these putative homologies hit the most ancient helical segments of the 19 fragments described in Fig. 8, especially the most ancient helices of fragments 3, 6, 7, 11, 13, and 15. These results support Model 2 of Fig. 7B in which structural modules were recruited to form complex structures and functions. With exceptions (fragments 3,8, 9 and 18), the oldest helices were in the most basal regions of the putative fragments. This facilitated outward growth of the originating modules. The fact that about 80% of fragments originated in basal positions closed to the termini of the fragmented molecules is significant and validates both the phylogenomic model and the ancestral fragment model of the ribosome. The exception could be explained by inward growth mechanisms such as helix reformation [7]. We note that unusual ribosomes exist in basal eukaryotes that are made of covalently non-continuous rRNA [65]. A crystallographic model of the large ribosomal subunit of *Euglena gracilis* for example revealed it is made of 14 discrete rRNA fragments that assemble non-covalently into the canonical subunit structure and harbors numerous segments [66]. This shows that in some lineages some of the fragments have not joined into standard subunit arrangements. In fact, 5S/5.8S rRNA may well represent a fragment that has remained separate from the central structure despite its close interactions with the CP structure of the LSU subunit. The idea of primordial rRNA being fragmented systems was already advanced three decades ago [67], but in general its significance has been neglected. It clearly demands further exploration.

Historical fragmentation of rRNA structure suggests a hierarchical structure driven by recruitment is embedded in the emergence of the molecular system. Mapping times of origin of the fragments in a chronology reveals significant temporal heterogeneities in the appearance of the fragments (Fig. 8C). In LSU, a clump of 4 fragments (mostly peripheral in the modern ribosomal structure) appears early and is followed by another clump of fragments during the early stages of ribosomal evolution. These initial structures are involved in ribosomal translocation, mRNA decoding, and helicase activities. In particular, the chronology highlights the late appearance of the PTC structure responsible for protein biosynthesis at the time of the major ribosomal transition (*nd* = 0.3). This occurred by joining fragments 1 and 10 to form an initial 4-way junction (Fig. 8B). Remarkably, the LSU rRNA structure of *Euglena gracilis* [66]

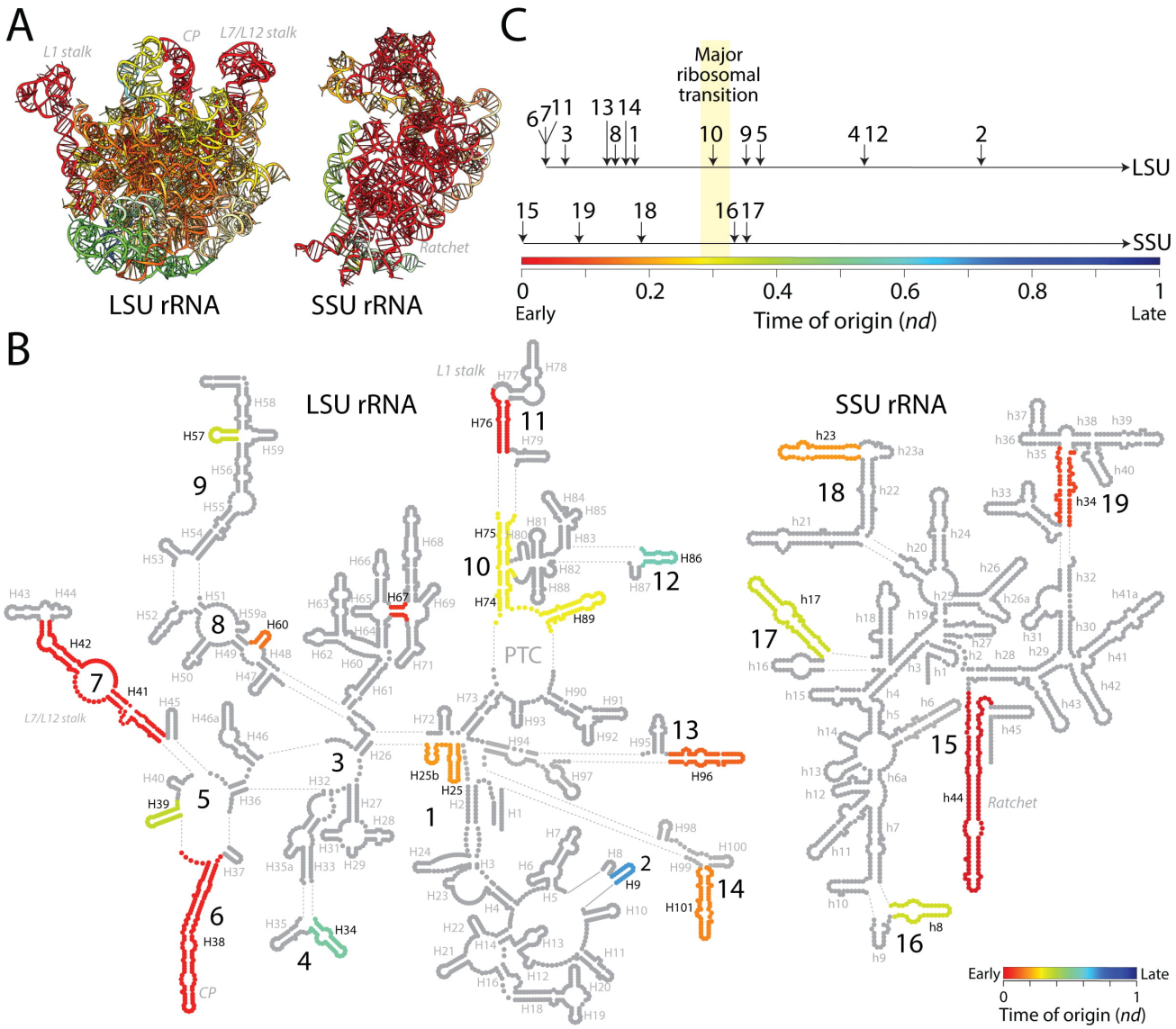

Fig. 8. Structural analysis of ribosomal RNA (rRNA) reveals putative fragments of different origins. (A) Crystallographic model of the large (LSU) and small (SSU) subunit rRNA (*Escherichia coli*; PDB entry 4V9D) with helical structures colored according to 19 segments reflecting distinct putative structural origins of the molecules. An analysis of junctions revealed 17 insertion fingerprints suggesting inward growth or grafting of 19 fragments of different origin. (B) Fragments dissected from secondary structure models of the LSU and SSU rRNA. The oldest helical structures of each fragment are colored according to a scale that reflects their time of origin (age, nd). The other helical structures have ages that span the entire temporal scale but are colored in gray to highlight only the oldest substructures. (C) Time of origins of individual fragments are mapped onto a chronology for LSU and SSU rRNA. Note the very early and heterogeneous appearance of most putative fragments.

shows that the 5-way junction subtending the PTC is also made of two fragments involving segments of fragments 1 and 10. The significance of hierarchy in the chronology of temporal parts will be made explicit below, but suggests parts can be structured in complex manner in evolving systems.

4.3 Identity and Evolutionary Change

The ribosome illustrates the centrality of parts when describing the evolutionary persistence of systems. It also illustrates the difficulties of defining the identity of the evolving whole. Can we even say there was a ribosome before the first major ribosomal transition? Perhaps the emerging ribosome at that time was only a loose ensemble of parts with a set of different functions. Chisholm [68] introduced a mereological theory of persistence that treated identity in a conventional ('loose and popular') sense. According to his theory, composites of parts such as those of the growing ribosome exist in their own right ('*ens per se*') at any moment. When these entities that have parts un-

fold in time, they become successive composites ('*ens successivum*') making up temporal parts. Entities also exist through other things by convention ('*ens per alio*'): "*They are ontological parasites that derive all of their properties from other things – from the various things that do duty for them.*" [68]. While they say nothing on their own, when *entia per alio* exist then *entia per se* must exist. Thus, the functions of evolving ribosomal fragments derived from ancestral similarities can in loose and popular sense tell us stories about ribosomal evolution. In addition, a succession of composites requires establishing how one evolves from another. For that purpose, Chrisholm introduces a mereological persistence view of direct evolution:

Persistence: '*X* evolves directly from *Y* (is a successor) if and only if (iff) either *X* is identical to *Y* or there is no time at which *X* and *Y* both exist but there is a *z* such that *z* is a part of *Y* at one time and *z* is a part of *X* at a later time'.

Such definition of successors follows Frege's 'fatherhood' (predecessor) relation of an ancestral relation [69] and is used by Chisholm to highlight how parts are important elements not only of identity but also of evolutionary change. The definition however explains persistence imperfectly and does not provide a directionality of change that would establish the flow of time. Remarkably, Hennig [70], the father of systematic biology, elaborated four years earlier a framework that could have addressed some of the difficulties of Chisholm's logical constructions. He proposed that 'shared and derived' properties, which he named 'synapomorphies', could test evolutionary hypothesis of history. These hypothesis of ancient origin of identities (homologies) have enabled retrodiction and phylogenetic analyses. In Caetano-Anolleś *et al.* [5], we modified Chisholm's definition of direct evolution to account for diversification (fission) processes that could explain temporal splitting in a tree-like ground plan of change and for unification (fusion) processes that could better explain molecular accretion, multiple origins (e.g., ribosomal fragments), and reticulation:

Fission: 'X and W evolve directly from *Y* iff either *X* is identical to *Y* and *W* or there is no time at which *X* and *Y* or *W* and *Y* both exist but there is a *z* such that *z* is a part of *Y* at one time and *z* is a part of *X* and *W* at a later time'.

Fusion: 'Y evolves directly from *X* and *W* iff either *Y* is identical to *X* and *W* or there is no time at which *Y* and *X* or *Y* and *W* both exist but there is a *z* such that *z* is a part of *X* and *W* at one time and *z* is part of *Y* at a later time'.

Under these general definitions, which are likely imperfect, parts and their properties can transmit change, and identity can be made strict or loose according to the demands of each system being studied. For example, Hull [71] considered individuality was constrained by mereological substance, patterns of interaction of parts, and continuous existence through time. Diachronic identity was preserved if the internal organization of successive states of an evolving system-maintained continuity, even in the presence of splits and mergers. Note that *z* is a part that is shared among successive composites. There is nothing that forbids these parts from being spatial, temporal or both. This sharing can become 'derived' when time is made explicit in the definition and useful for retrodiction if enough information is present in the evolving system.

5. Entanglement

Fusions and fissions describe two opposing forces, one of unification and the other of diversification. Both forces are at play in chronologies of temporal parts such as those of metabolism, proteins, molecular functions, viral change, or the ribosome, making any evolutionary ground plan a directed network. This intuition, which was already recounted in Empedocles' poems of the Strasbourg papyrus [4] and made explicit in the Apollonian and Dionysian forces of Nietzsche's *The Birth of Tragedy* [72] (in which order and individuation counteract chaos and dissolution), was recently made part of a theoretical framework of module generation that explains the emergence of hierarchical modularity in evolving networks [73,74]. The framework, which is backed by considerable explanatory evidence, elaborates a linkage-based *biphasic* (*bow-tie*) *model* that predicts both the evolutionary emergence of nested hierarchies of modules and the convergence under optimization or selection of those modules into tightly linked groups, which are then free to diversify and generate a new level of biological organization. In a first phase, parts of a biological system are initially weakly linked and associate variously. As these initial parts diversify, they compete with each other and are selected for performance. The emerging interactions constrain both their structure and organization. This causes parts to organize into modules with tight linkage. In a second phase, variants of the modules evolve and become new parts for a new generative cycle of higher-level organization. The concept of linkage can be formalized with networks and network chronologies, which can test the emergence of 'communities' and 'hierarchy' in the evolving networks. These communities are modules, groups of nodes that are more densely connected with themselves than with the rest of the network. They can be loosely or tightly linked with each other but often associate in biological networks to form a multi-level hierarchical structure.

5.1 Networks as Dynamic Patchworks of Temporal Parts

Evolving networks are *flow networks*, directed networks that describe flows of matter-energy, information, and/or time. Flows are described by using arcs (connecting arrows defining the direction of travel on a network) or by labeling nodes and/or links with time or other flow descriptors typical of a chronology (Fig. 9, Ref. [75]). Flow networks can be structured depending on how much they follow a typical hierarchical tree structure. Three characteristics describe the perfect hierarchical organization:

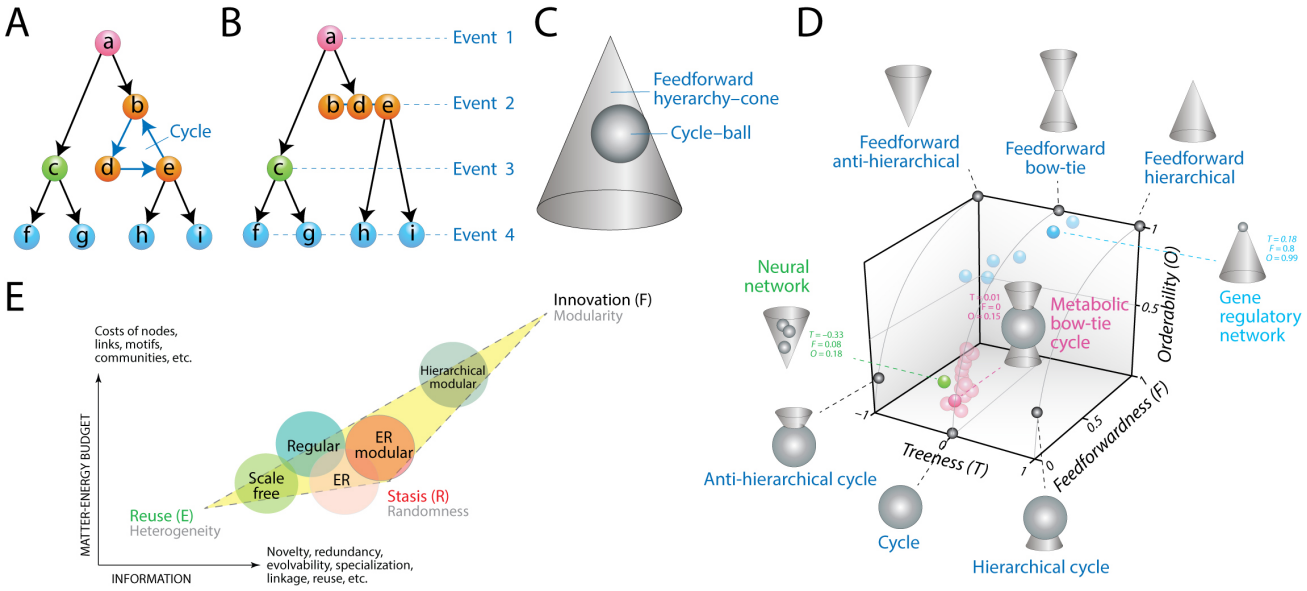

Fig. 9. The anatomy of network entanglement. (A) A feed-forward tree-like (hierarchical) network with a cycle. Arc connections (arrows) are links that flow in one direction (except for those in cycles). (B) When time flows through a network, the network can be annotated with time events that assign a date to individual nodes. (C) An icon of cones and balls can be used to describe the feedforward and cycle structure of networks of the type described in panel A. The placement of the ball in the icon describes the approximate placement of cyclic modules in the pyramid structure. (D) A morphospace of network hierarchies describes possible network structures existing in a phase pace of networks. Coordinates for some real biological networks are identified with colored spheres and live in an O-F plane region of the morphospace centered around T = 0 and enriched in bow-tie network structures. Data from [75]. (E) Networks with typical network topologies unfold in a persistence triangle finding optimal trade-off solutions given a matter-energy budget and information necessary to unfold novelty, redundancy, evolvability and other flexibility-inducing drivers. Trade-offs are mapped to a polytope (shaded in yellow), with circles describing locations of networks that are scale-free, regular, Erdös-Rényi (ER) random, ER modular, and hierarchical modular.

(i) *Order*: When *travelling* along the wiring diagram, paths (walks traversing nodes that are all different) that begin and end in a same node form 'cycles', return structures that interfere or slow down flows. Order describes a tendency of nodes to be arranged unambiguously without interference from cycles. Ordered networks have few cycles, disordered ones are almost complete tangles.

(ii) *Reversibility*: This property defines how much of the network follows the moto 'only one commander for any commanded'. A fully reversible network shows only one commander, an irreversible one shows none.

(iii) *Pyramidal structure*: This property describes the existence of a feedforward tree-like pattern (hierarchical or anti-hierarchical) resembling a pyramid in which a 'commander commands more than one node', there is only a single node that is not commanded by another node, and all downstream nodes of the pyramid are subjected to a chain of command of the same length. Deviations from these requirements result in non-pyramidal tangled structures.

A number of metrics can evaluate order, reversibility and pyramidal structure. For example, Corominas-Murtra *et al.* [75] used *orderability* (O), *feedforwardness* (F) and *treeness* (T) to define a morphospace of directed networks (Fig. 9D). O measures order by calculating the fraction of nodes that do not form cycles (O = 0 implies maximum cycle-induced disorder). F measures the impact of network modules that cannot be ordered on the feedforward structure. (F = 0 implies minimum order). Together, O and F (and the O-F plane in the morphospace) define the number and location of cycles in the pyramidal structure. Finally, T accounts for both reversibility and pyramidal structure by measuring how ambiguous is the chain of command. Sliding the O-F plane along the T axis transform an anti-hierarchical (T = –1) to a hierarchical (T = 1) structure by transition through bowtie (T = 0) symmetric structures. Note than the accretion process of the ribosome has a typical feedforward antihierarchical structure and a diversification process along the branches of a phylogenetic tree without reticulations follows a typical feedforward hierarchical structure. This transition is illustrated in the morphospace with icons of cones and balls (Fig. 9C). Remarkably, most biological networks are located in the O-F plane region of the morphospace that is centered around T = 0 and is enriched in bow-ties, including metabolic, neuronal, gene regulatory, genealogy, food web, social and language networks. Well balanced bow-tie hierarchical structures are

present in metabolism and social networks, but also in the products of mankind, including electronic circuits and software [75]. In all of these examples, what flows through the networks is either matter-energy or information. When time flows for example through networks of proteins (Fig. 2) or molecular functions (Fig. 3), recruitment processes have been shown to generate biphasic patterns typical of bow-tie network structure. These types of multilayered hierarchical entangled organizations result from the emergence of hierarchical modularity and are induced by the unification-diversification forces that operate in network evolution.

Real networks are heterogeneous, modular and strike a balance between order and randomness [76]. For that reason, the morphospace paradigm of Fig. 9 can be further used to dissect network entanglements. Indeed, measuring heterogeneity of link distributions with entropy measurements, modular architecture with correlations of degree distributions and information transfer, and the amount of network randomness, Solé and Valverde [77] established a zoo of possible networks and mapped real networks to this morphospace. Again, real networks expressed varying levels of heterogeneity (scalefreeness), modularity and randomness. Remarkably, these three strategies associate with the reuse, innovation and stasis vertices of the triangle of persistence we previously introduced to describe drivers of recruitment (Fig. 9E). While a matter-energy axis delimits the budget of nodes, links, motifs communities and even dynamic behaviors (establishing a matter-energy budgetary stratum for different network structures), drivers of novelty, redundancy, evolvability and specialization, linkage and reuse, to name a few, take advantage of information-processing mechanisms that rewire networks in response to changing environments. The two axes define a phase space in which reuse (economy) modulates network heterogeneity, innovation (flexibility) rewires networks by endowing them with hierarchy and modularity, and stasis (robustness) controls levels of network randomness. As we described above, phylogenomic-based chronologies have shown that hierarchical modularity of networks emerges in evolution within a stochastic background in evolving networks at different timescales and complexity levels [74]. Thus, the structure of networks is entangled by hierarchy, community structure, and noise and so does the systems of temporal parts those networks represent.

5.2 A Theory of Entanglement

The biphasic (bow-tie) model does not address the origin of the forces of unification and diversification responsible for entangled systems [73,74]. What causes nodes to get entangled? Within the framework of temporal parts described above and inspired by Verlinde's conjecture on the entropic origins of gravity for the idealized anti-de Sitter [78] and standard de Sitter [79] spacetime geometries, we recently proposed a theory of entanglement that would explain causal relationships responsible for the increasingly extended and complex makeup of biomolecules [5]. Entropic forces are macroscopic forces that originate in systems holding numerous degrees of freedom and tending to increase entropy by dissipation. Dissipation is the loss, dispersion or diffusion of matter-energy and information in natural systems. We previously discussed the connection between dissipation and the structure of those natural systems [44]. Under initial conditions however dissipation forces are minimal and changes in information (measured as an entropy) can lead to emergent properties. Because space is the first storage of information in the universe (holding positions and movements of particles and bodies), Verlinde unfolds gravity and spacetime in a holographic horizon taking advantage of the Maldacena duality. This holographic duality, also known as the AdS/CFT correspondence, is the ability to store information in surfaces (screens) as discrete quantum bits without detailing the microscopic dynamics of that storage. Note that surfaces can be stretched horizons separating microscopic data on one side from variables describing spacetime on another, a feature that establishes a direction in which space is emergent. Verlinde shows that entropic gravity arises when space has one emergent holographic direction holding entropic change, degrees of freedom are proportional to the area of the holographic screen, and energy is evenly distributed over degrees of freedom following the equipartition principle. Thus, the universal gravitational force of attraction results from a holographic principle of quantum information and its dissipation. When generalized to de Sitter spacetime typical of our universe, holography and the area law do not apply exactly. Instead, entanglement entropy arises from short-distance entanglements of neighboring degrees of freedom of the emergent bulk spacetime while de Sitter entropy arise from long-range entanglement of part of those microscopic degrees of freedom [79]. The mismatch results in an additional entropy that modifies emergent gravity and explains dark matter. Making use of Verlinde's proposal that entropic forces arise from information associated with the position of material bodies, we postulate that molecular growth results from an entropic-like attractive force driven by the interplay of short-distance entanglement of neighboring degrees of freedom (such as the greedy formation of helical structures in RNA) and long-distance entanglement of parts of those degrees of freedom (such as the long-range interactions forming RNA junctions). We view this putative entropic force as an emergent phenomenon that satisfies a tendency to increase entropy (broadly defined) not different from the type explaining the collapse of a polymer in a heat bath (as explained by Verlinde [78]), Brownian motion (beginning with Neumann [80]), polymer elasticity (recently computed for loops [81]), and the hydrophobic effect (derived from enthalpic and entropic contributions [82]). Note however that our conjectured entanglements project into a holographic-like horizon, in which diffusion of information produces structure from entropy bounds [44].

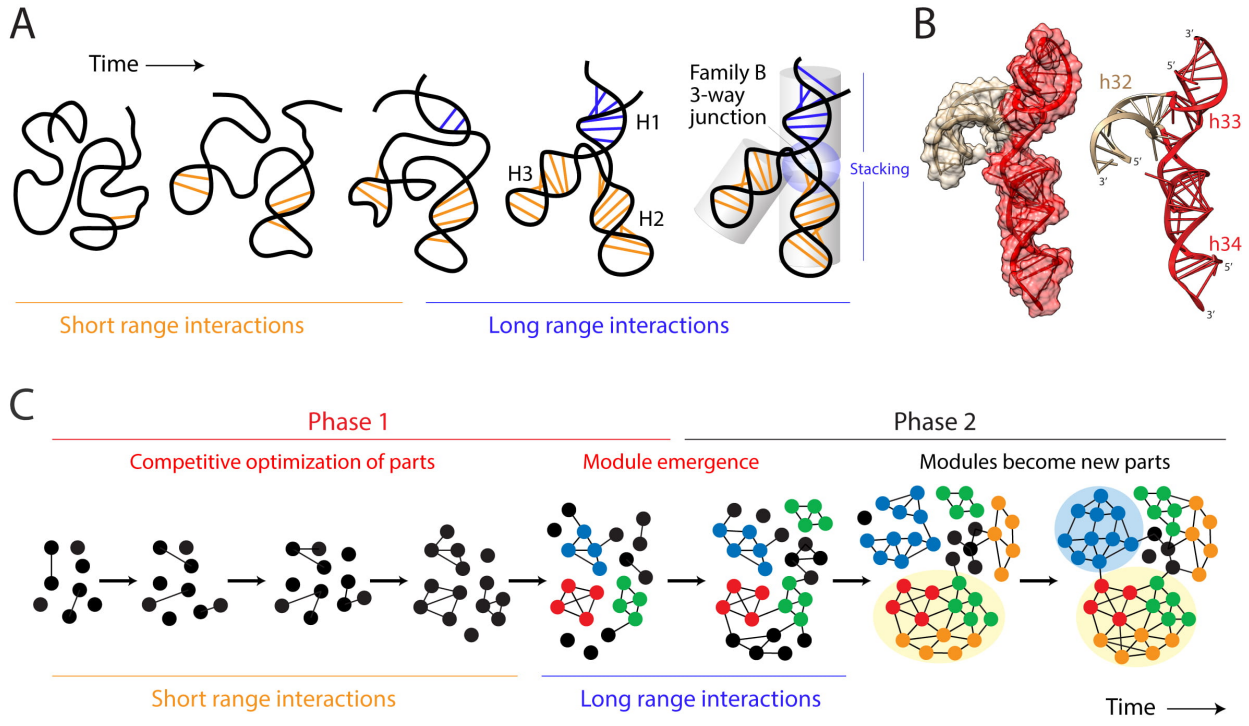

Fig. 10. Short and long-range entanglements are responsible for biological structure. (A) A model of RNA folding describes how short-range base pair interactions (orange lines) produce structural nuclei that lead to the formation of helical structures (H2 and H3), which later on establish long-range interactions (e.g., base pairing, formation of helix H1, coaxial helical stacking; blue lines and sphere) that stabilize a branched RNA structure with a 'family B' 3-way junction. (B) Atomic structural model of a typical 'family B' 3-way junction present in SSU rRNA in solid surface (left) and ribbon and ladder (right) depictions. Helices h33 and h34 (colored red) are coaxially stacked and make a trunk that holds functionally important pivot points of the ribosomal subunit. (C) The biphasic model of module emergence illustrates the emergence of network structure in evolution and the role of short-range and long-range entanglements. Nodes and links of a network describe parts and interactions in a biological system, respectively. The larger the number of links the more cohesive and stable is system's structure. The rise of hierarchical modularity during phase 1 result in highly connected communities (subnetworks), which become modules when their interactions stabilize. In phase 2, modules coalesce into higher level network substructures. The establishment of network communities in phase 1 involve at first short range interactions between neighboring nodes. When initial modules stabilize, they begin to establish range interactions with other emerging communities.

In the model, short and long-distance entanglements generate modules and hierarchy respectively, pushing growth through exploration of principled informational spaces within a spacetime dimension. The result of entanglements is plainly evident in the folding dynamics of the most important memory-holding macromolecules of the cell, nucleic acids and proteins. We highlight this fact with RNA. RNA molecules fold most fundamentally by satisfying hydrogen bonding interactions between base pairs that produce double helical duplexes separated by single stranded segments [83]. The process is highly frustrated as helical structures stabilize and nonpaired regions destabilize the molecules. The dynamics of folding and refolding has been visualized with advanced techniques, including fluorescence resonance energy transfer, atomic force microscopy, and optical tweezers [84]. These studies reveal folding pathways and insights into tertiary structure and molecular function. Folding kinetics is initiated by local base pair interactions between molecular regions that are physically close (Fig. 10A). The folding molecule explores a diverse set of folding configurations with free energies accessible to a given temperature. This 'plastic repertoire' represents an ensemble of conformations carefully constrained by processes of molecular evolution [85]. In fact, a free energy landscape constrains folding kinetics along a morphogenetic trajectory of 'canalized' sequences with low free energy barriers. Folding proceeds down a funnel to a single or small set of tertiary structures. Some of initial stems then extend to form helical segments, which stabilize by forming base pair stacking interactions. These flexible helical regions act as structural nuclei capable of bringing distant regions of the molecule together in space. Such molecular collapse results in the establishment of additional long-distance interactions through tertiary contacts and stabilization with counterions. Large RNA molecules form multibranched loops known as RNA junctions (Fig. 10B),

which foster structurally and evolutionarily constrained arrays of non-Watson-Crick base pairs, coaxial stacking, and spatial alignments of helical segments [86,87]. Long distance interactions in these helical communities still permit significant flexibility while constraining RNA dynamics [88].

Molecular entanglements are expected to become increasingly constrained with time as evolution proceeds. In fact, distance in short-range and long-range entanglements must be only considered within a spacetime dimension. Spatial distances along a nucleic acid or protein chain must be linked to timescales. In proteins for example, the nanosecond dynamics of loop structures is constrained by billions of years of evolutionary history [74]. Since the historical view of temporal parts allows dissection of constraints imposed by both evolution and biophysics, retrodiction is therefore necessary in any study. However, unanticipated connections arise because we do not know how entanglement evolves in time-dependent states. Most notably, there appears to be a surprising entropic force connection between the catenary function, which explains shape and static equilibrium of chains, cables and arches used in bridges and engineering constructs such as the Eiffel Tower, and the aggregate logistic Bass model of diffusion of innovations that propels evolutionary growth [89]. The local entropy function of the logistic distribution is proved to be catenary, and vice versa, establishing a connection with Verlinde's conjecture on the entropic origins of gravity. In addition, we note that the logistic S-shaped wavelets ('loglets') that are typical of paths of high performance in diffusion of innovation models account for sequential patterns of evolutionary accumulation we have observed in the growth helices and junctions in rRNA or domains in proteomes [73].

Our premise is that networks can effectively model evolving systems as dynamic patchworks of temporal parts. One way is to view networks as information processing devices capable of optimizing diffusion of matter-energy and information with time, especially when the modeled system changes in interaction with the suprasystem or the environment. For example, classical percolation methods of statistical mechanics can be used to establish maximal entangled states in quantum networks [90]. Similarly, spectral entropies can become information-theoretic devices to quantify information stored in systems, fit parametric network models, and compare networks describing multilayered organization [91]. Gibbsian-like density matrices can characterize networks as entanglements diffusing short-range and long-range signals along paths [92]. Even the analysis of fractal geometric characteristics with novel algorithms can quantify how structural complexity and heterogeneity of networks localize to space and scales [93]. If networks portray information processing in a system composed of a multiplicity of parts, then complex networks can connect models that explain the emergence and diversification of biological modules and recruitment with entropic entanglement. Tensor networks can be particularly useful to perform computations in high dimensional space by decomposing tensors representing multidimensional arrays into manageable parts [94]. The complexity of representations of many-body states can be reduced in this way by decomposition and reduction of tensor parameter space in the tensor network. The degrees of freedom that 'glue' tensors together (known as 'bond indices') describe an entanglement structure of states, which can be projected locally to some Hilbert space of smaller dimension. In particular, Tree Tensor Networks (TTN), with no loops and one extra dimension, and Multiscale Entanglement Renormalization Ansatz (MERA), with reticulations and constraints, provide extra dimensions that can be holographic and can recover a depiction of the many-body state ensemble at their boundaries. These tensor networks are particularly useful in describing criticality and scale invariance. In the context of the holographic constructs of dimensionality reduction of the MERA type, 'disentanglers' (unitaries) and 'coarse grainings'(isometries) are constraints that map matrices to matrices or vectors, respectively, compressing information in hierarchy and entanglement structures of the network. Note that entanglement entropies in MERA can correspond to an area-law in holographic space, with behaviors predicted by the Conformal Field Theory (CFT) of quantum mechanics. In addition, scale-invariant MERA can be understood as a collection of 'bulk' tensors spanning along one 'physical' and one 'holographic' normalization dimension, with properties that can be made to mimic the AdS/CFT correspondence. The challenge is to shift motivation from entanglements in quantum critical states of matter, which are built locally at every length scale, to long distance emerging entanglements in many-body states, while at the same time departing from spin-based quantum lattice systems.

Finally, the unification and diversification forces we have uncovered with chronologies and networks must also be taken into account. Fig. 10C describes how networks dissect the linkage-based *biphasic* (*bow-tie*) *model*, which predicts the evolutionary emergence of hierarchy and modularity in networks. Note than during the first phase of the model, emerging communities can be more or less fragmented from the rest of the network but that short distance interactions still materialize in those subnetworks. As long-distance interactions are stablished by perduring links connecting the emerging modules, the evolving network becomes more and more cohesive. Note that emergence of network communities within the specter of network fragmentation was plainly evident during discussion of evidence supporting the origins of the ribosome. The unification-diversification paradigm appears embedded at all time and spatial scales.

5.2 A Tendency towards Long-Lived Occurrents Increasingly Molded by Recruitment

When systems persist by extending their temporal parts in spacetime they do so within a framework of self-referential integration. However, perduring systems are not isolated entities and their integration manifests within the context of all suprasystems that can unambiguously embed them. In Caetano-Anolleś *et al*. [5], we predicted a tendency towards long-lived occurrents, which we find is supported by considerable evidence. We illustrated this tendency with the increasing age of a series of biological systems (ecosystems, communities, organisms, macromolecules, cells, metabolism) or modules of the protein world (proteomes, complexes, proteins, domains, loops, motifs). The lengthening of the life of these higher-level occurrents arising by their continued persistence (up to the present) was accompanied by an increase of granularity and abundance. We made this general trend evident when analyzing the network structure of metabolism [19]. Granularity refers to the continued structuring of an evolving biological system by forces of canalization. Older occurrents embodying '*ens per alio*' more basal levels of organization become more granular as causal relationships and entanglements with younger occurrents constrain their makeup. The problem that arises is that all occurrents are themselves part of suprasystems and the identity of all occurrents must be loose and popular. The persistence question often asks what must persist over time for X at time t1 be the same as Y at time t2. Once we are able to accept a loose instead of a strict concept of identity, the answer is spatiotemporal continuity of existence, i.e., shifting the question of persistence to that of existence [95]. This is particularly central, because all systems are ultimately integrated with each other into a single occurrent by forces of entanglement and recruitment. For example, all biological entities on our planet are unified by a common planetary origin that has given cells a genetic code and a ribosome. All novelties that have arisen since then have been unified, spread and homogenized in a diversifying world to form a recruitment-driven *tela vitae*. Describing ecosystems, organisms, or metabolic networks is therefore conventional: it simply defines elements of that common origin and subsequent unification-diversification for their better study. In fact, all entities of our universe can be unified and diversified in one way or another. This realization challenges the concept of temporal parts by prompting the 'bundling' of all occurrents into one. At the same time, it justifies generalizing holographic principles of quantum physics from atoms and subatomic particles to molecular, cellular and macroscopic structure and using them to understand the entangling forces and the drivers of recruitment.

6. Conclusions

Recruitment is a pervasive activity of life that is at the center of novelty generation. Novelties cannot effectively spread in the living world without recruitment. Recruitment should be also viewed as a central strategy of persistence. Because persistence is about how things behave across time, the problems of the two major philosophical accounts of persistence complicate any formal description of knowledge, including any effort of ontological modeling (e.g., GO database). To overcome difficulties, investigating persistence should be extended beyond the issue of systems having or lacking proper temporal parts [96]. Instead, the relationship between persistence, identity, change and causation is more relevant to scientific exploration [5]. Within a framework of fluxes of matter-energy and information and processing of information needed for a system to respond to internal and external challenges, we propose a 'triangle of persistence' in which reuse, innovation and stasis define a polytope in a phase space of trade-offs. Previously, we have shown that this landscape manifests in significant regularities, including patterns of reuse of novelties [39] or language-like laws of size-dependent decreasing returns [45]. The triangle offers the possibility to track trajectories of lineages during their evolution and of individual organisms during their lifetime, as these move in a landscape of performance. A focus on recruitment allows visualization of trajectories from a viewpoint of trade-offs between reuse, novelty and change. In addition, the concepts of occurrents and temporal parts embraced by the perdurantist school of thought can help dissect persistence by studying how temporal parts distribute when spatial parts change and evolve in time. This can be made explicit with evolutionary chronologies and evolving networks inferred for example with phylogenetic methodologies. To explain causal relationships responsible for the extended and complex *tela vitae* of occurrents we propose: (i) a theoretical framework and a biphasic (bow-tie) model of module generation in which forces of unification and diversification establish a frustrated landscape of change, and (ii) a 'theory of entanglement' that takes advantage of the dimensionality reduction offered by holographic principles that project N-dimensional spaces into N-1-dimensional boundaries to propose that short and long-distance interactions are responsible for structuring biological systems. Explanations offered by linkage and entanglement predict the emergence of hierarchy and modularity in complex networks that are used to model recruitment. We conclude by noting that time-varying networks offer foundations to connect structure and function across levels of biological organization [97] as well as information flows across timescales [98]. In addition, complex network theory has been applied to problems in quantum physics and is setting the stage for a theory of complexity that uses quantum information-inspired methods [99]. While explorations of network structure and dynamics are just beginning to be translated into concepts important for networked quantum systems, there is still much to learn. We need to further explore the power of time-varying multi-scale networks and develop mathe-

matical models, algorithmic tools and empirical approaches with which to understand the emerging structure, function and dynamics of biological systems.

Abbreviations

Ads/CFT, anti-de Sitter/conformal field theory; COVID-19, coronavirus disease 2019; DAG, direct acyclic graph; DAUP, doctrine of arbitrary undetached parts; EC, Enzyme Commission; ER, Erdös-Rényi; GO, gene ontology; Gy, billion year; Gya, billion years ago; LSU, large subunit; MANET, metabolic ancestry network; MERA, multiscale entanglement renormalization ansatz; NTD, N-terminal domain; PTC, peptidyl transferase center; RBD, receptor binding domain; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SSU, small subunit; TTN, tree tensor networks; VOC, variant of concern.

Author Contributions

GCA conceptualized the contribution. FM, MFA, IK, KCA and DCA contributed and analyzed data described in Figs. 1,2,3,4, and 6,7,8, respectively. All authors drafted, edited, improved, read and approved the manuscript. They have made substantial, direct and intellectual contributions to the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

GCA thanks the Office of Research and Office of International Programs in the College of Agricultural, Consumer and Environmental Sciences at the University of Illinois at Urbana-Champaign for supporting COVID-19 research.

Funding

Research was supported by grants from the National Institute of Food and Agriculture (ILLU-802-909 and ILLU-483-625), the Illinois Campus Cluster Program (ICCP), and Blue Waters supercomputing allocations from the National Center for Supercomputing Applications (NCSA) to GCA. DCA acknowledges support of fellowship award 1523549 from the National Science Foundation.

Conflict of Interest

The authors declare no conflict of interest. GCA is serving as the guest editor of this journal. We declare that GCA had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to GP.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/j.fbl2704128.

References

[1] Yčas M. On earlier states of the biochemical system. Journal of Theoretical Biology. 1974; 44: 145–160.
[2] Jensen RA. Enzyme recruitment in evolution of new function. Annual Review of Microbiology. 1974; 30: 409–425.
[3] Caetano-Anollés G, Aziz MF, Mughal F, Caetano-Anollés D. Tracing protein and proteome history with chronologies and networks: folding recapitulates evolution. Expert Review of Proteomics. 2021; 18: 863–880.
[4] Caetano-Anollés G, Janko R. Empedocles' On Nature, P. Strasb. Gr. Inv. 1665-6, a theory of networks and evolutionary growth 2,400 years before Darwin. In Caetano-Anollés G (ed.) Untangling Molecular Biodiversity (pp. 599–648). World Scientific: Singapore. 2021.
[5] Caetano-Anollés G, Mughal F, Caetano-Anollés K, Caetano-Anollés D. Temporal parts and biological change. In Caetano-Anollés G (ed.) Untangling Molecular Biodiversity (pp. 529–598). World Scientific: Singapore. 2021.
[6] Harish A, Caetano-Anollés G. Ribosomal history reveals the origin of modern protein synthesis. PLoS ONE. 2012; 7: e32776.
[7] Caetano-Anollés G, Caetano-Anollés D. Computing the origin and evolution of the ribosome from its structure — Uncovering processes of macromolecular accretion benefiting synthetic biology. Computational and Structural Biotechnology Journal. 2015; 13: 427–447.
[8] Wang M, Yafremava LS, Caetano-Anollés D, Mittenthal JE, Caetano-Anollés G. Reductive evolution of architectural repertoires in proteomes and the birth of the tripartite world. Genome Research. 2007; 17: 1572–1585.
[9] Wang M, Caetano-Anollés G. The Evolutionary mechanics of domain organization in proteomes and the rise of modularity in the protein world. Structure. 2009; 17: 66–78.
[10] Aziz MF, Caetano-Anollés G. Evolution of networks of protein domain organization. Scientific Reports. 2021; 11: 12075.
[11] Aziz MF, Caetano-Anollés K, Caetano-Anollés G. The early history and emergence of molecular functions and modular scale-free network behavior. Scientific Reports. 2016; 6: 25058.
[12] Koç I, Caetano-Anollés G. A natural history of molecular functions inferred from an extensive phylogenomic analysis of Gene Ontology data. PLoS ONE. 2017; 12: e0176129.
[13] Copley RR, Bork P. Homology among ($\alpha\beta$)8 barrels: implications for the evolution of metabolic pathways. Journal of Molecular Biology. 2000; 303: 627–641.
[14] Nagano N, Orengo CA, Thornton JM. One fold with many functions: the evolutionary relationship between TIM barrel families based on their sequences, structures and functions. Journal of Molecular Biology. 2002; 321: 741–765.
[15] Teichmann SA, Rison SC, Thornton JM, Riley M, Gough J, Chothia C. Small-molecule metabolism: an enzyme mosaic. Trends in Biotechnology. 2001; 19: 482–486.
[16] Teichmann SA, Rison SC, Thornton JM, Riley M, Gough J, Chothia C. The evolution and structural anatomy of the small molecule metabolic pathways in *Escherichia coli*. Journal of Molecular Biology. 2001; 311: 693–708.
[17] Kim HS, Mittenthal JE, Caetano-Anollés G. MANET: tracing evolution of protein architecture in metabolic networks. BMC Bioinformatics. 2006; 7: 351.
[18] Kim HS, Mittenthal JE, Caetano-Anollés G. Widespread recruitment of ancient domain structures in modern enzymes dur-

ing metabolic evolution. Journal of Integrative Bioinformatics. 2013; 10: 214.
[19] Mughal F, Caetano-Anollés G. MANET 3.0: Hierarchy and modularity in evolving metabolic networks. PLoS ONE. 2019; 14: e0224201.
[20] Caetano-Anollés G, Caetano-Anollés D. An evolutionarily structured universe of protein architecture. Genome Research. 2003; 13: 1563–1571.
[21] Caetano-Anollés G, Yafremava LS, Gee H, Caetano-Anollés D, Kim HS, Mittenthal JE. The origin and evolution of modern metabolism. The International Journal of Biochemistry & Cell Biology. 2009; 41: 285–297.
[22] Wetlaufer DB. Nucleation, rapid folding, and globular intrachain 1270 regions in proteins. Proceedings of the National Academy of Sciences of the United States of America. 1973; 70: 697–701.
[23] Janin J, Wodak SJ. Structural domains in proteins and their role in the dynamics of protein function. Progress in Biophysics and Molecular Biology. 1983; 42: 21–78.
[24] Han JH, Batey S, Nickson AA, Teichmann SA, Clarke J. The folding and evolution of 1275 multidomain proteins. Nature Reviews Molecular Cell Biology. 2007; 8: 319–330.
[25] Nasir A, Kim KM, Caetano-Anollés G. Global patterns of protein domain gain and loss in superkingdoms. PLoS Computational Biology. 2014; 10: e1003452.
[26] Wang M, Kurland CG, Caetano-Anollés G. Reductive evolution of proteomes and protein structures. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108: 11954–11958.
[27] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, *et al*. Gene Ontology: tool for the unification of biology. Nature Genetics. 2000; 25: 25–29.
[28] Shegogue D, Zheng WJ. Integration of the Gene Ontology into an object-oriented architecture. BMC Bioinformatics. 2005; 6: 113.
[29] Caetano-Anollés K, Hernandez N, Mughal F, Tomaszewski T, Caetano-Anollés G. The seasonal behavior of COVID-19 and its galectin-like culprit of the viral spike. Methods in Microbiology. 2021 (in press).
[30] Manrubia S, Lazaro E. Viral evolution. Physics of Life Reviews. 2006; 3: 65–92.
[31] Domingo E, Sabo D, Taniguchi T, Weissmann C. Nucleotide sequence heterogeneity of an RNA phage population. Cell. 1978; 13: 735–744.
[32] Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Challenges. 2017; 1: 33–46.
[33] Tomaszewski T, DeVries RS, Dong M, Bhatia G, Norsworthy MD, Zheng X, *et al*. New pathways of mutational change in SARS-CoV-2 proteomes involve regions of intrinsic disorder important for virus replication and release. Evolutionary Bioinformatics. 2020; 16: 117693432096514.
[34] Burra P, Soto-Díaz K, Chalen I, Gonzalez-Ricon RJ, Istanto D, Caetano-Anollés G. Temperature and latitude correlate with SARS-CoV-2 epidemiological variables but not with genomic change worldwide. Evolutionary Bioinformatics. 2021; 17: 117693432198969.
[35] Tomaszewski T, Gurtler V, Caetano-Anollés D, Caetano-Anollés G. The emergence of SARS-CoV-2 variants of concern in Australia by haplotype coalescence reveals a continental link to COVID-19 seasonality. Methods in Microbiology. 2021. (in press)
[36] Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, *et al*. Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. Cell. 2020; 182: 812–827.e19.
[37] Xu C, Wang Y, Liu C, Zhang C, Han W, Hong X, *et al*. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. Science Advances. 2020; 7: eabe5575.
[38] Gouwens NW, Sorensen SA, Badtizadeh F, Berg J, Murphy GJ, *et al*. Integrated morphoelectric and transcriptomic classification of cortical GABAergic cells. Cell. 2020; 183: P935–P953.
[39] Yafremava LS, Wielgos M, Thomas S, Nasir A, Wang M, Mittenthal JE, *et al*. A general framework of persistence strategies for biological systems helps explain domains of life. Frontiers in Genetics. 2013; 4: 16.
[40] Garland T. Trade-offs. Current Biology. 2014; 24: R60–R61.
[41] von Uexküll J. Umwelt und Innenwelt der Tiere. Springer: Berlin, Germany. 1909.
[42] Miller JG. Living Systems. McGraw-Hill: New York, USA. 1978.
[43] Sheftel H, Shoval O, Mayo A, Alon U. The geometry of the Pareto front in biological phenotype space. Ecology and Evolution. 2013; 3: 1471–1483.
[44] Caetano-Anollés G, Yafremava L, Mittenthal JE. Modularity and dissipation in evolution of macromolecular structures, functions, and networks. In Caetano-Anollés G (ed.) Evolutionary Bioinformatics and Systems Biology (pp. 443–449). Wiley-Blackwell: Hoboken, New Jersey, USA. 2010.
[45] Shahzad K, Mittenthal JE, Caetano-Anollés G. The organization of domains in proteins obeys Menzerath-Altmann's law of language. BMC Systems Biology. 2015; 9: 44.
[46] Twaddell WF, Menzerath P. Die Architektonik des Deutschen Wortschatzes. Dümmler: Bonn, Germany. 1954.
[47] Lewis D. On the plurality of worlds. Blackwell: Oxford, UK. 1986.
[48] Hawley K. Temporal parts. In Zalta EN (ed.) The Stanford Encyclopedia of Philosophy, Spring Edition. 2018. Available at: https://plato.stanford.edu/archives/spr2018/entries/temporal-parts/ (Accessed: 9 April 2022).
[49] Sider T. Four-Dimensionalism: An Ontology of Persistence and Time. Clarendon Press: Oxford, UK. 2001.
[50] Moyer M. Why we shouldn't swallow worm slices: a case study in semantic accommodation. Nous. 2008; 42: 109–138.
[51] Balashov Y. Persistence. In Callender C (ed.) The Oxford Handbook of Philosophy of Time. Oxford University Press: Oxford, UK. 2011.
[52] McTaggart JE. The unreality of time. Mind: A Quarterly Review of Psychology and Philosophy. 1908; 17: 456–473.
[53] Jaszczolt KM. Human Imprints of Real Time: from Semantics to Metaphysics. Philosophia. 2020; 48: 1855–1879.
[54] Ship of Theseus paradox (recounted by Plutarch in *The Life of Theseus*; 75 AD): "The ship wherein Theseus and the youth of Athens returned had thirty oars, and was preserved by the Athenians down even to the time of Demetrius Phalereus, for they took away the old planks as they decayed, putting in new and stronger timber in their place, insomuch that this ship became a standing example among the philosophers, for the logical question of things that grow; one side holding that the ship remained the same, and the other contending that it was not the same."
[55] Thomson JJ. Parthood and identity across time. The Journal of Philosophy. 1983; 80: 201.
[56] Bapteste E, Dupré J. Towards a processual microbial ontology. Biology & Philosophy. 2013; 28: 379–404.
[57] Van Inwagen P. Doctrine of arbitrary undetached parts. Pacific Philosophical Quarterly. 1981; 62: 123–137.
[58] Knauff M. Abstract objects and the doctrine of arbitrary undetached parts. Aporia. 2007; 2: 11–20.
[59] Chrysippus' paradox (recounted by Philo of Alexandria in De aeternitate mundi; 30 BC–45 AD): "Chrysippus, the most distinguished of their school, in his work On the Growing Argument, creates a freak of the following kind. Having first estab-

lished that it is impossible for two peculiarly qualified individuals to occupy the same substance jointly, he says: 'For the sake of argument, let one individual be thought of a whole-limbed; the other as minus one foot. Let the whole-limbed one be called Dion, the defective one Theon. Then let one of Dion's feet be amputated.' The question arises which one of them has perished, and his claim is that Theon is the stronger candidate. These are the words of a paradox-monguer rather than a speaker of truth. For how can it be that Theon, who had no part chopped off, has been snatched away, while Dion, whose foot has been amputated, has not perished? 'Necessarily', says Chrysippus. 'For Dion, the one whose foot has been cut off, has collapsed into the defective substance of Theon. And two peculiarly qualified individuals cannot occupy the same substrate. Therefore is is necessary that Dion remains while Theon has perished'."

[60] Parsons J. Dion, Theon, and DAUP. Pacific Philosophical Quarterly. 2004; 85: 85–91.

[61] Bokov K, Steinberg SV. A hierarchical model for evolution of 23S ribosomal RNA. Nature. 2009; 457: 977–980.

[62] Petrov AS, Bernier CR, Hsiao C, Norris AM, Kovacs NA, Waterbury CC, *et al*. Evolution of the ribosome at atomic resolution. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111: 10251–10256.

[63] Demongeot J, Seligmann H. Accretion history of large ribosomal subunits deduced from theoretical minimal RNA is congruent with histories derived from phylogenetic and structural methods. Gene. 2020; 738: 144436.

[64] Caetano-Anollés D, Caetano-Anollés G. Commentary: History of the ribosome and the origin of translation. Frontiers in Molecular Biosciences. 2017; 3: 87.

[65] Schnare MN, Gray MW. Sixteen discrete RNA components in the cytoplasmic ribosome of *Euglena gracilis*. Journal of Molecular Biology. 1990; 215: 73–83.

[66] Matzov D, Taoka M, Nobe Y, Yamauchi Y, Halfon Y, *et al*. Cryo-EM structure of the highly atypical cytoplasmic ribosome of Euglena gracilis. Nucleic Acids Research. 2020; 48: 11750–11761.

[67] Gray MW, Schnare MN. Evolution of rRNA gene organization. In Zimmermann RA, Dahlberg AE (eds.) Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis (pp. 49–69). CRC Press: Boca Raton, FL, USA. 1996.

[68] Chisholm R. Identity through time. In Keifer H, Munitz M (eds.) Language, belief and metaphysics. SUNY Press: Albany, New York, USA. 1970.

[69] Frege G. Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens. Louis Nebert Verlag: Halle, Germany.1879.

[70] Hennig W. Phylogenetic systematics. University of Illinois Press: Urbana, Illinois, USA. 1966.

[71] Hull DL. A matter of individuality. Philosophy of Science. 1978; 45: 335–360.

[72] Nietzsche F. Die Geburt der Tragödie aus dem Geiste der Musik. Verlag Von W.W. Fritzsch: Leipzig, Germany. 1872.

[73] Caetano-Anollés D, Caetano-Anollés K, Caetano-Anollés G. Evolution of macromolecular structure: a 'double tale' of biological accretion and diversification. Science Progress. 2018; 101: 360–383.

[74] Caetano-Anollés G, Aziz MF, Mughal F, Gräter F, Koç I, Caetano-Anollés K, *et al*. Emergence of hierarchical modularity in evolving networks uncovered by phylogenomic analysis. Evolutionary Bioinformatics Online. 2019; 15: 1176934319872980.

[75] Corominas-Murtra B, Goñi J, Solé RV, Rodríguez-Caso C. On the origins of hierarchy in complex networks. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110: 13316–13321.

[76] Watts DJ. Small worlds: the dynamics of networks between order and randomness. Princeton Univ Press: Princeton, New Jersey, USA. 1999.

[77] Solé RV, Valverde S. Information theory of complex networks: on evolution and architectural constraints. Lecture Notes in Physics. 2004; 650: 189–207.

[78] Verlinde E. On the origin of gravity and the laws of Newton. Journal of High Energy Physics. 2011; 2011: 29.

[79] Verlinde, A. Emergent gravity and the dark universe. SciPost Physics. 2017; 2: 016.

[80] Neumann RM. Entropic approach to Brownian movement. American Journal of Physics. 1980; 48: 354–357.

[81] Waters JT, Kim HD. Force distribution in a semiflexible loop. Physical Review E. 2016; 93: 043315.

[82] Schauperl M, Podewitz M, Waldner BJ, Liedl KR. Enthalpic and entropic contributions to hydrophobicity. Journal of Chemical Theory and Computation. 2016; 12: 4600–4610.

[83] Schuster P. Genotypes and phenotypes in the evolution of molecules. In Caetano-Anollés G (ed.) Evolutionary Bioinformatics and Systems Biology (pp. 123–152). Wiley-Blackwell: Hoboken, New Jersey, USA. 2010.

[84] Li PTX, Vieregg J, Tinoco I. How RNA unfolds and refolds. Annual Review of Biochemistry. 2008; 77: 77–100.

[85] Fontana W. Modelling 'evo-devo' with RNA. BioEssays. 2002; 24: 1164–1177.

[86] Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. RNA. 2001; 7: 499–512.

[87] Lescoute A, Westhof E. Topology of three-way junctions in folded RNAs. RNA. 2006; 12: 83–93.

[88] Besseová I, Réblová K, Leontis NB, Sponer J. Molecular dynamics simulations suggest that RNA three-way junctions can act as flexible RNA structural elements in the ribosome. Nucleic Acids Research. 2010; 38: 6247–6264.

[89] Guseo R. Diffusion of innovations dynamics, biological growth and catenary function. Physica A. 2016; 464: 1–10.

[90] Acín A, Cirac JI, Lewenstein M. Entanglement percolation in quantum networks. Nature Physics. 2007; 3: 256–259.

[91] De Domenico M, Biamonte J. Spectral Entropies as Information-Theoretic Tools for Complex Network Comparison. Physical Review X. 2016; 6: 041062.

[92] Ghavasieh A, Stella M, Biamonte J, De Domenico M. Unraveling the effects of multiscale network entanglement on empirical systems. Communications Physics. 2021; 4: 129.

[93] Xue Y, Bogdan P. Reliable Multi-Fractal Characterization of Weighted Complex Networks: Algorithms and Implications. Scientific Reports. 2017; 7: 7487.

[94] Orús R. Advances on tensor network theory: symmetries, fermions, entanglement, and holography. The European Physical Journal B. 2014; 87: 280.

[95] Nielsen L. We are all loose and popular. An argument in support of loose and popular identity. Ostium. 2020; 16: 1–12.

[96] Toyoshima F. Foundations for ontology of persistence: Beyond talks of temporal parts. In Brodaric C, Neuhaus F (eds.) Formal ontology in information systems (pp. 17–31). IOS Press: Amsterdam, The Netherlands. 2020.

[97] Bogdan P, Caetano-Anollés G, Jolles A, Kim H, Morris J, Murphy CA, *et al*. Biological networks across scales. Integrative and Comparative Biology. 2021; 61: 1991–2010.

[98] Alicea B, Yuan C. Complex Temporal Biology: towards a Unified Multi-Scale Approach to Predict the Flow of Information. Integrative and Comparative Biology. 2022; 61: 2075–2081.

[99] Biamonte J, Faccin M, De Domenico M. Complex networks from classical to quantum. Communications Physics. 2019; 2: 53.