

Review

Not peer-reviewed version

A Guide to the Multidimensional Nature of Biological Information: Of Maxwell's and Other Demons

[Bharath Srinivasan](#) *

Posted Date: 11 December 2024

doi: 10.20944/preprints202410.0876.v2

Keywords: Maxwell's demon; Information; macromolecules; sequence; structure



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Review

A Guide to the Multidimensional Nature of Biological Information: Of Maxwell's and Other Demons

Bharath Srinivasan

Department of Chemistry, College of Arts and Sciences, Stony Brooks University, NY;
bharath.srinivasan@stonybrook.edu

Abstract: The intricate orchestration of molecular events that constitutes life, as we know it, is oftentimes ascribed to an intelligent design motive. It is portrayed as if a paranormal watch maker has engineered the assemblage of parts and events that eventually manifests as the clockwork precision with which ontogeny and evolution operate within the boundary conditions imposed by what we know as life. The debate between those who advocate evolution versus intelligent design as the shaping force for modern life forms persists. However, that is not the motive for this essay. Though there has been several studies and scholarly articles that aspire to decipher and present the complexity of the molecular events that constitutes life in an insular fashion, there is a paucity of resources for undergraduates that summarizes the complexity cogently, as parts of a whole, for appreciating the enormity of the multidimensional nature of biological information and the inherent contradictions that are embedded therein. The fluidity inherent in the way information flows through signalling networks, the way biomolecular architectures maintain homeostasis in the ying-yang of life and the way multicellular complexity evolves through precise cell-cycle control is inherent in the evolution, maintenance and reproduction of biological systems. The matrix of events that encode information at the sequential and structural level across different macromolecules is a fascinating jungle of information. Combining these with how small-molecule metabolites interact with macromolecules in a spatially and temporally coordinated manner to evolve signal over random noise generates a staggering picture of informational organization and relays that, most often, defies human comprehension. This perspective is an attempt to array the various sources of information and put them together in context to appreciate the nuances of how information is stored and relayed within the biological systems. Further, it speculates on how these multitiered organization of biological information could be adapted to guide synthetic biology approaches and the future of information storage and organization.

Keywords: Maxwell's demon; Information; macromolecules; sequence; structure

Introduction:

Information is an abstract concept that is non-random in nature. Its relevance and import are dependent on its magnitude, organization and/or interpretation with reference to random noise. Also, the organization of information could be both hierarchical and interdependent in a matrix setting depending on the context of how the information is processed to derive value. For instance, the value of a book in terms of the information content embedded therein is judged based on its comparison with other books published in that genre, the value of a chapter is inherent in its relationship with other chapters constituting a book, the value of a paragraph is dependent on its relationship with other paragraphs that make up a chapter, the value of a sentence is contextualized within those of other sentences that make up a paragraph, the value of a word is evaluated within the framework of other preceding and succeeding words within a sentence and the value of an alphabet is embedded within the neighbouring alphabets in a word. Also, the information content is highly dependent on the abilities and training of the receiver to interpret and assess it and is encapsulated within several layers. However, care would have to be exercised to differentiate information perception and pattern detection by conscious mind as opposed to information that is

required to maintain aspects of biological order, replication, mutation and transmission independent of its interpretation by the conscious mind. Our interest in this essay would be to analyse and interpret the latter source of information within the ambit of human cognition for greater acknowledgement of what remains to be understood and a better appreciation of what we know already within a holistic framework of cognition.

Life is an unending marvel. Though polymaths and scientists have been toiling incessantly for centuries to tease apart the various bits of this phenomenon, a holistic appreciation of life in all its manifestations is an extremely arduous undertaking. The way life operates and distinguishes itself from its surrounding in terms of its ontogeny over the timespan of its biological cycle from birth to death, the way evolution attempts adapting and assimilating with the everchanging ecosystem that an organism inhabits employing a reserve pool of mutations over millions of years, the way the various intracellular events and players are orchestrated to play their part in a spatially and temporally controlled manner for genotype to reveal itself as phenotype and the way the animate entity replicates and reproduces transferring the complexity of information from one generation to the next all sound like chapters from a fairy tale textbook rather than aspects embedded in rational thought process and scientific reasoning. Life's unique way of encoding, manipulating and transmitting information is without any parallel and requires a lot more work to appreciate it in its totality (Nurse, 2008; Farnsworth *et al*, 2013).

Thermodynamically speaking, the creation, organization or manipulation of information (or order) in an isolated system entails an energy cost. Creating order entails reducing entropy. However, as per the second law of thermodynamics, the entropy of an isolated system allowed to evolve spontaneously cannot decrease. However, James Clerk Maxwell conceived of a thought experiment that set out to test the limits of the second law of thermodynamics. For an isolated system at thermodynamic equilibrium (and hence maximal entropy), he discussed of a "*finite being*" whose act of selective sieving of gas molecules caused thermal asymmetries taking the system away from equilibrium and hence, to a state of lesser entropy. However, time and again, it has been argued that this scenario doesn't, in fact, violate the second law of thermodynamics because the information assimilating act of the "*finite being*" or "*demon*" was excluded from the way the final entropy was calculated. Without getting involved in the semantics of the problem, one can wonder whether such exceptions to the second law of thermodynamics do happen in the way information is organized and manipulated under the open thermodynamic nature of biological systems, where matter and energy are constantly exchanged and where free energy is widely available. This essay is one among many attempts at the conception and possible comprehension of biological Maxwell's demons (Mizraji, 2021; Boël *et al*, 2019) especially as it pertains to the way information is managed in biological system.

This essay is an attempt to draw appropriate parallels between the way cognition treats information as acquisition, storage, manipulation and retrieval and the way biological systems treat information (Draganeacu's philosophic concepts). The aim being quantitation of biological information channels and the ability to leverage this knowledge for therapeutic solutions to ailments and synthetic biology applications. The cognitive information theory deals with the quantitation, storage, interoperability and communication of information. It uses the tools of statistics, statistical mechanics, computer sciences, probability theory and information/electric engineering and relies on the logical gates to create the framework for understanding information. Likewise, linguistic knowledge and information embedded in the syntactic (Form: grammatical rules that govern the structure and order of words for the right expression), semantic (Meaning: context dependent connotation of words, signs, tone and sentence structure in its entirety), pragmatic (utility) and esoteric (occult symbolism of conveying and interpreting feelings & signs) frameworks can guide us in understanding the multilayered and multidimensional nature of biological information (Zhong, 2017). A careful look at the basic tenants and structures of information theory and linguistic theory and the ability to frame the current state of knowledge in biological systems may help us evolve more and better interpretation of how the parts operate to evolve the whole (Koonin, 2016; Keller, 2009; Pharoah, 2020; Diniz & Canduri, 2017; Rutten *et al*, 2018; Walker *et al*, 2016; Binder & Danchin, 2011; Boël *et al*, 2019). A wonderful treatment of this concept is presented in literature where Shannon's

“entropy/uncertainty” is cogently defined as an aspect that is conveyed with respect to a particular observer and demonstrates its relationship to the Boltzmann-Gibbs thermodynamic entropy. Further, the subjective nature of information is reiterated in quantifying it as the amount of correlation between two systems (i.e. measurement of the quanta of entropy shared mutually between two systems, signifying the information that one system has about the other)(Adami, 2004; Fabris, 2009).

Since we would be discussing about several macromolecules and their crosstalk as part of this narrative, I would like to reiterate that the purpose of this essay is not to revisit the contents of a conventional biochemistry textbook by going into the details of the concepts discussed. There are several exceptional resources that an interested reader can turn to for the details. Additionally, during the discussion of macromolecules, we wouldn't necessarily discuss the hierarchical nature of 1°, 2° and 3° structure in that order as is usually done in the traditional textbooks. Rather, this essay would attempt brevity in summarizing the parts under a single rubric focussing on emphasizing their role within the broader information matrix. It would also describe the organization of the structural units as the authors understands happens in the biological systems *in situ*. Further, it will aspire to provide a holistic perspective on the several ways in which biological systems hoard information in a manner that lets an aspiring biochemist appreciate the enormity of the challenge and the opportunities that studying biochemical and biological systems pose. Given this, I would request the indulgence of the reader in understanding that not all concepts and parts are discussed at the length that they deserve in proportion to their perceived importance within the information scheme and some concepts may have been altogether missed out in this narrative. The treatment is biased by the author's perceived importance of a particular biological part in the holistic information scheme and is an attempt at balancing the treatment for an overview of this emergent aspect of inquiry.

1.1. Nucleic acid and information

Deoxyribonucleic acid Sequence: Nucleic acid is the gold currency of hereditary information flow from parents to offsprings (Dickinson *et al*, 2021). From the time Friedrich Miescher discovered it in the nucleus of white blood cells (thus named nucleic acid for its acidic nature), it was clear that nucleic acid, especially deoxyribonucleic acid (DNA), housed the blueprint of an organisms within its viscous strands. DNA is a polymer of four nucleotides (in turn, made of four nitrogenous bases adenine, guanine, thymine and cytosine, a deoxyribose sugar and phosphate). The sugar and the phosphate moiety form the backbone while the nitrogenous bases are free to base pair with another strand of nucleic acid and are the original source of information storage. Though there are exceptions, the most common base pairing is that suggested by James Watson and Francis Crick, where adenine base pairs with thymine and guanine base pairs with cytosine. Immediately after postulating this base pairing, the authors famously commented that “*It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material*” which forms one of the first concrete instances of how biological entities was equated with an information replicating device. The exact nature of genetic information is housed in the sequence of the four nucleobases organized along the DNA polymer and the replication machinery faithfully copies this sequence for transmission to the progeny using a network of molecular players (Chagin *et al*, 2010). This machinery is quite different across lower and higher organisms and forms the backbone of fidelity in information retention and mutability as a fodder for evolution and increased fitness. For the greater part of 20th century, investigators worked assiduously to uncover the molecular basis of how the information content of the genome was eventually revealed in the phenome of an organism. The work of Francis Crick, Har Gobind Khorana, Robert Holley and Marshall Nirenberg undeciphered that the permutation and combination of the four alphabets in the DNA is read in triplets for the genotype to reveal itself as the phenotype. The central dogma of information transmission (Cobb, 2017) dictates that the DNA is transcribed into RNA and the RNA is translated into proteins. A total of 64 combinations of degenerate triplet codons can form from the four alphabets of genetic code, 61 of which code for 20 amino acids while the remaining 3 code for cessation of translation. The exact nature of degeneracy and its evolution is still a matter of debate

(Gonzalez *et al*, 2019) and further work may be required to appreciate the true implication of how this degeneracy contributes to the possible parsimony of information storage.

Depending on the percentage AT or GC content of an organism's genome, the exact preponderance of a codon coding for a particular amino acid may vary dramatically resulting in aspects of codon bias (Bernardi, 1993; Bulmer, 1991; Hershberg & Petrov, 2009). The DNA is a highly protected macromolecule housed within the nuclear membrane which signifies dense packing of information, high stability and durability to withstand the onslaught of elemental perturbation and energy efficiency. The information content of this polymer has and continues to be a matter of debate among scientists given the fraction of DNA that has no known biological function and, thus, is known by the moniker "*junk DNA*". A few prominent examples of junk DNA include introns, repetitive DNA (short tandem repeats, variable number tandem repeats and satellite DNA enriched at telomeres and centromeres) and non-coding DNA. The fact that the proportion of "junk DNA" increases with the developmental complexity of the organism is an important indicator that it may potentially have important roles. Attesting to this line of thought, recent research is increasingly highlighting the critical role that this DNA plays in aspects of regulation, signalling and gene-expression. Moreover, opinions have been expressed that this DNA could act as sponge to absorb lethal mutations and can be potential raw material for evolutionary innovation. However, care should be exercised in interpreting these statements within the ambit of the C-value paradox which, paradoxically, states that the DNA content in an organism's genome (the C-value) doesn't always correlate with the organism's complexity (Lakhotia, 2023). To add to the complexity, and thus the confusion, it is dawning that almost 8% of the human genome and 46-54% of the gut bacterial genomes are viral remnants of lysogenic phage (Blinov *et al*, 2017; Kim & Bae, 2018).

Mutations are the fodder for evolution by natural selection. The sequence of nucleotides arrayed on the genetic material also house the repertoire of non-lethal mutations that fuel evolution and represent the long-term temporal revelation of information (Loewe & Hill, 2010; Hofkin, 2021) (Figure 1).

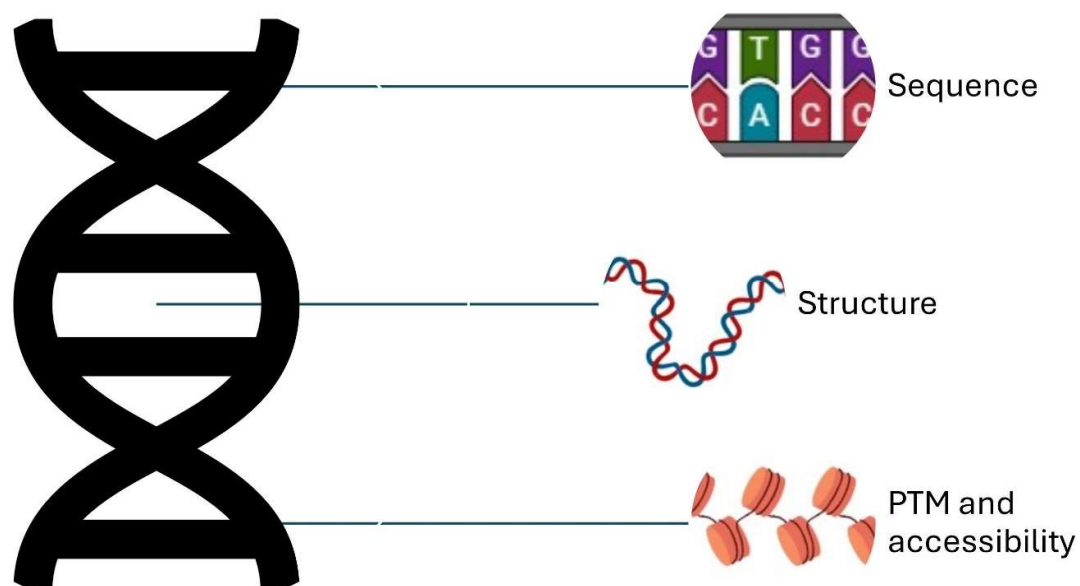


Figure 1. The hierarchical information content embedded in the nucleic acid polymer in terms of permutation and combination of the sequence, modification of the monomers governing accessibility and structural organization (secondary, tertiary and quaternary).

Given the primacy of the nucleic acid polymer as repositories of information, we often forget how critical the free nucleotide pool is in the information baton relay. Triphosphates of nucleotides

are very critical for energy and signalling. For instance, ATP is usually maintained at high concentrations intracellularly (from 1 mM-5 mM) and ensures the availability of stores of free energy to drive reactions that are endergonic. Likewise, activation of metabolites for anabolic processes (in carbohydrate and lipid biosynthesis) also has a critical need for nucleotides (Lane & Fan, 2015). However, we will be discussing more on this in our section on “metabolites”

DNA Structure and Accessibility: The previous section dealt with information embedded in the nucleic acid sequence. This section will briefly discuss the information embedded in the structure and organization of nucleic acid polymers (**Figure 1**). The double stranded helical nature of DNA imposes several unique symmetry constraints on the macromolecule. Further, topologically closed DNA molecules are negatively supercoiled and this property, in tandem with aspects of twisting and writhing, makes this macromolecule highly dynamic and prone to adapt different conformations (Garcia *et al*, 2007). The predominant conformation of DNA, predicted by James Watson and Francis Crick, is the right-handed helix annotated as B-DNA (Watson & Crick, 1953). There are other conformations of DNA that, depending on the handedness and rise per base, can be either classified as A- or Z- form. While the A-form is right-handed, the Z-conformation of nucleic acid is left-handed with a Pu-Pyr dinucleotide repeat as monomer. The latter is specifically recognized by the Z- α domains of some proteins (Srinivasan *et al*, 2022). It is often seen that local transition of B-conformation to either A or Z-conformation occurs frequently during transcription upon protein binding and induction of superhelicity. This, it is emerging, is tightly coupled to the recognition, replication, transcription, translation, recombination and mutation of information from the nucleic acid polymer at a given locus. Other structural motifs, without going into this exhaustively, include A/B backbone deformed conformations, G-quadruplexes, cruciform structure for palindromic sequences, triple helical structures with Hoogsteen base pairing for the third strand (H-DNA) and so forth (Rich, 1993; Wang & Vasquez, 2023; Potaman & Sinden, 2005).

From lower life forms to higher organisms, the DNA pool of a cell is variedly organized. For bacteria, the genetic material is segregated into nucleoid and plasmids while higher life forms have more complex organization. During the life cycle of a cell, the entire genome is copied and organized into discreet entities called chromosomes to facilitate mistake-free transmission of the genetic information to the progeny during cell-division. The concept of ploidy adds to the rich complexity of information management in nucleic acid and refers to the number of copies of chromosomes. Lower organisms are haploid with a single copy while most higher life forms are diploid with two copies of chromosomes. Following on from there, humans contain 46 pairs of chromosomes that are made of 22 pairs of autosomes and 1 pair of sex chromosomes. All autosomal chromosomes are diploid while the x-sex chromosome is diploid in females and haploid in males. Further, a lot of fungal and plant species show polyploidy conferring on them traits of heterosis, higher gene redundancy and improved protection from undesirable mutations and advantages of asexual reproduction. It has been argued extensively that ploidy is a means of introducing adaptation, plasticity, and stress resilience in biological system by introducing a layer of horizontal information retention apart from the vertical information transmission signified by cell division (Anatskaya & Vinogradov, 2022).

Epigenetic modifications are heritable changes that do not alter the underlying DNA sequence, exclusively resulting in altered accessibility, and thus, information relay with implications in gene expression, development and differentiation (**Figure 1**). There are several different modifications that have been studied on DNA with the most predominant one being 5-methylcytosine (5mC). This modification is brought about by an enzyme called DNA methyltransferases (DNMTs) that transfers a methyl group from S-adenosyl methionine (SAM) to the cytosine of a palindromic 5'-CpG-3' sequence and represents 3-5 % of the total cytosine pool in vertebrates. DNMT1 prefers hemimethylated substrate and hence helps in the faithful transmission of methylated mark to the daughter strand while DNMT3A and 3B helps establish new methylation patterns (Moen *et al*, 2015). The principal role of methylation is thought to be heterochromatinization mediated by methyl-binding protein (MBP). MBP binding results in highly condensed form of DNA that can sequester DNA and prevent access by transcription factors to cause gene repression (Gibney & Nolan, 2010). This repression as a currency of information is regulated temporally within an organism's lifespan.

A prominent example being the enhanced expression of genes coding for primary and secondary sexual traits post-puberty. Additional critical roles of 5mC mark include selective expression of genes in a tissue specific manner, X-chromosome inactivation and genomic imprinting (refers to the selective epigenetic silencing of one copy of a gene during egg or sperm formation). Other DNA modifications include 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), though these are very sparsely detected compared to 5mC (Zhao *et al*, 2020).

Another layer of information in-built into the system for accessing the information coded by nucleic acid is the epigenetic modification of DNA interacting proteins. As we will see in a later section, individual amino acids on a protein can get variably modified resulting in structural alterations and, thus, altered binding preferences. The enzymes bringing about these epigenetic modifications can be broadly classified as writers, erasers, readers and remodellers. If one zooms into the chromosomes (the compacted nucleic acid during cell division), one will observe the tight packing of negatively charged DNA onto positively charged octameric histone proteins having lysine and arginine rich tails. Epigenetic alterations of these positively charged residues on histone tails have been shown to play critical roles in the interaction of histones with DNA and thus in gene expression and/or repression. The most prominent mark includes acetylation of lysine residues on all four histones (H2A, H2B, H3 and H4) associated with transcriptional activation. The mark is dynamic balanced by two enzymes with opposing action: histone acetyltransferases (HATs) and histone deacetylases (HDACs). HATs transfer an acetyl group from acetyl-CoA onto the ϵ -group of a lysine residue neutralizing the positive charge on lysine and thus weakening the interaction between the histone and the DNA, facilitating access of the genetic information encoded in the nucleic acid to factors involved in transcription (Wu *et al*, 2009; Bannister & Kouzarides, 2011; Su & Denu, 2016). Other prominent marks include lysine mon-, di- or tri-methylation (catalysed by histone lysine methyltransferases, HKMTs, with a SET domain), arginine mono-, symmetrical di- or asymmetrical di-methylation (catalysed by Protein arginine methyltransferases, PRMTs) and phosphorylation of serine and/or threonine (catalysed by kinases). All these marks are reversible with a cognate eraser acting in tandem with its writer, both spatially and temporally regulated. Less prominent alterations include K-48 and K-63 linked ubiquitination (see below) and SUMOylation, mono- and poly-ADP ribosylation on arginine and glutamate amino acids, deamination (arginine to citrulline neutralizing the positive charge), β -N-acetylglucosamine (O-GlcNAc) sugar modification on serine/threonine residues, proline isomerization and so forth (Kouzarides, 2007). Though the epigenetic code looks simple and straightforward upon first perusal, the combinatorial aspect of the crosstalk among the various marks across different residues can lead to exponential degree of novel ways in which it can be interpreted. The intricate nature of information channelling by writer, eraser, reader and remodeler protein/protein-complexes leads to a complex multidimensional terrain of chemical language with important implication in how this information is translated for organismal growth and survival (Furth & Shema, 2022). Increasing implementation of approaches such as ChIP-seq, Cut&Run, Cut&Tag, mass-spectrometry, CyTOF and single molecule TIRF microscopy imaging is revealing the true scale and complexity of this combinatorial histone modification language with ever increasing resolution. We will also see that such combinatorial outburst of information will be a repeating motif across biological information processing as will become clearer when we look at the ubiquitination code.

Apart the roles of the typical circuit and control elements (RNA polymerases, transcription factors, repressor proteins, silencer elements and promoters) that modulate the extent and pace of gene expression (Phillips, 2014), i.e. transcription of genetic information to the intermediary RNA form, there are other very interesting modes by which transcription is controlled. One such mechanism that determines the extent of transcription is regulatory squelching. Through this process, transcription factors control the expression of certain genes in trans by sequestering and siphoning off limiting components, such as coactivators or corepressors, away from the respective promoter (Cahill *et al*, 1994; Natesan *et al*, 1997).

Ribonucleic acid sequence, structure and accessibility: The primacy of Ribonucleic acid polymer as the genetic material of choice has been gaining traction. In the central dogma of

information flow, DNA is transcribed to RNA followed by translation of the ribonucleotide polymer to protein, the latter showing catalytic prowess for incredible rate enhancement of biologically essential reactions. The proponents of the RNA world hypothesis rightfully argue that the genomic material of several viruses (a ribonucleoprotein complex that can replicate only inside a living organism and considered to be primitive precursor for more advanced life forms that subsequently emerged) is RNA. This combined with discovery of catalytic activity in RNA (ribozyme) reinforced the belief in an RNA first world (Engelhart & Hud, 2010; Wächtershäuser, 2014; Dworkin *et al*, 2003; Guerrier-Takada *et al*, 1983). The discovery of reverse transcriptase activity further strengthened this belief by providing evidence for the synthesis of the more stable DNA from RNA templates. This, combined with the lack of uracil in the DNA (see below) and the fact that group transfer coenzymes are mostly ribonucleotides or their derivatives (see sections on proteins), provides strong evidence for a world initially dominated by RNA as the information currency of choice.

RNA as a polymer differs from DNA in that the latter contains thymine while the former contains uracil as one of the pyrimidine bases. The reason for this swap, in spite of the fact that both thymine and uracil contain essentially the same information content (as evident in the DNA of the bacteriophage PBS 1/2 that lacks thymine), is rationalized based on the following observations: (1) The biosynthesis of uracil is less expensive thermodynamically making it ideal for being a component of the transitory and ever-changing capsules of information encoded in RNA, (2) Uracil has weakly acidic properties making it more resilient to oxidative stress in the cytosolic milieu, and, (3) presence of native uracil in DNA might complicate the removal of cytosine deaminated uracil by uracil-DNA glycosylase (the latter is essential for recognizing and removing an aberration that might pose danger for DNA stability and durability) (Vértessy & Tóth, 2009).

Oftentimes, it isn't straightforward to understand the primary sequence of RNA from that of the DNA because of presence of modified bases, splicing, and other deletions/excisions. A comprehensive list of posttranscriptional RNA modifications that can compound the problem of interpreting the primary structure of RNA include N6-methyladenosine (m6A), N1-methyladenosine(m1A), 5-methylcytidine (5mC), N7-methylguanosine (m7G), 2'-O-methylation (2'-O-me), pseudo uracil, dihydrouridine and adenosine-to-inosine (A-to-I) editing (Delaunay *et al*, 2024). The A-to-I editing is catalysed by adenosine deaminases acting on RNA (ADAR), some of which contain the Z- α domain that is enriched in liquid-liquid phase separated (LLPS) bodies called stress granules (see below for further details)(Gabriel *et al*, 2021). Discovery of reverse transcriptase not only heralded a unique perspective on the "RNA-first" world but also helped facilitate the determination of the primary sequence of RNA with ease and economy in a high-throughput manner. RNA is also capable of adopting secondary (duplexes, single-stranded regions, hairpins, bulges, internal loops, and junctions)(Chastain & Tinoco, 1991) and tertiary structural formations similar to proteins (Batey *et al*, 1999). Determinants of tertiary structure formation in RNA include pseudoknots/kissing loops, sequence-specific interactions such as triple-helices/tetraloops and stacking and backbone interactions (Butcher & Pyle, 2011). The tertiary structural organization of transfer RNA (tRNA) is the key to its ability in assisting the translation of the information embedded in messenger RNA (mRNA) to proteins. The tertiary structure of RNA also helps it interact with proteins in a sequence-specific manner as is evident in the structure of the ribosomal RNA (rRNA) which is a huge ribonucleoprotein (RNP) complex. Other notable examples of RNA tertiary and quaternary structure include the hammerhead ribozyme and the P4-P6 domain of the self-splicing intron from *Tetrahymena thermophila* to name but a few. Approaches such as CLIP-seq, sono-seq and PAR-CLIP (Danan *et al*, 2022; Hafner *et al*, 2021; Weidmann *et al*, 2021) are revolutionizing the way RNA protein interactions are deciphered to understand the way their mutual interactions embody information. Another very exciting find pertains to tertiary structural formations of RNA called riboswitches. These structures can recognize small molecules and ions and modulate gene expression (Edwards & Batey, 2010; Kavita & Breaker, 2023) and thus behave like protein.

Multivalent interactions mediated by RNA plays critical roles in the biogenesis and physicochemical properties of membrane-less biomolecular condensates (Bevilacqua *et al*, 2022). Noncoding RNA (ncRNA) recruits various proteins (containing intrinsically disordered regions) to

drive phase separation (Mittag & Parker, 2018; Sfakianos *et al*, 2016). Though it is emerging that these LLPS condensates play important roles in RNA transcription, transport, and metabolism, the full extent of the information content that they encode is not yet clear and may require further investigation.

In higher organisms, another two very important classes of RNA molecules are the small RNAs and the long non-coding RNA. The small RNAs are short non-coding RNA molecules ~18-30 nucleotides long that modulates target gene expression via imperfect or perfect complementary base pairing. They can either inhibit (post-transcriptional gene silencing, PTGS, or chromatin-dependent gene silencing, CDGS) or activate (RNA activation, RNAa) the expression of the target gene and thus, resemble transcription factors in their action (Zhang, 2009). These RNAs regulate biological information in ways that are still emergent and not fully appreciated within the context of the overall information flow. This includes modulation of information at both the genotype to phenotype level and at the epigenetics of transgenerational level (Duempelmann *et al*, 2020). The most well known small RNAs are microRNA (miRNA), siRNA and, to a slightly lesser extent, Piwi-interaction RNAs (piRNAs). Both miRNA and siRNA act through the RNA-induced silencing complex (RISC), a nucleoprotein complex, in recognizing the complementary RNA substrate to be cleaved. However, siRNAs are derived from long double-stranded RNAs and are specific to one target RNA, while miRNAs are derived from short RNA hairpins and are degenerate in their target substrate specificity (Zhang, 2009). Additionally, the eucaryotic genome also encodes several 100s of long non-coding RNA (lncRNA) and circular RNAs (circRNA) that have critical functions in maintaining homeostasis and development (Mattick *et al*, 2023; Chekulaeva & Rajewsky, 2019).

The half-life of the RNA macromolecule is yet another level of information that is encoded within the biological systems. Among the different RNA species, the half-life of rRNA is the longest followed by intermediate half-life for mRNA. However, tRNA is highly unstable under all conditions (Abelson *et al*, 1974). The deadenylation of the polyA containing mRNA is a rate limiting step in the decay of mRNA (Chen *et al*, 2008). However, approaches like TAIL-seq have demonstrated the presence of non-A nucleotides in the tail of mRNA requiring further work to understand the true implication of this doping on the half-life of mRNA (Lee *et al*, 2024). The primary sequence, structure, modification and susceptibility to endo and exo-nucleases all contribute to the instability of the RNA species intracellularly (Yang *et al*, 2003).

We have been discussing the importance and mainstay of genetic inheritance as repositories of biological information. However, we would take a break here and discuss about a unique way of information propagation. Maternal inheritance is another unique form of inheritance that is distinct from the way conventional mode of meiotic mixing of sex chromosomes to transmit the genomic information is undertaken by biological systems. In this form, mother transmits the mitochondrial genome to her progeny directly. Maternal inheritance can also include transmission of small-peptides and/or metabolites such as hormones, antioxidants, antibodies, and other immunological factors. These can have a wide range of effects on the development and functioning of the offspring.

As we transition from DNA to RNA to proteins, we must transit the information rich territory of translation. An amazing ingenious molecular assembly of disparate machines for the sole purpose of orchestrating information transfer from an mRNA molecule to a protein molecule whose three-dimensional structure represents the global energy minima of the system and encodes various nuanced functional information (Liquori, 1969). However, the amazement and information enrichment of this step would go unappreciated without first appreciating the inherent information content embedded in the protein polymer. Hence, this section on translation would be discussed after and alongside discussing the hierarchical organization of information inherent in a protein's structure. Suffice it to say that the temporally and spatially regulated translation and translocation of the proteins forms the molecular basis of fluxes through the metabolic circuitry that is the currency of how information is delivered as eventual phenotypes (see "**metabolites**").

1.3. Protein and information:

Protein sequence, structure and accessibility: Protein macromolecules constitute the pinnacle of information enrichment. These are molecular machines that make reactions feasible on timescales that are compatible with life, form structural scaffolds, function as conveyance motors transporting cargo, operate as sentinels of entry and exit on the portals of life's sacrosanct boundaries, create signals by impeccable molecular recognition of its partners in an extremely crowded intracellular milieu and the list can extend unendingly (Srinivasan, 2023). Proteins are polymers of 20 naturally occurring L-amino acids and can create an infinite variety of polymeric permutations and combinations dependent solely on the length of the protein sequence. However, not all permutations and combinations are found in biologically occurring proteins indicating a non-randomness to the way amino acids are arrayed on the polypeptide chains. This non-randomness is likely dictated by aspects of self-assembly and self-organization as we will appreciate later. It is also not clear why evolution decided to restrict the choice of amino acids to the 20 that we know proteins are made of despite the large diversity from which it could have chosen to operate neither do we know the reason for the preference of the L enantiomer of amino acid. Though one can speculate on various reasons to rationalize these choices, the most compelling ones indicate to possible information parsimony with potential for self-assembly and organization. Suffice it to say that the amino acid pool represents a skewed distribution of monomers that can help facilitate different kinds of inter molecular interactions spanning hydrophobic, ionic and hydrogen-bonds. A detailed treatment of the chemi-information content of each amino acid is beyond the scope of this perspective and the reader is referred to exceptional resources elsewhere.

Two amino acids undergo a condensation reaction and form the amide (peptide) bond with the concomitant release of water. A casual perusal of the primary sequence of the polypeptide shows three degrees of freedom for each amino acid (φ is the dihedral angle about the N-C α bond, ψ is the dihedral angle about the C-C α bond and ω is the amide bond, C-N, dihedral angle). It was G.N. Ramachandran and coworkers who experimented with the allowed angles for φ and ψ based on the principle of Van der Waals exclusion and steric clash and proposed the Ramachandran map as a benchmark to assess the correctness or otherwise of a theoretically proposed or experimentally modelled protein structure (Ramakrishnan & Ramachandran, 1965; Ramachandran *et al*, 1966). Depending on the distribution of the φ and ψ angles (note that ω is restrained to either 0° or 180° given the partial double bonded nature of the peptide bond due to the delocalized lone pair of electrons which imposes an expensive energy barrier for rotation around this bond), polypeptides can adopt two distinct secondary structural adaptation: α -helix ($\varphi = -60^\circ$ and $\psi = -45^\circ$) and β -sheets ($\varphi = -140^\circ$ and $\psi = 130^\circ$). These secondary structural elements combine in various sequences (i.e. α/β , all α , all β and $\alpha+\beta$) to evolve folding domains that subsequently collapse to the global minima of tertiary structure. This gives rise to the now-famous Levinthal's paradox that says that for a polypeptide of 101 residues with three degrees of freedom (φ , ψ and ω) about each peptide unit, it will take sampling of 3^{100} ($\sim 10^{47}$) different configurations to find the one configuration that represents the global minima of the tertiary structure. Even if that sampling is done at a phenomenally fast rate (i.e. 10^{13} /sec or 10^{20} /year), it will take 10^{27} years to sample them all (age of planet earth is a mere 10^9 years!) (Zwanzig *et al*, 1992).

However, the conception of primary, secondary and tertiary structures evolving hierarchically exploring the full configurational space as a polypeptide chain emerges from the translational machinery is misplaced. The collapse of the fully extended unfolded polypeptide to the fully folded protein tertiary structure never happens within the biological systems. As the nascent polypeptide chain exits the translational machinery, it starts exploring the energy landscapes in tandem (occasionally with the help of chaperone proteins) making it adopt its three-dimensional state (also known as co-translational elongation and folding). The folding code which codes for the information in the protein's three-dimensional structure is intricately intertwined not only with the primary sequence of DNA coding for it but by other parameters such as synonymous codon variants, charged patches in the nascent polypeptide chain and modification of bases in tRNA, that causes fold pausing and change the rate at which protein folds (Komar *et al*, 2024; Nedialkova & Leidel, 2015). Further, multivariate cues in the sequence of mRNAs can have important implication for translation initiation,

open reading frame (ORF) selection, velocity of nascent chain elongation, and the folding pathway that the polypeptide chooses (Rodnina, 2016). It has been demonstrated that the speed of translation has oftentimes been shown to have a large impact on expression level, tertiary three-dimensional structure of the resultant protein and stability! (Lampson *et al*, 2013; Stein & Frydman, 2019). Additional solution to this paradox has been provided, even for *in-vitro* unfolding and folding of polypeptides, largely by the introduction of the folding funnel where the polypeptide samples configurations in a progressively funnel like energy landscape rather than expansively as it would have done at the base of the funnel enabling it to fold within fractions of a second or, oftentimes, minutes. The principal driver for the folding funnel is hydrophobicity that coalesces water repelling aliphatic amino acids to form the core of the protein while the water-loving polar sidechain containing amino acids are organized in the outer shell in proximity with the solvent. This leads to a compact, globular and desolvated structure. The oily interior decreases the dielectric constant of the core substantially, leading to charge amplification. The need to neutralize this amplified charge in the core of the protein in order not to destabilize the energetics of the hydrophobic collapse is the principal driver for local secondary structure formation (Marín *et al*, 2017)

These conformational restraints on the polymer result in structural strains, Though the exact import and relevance of this strains as units of information is evolving and is still underappreciated, it is still essential to discuss it within the context of biological information relay. Conformons are *Sequence-specific conformational strains* (SSCS) within biopolymers that have been argued to encapsulate free-energy and genetic information and are the basic minimal unit (quanta) responsible for all biological action. The free energy of a nucleic acid conformons is estimated to be 500–2500 kcal/mol with an information content of 200–600 bits compared to 8–16 kcal/mol and 40–200 bits for proteins (Ji, 2000).

The tertiary structure of the protein interacts with either itself or other proteins in a process called as homooligomerization and heterooligomerization, respectively. These interactions lead to the quaternary structure of the protein. Homooligomerization is a special case of information parsimony whereby one genetic piece of information can code for several copies of the protein that eventually oligomerize to form homooligomeric complexes that have traits of stability, increased resistance to mutability and decreased tendency for error propagation, apart from any context-dependent functional significance. Symmetry of such homooligomeric interactions adds another layer of information with implications for stability and durability (Goodsell & Olson, 2000). A notable example of such huge homooligomeric assemblies include the way viral capsids are assembled from the parsimonious information encoded in the genome of a virus. Both homooligomeric and heterooligomeric organization also give rise to interaction surfaces and cavities at the interface (more about the role and importance of pockets in the subsequent paragraphs). Oligomerization has evolved during evolution as a means of creating scaffolding for supramolecular assemblies serving a precise function (channels, gates) or for embedding regulation control of information relay across the monomers constituting the oligomer. Oligomerization is also a means for sequestering proteins in inactive complexes that can have important roles in generating ultrasensitivity or “all-or-none” responses as is discussed for negative cooperativity below. Titration of these macromolecules can make them rate limiting for their native function thus bringing tight regulation of information flow.

As we saw with nucleic acid, yet another level of information as it pertains to protein structure represents the heritable influence of environment (epigenetics) on the way a protein functions without unduly affecting the underlying order of the amino acid sequence. The chemical space accessible to a protein is vastly expanded by means of various post-translational modifications (PTMs). This in turn expands the functional quanta of information encoded within a protein's structure by bringing about changes in protein interaction networks, protein's function as catalyst, protein's subcellular localization, to mention but a few. Prominent PTMs, as was discussed for histone proteins above, include (1) Reversible phosphorylation of serine, threonine and tyrosine residues with prominent roles in signalling, cell-cycle regulation, growth and cell-death, (2) Glycosylation of asparagine (N-linked) and/or serine/threonine (O-linked) residues with prominent roles in protein folding, stability of the tertiary structure, conformational dynamics, distribution, half-

life and activity. Glycosylated proteins are modified both co-translationally and/or posttranslationally, are predominantly secreted and are modified with both simple monosaccharide sugars and/or complex branched oligosaccharides. Glucose, galactose, N-acetylglucosamine (GlcNAc), N-acetyl galactosamine (GalNAc), fucose, xylose, and mannose are some of the sugars that are shown to modify proteins. The combinatorial expansion of the glycan code is discussed in detail in the next section and plays a huge role in how proteins and lipids serve as informational strings within the broader information meshwork (Williams & Davies, 2001; Yu & Chen, 2007; Muthana *et al*, 2012; Reily *et al*, 2019). (3) Ubiquitination of proteins is usually associated with priming them for proteasomal degradation. Ubiquitin is a 76 amino acid protein appended to lysine residues (discussed further below). (4) S-nitrosylation is a very critical PTM of the cysteine residue that has important roles in regulating protein stability and cellular homeostasis. This modification is highly evanescent because of a host of reducing enzymes and metabolites (glutathione and thioredoxin). Proteins that are S-nitrosylated are thus protected from denitrosylation by secluding them within membranes, interstitial space or vesicles (caspase is a prominent example that, upon denitrosylation, gets activated and triggers apoptosis). (4) Methylation and N-acetylation are two additional PTMs that have been discussed in the context of histone modifications for gene regulation in chromatin. (5) SUMOylation is a modification whereby small ubiquitin-like modifier (SUMO) proteins are tethered to lysine residues on proteins with varied functional outcomes. (6) Like glycosylation, another prominent modification of proteins involves their lipidation to enable them to localize to lipidic membranes. Prominent modifications that increase the hydrophobicity of protein molecules to facilitate membrane localization include attaching glycosyl phosphatidylinositol anchor to the C-terminal of proteins, N-terminal or S-myristoylation and S-prenylation. Details about some of these will be discussed in the next section (7) Poly (ADP-ribosyl)ation (PARylation) is yet another prominent PTM of proteins whereby linear or branched chains of ADP-ribose units are tethered onto the target protein by poly(ADP-ribose) polymerase 1 (PARP1) assisting by the cofactor NAD⁺. This modification has important roles in DNA-damage repair and regulation of other PTMs (Alemasova & Lavrik, 2019).

Cysteine is an important proteinogenic amino acid that plays very critical roles in protein structure. The oxidation of cysteine residues has been equated with disulphide bond that, oftentimes, determines the tertiary fold of the protein. However, what is lacking appreciation is the multiple oxidation states of cysteine (sulfenic, sulfinic, and sulfonic acid) that has immense effect on protein structure and function and, thus, serves as an important link in the information quanta represented by proteins. Emergence of mass-spectrometry has played a significant role in deciphering this level of information. Cysteine oxidation state can coordinate a combinatorial seeding of other PTMs discussed above in response to oxidative stress and changes in redox environment (Jacob *et al*, 2003; Garrido Ruiz *et al*, 2022).

A combination of sequence guided, oligomerization facilitated and PTM influenced tertiary/quaternary structure of protein represents a dense packet of information encoding critical details at multiple levels. The plasticine model of myoglobin built by John Kendrew, the first ever to be built of any protein, shows an irregular shaped entity twisting through space on wooden pegs (de Chadarevian, 2018). However, this disenfranchising structure hid beneath its simple shell the secret that would revolutionize the way we understand and interpret molecular biology getting into the 21st century. A careful perusal of the molecular envelope of this structure reveals pockets and flat featureless surfaces. It immediately dawned upon investigators from across the globe that these, likely desolvated, pockets and surfaces and the precise orientation of amino acid residues within these physicochemical and geometrical abstractions house the knowledge of how small-molecules and other macromolecules interact with proteins to bring about the orchestration of information. Skolnick and colleagues have extensively studied the nature and distribution of pockets on protein surfaces and protein-protein interfaces for oligomeric assemblies (Di Rienzo *et al*, 2020; Gao & Skolnick, 2012, 2013; Skolnick *et al*, 2015). Their studies have given detailed information on the total number of representative geometrical pockets, the ability to predict ligand binding propensity across cavities with substantial pocket similarity and the promiscuity of ligand binding (Srinivasan *et al*,

2016a). Their studies provided the first molecular rationalization for the widespread promiscuity of drug target interactions representing safety pharmacology liabilities that is often seen in drug development programs (Skolnick *et al*, 2015; Hu *et al*, 2014; Paolini *et al*, 2006; Bowes *et al*, 2012). They also provided the first rationalization for pockets that are exclusively made up of protein component and pockets that were composite of proteins and another bound small-molecule (Tonddast-Navaei *et al*, 2017). This assessment of promiscuity brings the recognition versus discrimination debate back to the centre stage and poses the question of how living organisms, reliant on accurate detection of signal within the noisy intracellular environment, manage promiscuity. Suffice it to say that likely (1) the cellular information content is in fact embedded within the promiscuous interactions of macromolecules with other small- or macromolecule, i.e. promiscuity is the signal rather than the noise. This could reveal itself as the polypharmacology of beneficial versus deleterious promiscuous interactions. (2) There are other constraints inherent within the biological system and the design of the study (i.e. aspects of chemical discrimination, subcellular localization and segregation and the likely non-representative nature of the PDB biased structural dataset were not taken into consideration), which might indicate that the promiscuity index computed using geometrical complementarity metrics alone might not be sufficient to speculate widespread promiscuity (again leaving an open question on the widespread secondary pharmacology interactions seen). The answer will become clearer with further studies. A group of pockets that are transitory revealed under specific conditions in the dynamic life cycle of an enzyme and/or are stabilized by ligand binding are referred to as cryptic pockets and are emerging as an important subgroup of pockets, thus housing additional information content within protein molecules.

It was known for some time that, some of these pockets have evolved an ability to reduce the activation energy barrier for reaction (thus enabling faster approach towards equilibrium) facilitating the turnover of substrate to product (chemical reactions) on timescales that are compatible with life (Wolfenden & Snider, 2001). These special pockets are housed within proteins called enzymes and are known as the active site. These pockets are often endowed with aspects of specificity (except when promiscuity is seen) and rate enhancements. Aspects of specificity is governed by the nature and composition of the active site (acid/base or nucleophilic functional groups) and the formation of specific bonds between substrate and enzyme (interactions such as covalent, electrostatic, ion-dipole, dipole-dipole, hydrogen, hydrophobic and Van der Waals)(SILVERMAN, 2000). Aspects discussed above are modulated by the pH of the microenvironment that an enzyme inhabits and the pKa of the amino acid, the latter a property of the near neighbour composition of a residue in the primary, secondary and tertiary structural environment. Further, enzymes are also known to expand the chemical armament at their disposal by recruiting coenzymes (organic or organo-metallic compounds) and cofactors (inorganic such as iron-sulphur cluster or metal ions). This constitutes yet another layer of information embedded with the structure of proteins. The presence of nucleotides or nucleotide derivatives as group-transfer coenzymes, prominent examples include Adenosine triphosphate (ATP), S-adenosyl methionine (SAM), Coenzyme A(CoA) and nicotinamide adenine dinucleotide (NAD⁺), is interesting and has given rise to speculation about the retention of RNA facilitated catalysis way after proteins have taken primacy is the information relay (see previous section) and the co-evolution of coenzyme/protein pair (Goldman & Kacar, 2021; Kirschning, 2021).

Wolfenden and colleagues have demonstrated that the magnitude of rate enhancements delivered by active sites are huge reaching a whopping 10¹⁸-fold for some enzymes (Radzicka & Wolfenden, 1995; Wolfenden & Snider, 2001; Miller & Wolfenden, 2002; Srinivasan, 2022, 2021a). Several different postulates and mechanisms have been proposed, and are evolving continuously, to explain these unimaginable rate enhancements (lock and key, induced fit, proximity and orientation effect, entropic constraints, orbital steering, stereopopulation control, substrate anchoring, distortion/rack mechanism, transition state stabilization, covalent catalysis, general acid/base catalysis, electrostatic catalysis, desolvation, and so forth)(Segel, 1976; Richard, 2013; Åqvist *et al*, 2017). Recently protein dynamics has been shown to be a critical factor that contributes to the rate enhancement (Agarwal, 2005; Rajagopalan & Benkovic, 2002). However, it is likely that the eventual true solution to the rate enhancement conundrum could be a combination of the different postulates

stated above contributing to various degrees. This ability by an array of protein catalysts to change a latent (thermodynamically unstable but kinetically stable) metabolite to something else that can fuel the traffic through metabolic circuits eventually results in either the generation of energy through exergonic reactions storing them in the bonds of energy rich molecules such as ATP or in building structure through coupling of energy from these metabolites for driving endergonic reactions. This is yet another nuance on the presence of a Maxwell's demon who operates within biological information circuits until we exactly understand the basis for the rate enhancement and the reduction in the activation energy facilitating us to build artificial protein-based catalysts (Lovelock *et al*, 2022).

The kinetics of an enzyme could be quantified by the framework provided to us by Leonor Michaelis and Maud Menten (Srinivasan, 2022). However, with increasing studies of more complex enzymes, it is emerging that a lot of enzymes do not follow the tenants and assumptions that led to the Michaelis-Menten kinetics (Srinivasan, 2021b). These enzymes display complex behaviours, notable among which are aspects of cooperativity, self-catalysis and non-equilibrium kinetics. In this context, another crucial feature of pockets is their ability for crosstalk. Oftentimes, within a single monomer, orthosteric pocket and the allosteric pocket interacts upon ligand binding at either of the sites. These interactions and their crosstalk often result in complex logical circuits and gates that convey valuable information. For instance, in a homooligomeric organization with positive homotropic cooperativity, binding of a small-molecule to the first subunit results in increased affinity of binding for the same small-molecule for the second and subsequent subunits. This is a special instance of linked equilibrium where the equilibrium dissociation constant (and thus the underlying association and dissociation rate constants) are modulated by an oligomeric coupling constant for all subsequent binding events after the first binding event. Likewise, in the case of negative cooperativity, the first binding event to subunit one reduces the binding affinity of binding for all subsequent subunits in a linked equilibrium (Bush *et al*, 2012). This, it has been shown, can create threshold effects and ultrasensitivity to ligand concentration perturbation conferring a switch-like property to signalling flux and the traffic through these networks (Ha & Ferrell, 2016). This aspect of regulation becomes exponentially complex mathematically for heterotropic interactions (where the identity of the small molecules is different) with disparate substrates and products, regulatory small-molecules and multi-subunit organization leading to several different permutations and combinations of interactions (Koshland *et al*, 1982; Levitzki & Koshland, 1976). To further compound this complexity, these regulatory circuits are implemented at the level of pathways (an array of enzymes) rather than single enzymes. At the network level, finely balanced perturbation of repression and activation, using feedback inhibition of anabolic pathways and feedback activation of catabolic pathways, can cause amplification and control that might be essential for effective information relay and parsimony of how information is converted into useful work.

For long it was believed that a unique sequence, aided by the PTMs it has acquired, folds into a unique three-dimensional structure. While this may still hold true for most of the proteins, it is emerging that a lot of proteins are metamorphic, adopting more than one unique fold in response to unique environmental cues (cofactor starvation, redox challenge, unique PTM and, in some cases, a single mutation) with distinct functional roles for each fold (Dishman & Volkman, 2022; Alexander *et al*, 2009; Porter & Looger, 2018; Murzin, 2008; Schafer & Porter, 2023). It is speculated that almost 5 % of the structures deposited in PDB could be metamorphic (and no quantification for those available in Uniprot!). The degeneracy of protein folds, and thus function, tied to a unique sequence is changing dramatically in turn increasing the information content embedded therein. Though there are several examples in literature, two prominent examples include the fold switching in a small chemokine protein XCL1 and a large iron-regulatory protein/aconitase. The latter is a specific example wherein, under conditions of iron starvation, the iron-sulphur cluster of the globular aconitase protein disassembles and the protein metamorphoses into an elongated fold representing iron regulatory protein that acts to facilitate the iron homeostasis (Dupuy *et al*, 2006).

Other notable exceptions to the one sequence-one structure hypothesis include the widespread occurrence of (1) Intrinsically disordered proteins (IDP) that undergo ligand induced structural organization (Oldfield & Dunker, 2014; Tompa, 2012) and play prominent roles in several important

cellular processes including protein-protein interaction, homeostasis of critical supramolecular complexes, signalling and liquid-liquid phase separation (LLPS) (Wright & Dyson, 2015; Tesei *et al*, 2021), (2) chameleonic sequences, that are sequences in proteins which have equal propensity to fold into different secondary structures elements (i.e. α -helix, β -sheets and/or coils) and have prominent roles in neurodegenerative diseases (Li *et al*, 2015), and (3) morpheins that are alternate quaternary structural assemblies whose association is determined by distinct conformational states of the dissociated state (Jaffe & Lawrence, 2012). All these emerging structural polymorphisms signifies an enriched and hidden layer of information that hasn't yet fully revealed itself to the decoders. Apart from these structural polymorphic proteins, a substantial fraction of proteome codes for protein of unknown function (PUF) (Niehaus *et al*, 2015; Shumilin *et al*, 2012) and a substantial fraction of genome that is expressed has the potential to code for orphan hypothetical proteins (OHP) (Sundararajan *et al*, 2018). These proteins likely represent repositories for future function evolution or protein sequestering pools that can sequester, or release, functionally important proteins based on environmental cues (Srinivasan *et al*, 2015; Méheust *et al*, 2022).

Bulk of what was discussed above pertains to the proteins synthesized by the translational machinery as the last step in the central dogma of information transmission. However, there are peptidic polymers that are synthesized by nonribosomal peptide synthesis (NRPS). The field of NRPS is fascinating and specifies a multienzyme process that synthesizes peptides outside the remit of the central dogma. These peptides have been shown to possess biological activities, such as antibiotic and antifungal properties. The process involves the use of nonribosomal peptide synthetases (NRPSs), which are multidomain enzymes that assemble peptides from both proteinogenic and nonproteinogenic amino acids (Felnagle *et al*, 2008). These are modular enzymes that use assembly line synthesis to synthesize peptides and can create combinatorial diversity in the way they are organized.

The traditional view of subsequent reactions in a metabolic pathway being spatially proximal was hard to understand in a conception of the cytosol that was homogenous and dilute where metabolites were like batons passed between the enzymes (the runners) only limited by the speed of diffusion. However, the heterogeneous nature of cytosol compressed and segregated by membranous organelles and membrane less phase separated entities with variable diffusion rates within the intracellular environment needing guidance and chaperoning to reach the next enzyme in the cascade of transformation has led to the resurgent interest in spatiotemporal segregation and clustering of functionally related proteins within a pathway called metabolons (Møller, 2010). This co-localization of multi-enzyme complexes based upon environmental cues (not inheritable epigenetics!) facilitates the prevention of loss in metabolites, stabilization of labile intermediates, sequestration of the metabolic pool from competing pathways and enables maximal requisite flux (maximal information) through a particular pathway. Prominent examples of metabolons include colocalization of enzymes from glycolysis and TCA cycle (Fuller & Kim, 2021; Zhang *et al*, 2020), purinosome (Pedley *et al*, 2022), glycosome (Wang *et al*, 2023) and so forth. Though definite proof for the presence of these entities is a matter of ongoing debate, this represents yet another layer of complexity in information organization. On similar lines, clustering of enzymes on membranes by lipidic modification or their assembly in the cytosol due to "loose" electrostatic interactions mirrors the concepts of metabolon in some sense.

Just like in the case of RNA molecules, the half-life and the turnover of the protein encodes important informational content. Proteins are continually synthesized and degraded to maintain proteostasis. However, under conditions of external stimuli, the synthesis and degradation rates undergo dramatic changes causing pool expansion and/or contraction. Turnover of proteins is an indelible signature reporting on the physiological state of a cell and the flux through its metabolic pathways. In our previous sections, we have discussed the various ways in which proteins and peptides are synthesized by living organisms. There are several different mechanisms of protein degradation, prominent among which are Ubiquitin (Ub) proteasomal system and lysosomal pathways.

In Ub-proteasomal system, Ub moieties are attached to a lysine residue on the protein marking it for degradation. Ubiquitin is a small, 76-amino acid, ubiquitously present regulatory protein. Three distinct enzymes are involved in the process of ubiquitination of the target protein. E1 (Ub-activating enzyme) activates ubiquitin through an ATP-dependent reaction, E2 (Ub-conjugating enzyme) transfers activated ubiquitin to the target protein of interest (with help from E3)(Stewart *et al*, 2016) and E3 (Ub protein ligase) determines the substrate specificity for ubiquitination. There are more than 700 E3-ligases belonging to two distinct classes (RING-type and HECT-type) that are involved in determining the substrate specificity of ubiquitination and hence, protein degradation(George *et al*, 2018). Ubiquitination is a reversible process whereby deubiquitinases (DUBs) are involved in removing the Ub from the target protein of interest. Defying initial understanding, the ubiquitination code is both complex and multitiered and has roles beyond protein degradation. Broadly, the following main types of ubiquitination are seen: (1) Monoubiquitination and multi-monoubiquitination has important roles in stabilization, distinct subcellular localization, altered binding properties, proteasomal degradation and enzymatic activity of the target proteins. (2) homotypic polyubiquitination of K48 and K11 directs proteins to 26S proteasomal degradation and homotypic polyubiquitination of K63 and M1 mediate complex assembly. Similarly, homotypic polyubiquitination of K27/29, K33, K6 has varied cellular roles and functions, (3) heterotypic ubiquitination involves mixed and branched organization of the ubiquitin monomers and can have complex outcomes in terms of protein fate and role (Yau & Rape, 2016). At the time of this writing, understanding on the evolution of Ub code and its myriad roles are still underappreciated and evolving.

In lysosomal pathway mediated degradation, proteins are engulfed into lysosomal particles and recycled by processes such as endocytosis, phagocytosis and autophagy. These processes have important roles in recycling of misfolded and aggregated proteins.

1.4. Lipids and carbohydrates as information components:

Lipids: In the primordial soup, spontaneous assembly of organic small molecules to form droplets that enclosed solute and other organic molecules might have been the first instance of how information was organized during the evolution of life. This housing of the primordial soup into discrete coacervated compartments served to concentrate reactants, facilitate enzyme reactions by sequestering substrates, and exchange ions and small molecules with their surroundings. This facilitated the coevolution of information molecules like DNA, RNA and proteins and assisted away-from equilibrium states that were conducive to the anabolic reactions to build elaborate intracellular compartmentalization and structure. These were the first entities that served the transition from membrane-less life to life housed within the boundaries of membranes. In the evolved cells, all membranes are lipids and are critical for cellular compartmentalization by serving as important components of the plasma membrane and other subcellular boundaries such as nuclear membrane, endoplasmic reticulum, golgi apparatus, mitochondrial membrane and vesicles such as endosomes and lysosomes with prominent roles in intracellular trafficking (Muro *et al*, 2014). This brief introduction serves to establish the primacy of lipids as the most critical information currency within the biological information matrix.

Biological lipids are a hugely diverse group of macromolecules (competing with the diversity embedded in proteins) and encompass molecules like fats/oils, phospholipids, waxes and steroids. Fats/oils are esters of a 3-carbon polyol glycerol bonded to 3 fatty acids. The diversity of the lipids originates from the diversity of the hydrocarbon chains of the fatty acid that can vary in length (typically even from 14-24 long) and unsaturation (with aspect of *cis* and *trans* organization) and carboxylic acid groups at their terminus. Glycerophospholipids (GPLs), which are the principal components of membranes, are slight variation on the structure of fats/oils by substituting one fatty acid on the glycerol with a phosphate moiety with variable headgroup substituents. Other components of membranes include sphingolipid and cholesterol. The diversity embedded within the constituents of membrane lipids is enormous depending on the type of fatty acid used (Palmitic acid,

stearic acid, oleic acid, arachidonic acid, linoleic acid, γ -linolenic acid, α -linolenic acid and docosahexaenoic acid), the diversity of headgroups based on phosphoinositide substituents and diversity of the oligosaccharide substitutions (with different constellation of monosaccharides stitched together). A full description of this combinatorial expansion of the lipid code is beyond the scope of this perspective and readers are referred to other exceptional sources (Harayama & Riezman, 2018). The amphipathic nature of the phospholipids with a polar headgroup (made of glycerol and the phosphate moiety and the two hydrophobic tails) allows them to form lipid bilayers thus acting as barrier between two aqueous compartments. The heterogeneity and distribution of the several phospholipids determine unique properties of membranes. For instance, cardiolipins have repeating units of phosphoryl and glycerol moieties i.e. polyglycerophospholipids (PtdGro) and are enriched at membrane curvatures associated with cell poles (and thus implicated in replication) and mitochondrial cristate (Schlame, 2008). Another interesting lipid is the phosphatidylinositol 4,5-bisphosphate (PIP₂) that are enriched in plasma membranes around membrane ruffles and nascent phagosomes (McLaughlin *et al*, 2002).

Lipidation of carbohydrates and proteins constitutes another large fraction of potentially information rich roles of lipid molecules by modulating the structure and, thus, information content of these other macromolecules. We briefly saw this role of lipids while discussing the PTMs on proteins. We will discuss a few additional examples of them here. (1) A prominent mode of protein anchoring onto membrane is the addition of reversible protein conformation specific addition of glycosylphosphatidylinositol (GPI) anchors to nascent polypeptide chains in the endoplasmic reticulum (ER). In this type of linkage, the carboxy terminus of a protein is linked via a phosphodiester bond to a trimannosyl-glucosamine (Man3-GlcN) core tethered phosphoethanolamine. The glucosamine is attached to a phosphatidylinositol that is business end of the molecule anchored to the membrane. GPI-anchored proteins are enriched in the cholesterol- or sphingolipid-enriched regions of membranes called lipid rafts and play important roles in signalling, proliferation and homeostasis (Kinoshita, 2016; Reily *et al*, 2019). (2) N-myristoylation is addition of C14 saturated fatty acids onto the N-terminus of proteins (after methionine cleavage) for their reversible localization onto membranes. N-myristoylated proteins play important roles in signalling (e.g. Src-family kinases) (3) S-palmitoylation is a PTM whereby C16 palmitoyl group is added to the thiolate side chain of cysteine reversibly and is usually present in combination with other modifications such as myristoylation and farnesylation. Like GPI anchors, proteins that are S-palmitoylated are selectively enriched on lipid rafts. (4) S-prenylation is the irreversible covalent addition of either C15 farnesyl or C20 geranylgeranyl isoprenoid groups (or a combination of both) to the cysteine residues at the carboxy-terminus of proteins. This modification is tethered in the ER and is dependent on a 4-amino acid cysteine proximal motif. Prominent examples of proteins that undergo S-prenylation include Ras family GTPases and heterotrimeric G-proteins (Palsuledesai & Distefano, 2015). We would be looking at glycolipidation (addition of lipid modifications to carbohydrates) in the next paragraph. Apart from roles in structure and membrane, lipids also serve as important energy storage molecules (triglycerides and sterol esters) and informational molecules in the endocrine signalling (cortisol, aldosterone, estrogen, progesterone, and testosterone). With increasing means of characterizing the diversity and functional versatility of lipids, it is emerging that they are one of the most important informational molecules, oftentimes rivalling the role proteins and nucleic acids play.

Carbohydrate: Carbohydrates (glycans), the most abundant organic material in nature, are one of the most underappreciated yet emerging to be one of the most important constituents of cellular informational currency. The misleading simplicity of a carbohydrate polymers, as written on a piece of paper, defies the configurational, conformational and combinatorial diversity that carbohydrates generate to maintain cellular homeostasis and propagate information. Cells communicate with their surrounding via the nuanced diversity of the carbohydrate codes. Stated differently, cells assume their identity exclusively dependent on their carbohydrate signatures. Stultifyingly complex, extremely dense and combinatorially diverse carbohydrates decorate the cell surface enabling cell-cell recognition, communication, defence and host-pathogen interactions (Varki *et al*, 2017). A

prominent example of this combinatorial diversity is seen in the histo-blood group antigens that helps classify blood-group type as A+, A-, B+, B-, O+, O-, AB+ and AB-. Sugars are also recognized by proteins called lectins that facilitate adhesion of cellular entities and, in turn, elicit specific responses (specific examples include fertilization of eggs by sperms, bacterial or viral infection, immune recognition and so forth) (Brandley & Schnaar, 1986). Similarly, most of the lipids, hormones and proteins in the human body are glyco-conjugated. This conjugation helps facilitate the adoption of the correct tertiary structure and stability which are essential for the functioning of these macromolecules.

Oligosaccharides and polysaccharides are polymers of monosaccharide units linked by glycosidic bonds. The physicochemical properties of carbohydrates are diverse and are determined by factors such as the combinatorial arrangement of monosaccharide units, length of the polymer chain, linear versus branched chain arrangement, stereochemistry (configurational and conformational), molecular weight and potential modifications. Borrowing the vocabulary used for understanding protein structure, the primary structure of a glycan polymer is understood in terms of the linear sequence of the monosaccharides. The relative orientation of the two monosaccharide units connected via the glycosidic bond is assessed by the torsion angles Φ ($H_1-C_1-O_x-C_x$) and Ψ ($C_1-O_x-C_x-H_x$). ω ($O_6-C_6-C_5-O_5$) defines the additional torsion angle for the 1,6 glycosidic bonds. These torsion angles are affected by exo-anomeric effect, steric interactions, electronic effect, hydrogen bonding and the effect of solvation and play a large part in understanding the glycan conformation. The structural diversity of carbohydrates can be further increased by modification of their monosaccharide units by acylation, methylation, sulfation, epimerization, and phosphorylation (Muthana *et al*, 2012).

A substantial fraction of the human glycome is invested in modifying the proteome (Schjoldager *et al*, 2020). N-glycosylation is one of the most prominent carbohydrate modifications of both secreted and membrane localized proteins on its asparagine residue. In this type, the carbohydrate moiety is covalently attached through the anomeric carbon of a sugar molecule to the nitrogen atom of an amine. The core-motif of the N-glycan modification contains a pentasaccharide core motif (a chitobiose ($GlcNAc\beta 1\rightarrow 4GlcNAc$) and three mannose). N-glycans can be broadly classified into oligomannose, complex and hybrid depending on the residue/s that are attached to the core motif. For oligomannose, either linear or branched mannose chains are added to the core motif, while a combination of mannose, sialic acid, galactose and fucose chains are added to the complex N-glycans. In the hybrid N-glycosylation, a combination of mannose chains and heterogenous chains are added simultaneously on the core motif (Yu & Delbianco, 2020). N-glycanases are enzymes that remove glycan groups from proteins to facilitate their degradation by proteasomal machinery. Deficiency of N-glycanase function can lead to disorders where misfolded proteins aggregate intracellularly leading to different physiological outcomes that are increasingly appreciated (Srinivasan *et al*, 2016b).

Of higher complexity and informational content is the O-glycosylation pathway. This involves the attachment of activated carbohydrate chains to serine/threonine residues in proteins in the ER and Golgi apparatus of eukaryotic cells. O-glycans are classified based on the first sugar tether added to the protein followed by subsequent sugars added to the initial glycan. e.g. mucin-type O-glycosylation. Proteins containing the epidermal growth factor (EGF) repeats or thrombospondin type I repeat (TSR) have also been shown to be O-glycosylated with O-linked fucose and/or O-linked mannose. Both O- and N- linked glycans are capped with negatively charged sialic acid.

Other carbohydrate modifications are GPI anchored proteins (discussed above), proteoglycans, and Glycosphingolipids. In proteoglycan modification of proteins, long glycosaminoglycan (GAG) chains are attached to the hydroxyl group of serine residues (embedded within the Ser-Gly-X-Gly motif) on target proteins through a tetra-saccharide core consisting of [glucuronic acid (GlcA)]-[galactose(Gal)]₂-[xylose(Xyl)]; This modification is quite diverse and depending on the number, composition and degree of sulfation on the repeating disaccharide, proteoglycan GAGs can be classified into several different classes; prominent examples include heparan sulfate, chondroitin sulfate and dermatan sulfate. Apart from GAG addition to proteins, free GAGs like hyaluronan, are

synthesized at the plasma membrane by sequential addition of GlcA and GlcNAc and have several important roles in cellular information relays.

Glycosphingolipid conjugation of glycans with cellular membrane lipids is another major class of modifications where galactose and glucose moieties are added on the lipids. Similar to our discussions for other macromolecules, there is information content embedded in turnover rate of lipid and carbohydrates.

1.4. Small-molecules, ions and Metabolites:

If composition of a bacterial cell is assessed, it is 74 % small-molecules (ions and inorganic small molecules, sugars, fatty acids, amino acids, nucleotides and water) and 26 % macromolecules. Of the 74 % small molecules, the bulk of it (70%) is water leaving the rest 4 % to be of other small molecules. It comes down eventually to the flux and signalling potential of small molecules, importantly water, that dictates biological responses.

Water is a solvent that borders on being magical in its ability to sustain life. The polarity of water molecule, with spatial separation of its positive and negative poles, and the asymmetry confers upon it the ability to interact with itself and other polar molecules (Ball, 2017; Dargaville & Hutmacher, 2022). This, in turn, enables cohesion and formation and stabilization of polymeric structures of macromolecules. For instance, water's ability to exclude the hydrophobic part of the amphipathic phospholipid facilitates their clustering to form lipidic membranes that compartmentalize cells. Similarly, the hydrophobic collapse witnessed during protein folding is a direct consequence of aliphatic amino acids coalescing together to exclude water. On similar lines, water is critical for stabilization of the tertiary structure of nucleic acids. Water, given its high dielectric constant of 81, functions to mute the magnitude on charged particles making them compatible with the soft nature of biological reactivity. Water also serves as acid and/or base in important biochemical reactions and buffers the medium. Furthermore, selective exclusion of water from micro pockets within biological systems results in charge magnification and lowering of activation energy for enzyme catalysed reactions and other applications. Most importantly, give the amphoteric, protic and polar nature of this molecule, water is rightly termed the universal solvent making the information relays within biological systems feasible. Its role is aptly summarized as being very critical for "*refining and conditioning intermolecular information transfer*" (Ball, 2017).

There are several metal ions that are important for aiding structure, function and signalling within intracellular information relays. Notable among these are monovalent ions such as Na^+ , K^+ , redox-inert divalent ions such as Mg^{2+} , Ca^{2+} , Zn^{2+} and redox-active metal ions such iron, manganese, cobalt, molybdenum, copper and nickel. These metal ions are predominant players in signalling cascades and are integral constituents of metalloenzymes and metal-dependent enzymes. Further, they function by neutralizing negative charges in enzyme pockets by forming coordination bonds with amino acid residues or adducting with negatively charged substrates/coenzymes (i.e. nucleotide- Mg^{2+} complex), activating substrates through their Lewis acid properties, regulating redox biology and acting as antioxidants (Andreini *et al*, 2008).

Most of the metabolites within mammalian systems are either intermediates or end-products of biological transformations of substrates brought about by enzymes as part of metabolic cascades. Prominent metabolic pathways within mammalian systems include glycolysis, pentose phosphate pathway, glycogen metabolism, amino acid biosynthesis/degradation and urea cycle, tricarboxylic acid cycle, gluconeogenesis, pyruvate decarboxylation, ketogenesis, fatty acid biosynthesis and fatty acid beta oxidation. Other pathways of note include salvage pathways for purine nucleoside/nucleobase, pyrimidine nucleoside and nucleotide reuptake and *de-novo* biosynthetic pathways for nucleotides and thymidylates. It is estimated that there are more than 8000 metabolites (mol. w. <1500) whose flux through cellular metabolic cascades is tightly controlled by several hundred enzymes, pathway branching nodes and nuanced regulation effected by feedback modulation, rate-limiting enzymes and gatekeepers (Walsh *et al*, 2018). This makes an appreciation of the complexity of information encoded in small molecules challenging. The complexity of the metabolite pool is not encoded by the genome directly, but the enzymes encoded in the genome are

critical in the biosynthesis, degradation and transformation of these small-molecule pools that create the palette of varied chemistries that encodes life. Simultaneous use of bottom-up and top-down approaches to establish the flow of causality and the assembly of information in biological pathways is commonplace but the way non-genome encoded metabolic pool coalescences and is-controlled by genomically encoded protein pool is a beautiful example of bottom-up approach in biological information organization. The ability to assess the steady state concentration of 100s of metabolites has revolutionized the way metabolome has been thrust on the centre stage of cellular information relays (Siuzdak *et al*, 2012).

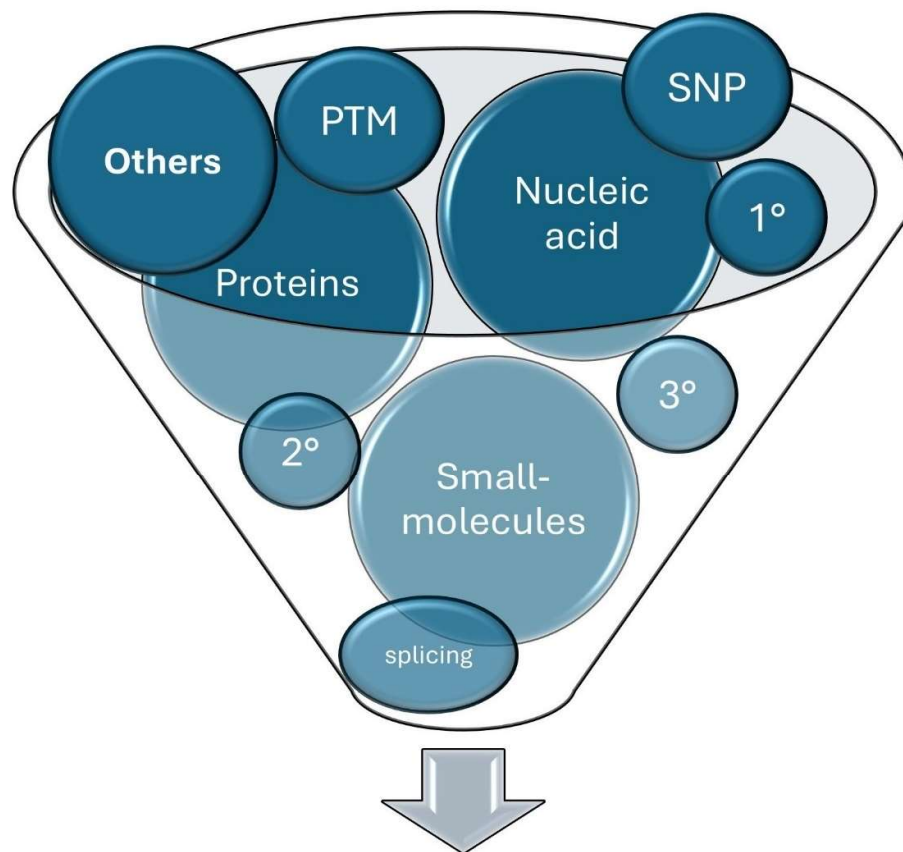
Metabolites can be classified broadly into *primary* and *secondary* depending on their physiological utility in the context of an organism's metabolism (James, 2017; Torres & Schmidt, 2019). Primary metabolites include monomers of carbohydrates, amino acids, fatty acids and organic acids. Though extensive studies have been carried out on the secondary metabolite pool of fungal, bacterial and plant systems (with extensive discussion done elsewhere), we would attempt focussing here on the most important classes of secondary metabolites that are broadly divided into NRPS and alkaloids, terpenes, shikimate-like and polyketides. Secondary metabolites are usually involved in protective or offensive roles, and/or involved in communication (pheromones and hormones).

Among primary metabolites, the group transfer potential of seven small molecules is called out as the most critical for dictating the intertwined logic of biological circuitry. ATP, NADH and acetyl-CoA are generated during catabolic processes that, in turn, is used to drive anabolic pathways of small molecules and macromolecules. These three molecules are important because of their abilities to transfer phosphoryl- or nucleotide groups by ATP, electron donation by NADH along the respiratory chain components to molecular O₂ and acyl-group transfer by acetyl-CoA facilitating the anabolism of carbon scaffolds. These group transfers play critical roles in the biosynthesis of small-molecules and macromolecules such as proteins, nucleic acids and oligonucleotides. The other four metabolites with equivalent group transfer potential are S-adenosyl methionine (SAM) as a universal donor of methyl groups, UDP-glucose as a donor of the glycosyl ring for oligosaccharide biosynthesis, carbamoyl phosphate as a means of capturing ammonia and channelling it into biosynthetic nitrogen metabolism (along with glutamine and glutamate) and Δ^2 -isopentenyl PP with critical roles in building isoprenoid frameworks (Walsh *et al*, 2018).

Given the emerging technological prowess to estimate the dynamic flux of metabolites, their role and import on information relay will become prominent with further studies.

1.4. At the intersection of small and large molecules

A repeated motif that was seen all throughout this essay was the recurrence of how small molecules interacted with macromolecules and how one macromolecule interacts with other macromolecules to generate the combinatorial possibility of generating and controlling information transmission (**Figure 2**). Multi-component Signal transduction pathways (controlling aspects of cell-cycle regulation, apoptosis and differentiation) and regulatory circuits (that control replication, gene expression, translation) are instances of complex and well-orchestrated crosstalk between several different macromolecules and small-molecules in-tandem to elicit integrated cellular responses to perturbations (Zhang *et al*, 2013; Marijuán & Navarro, 2021). The interactions of all these components enacted in space and time constitutes the way the embedded information of the biological system is revealed eventually.



Cellular Homeostasis

Figure 2. Information on cellular homeostasis is encoded in the multidimensional space of macromolecular sequence, structure, post-translational modification (PTMs) and their interaction with small molecules.

A very well studied “*logic module*” that has been discussed extensively in conventional biochemistry textbooks and that signifies the intricate interaction of protein, nucleic acid and small-molecule players is the regulation of Lac operon. Though we wouldn’t be getting into the details of this operon’s function, it would have to be emphasized that this regulation represents how a small-molecule, depending on its increased steady-state level, interacts with an oligomeric protein to change its conformation in turn making it lose its affinity of binding for the DNA partner, rendering the DNA accessible to transcriptional and translational machinery that codes for a series of enzymes that can hydrolyse the metabolite bringing down its steady-state concentration reinstating the blockade. This module, often referred to as feedback cyclical loop, operates in a homeostatic manner and is a common feature of several metabolic pathways and their crosstalk (i.e. transition from glycolysis to gluconeogenesis under conditions of high energy charge, inhibition of threonine deaminase by isoleucine and so forth). A clear integrative understanding of all these modules and their potential linkages can help one assemble the motherboard housing all these circuits (Piazza *et al*, 2018). This requires understanding the sources from which information is harnessed, means of putting the harnessed information within a unified framework and processing it, and how, eventually, is it leveraged, stored or discarded to generate the signal over the cellular noise (Nurse, 2008).

The initial perception of cytoplasm as a dilute medium with diffusion limited Brownian approach of interacting partners has been convincingly proven wrong (**Figure 3A**). The beautiful and scientifically accurate illustrations of intracellular crowding by Dave Goodsell have given this aspect strong visual reinforcement for any lingering doubts that isolated pockets of investigators might still harbour. With this added complexity, one would have to be conscious of the fact all these multicomponent interactions are happening within the highly heterogenous and crowded interiors of the intracellular environment with considerable influence on the reactivity and distribution of macromolecules/small molecules (Collette *et al*, 2023; Rivas & Minton, 2016). In this context, crowding should be understood as high concentration (weight/unit volume) of non-interacting soluble entities that are functionally unrelated (**Figure 3B**). Crowding results in increased effective viscosity of the medium thus reducing diffusion rates. Further, crowding facilitates depletion effects by segregating macromolecules by size thus increasing the free volume for solutes. It can also have important roles in modulating the reactivity, diffusion rates of macromolecules (sieving), folding, molecular recognition, discrimination and assembly in a multimolecular complex. The fact that a small molecule finds its macromolecular interacting partner, or one macromolecule finds another macromolecule, for generating specific interaction (and hence signal) within a reasonable timeframe (from seconds to minutes) must be potentially facilitated by informational highways operating within cells. Very less is known or appreciated about these informational highways (**Figure 3C**).

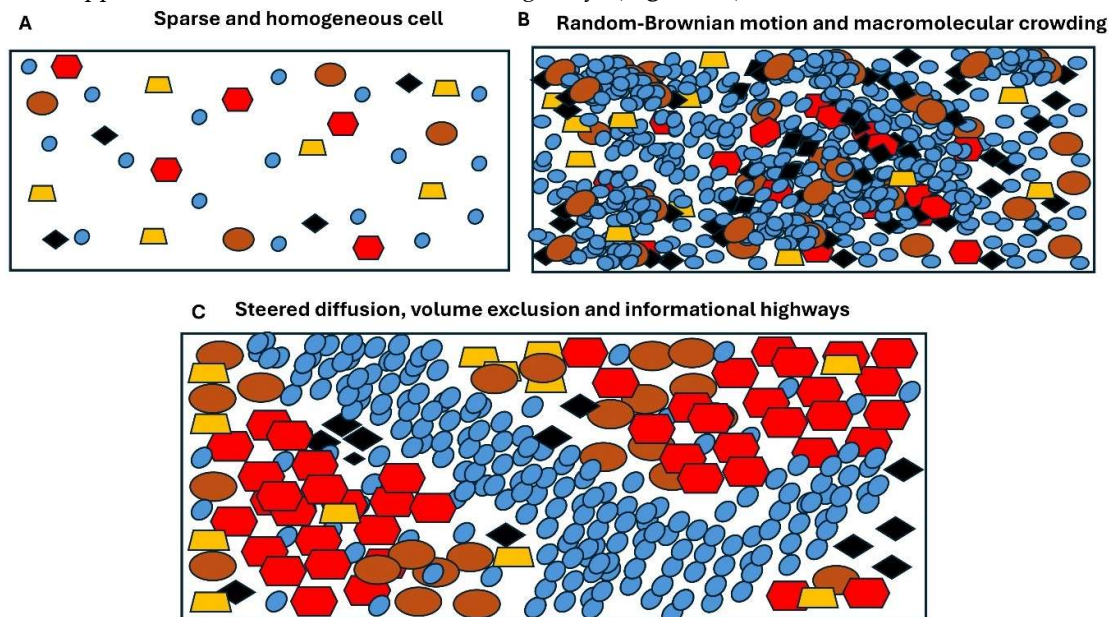


Figure 3. Evolving knowledge on the nature and density of intracellular milieu. (A) Initial assumption that intracellular space was sparse and homogeneous with bimolecular second order interactions constrained by the limit of diffusion. (B) Intermediate understanding that intracellular spaces are highly crowded and heterogenous slowing the rate of bimolecular interactions. (C) Evolving understanding that, though intracellular space is crowded and heterogenous, there are micro pockets of homogeneity and bimolecular interactions are facilitated by steered diffusion and volume exclusion, similar to highways regulated by traffic signals.

A few prominent examples of such events that defy the limit of diffusion (10^8 - 10^9 $M^{-1} s^{-1}$), include (1) DNA-binding proteins find their cognate DNA sequence at rates that are an order of magnitude higher than that expected with mere diffusion limit. This, it is speculated, is facilitated by 1-dimensional and 3-dimensional search carried out in-tandem where facilitated diffusion mediated by sliding and intermittent hopping quickens the search for the cognate DNA sequence (Halford & Marko, 2004; Von Hippel & Berg, 1989) and, (2) Enzymes that catalyse reactions with catalytic efficiencies (k_{cat}/K_m) that surpass the diffusion limit (i.e. $> 10^{10}$ $M^{-1} s^{-1}$). These high catalytic efficiencies have been ascribed to aspects of dipolar electrical fields and electrostatic interactions.

On similar lines, it came as a surprise to many investigators that the steady state concentration of metabolites intracellularly remains invariable (homeostatic) over large scale changes in the flux through various pathways. This is known as the stability paradox which is highly counterintuitive to explain if one assumes homogenous sparse cytosol with uniform rates of diffusion. However, increasing understanding of molecular systems as machines that are conducted across cytosol at variable speed depending on the necessity (i.e. myosin and kinesin motors running on actin/tubulin filaments at variable speed) has given rise to the concept of “*intracellular perfusion or convection*”. This explains why change in convection rates can change the rate at which metabolites can increase with their cognate protein and thus bring pathway flux change without effecting the steady state metabolite concentration (Hochachka, 1999).

For simplicity's sake, it would be wrong to assume that intracellular traffic is the jostling and bumping of a molecule within a swarm of entities in Brownian motion to locate its partner solely by chance encounter at any respective steady state level. In fact, molecular search and recognition are highly orchestrated events guided by highways of information relay with signposts and lane markings guiding that traffic. Additionally, if demand arises, the transportation rate can be selectively accelerated or decelerated to the probability of macromolecular interactions without increasing the concentration of the steady state level of a metabolite (**Figure 3C**).

1.4. Epilogue:

This brings me to the epilogue of this essay. Despite best attempt to use this space to summarize the perspective, I have a keen awareness that I might be leaving this space with further questions that compound the already daunting mass of information and their interconnectivity that biological systems represent. Recognition and discrimination are the core of how information is treated within the biological system. As alluded to above, most of what we know as information within the biological system is firstly reliant on the unique properties of water and how it shapes and evolves the multilayered, multitiered, multicomponent and modular nature of biological information. The intricate nature and complexity of this information would be non-existent if the solvent were not water.

This interplay in the aqueous broth sustained by water is dependent on equilibrium and non-equilibrium events happening on the steady-state and pre-steady state timescales of macromolecular and small-molecular interactions that are either exergonic or endergonic competing among thousands of such interactions happening simultaneously. This unfurling spatiotemporal theatre is orchestrated in a precise fashion by the thermodynamic sentinels who ration everything based upon the principles of staunch thermodynamic parsimony. It is also coming to our attention more that, under the open thermodynamic settings of the cellular system that enables exchange of both matter and energy, non-equilibrium modalities of interaction with insurmountable kinetic barriers predominate the intra- and inter-molecular dialogues among the macromolecules and small molecules. Recognition versus discrimination in a broth steeped with aspects of promiscuity and the precise spatiotemporal regulation of these information highways in the crowded intracellular environment does allude to smoke making one suspect the fire of invisible demons unwinding and winding the information circuits often hoodwinking the thermodynamic sentinels. The full effect and implication of kinetic versus thermodynamic control under non-equilibrium conditions within cells is evolving and will likely mature in the decades to come.

As I alluded to above, despite the huge oeuvre of work, we are yet to have definitive answers for a lot of “*chicken and egg*” scenarios within biological information channelling that, likely, are operated by Maxwell's and other demons. Why did evolution choose water as the medium, why of all the elements carbon/oxygen/nitrogen/sulphur and phosphorus are enriched in biopolymers, why was nucleic acid the choice as the hereditary material, why are sugars D-enantiomers, why are proteinaceous amino acids the L-enantiomers, why are most helices right handed, what is the true implication of the combinatorial codes for histone modification/ubiquitin decorations/and other biological information repositories, what is the true extent of proteoforms space (Aebersold *et al*, 2018), why did nature create such profound gap in the depth of chemical time, how much of the

information content within a cell is “dark”, why is the diester bond and phosphate happens to be such an important component of biology, does cellular memory exist and the questions never cease.

In this essay, we tried understanding the various levels at which information is organized intracellularly at the molecular level, However, even at the cellular level, we are far from appreciating the true scale and magnitude in the regulation of events like cellular homeostasis, structural coherence, spatiotemporal organization, signalling relays both within and among cells, reproduction and cellular memory as a distinct entity compared to body memory (BM). Complexity and organization beyond the cell are beyond the scope of this perspective and includes aspects like organization of tissues, organs, organ systems, organisms, inter-organismal interactions, an organism's or community's interaction with its abiotic environment and so forth. This involves modelling and understanding of complex phenomenon such as respiration, circulation, digestive, nervous and immune system to mention but a few and the intricate way in which the framework of developmental biology operates for ontogeny.

In conclusion, this perspective was an attempt at understanding and appreciating the multifaceted and complex nature of information embedded within the cellular boundaries and their nuanced spatiotemporal regulation giving rise to the phenomenon of life. Much needs to be understood and this framework is dynamic and ever evolving. I earnestly hope that this perspective provides a unified and simplified framework for an aspiring student of biology to appreciate the enormity of the information hoard that biological systems represent.

Acknowledgements: The author would like to acknowledge Dr. Rachel Grimley and Prof. Peter Tonge for their constant support and encouragement. The author would also like to acknowledge the following individuals for their exceptional teaching skills and resource materials they have generated that are accessible to both experts and lay-audience alike: Robert Copeland, Philip Nelson, P. Balaram, M.R.N.Murthy, Charles Cantor, Paul Schimmel, Carl Ivar Branden, John Tooze, Richard Wolfenden, Christopher Walsh, Richard Bruce Silverman, Thomas Creighton, Francis Crick, John Desmond Bernal, David Goodsell, G.N. Ramachandran, Louis Lyons and the innumerable other authors who have helped me comprehend this phenomenon that we know as life.

Conflict of interest: The author is affiliated with the Robert Gordon University as an associate senior lecturer and declares no conflict of interest.

Abbreviations: LLPS, Liquid-liquid phase separation; NRPS, nonribosomal protein synthesis; ORF, open reading frame; IDPs, intrinsically disordered proteins; OHPs, orphan hypothetical proteins; PUF, Proteins of unknown function; PTMs, posttranslational modifications; RISC, RNA-induced silencing complex; PTGS, post-transcriptional gene silencing; CDGS, chromatin-dependent gene silencing; RNP, ribonucleoprotein; 5mC, 5-methyl cytosine; HATs, histone acetyltransferases; HDACs, histone deacetylases (HDACs); HKMT, histone lysine methyltransferases; DNMT, DNA methyltransferases.

References

1. Abelson HT, Johnson LF, Penman S & Green H (1974) Changes in RNA in relation to growth of the fibroblast: II. The lifetime of mRNA, rRNA, and tRNA in resting and growing cells. *Cell* 1
2. Adami C (2004) Information theory in molecular biology. *Phys Life Rev* 1 doi:10.1016/j.plrev.2004.01.002 [PREPRINT]
3. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, Costello CE, Cravatt BF, Fenselau C, Garcia BA, *et al* (2018) How many human proteoforms are there? *Nat Chem Biol* 14 doi:10.1038/nchembio.2576 [PREPRINT]
4. Agarwal PK (2005) Role of protein dynamics in reaction rate enhancement by enzymes. *J Am Chem Soc* 127
5. Alemasova EE & Lavrik OI (2019) Poly(ADP-ribosyl)ation by PARP1: Reaction mechanism and regulatory proteins. *Nucleic Acids Res* 47: 3811–3827 doi:10.1093/nar/gkz120 [PREPRINT]
6. Alexander PA, He Y, Chen Y, Orban J & Bryan PN (2009) A minimal sequence code for switching protein structure and function. *Proc Natl Acad Sci U S A* 106
7. Anatskaya O V. & Vinogradov AE (2022) Polyploidy as a Fundamental Phenomenon in Evolution, Development, Adaptation and Diseases. *Int J Mol Sci* 23 doi:10.3390/ijms23073542 [PREPRINT]

8. Andreini C, Bertini I, Cavallaro G, Holliday GL & Thornton JM (2008) Metal ions in biological catalysis: From enzyme databases to general principles. *Journal of Biological Inorganic Chemistry* 13
9. Åqvist J, Kazemi M, Isaksen GV & Brandsdal BO (2017) Entropy and Enzyme Catalysis. *Acc Chem Res* 50
10. Ball P (2017) Water is an activematrix of life for cell and molecular biology. *Proc Natl Acad Sci U S A* 114 doi:10.1073/pnas.1703781114 [PREPRINT]
11. Bannister AJ & Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* 21: 381–395 doi:10.1038/cr.2011.22 [PREPRINT]
12. Batey RT, Rambo RP & Doudna JA (1999) Tertiary motifs in RNA structure and folding. *Angewandte Chemie - International Edition* 38 doi:10.1002/(SICI)1521-3773(19990816)38:16<2326::AID-ANIE2326>3.0.CO;2-3 [PREPRINT]
13. Bernardi G (1993) The vertebrate genome: Isochores and evolution. *Mol Biol Evol* 10
14. Bevilacqua PC, Williams AM, Chou HL & Assmann SM (2022) RNA multimerization as an organizing force for liquid–liquid phase separation. *RNA* 28
15. Binder PM & Danchin A (2011) Life's demons: information and order in biology. *EMBO Rep* 12
16. Blinov VM, Zverev V V., Krasnov GS, Filatov FP & Shargunov A V. (2017) Viral component of the human genome. *Mol Biol* 51 doi:10.1134/S0026893317020066 [PREPRINT]
17. Boël G, Danot O, de Lorenzo V & Danchin A (2019) Omnipresent Maxwell's demons orchestrate information management in living cells. *Microb Biotechnol* 12
18. Bowes J, Brown AJ, Hamon J, Jarolimek W, Sridhar A, Waldron G & Whitebread S (2012) Reducing safety-related drug attrition: The use of in vitro pharmacological profiling. *Nat Rev Drug Discov* 11 doi:10.1038/nrd3845 [PREPRINT]
19. Brandley BK & Schnaar RL (1986) Cell-surface carbohydrates in cell recognition and response. *J Leukoc Biol* 40 doi:10.1002/jlb.40.1.97 [PREPRINT]
20. Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129
21. Bush EC, Clark AE, DeBoever CM, Haynes LE, Hussain S, Ma S, McDermott MM, Novak AM & Wentworth JS (2012) Modeling the Role of Negative Cooperativity in Metabolic Regulation and Homeostasis. *PLoS One* 7
22. Butcher SE & Pyle AM (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc Chem Res* 44
23. Cahill MA, Ernst WH, Janknecht R & Nordheim A (1994) Regulatory squelching. *FEBS Lett* 344
24. de Chadarevian S (2018) John Kendrew and myoglobin: Protein structure determination in the 1950s. *Protein Science* 27: 1136–1143
25. Chagin VO, Stear JH & Cardoso MC (2010) Organization of DNA replication. *Cold Spring Harb Perspect Biol* 2 doi:10.1101/cshperspect.a000737 [PREPRINT]
26. Chastain M & Tinoco I (1991) Structural Elements in RNA. *Prog Nucleic Acid Res Mol Biol* 41
27. Chekulaeva M & Rajewsky N (2019) Roles of long noncoding rnas and circular rnas in translation. *Cold Spring Harb Perspect Biol* 11
28. Chen CYA, Ezzeddine N & Shyu A Bin (2008) Chapter 17 Messenger RNA Half-Life Measurements in Mammalian Cells. *Methods Enzymol* 448 doi:10.1016/S0076-6879(08)02617-7 [PREPRINT]
29. Cobb M (2017) 60 years ago, Francis Crick changed the logic of biology. *PLoS Biol* 15
30. Collette D, Dunlap D & Finzi L (2023) Macromolecular Crowding and DNA: Bridging the Gap between In Vitro and In Vivo. *Int J Mol Sci* 24 doi:10.3390/ijms242417502 [PREPRINT]
31. Danan C, Manickavel S & Hafner M (2022) PAR-CLIP: A Method for Transcriptome-Wide Identification of RNA Binding Protein Interaction Sites. In *Methods in Molecular Biology*
32. Dargaville BL & Hutmacher DW (2022) Water as the often neglected medium at the interface between materials and biology. *Nat Commun* 13
33. Delaunay S, Helm M & Frye M (2024) RNA modifications in physiology and disease: towards clinical applications. *Nat Rev Genet* 25 doi:10.1038/s41576-023-00645-2 [PREPRINT]
34. Dickinson GD, Mortuza GM, Clay W, Piantanida L, Green CM, Watson C, Hayden EJ, Andersen T, Kuang W, Graugnard E, et al (2021) An alternative approach to nucleic acid memory. *Nat Commun* 12
35. Diniz WJS & Canduri F (2017) Bioinformatics: An overview and its applications. *Genetics and Molecular Research* 16 doi:10.4238/gmr16019645 [PREPRINT]
36. Dishman AF & Volkman BF (2022) Design and discovery of metamorphic proteins. *Curr Opin Struct Biol* 74 doi:10.1016/j.sbi.2022.102380 [PREPRINT]
37. Duempelmann L, Skribbe M & Bühler M (2020) Small RNAs in the Transgenerational Inheritance of Epigenetic Information. *Trends in Genetics* 36 doi:10.1016/j.tig.2019.12.001 [PREPRINT]
38. Dupuy J, Volbeda A, Carpentier P, Darnault C, Moulis JM & Fontecilla-Camps JC (2006) Crystal structure of human iron regulatory protein 1 as cytosolic aconitase. *Structure* 14
39. Dworkin JP, Lazcano A & Miller SL (2003) The roads to and from the RNA world. *J Theor Biol* 222
40. Edwards AL & Batey RT (2010) Riboswitches: A common RNA regulatory element. *Nature Education* 3

41. Engelhart AE & Hud N V. (2010) Primitive genetic polymers. *Cold Spring Harb Perspect Biol* 2 doi:10.1101/cshperspect.a002196 [PREPRINT]
42. Fabris F (2009) Shannon information theory and molecular biology. *Journal of Interdisciplinary Mathematics* 12 doi:10.1080/09720502.2009.10700611 [PREPRINT]
43. Farnsworth KD, Nelson J & Gershenson C (2013) Living is Information Processing: From Molecules to Global Systems. *Acta Biotheor* 61
44. Felnagle EA, Jackson EE, Chan YA, Podevels AM, Berti AD, McMahon MD & Thomas MG (2008) Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol Pharm* 5 doi:10.1021/mp700137g [PREPRINT]
45. Fuller GG & Kim JK (2021) Compartmentalization and metabolic regulation of glycolysis. *J Cell Sci* 134 doi:10.1242/jcs.258469 [PREPRINT]
46. Furth N & Shema E (2022) It's all in the combination: decoding the epigenome for cancer research and diagnostics. *Curr Opin Genet Dev* 73 doi:10.1016/j.gde.2022.101899 [PREPRINT]
47. Gabriel L, Srinivasan B, Kus K, Mata JF, Amorim MJ, Jansen LET & Athanasiadis A (2021) Enrichment of α domains at cytoplasmic stress granules is due to their innate ability to bind to nucleic acids. *J Cell Sci* 134
48. Gao M & Skolnick J (2012) The distribution of ligand-binding pockets around protein-protein interfaces suggests a general mechanism for pocket formation. *Proc Natl Acad Sci U S A* 109
49. Gao M & Skolnick J (2013) A Comprehensive Survey of Small-Molecule Binding Pockets in Proteins. *PLoS Comput Biol* 9
50. Garcia HG, Grayson P, Han L, Inamdar M, Kondev J, Nelson PC, Phillips R, Widom J & Wiggins PA (2007) Biological consequences of tightly bent DNA: The other life of a macromolecular celebrity. *Biopolymers* 85 doi:10.1002/bip.20627 [PREPRINT]
51. Garrido Ruiz D, Sandoval-Perez A, Rangarajan AV, Gunderson EL & Jacobson MP (2022) Cysteine Oxidation in Proteins: Structure, Biophysics, and Simulation. *Biochemistry* 61 doi:10.1021/acs.biochem.2c00349 [PREPRINT]
52. George AJ, Hoffiz YC, Charles AJ, Zhu Y & Mabb AM (2018) A comprehensive atlas of E3 ubiquitin ligase mutations in neurological disorders. *Front Genet* 9 doi:10.3389/fgene.2018.00029 [PREPRINT]
53. Gibney ER & Nolan CM (2010) Epigenetics and gene expression. *Heredity (Edinb)* 105 doi:10.1038/hdy.2010.54 [PREPRINT]
54. Goldman AD & Kacar B (2021) Cofactors are Remnants of Life's Origin and Early Evolution. *J Mol Evol* 89 doi:10.1007/s00239-020-09988-4 [PREPRINT]
55. Gonzalez DL, Giannerini S & Rosa R (2019) On the origin of degeneracy in the genetic code. *Interface Focus* 9
56. Goodsell DS & Olson AJ (2000) Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* 29: 105–153 doi:10.1146/annurev.biophys.29.1.105 [PREPRINT]
57. Guerrier-Takada C, Gardiner K, Marsh T, Pace N & Altman S (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35
58. Ha SH & Ferrell JE (2016) Thresholds and ultrasensitivity from negative cooperativity. *Science* (1979) 352: 990–993
59. Hafner M, Katsantoni M, Köster T, Marks J, Mukherjee J, Staiger D, Ule J & Zavolan M (2021) CLIP and complementary methods. *Nature Reviews Methods Primers* 1 doi:10.1038/s43586-021-00018-1 [PREPRINT]
60. Halford SE & Marko JF (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res* 32 doi:10.1093/nar/gkh624 [PREPRINT]
61. Harayama T & Riezman H (2018) Understanding the diversity of membrane lipid composition. *Nat Rev Mol Cell Biol* 19 doi:10.1038/nrm.2017.138 [PREPRINT]
62. Hershberg R & Petrov DA (2009) General rules for optimal codon choice. *PLoS Genet* 5
63. Von Hippel PH & Berg OG (1989) Facilitated target location in biological systems. *Journal of Biological Chemistry* 264 doi:10.1016/s0021-9258(19)84994-3 [PREPRINT]
64. Hochachka PW (1999) The metabolic implications of intracellular circulation. *Proc Natl Acad Sci U S A* 96 doi:10.1073/pnas.96.22.12233 [PREPRINT]
65. Hofkin B (2021) Mutations are the raw material of evolution. In *Living in a Microbial World*
66. Hu Y, Gupta-Ostermann D & Bajorath J (2014) Exploring compound promiscuity patterns and multi-target activity spaces. *Comput Struct Biotechnol J* 9 doi:10.5936/csbj.201401003 [PREPRINT]
67. Jacob C, Giles GI, Giles NM & Sies H (2003) Sulfur and Selenium: The Role of Oxidation State in Protein Structure and Function. *Angewandte Chemie - International Edition* 42 doi:10.1002/anie.200300573 [PREPRINT]
68. Jaffe EK & Lawrence SH (2012) The morpheein model of allostery: Evaluating proteins as potential morpheeins. *Methods in Molecular Biology* 796

69. James KD (2017) Animal Metabolites: From Amphibians, Reptiles, Aves/Birds, and Invertebrates. In *Pharmacognosy: Fundamentals, Applications and Strategy*
70. Ji S (2000) Free energy and information contents of Conformons in proteins and DNA. *BioSystems* 54
71. Kavita K & Breaker RR (2023) Discovering riboswitches: the past and the future. *Trends Biochem Sci* 48 doi:10.1016/j.tibs.2022.08.009 [PREPRINT]
72. Keller EF (2009) Rethinking the Meaning of Biological Information. *Biol Theory* 4
73. Kim MS & Bae JW (2018) Lysogeny is prevalent and widely distributed in the murine gut microbiota. *ISME Journal* 12
74. Kinoshita T (2016) Glycosylphosphatidylinositol (GPI) anchors: Biochemistry and cell biology: Introduction to a thematic review series. *J Lipid Res* 57 doi:10.1194/jlr.E065417 [PREPRINT]
75. Kirschning A (2021) The coenzyme/protein pair and the molecular evolution of life. *Nat Prod Rep* 38 doi:10.1039/d0np00037j [PREPRINT]
76. Komar AA, Samatova E & Rodnina M V. (2024) Translation Rates and Protein Folding. *J Mol Biol* 436 doi:10.1016/j.jmb.2023.168384 [PREPRINT]
77. Koonin E V. (2016) The meaning of biological information. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374
78. Koshland DE, Goldbeter A & Stock JB (1982) Amplification and adaptation in regulatory and sensory systems. *Science* (1979) 217
79. Kouzarides T (2007) Chromatin Modifications and Their Function. *Cell* 128: 693–705 doi:10.1016/j.cell.2007.02.005 [PREPRINT]
80. Lakhota SC (2023) C-value paradox: Genesis in misconception that natural selection follows anthropocentric parameters of ‘economy’ and ‘optimum’. *BBA Advances* 4
81. Lampson BL, Pershing NLK, Prinz JA, Lacsina JR, Marzluff WF, Nicchitta C V., MacAlpine DM & Counter CM (2013) Rare codons regulate KRas oncogenesis. *Current Biology* 23
82. Lane AN & Fan TWM (2015) Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Res* 43 doi:10.1093/nar/gkv047 [PREPRINT]
83. Lee YS, Levdansky Y, Jung Y, Kim VN & Valkov E (2024) Deadenylation kinetics of mixed poly(A) tails at single-nucleotide resolution. *Nat Struct Mol Biol* 31
84. Levitzki A & Koshland DE (1976) The Role of Negative Cooperativity and Half-of-the-Sites Reactivity in Enzyme Regulation. In *Current Topics in Cellular Regulation*
85. Li W, Kinch LN, Karplus PA & Grishin N V. (2015) ChSeq: A database of chameleon sequences. *Protein Science* 24
86. Liquori AM (1969) The stereochemical code and the logic of a protein molecule. *Q Rev Biophys* 2
87. Loewe L & Hill WG (2010) The population genetics of mutations: Good, bad and indifferent. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 doi:10.1098/rstb.2009.0317 [PREPRINT]
88. Lovelock SL, Crawshaw R, Basler S, Levy C, Baker D, Hilvert D & Green AP (2022) The road to fully programmable protein catalysis. *Nature* 606 doi:10.1038/s41586-022-04456-z [PREPRINT]
89. Marijuán PC & Navarro J (2021) From molecular recognition to the “vehicles” of evolutionary complexity: An informational approach. *Int J Mol Sci* 22 doi:10.3390/ijms22211965 [PREPRINT]
90. Marín M, Fernández-Calero T & Ehrlich R (2017) Protein folding and tRNA biology. *Biophys Rev* 9 doi:10.1007/s12551-017-0322-2 [PREPRINT]
91. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen LL, Chen R, Dean C, Dinger ME, Fitzgerald KA, *et al* (2023) Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol* 24
92. McLaughlin S, Wang J, Gambhir A & Murray D (2002) PIP2 and proteins: Interactions, organization, and information flow. *Annu Rev Biophys Biomol Struct* 31 doi:10.1146/annurev.biophys.31.082901.134259 [PREPRINT]
93. Méheust R, Castelle CJ, Jaffe AL & Banfield JF (2022) Conserved and lineage-specific hypothetical proteins may have played a central role in the rise and diversification of major archaeal groups. *BMC Biol* 20
94. Miller BG & Wolfenden R (2002) Catalytic proficiency: The unusual case of OMP decarboxylase. *Annu Rev Biochem* 71: 847–885 doi:10.1146/annurev.biochem.71.110601.135446 [PREPRINT]
95. Mittag T & Parker R (2018) Multiple Modes of Protein–Protein Interactions Promote RNP Granule Assembly. *J Mol Biol* 430 doi:10.1016/j.jmb.2018.08.005 [PREPRINT]
96. Mizraji E (2021) The biological Maxwell’s demons: exploring ideas about the information processing in biological systems. *Theory in Biosciences* 140
97. Moen EL, Mariani CJ, Zullo H, Jeff-Eke M, Litwin E, Nikitas JN & Godley LA (2015) New themes in the biological functions of 5-methylcytosine and 5-hydroxymethylcytosine. *Immunol Rev* 263
98. Møller BL (2010) Dynamic metabolons. *Science* (1979) 330 doi:10.1126/science.1194971 [PREPRINT]

99. Muro E, Ekin Atilla-Gokcumen G & Eggert US (2014) Lipids in cell biology: How can we understand them better? *Mol Biol Cell* 25 doi:10.1091/mbc.E13-09-0516 [PREPRINT]
100. Murzin AG (2008) Biochemistry: Metamorphic proteins. *Science* (1979) 320 doi:10.1126/science.1158868 [PREPRINT]
101. Muthana SM, Campbell CT & Gildersleeve JC (2012) Modifications of glycans: Biological significance and therapeutic opportunities. *ACS Chem Biol* 7 doi:10.1021/cb2004466 [PREPRINT]
102. Natesan S, Rivera VM, Molinari E & Gilman M (1997) Transcriptional squelching re-examined [8]. *Nature* 390 doi:10.1038/37019 [PREPRINT]
103. Nedialkova DD & Leidel SA (2015) Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity. *Cell* 161
104. Niehaus TD, Thamm AMK, De Crécy-Lagard V & Hanson AD (2015) Proteins of unknown biochemical function: A persistent problem and a roadmap to help overcome it. *Plant Physiol* 169
105. Nurse P (2008) Life, logic and information. *Nature* 454 doi:10.1038/454424a [PREPRINT]
106. Oldfield CJ & Dunker AK (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* 83
107. Palsuledesai CC & Distefano MD (2015) Protein prenylation: Enzymes, therapeutics, and biotechnology applications. *ACS Chem Biol* 10 doi:10.1021/cb500791f [PREPRINT]
108. Paolini G V., Shapland RHB, Van Hoorn WP, Mason JS & Hopkins AL (2006) Global mapping of pharmacological space. *Nat Biotechnol* 24
109. Pedley AM, Pareek V & Benkovic SJ (2022) The Purinosome: A Case Study for a Mammalian Metabolon. *Annu Rev Biochem* 91 doi:10.1146/annurev-biochem-032620-105728 [PREPRINT]
110. Pharoah M (2020) Causation and Information: Where Is Biological Meaning to Be Found? *Biosemiotics* 13
111. Phillips T (2014) Regulation of Transcription and Gene Expression in Eukaryotes. *Nature Education* 1
112. Piazza I, Kochanowski K, Cappelletti V, Fuhrer T, Noor E, Sauer U & Picotti P (2018) A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* 172
113. Porter LL & Looger LL (2018) Extant fold-switching proteins are widespread. *Proc Natl Acad Sci U S A* 115
114. Potaman VN & Sinden RR (2005) CHAPTER 1 DNA: Alternative Conformations and Biology. *DNA Conformation and Transcription*
115. Radzicka A & Wolfenden R (1995) A proficient enzyme. *Science* (1979) 267: 90–93
116. Rajagopalan PTR & Benkovic SJ (2002) Preorganization and protein dynamics in enzyme catalysis. *Chemical Record* 2
117. Ramachandran GN, Venkatachalam CM & Krimm S (1966) Stereochemical Criteria for Polypeptide and Protein Chain Conformations: III. Helical and Hydrogen-Bonded Polypeptide Chains. *Biophys J* 6: 849–872
118. Ramakrishnan C & Ramachandran GN (1965) Stereochemical Criteria for Polypeptide and Protein Chain Conformations: II. Allowed Conformations for a Pair of Peptide Units. *Biophys J* 5: 909–933
119. Reily C, Stewart TJ, Renfrow MB & Novak J (2019) Glycosylation in health and disease. *Nat Rev Nephrol* 15 doi:10.1038/s41581-019-0129-4 [PREPRINT]
120. Rich A (1993) DNA comes in many forms. *Gene* 135
121. Richard JP (2013) Enzymatic rate enhancements: A review and perspective. *Biochemistry* 52 doi:10.1021/bi3017119 [PREPRINT]
122. Di Rienzo L, Milanetti E, Alba J & D'Abramo M (2020) Quantitative Characterization of Binding Pockets and Binding Complementarity by Means of Zernike Descriptors. *J Chem Inf Model* 60
123. Rivas G & Minton AP (2016) Macromolecular Crowding In Vitro, In Vivo, and In Between. *Trends Biochem Sci* 41 doi:10.1016/j.tibs.2016.08.013 [PREPRINT]
124. Rodnina M V. (2016) The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Science* doi:10.1002/pro.2950 [PREPRINT]
125. Rutten MGTA, Vaandrager FW, Elemans JAAW & Nolte RJM (2018) Encoding information into polymers. *Nat Rev Chem* 2 doi:10.1038/s41570-018-0051-5 [PREPRINT]
126. Schafer JW & Porter LL (2023) Evolutionary selection of proteins with two folds. *Nat Commun* 14
127. Schjoldager KT, Narimatsu Y, Joshi HJ & Clausen H (2020) Global view of human protein glycosylation pathways and functions. *Nat Rev Mol Cell Biol* 21 doi:10.1038/s41580-020-00294-x [PREPRINT]
128. Schlame M (2008) Thematic Review Series: Glycerolipids. Cardiolipin synthesis for the assembly of bacterial and mitochondrial membranes. *J Lipid Res* 49
129. Segel IH (1976) Biochemical Calculation
130. Sfakianos AP, Whitmarsh AJ & Ashe MP (2016) Ribonucleoprotein bodies are phased in. *Biochem Soc Trans* 44

131. Shumilin IA, Cymborowski M, Chertihin O, Jha KN, Herr JC, Lesley SA, Joachimiak A & Minor W (2012) Identification of unknown protein function using metabolite cocktail screening. *Structure* 20
132. SILVERMAN RE (2000) The organic chemistry of enzyme-catalysed reactions
133. Siuzdak G, Patti GJ & Yanes O (2012) Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* 13
134. Skolnick J, Gao M, Roy A, Srinivasan B & Zhou H (2015) Implications of the small number of distinct ligand binding pockets in proteins for drug discovery, evolution and biochemical function. *Bioorg Med Chem Lett* 25: 1163–1170
135. Srinivasan B (2021a) Words of advice: teaching enzyme kinetics. *FEBS Journal* 288: 2068–2083
136. Srinivasan B (2021b) Explicit Treatment of Non-Michaelis-Menten and Atypical Kinetics in Early Drug Discovery**. *ChemMedChem* 16: 899–918
137. Srinivasan B (2022) A guide to the Michaelis–Menten equation: steady state and beyond. *FEBS Journal* 289: 6086–6098 doi:10.1111/febs.16124 [PREPRINT]
138. Srinivasan B (2023) Words of advice: teaching macromolecular crystallography. *FEBS Journal* 290
139. Srinivasan B, Kempaiah Nagappa L, Shukla A & Balaram H (2015) Prediction of substrate specificity and preliminary kinetic characterization of the hypothetical protein PVX_123945 from *Plasmodium vivax*. *Exp Parasitol* 151–152: 56–63
140. Srinivasan B, Kus K & Athanasiadis A (2022) Thermodynamic analysis of Z α domain-nucleic acid interactions. *Biochemical Journal* 479
141. Srinivasan B, Marks H, Mitra S, Smalley DM & Skolnick J (2016a) Catalytic and substrate promiscuity: distinct multiple chemistries catalysed by the phosphatase domain of receptor protein tyrosine phosphatase. *Biochemical Journal* 473: 2165–2177
142. Srinivasan B, Zhou H, Mitra S & Skolnick J (2016b) Novel small molecule binders of human N-glycanase 1, a key player in the endoplasmic reticulum associated degradation pathway. *Bioorg Med Chem* 24: 4750–4758
143. Stein KC & Frydman J (2019) The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis. *Journal of Biological Chemistry* 294 doi:10.1074/jbc.REV118.002814 [PREPRINT]
144. Stewart MD, Ritterhoff T, Kleivit RE & Brzovic PS (2016) E2 enzymes: More than just middle men. *Cell Res* 26 doi:10.1038/cr.2016.35 [PREPRINT]
145. Su Z & Denu JM (2016) Reading the Combinatorial Histone Language. *ACS Chem Biol* 11 doi:10.1021/acscchembio.5b00864 [PREPRINT]
146. Sundararajan VS, Malik G, Ijaq J, Kumar A, Das PS, P.R. S, Nair AS, Dhar PK & Suravajhala P (2018) HYPO: A Database of Human Hypothetical Proteins. *Protein Pept Lett* 25
147. Tesei G, Schulze TK, Crehuet R & Lindorff-Larsen K (2021) Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc Natl Acad Sci U S A* 118
148. Tompa P (2012) Intrinsically disordered proteins: A 10-year recap. *Trends Biochem Sci* 37 doi:10.1016/j.tibs.2012.08.004 [PREPRINT]
149. Tonddast-Navaei S, Srinivasan B & Skolnick J (2017) On the importance of composite protein multiple ligand interactions in protein pockets. *J Comput Chem* 38: 1252–1259
150. Torres JP & Schmidt EW (2019) The biosynthetic diversity of the animal world. *Journal of Biological Chemistry* 294 doi:10.1074/jbc.REV119.006130 [PREPRINT]
151. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, *et al* (2017) Essentials of glycobiology, third edition
152. Vértessy BG & Tóth J (2009) Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases. *Acc Chem Res* 42
153. Wächtershäuser G (2014) The place of RNA in the origin and early evolution of the genetic machinery. *Life* 4
154. Walker SI, Kim H & Davies PCW (2016) The informational architecture of the cell. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374
155. Walsh CT, Tu BP & Tang Y (2018) Eight Kinetically Stable but Thermodynamically Activated Molecules that Power Cell Metabolism. *Chem Rev* 118 doi:10.1021/acs.chemrev.7b00510 [PREPRINT]
156. Wang G & Vasquez KM (2023) Dynamic alternative DNA structures in biology and disease. *Nat Rev Genet* 24 doi:10.1038/s41576-022-00539-9 [PREPRINT]
157. Wang H, Vant J, Wu Y, Sánchez R, Micou ML, Zhang A, Luczak V, Yu SB, Jabbo M, Yoon S, *et al* (2023) Functional Organization of Glycolytic Metabolon on Mitochondria. *bioRxiv*
158. Watson JD & Crick FHC (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* 171: 737–738
159. Weidmann CA, Mustoe AM, Jariwala PB, Calabrese JM & Weeks KM (2021) Analysis of RNA–protein networks with RNP-MaP defines functional hubs on RNA. *Nat Biotechnol* 39

160. Williams SJ & Davies GJ (2001) Protein-carbohydrate interactions: Learning lessons from nature. *Trends Biotechnol* 19 doi:10.1016/S0167-7799(01)01699-7 [PREPRINT]
161. Wolfenden R & Snider MJ (2001) The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res* 34: 938–945
162. Wright PE & Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* 16 doi:10.1038/nrm3920 [PREPRINT]
163. Wu JI, Lessard J & Crabtree GR (2009) Understanding the Words of Chromatin Regulation. *Cell* 136 doi:10.1016/j.cell.2009.01.009 [PREPRINT]
164. Yang E, van Nimwegen E, Zavolan M, Rajewsky N, Schroeder M, Magnasco M & Darnell JE (2003) Decay rates of human mRNAs: Correlation with functional characteristics and sequence attributes. *Genome Res* 13
165. Yau R & Rape M (2016) The increasing complexity of the ubiquitin code. *Nat Cell Biol* 18 doi:10.1038/ncb3358 [PREPRINT]
166. Yu H & Chen X (2007) Carbohydrate post-glycosylational modifications. *Org Biomol Chem* 5
167. Yu Y & Delbianco M (2020) Conformational Studies of Oligosaccharides. *Chemistry - A European Journal* 26 doi:10.1002/chem.202001370 [PREPRINT]
168. Zhang C (2009) Novel functions for small RNA molecules. *Curr Opin Mol Ther* 11 [PREPRINT]
169. Zhang Q, Bhattacharya S & Andersen ME (2013) Ultrasensitive response motifs: Basic amplifiers in molecular signalling networks. *Open Biol* 3 doi:10.1098/rsob.130031 [PREPRINT]
170. Zhang Y, Sampathkumar A, Kerber SML, Swart C, Hille C, Seerangan K, Graf A, Sweetlove L & Fernie AR (2020) A moonlighting role for enzymes of glycolysis in the co-localization of mitochondria and chloroplasts. *Nat Commun* 11
171. Zhao LY, Song J, Liu Y, Song CX & Yi C (2020) Mapping the epigenetic modifications of DNA and RNA. *Protein Cell* 11 doi:10.1007/s13238-020-00733-7 [PREPRINT]
172. Zhong Y (2017) A theory of semantic information. *China Communications* 14
173. Zwanzig R, Szabo A & Bagchi B (1992) Levinthal's paradox. *Proc Natl Acad Sci U S A* 89

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.