Review/Feature Article

Reciprocally-coupled Gating: Strange Loops in Bioenergetics, Genetics, and Catalysis

Charles W. Carter[1], Jr & Peter R. Wills[2]

[1]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7260

ORCID ID (Carter): /0000-0002-2653-4452

0000-0001-6765-3813

[2]Department of Physics and Te Ao Marama Centre for Fundamental Inquiry, University of Auckland, PB 92019, Auckland 1142, New Zealand

ORCID ID (Wills): /0000-0002-2670-7624

Abstract

Bioenergetics, genetic coding, and catalysis are all difficult to imagine emerging without pre-existing historical context. That context is often posed as a "Chicken and Egg" problem; its resolution is concisely described by de Grasse Tyson: "the egg was laid by a bird that was not a chicken". The concision and generality of that answer furnish no details—only an appropriate framework from which to examine detailed paradigms that might illuminate paradoxes underlying these three life-defining biomolecular processes. We examine experimental aspects here of five examples that all conform to the same paradigm. The paradox in each example is resolved by coupling if, and only if, conditions for two related transitions between levels. One drives, and each restricts fluxes through, or "gates" the other. That reciprocally-coupled gating, in which two gated processes constrain one another, maps onto the formal structure of "strange loops". That mapping may help unite the axiomatic foundations of genetics, bioenergetics, and catalysis. As a physical analog for Gödel's logic, biomolecular strange-loops provide a natural metaphor around which to organize these data, linking biology to the physics of information, free energy, and the second law of thermodynamics.

Keywords

Genetic coding; free energy transduction; non-equilibrium thermodynamics; transition-state stabilization; conformational change; aminoacyl-tRNA synthetases; emergent phenomena

Abbreviations

XNOR, Exclusive NOR, output is true if all inputs are true; AaRS, aminoacyl-tRNA synthetase(s); TrpRS, (*Bacillus stearothermophilus*) tryptophanyl-tRNA synthetase; CP1, connecting peptide 1; ABD, anticodon-binding domain; PreTS, Pre-transition-state)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

1. Introduction

This review is motivated by a recent effort to define the physical forces leading to the origin of life [1]. It is a bold conjecture indeed that physico-chemical forces drove production and selection of the biomolecules that enabled nature to (i) invent protein catalysts and heredity, (ii) store symbolic representations of nature in two distinct biopolymers, (iii) separate genotype from phenotype, and (iv) implement efficient mechanisms for capturing, storing, and transforming the chemical free energy necessary to sustain far from equilibrium states. We attempt here to build on that foundation by describing formally-coupled structures that arise from experimental studies of each of these processes and distilling shared characteristics that appear to define two novel kinds of forces, one constraining dissipative losses, the other driving creativity. These new forces incorporate and enhance the creativity potentiated by the distinction between hydrophobic and polar properties of matter described by Dill and Agozzino.

The examples are these:

(i) Coupling between ATP utilization and protein conformational change as Tryptophanyl-tRNA synthetase (TrpRS) activates tryptophan [2-6] creates an escapement or ratchet: Mg^{2+} ion accelerates catalysis if and only if the conformation changes, but the conformation change is thermodynamically favorable if and only if the resulting PPi product is released to solvent.

(ii) Studies of aminoacyl-tRNA synthetase (aaRS) evolution [7] led us to recognize that both protein folding and the genetic coding table depend intimately on amino acid side chain behavior [8-10]. Two sets of rules: folding—activity arises if and only if amino acid sequences fold—and coding—amino acid sequences arise if and only if the coding rules are obeyed—form a self-referential feedback loop accelerating the evolutionary search for polypeptides whose substrate recognition properties allow them to enforce the coding rules according to which they, themselves, were assembled [11, 12].

(iii) AARS Urzymes—130-residue excerpts that accelerate both amino acid activation and tRNA acylation—appear to be catalytically active molten globules [13]. Catalysis occurs if and only if the catalytic structure of the molten globule assembles, and that conformation assembles if and only if the substrate is presented in its rare transition state. Two rare species—folded Urzyme and reaction transition state—must thus occur simultaneously to achieve catalysis.

(iv) Expansion of the genetic code from a binary alphabet to the current 20 letters likely obeyed a variational principle minimizing dissipation of information and free energy, and aaRS evolution converted the former type of dissipation into the latter to enhance fidelity [14]. Errors could be reduced if and only if they became more energetically costly, but making errors energetically costly could be achieved if and only if errors were less frequent.

(v) The coding alphabet itself could expand if and only if there were enhancements in aaRS specificity; yet aaRS specificity could increase if and only if the alphabet size expanded [14].

Each example has been cast in identical format (Fig. 1C) in which two distinct gated transitions are coupled by the fact that one filters, or gates the other. Our purpose here is to review these examples, which to our knowledge have never been considered as of a piece, and to examine why the associated logical operators function so powerfully.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

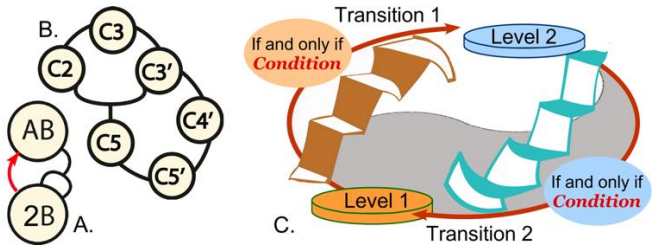

Figure 1. Biological process control blocks. **A.** Simple autocatalysis. Reaction of molecules A and B gives rise to two B molecules, which therefore increase the rate at which B is produced. **B.** A reflexive autocatalytic set, often called a Reflexive Autocatalytic Food Set (RAFS). The network is closed because the product of catalyst 5 is the substrate for catalyst 2. **A** and **B** were adapted from [15]. **C.** Elements of reciprocally coupled gating, presented as a strange loop [16]. White stairs on left and right sides are filtered by an "if and only if" condition as both ascend from level to level, and each is. The antecedent of each staircase is the consequent of the other.

2. Biological process control: feedback, autocatalysis, hypercycles, and strange loops

Efforts to understand the origin of biology invariably invoke feedback. Feedback—autocatalysis—can be seen as a reflexive force by virtue of the fact that, even in its simplest form, as a product increases, its concentration induces change in its rate of formation, Fig. 1A. Feedback also introduces a minimal historical context because the change in rates of synthesis connect present to past events. Kauffman [17] and Manfred Eigen [18, 19] argued at about the same time (1971) that autocatalysis alone could not explain the origin of life, which, both argued, also required some form of integration. Kauffman introduced linked "autocatalytic sets" [20]; Eigen introduced "hypercycles", a closely related concept which, though conceptually similar, he advocated for different reasons having to do with limitations on the amount information that could be replicated in an error-ridden replicative regime [21, 22].

An important recent contribution derives a formal structure of autocatalysis from the matrix of stoichiometries, with reactions as columns and species as rows [15]. That matrix representation provided a unified framework for distinguishing their properties into only five distinct networks, which therefore compose a basis set for identifying autocatalytic subnetworks in arbitrarily large autocatalytic networks. That afforded a means to systematize the bewildering array of networks for modeling the emergence of life-like properties built by computational theorists as "***Reflexive*** Autocatalytic Food Sets" (RAFS [23-29]) or GARDs [30], and show that they have the same formal structure [15]. Reflexive in this context simply refers to a closure property in which the product of the final catalyst provides the substrate for the first. The relevance of these control networks is almost certainly validated by their homologies to various chemical cycles, such as the formose cycle [31], the Krebs cycle [32], and related core metabolisms [33]. The more recent studies strengthen arguments favoring metabolism first, autotrophic pre-biology forms.

There are several reasons to believe that autocatalysis alone is insufficient and that other process control mechanisms may be necessary to account for the emergence of biology. The first of these is a problem known variously as the error catastrophe [34], the "paradox of specificity" [35], or Eigen's Cliff [36]: high specificity conflicts with survival in error-prone systems, because side reactions (parasites) develop, leading to extinction. Solutions to this paradox invoked to rescue autocatalysis include compartmentation [15] and reaction-diffusion coupling [37], but these are unconvincing. Identifying a more fatal defect in autocatalytic sets, Wills argued that RAFS cannot, by themselves, account for the embedding of symbolic meaning into biomolecules [38-42] and hence that they cannot suffice to account for the emergence of biology.

It is possible that qualitatively different forms of process control can navigate the error catastrophe more definitively. Examples (i-v) described in the Introduction share three formal properties that suggest a new class of process control structures with different and substantially more robust properties, Fig. 1C. (i) They

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

connect distinct realms (levels). (ii) Gating is represented by "if and only if" tests that filter what passes from realm to realm. These complementary bi-conditional logical operators are logically equivalent to XNOR (Exclusive NOR; true if all inputs are true) gates in computer architecture. (iii) Coupling is effected by interchanging the respective antecedent and consequent of the two logical connectives. Interchanging creates a paradoxical, but quite robust, coupling which the five examples suggest underlies many of the most interesting biological phenomena. We called this process control structure "reciprocally-coupled gating".

Hofstadter [16] introduced the idea of strange loops to link such paradoxes, familiar from M. C. Escher's intertwined staircases, to the Gödel incompleteness theorem. Extending self-reference from mathematics to material systems, he argued that when two conditional—if and only if—processes control one another, the resulting reciprocally-coupled gating can generate ever-expanding novelty. Examples from experimental biochemistry and biophysics suggest that this logic of incompleteness may underlie the generativity of non-equilibrium thermodynamic steady-states.

3. Biogenetics: free energy transduction requires reciprocally-coupled gating.

Living things sustain themselves far from equilibrium by capturing the free energy of NTP hydrolysis and converting that free energy source efficiently into mechanical work and/or information. Detailed molecular mechanisms of that process, however, remained puzzling and incomplete for many decades [43-45], despite substantial theoretical advances [46-50]. It is generally accepted that molecular mechanisms that utilize that free energy are closely related to those necessary to convert the free energy of ion gradients into ATP synthesis, by which NTPs are then regenerated. We previously reported that catalysis of tryptophan activation by tryptophanyl-tRNA synthetase, TrpRS, requires relative domain motion to re-position the catalytic Mg^{2+} ion for ATP utilization, noting the analogy between that conditional hydrolysis of ATP and the escapement mechanism of a mechanical clock. The escapement allows the time-keeping mechanism to advance discretely, one gear at a time, if and only if the pendulum swings, thereby converting energy from the weight or spring driving the pendulum into rotation of the hands.

Coupling of catalysis to domain motion, however, mimics only half of the escapement mechanism, suggesting that domain motion should be reciprocally coupled to catalysis by a complementary if and only if condition, completing the escapement metaphor. Computational studies of the ligand-dependence of the free energy surface restraining domain motion later confirmed that reciprocal coupling: the catalytic domain motion is thermodynamically unfavorable unless the PPi product is released from the active site. These two conditional phenomena—demonstrated together only for the TrpRS mechanism—are ultimately driven by ATP hydrolysis, and function as reciprocally-coupled gates. The experimental data that underlie this novel allosteric mechanism arose from attempts to understand the controversial question of how domain motion can contribute to catalysis, given that such motions are so much slower than transition-state formation and breakdown [51-59].

The answer to that question—that transition-state complementarity develops only transiently while the active site is being reconfigured by domain motion—emerged in pieces (Fig. 2B). (i) The first piece was to identify the core amino acid side chains that mediated the shear forces preventing domain motion [60], which identified the D1-switch residues, whose configuration changed most dramatically during catalysis. (ii) Four of the seven side chains involved in that switching motif were then subjected to carefully designed combinatorial mutagenesis together with substitution of Mg^{2+} with Mn^{2+} as the catalytic divalent cation (Fig. 2A [5, 61, 62]). (iii) Complementary computational studies of the conformational transition by minimum action path analysis revealed that the mutated D1 switch residues, three of which are aromatic, change positions in the conformational transition state encountered during induced-fit, and that similar configurational changes of aromatic residues also defined the conformational transition states in unrelated

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

dynamic proteins myosin and calmodulin [4]. (iv) A decisive new piece to the puzzle emerged when we found that single turnover kinetic measurements of the pre-steady state rate, k_{chem}, for the combinatorial mutants exhibited the same pattern observed for k_{cat} with steady-state kinetics, and that k_{chem} (625/s) was itself actually comparable to timescale (ms) expected for domain motion [2]. (v) Finally, free energy surfaces computed by replica exchange discreet molecular dynamics validated the structure of the catalytic conformational transition state along the minimum action pathway [4] and showed that the conformational equilibrium shifted to favor the products conformation only after the PPi product was released (Fig. 2C [63]).

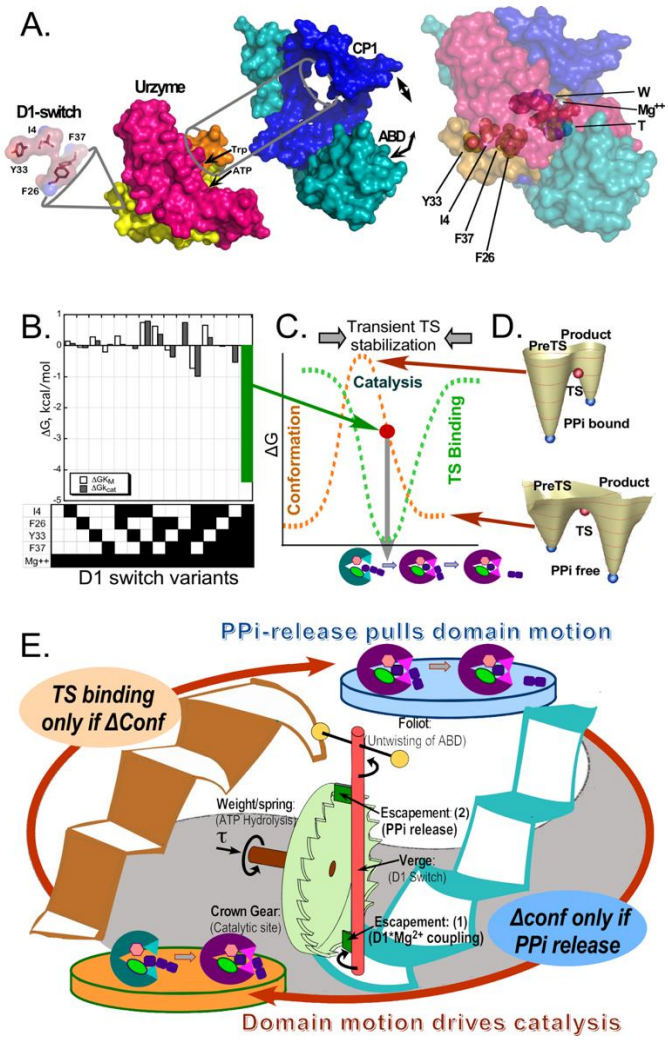

Figure 2. Reciprocally-coupled gating in the TrpRS escapement mechanism. **A.** Structural biology of the TrpRS monomer. The catalytic machinery is located within the Urzyme and is modified by its interaction with two accessory domains, including the CP1 (Connecting Peptide 1 [64, 65]; blue) and the anticodon-binding domain (ABD; teal), which move relative to the active-site. The D1 switch is located within the Protozyme (yellow), which is the first crossover connection of the Rossmann fold within the Urzyme. The specificity helix (sand) connects the first and second halves of the Urzyme (yellow, red). The D1 switch is ~20 Å from the active-site metal ion. **B.** Combinatorial mutagenesis of the TrpRS D1 switch revealed that the entire catalytic contribution of the catalytic Mg^{2+} ion (~5 Kcal/mole) can be attributed to the 5-way interaction between it and the four D1 residues. **C.** Multiple studies [60, 66-68] showed that the PreTS state TrpRS conformation is an excited state ~3 kcal/mole higher in energy than either the open or Products ground states (brown dashed curve). Data from **A** show that the conformation complementary to the transition state for amino acid activation develops transiently, during domain motion from the PreTS to the Produccts conformation

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

(green dashed curve). **CD.** Computational free energy surfaces of the transition between PreTS and Products states [4] with and without the product PPi show that the conformational equilibrium shifts to favor the Products conformation if and only if the PPi is absent from the active site. **E.** Schematic representation of the reciprocally coupled gating as a strange loop, emphasizing the parallel between the data in in **B-D** and the two green blades (Escapement (1) and (2)) that make the rotation of the crown gear in the escapement mechanism of a mechanical clock conditional on the rotation cycle of the pendulum or Foliot.

As we and others [69, 70] have noted, such an escapement mechanism is essential for efficient transduction of NTP hydrolysis free energy into other useful forms of mechanical or chemical work and/or information. Some implementation of both gating mechanisms—catalysis by domain motion and domain motion by catalysis—will thus likely be found in many other systems. In the present context, this observation provides an important clue regarding the general relevance of reciprocally-coupled gating to biology. By definition, efficiency means that a high proportion of the ATP hydrolysis free energy during amino acid activation is captured and employed to ensure aminoacylation of tRNA. Gating therefore greatly reduces the proportion of ATP that, from a system perspective, is wasted on non-productive side reactions.

4. Amino acid physical chemistry drove the origin of genetics.

Dill and Agozzino [1] rightly attribute creative force to the physical properties of amino acid side chains and their behavior in water. Forces are gradients of energy with respect to distance, which in this context refers to changes in distribution constants between different environments. From the work of Wolfenden [71-74], we can position each amino acid side chain precisely in a two-dimensional coordinate system whose axes are free energies of transfer from cyclohexane to water (polarity) and to the vapor phase (size). Moreover, those two free energies are necessary and sufficient to estimate quantitatively the mean exposure to solvent in folded proteins, and to characterize the specific identity elements recognized by aaRS in cognate tRNAs [8-10].

We have argued [75] that genetic coding arose from an underlying duality in aaRS•tRNA cognate pairs first identified by Eriani [76]. That duality rests on impressive and comprehensive experimental data. (i) Primary [76] and tertiary [77-80] structural differences between Class I and II aaRS have been widely recognized as fundamental. (ii) The Class partitioning of amino acid substrates by contemporary aaRS appears to be according to their side-chain volume [8-10]. (iii) Class-dependent discrimination between cognate and non-cognate tRNA [81, 82] and amino acid [83] substrates has been attributed to secondary structural differences between the two aaRS Classes, and does not appear to depend on specific side chains. One can thus readily imagine quite deeply-based ancestry of the rudimentary distinctions necessary for the initial differentiation between coding letters.

It is nearly certain that the aaRS Class duality arose in the form of a bidirectional ancestral gene encoding an ancestral Class I aaRS on one strand and an ancestral Class II aaRS on the opposite strand [84, 85]. (i) Experimental deconstruction of aaRS from both Classes confirmed that all essential catalytic activities required for genetic coding are retained in excerpts containing only the portions capable of antiparallel alignment of the corresponding coding sequences, and which have been called Urzymes on that basis [86-90]. (ii) Protozymes, representing ~40% of the Urzyme sequences, have been encoded in a single bidirectional gene whose two products both accelerate amino acid activation 10^6-fold [91]. (iii) Phylogenetic metrics derived from excerpts comparable to those examined experimentally differ by amounts that are statistically significant and which implicate earlier genetic origins for the Urzyme and protozyme excerpts [92]. Moreover, those metrics track linearly with catalytic proficiency. (iv) Antiparallel alignments of Class I and II middle codon-bases within the region of putative bidirectional coding exhibit statistically significant base pairing in excess of that observed within Classes [93].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

Making the enforcement of genetic coding rules (recognition of both cognate amino acids and cognate tRNAs) conditional on protein folding coordinated nature's exploration of both protein folding and genetic coding rules (Fig. 3A). As with the bioenergetic escapement mechanism, the antecedent and consequent levels for the two sets of rules are interchanged, and the rules themselves can thus be viewed as reciprocally-coupled gates (Fig. 3B). We encounter here a novel depth of reflexivity: the aaRS sequences that fold can, collectively, enforce the coding rules according to which they were assembled.

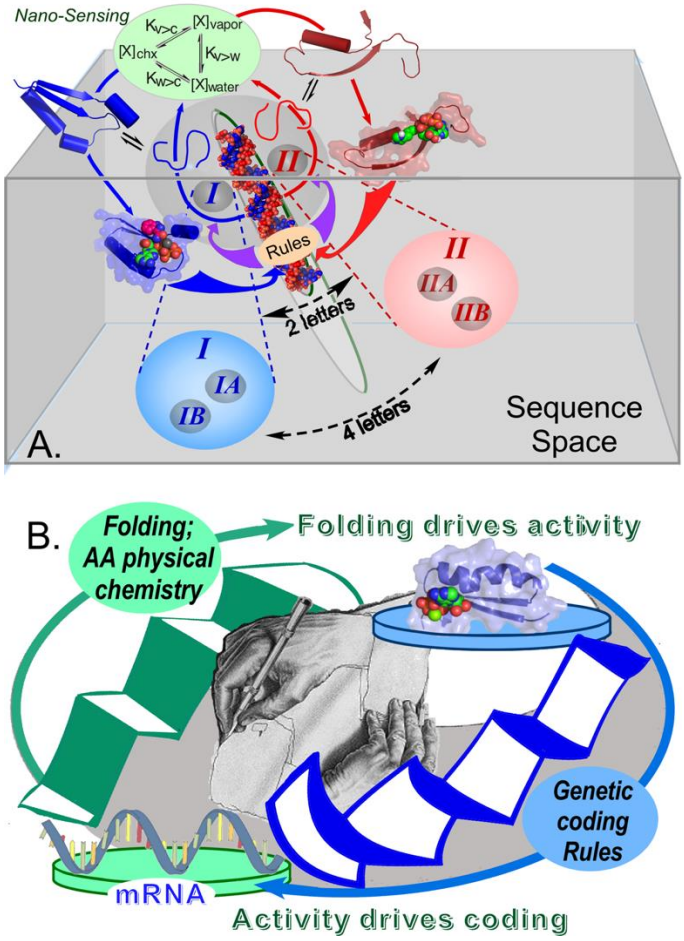

Figure 3. Reciprocally-coupled gating in the origin of genetic coding. **A.** A single bidirectional gene encoding ancestral genes for Class I (blue) and Class II (red) aaRS on opposite strands combines with nanosensing arising from the equilibrium distributions of amino acid side chains between vapor, water, and cyclohexane [8-10, 71-74] provide a boot-block for installing biology's operating system by furnishing the minimal instruction set necessary to launch coded peptide-bond formation [11, 12] based on two distinct amino acid types. **B.** The underlying process control structure of that boot-block. Sequences are active if and only if they fold; and active sequences fold if and only if they obey the relevant coded sequence, which the aaRS collectively themselves enforce.

Many details remain to be tested experimentally about this idea. (i) Can a bidirectional gene based on a two-letter alphabet produce active amino acid activation catalysts homologous to Class I and II protozymes [91]? (ii) Can Class I and II protozymes discriminate between two disjoint sets of amino acids? (iii) Can protozymes recognize and acylate cognate tRNAs with sufficient specificity? (iv) If not, can protozymes participate with suitable ribozymal catalysts to acylate cognate tRNAs [94]? How do protozyme specificities improve as the size of the coding alphabet increases? Notwithstanding, our previous work furnishes the experimental tools necessary to answer these questions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

5. Enzyme catalysis likely evolved via reciprocally-coupled gating.

Studies by Hilvert [95, 96] reveal that enzymatic activity does not require properly folded proteins, because a variant of chorismate mutase with identical catalytic activity to that of the native enzyme, is actually a molten globule. Experimental [97] and computational [98] studies of that system confirmed that the molten globular variant had a more negative entropy of activation than the wild type enzyme, and therefore also required a more negative activation enthalpy to achieve the same catalytic proficiency. Remarkably, the intrinsically disordered molten globular structure of the monomeric chorismate mutase variant assumes a highly ordered tertiary structure in the presence of a transition-state analog inhibitor [96]. Recent work in our lab [13] showed that the TrpRS Urzyme may also be a catalytically active molten globule.

A key inference of these studies is that protein "[protein] folding can be coupled to catalysis with minimal energetic penalty" and that "many modern enzymes might have evolved from molten globule precursors." [96]. Unpacking these remarkable statements, we discover a third form of reciprocally coupled gating. A chemical reaction catalyzed by a molten globule implies the simultaneous formation of two extremely rare species. The chemical transition state is, by consensus, an extremely rare species, and the low dispersion of the NMR HSQC spectra of a molten globule means that only a tiny fraction of molecules in that population have properly formed active-site configurations for catalytic activity. A simple re-formulation furnishes this description of what happens during catalysis by a molten globule: the concentration of the chemical transition state increases if and only if the active-site is properly folded, whereas that of the properly-folded active site increases if and only if the substrate is in its, rare, transition-state configuration (Fig. 4A). In this case, the experimental data—minimally dispersed HSQC spectra along the proton dimension, except in the presence of a transition state analog—provide a vivid demonstration of the paradox: the apparent transition-state dissociation constant, given by the rate enhancement, is orders of magnitude greater than the ratio of the product to the reactant concentrations in isolated solution. The apparent resolution to this paradox is that: (i) non-productive complexes between substrate and molten globule must exist at a much higher concentration than the E•TS complex, and (ii) there must be important correlations between folding and catalysis by which the presence of the ground-state substrate can induce the ready formation of the catalytic configuration. That rationalization provides another clue to the relevance of the reciprocally-coupled gating (see §7).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

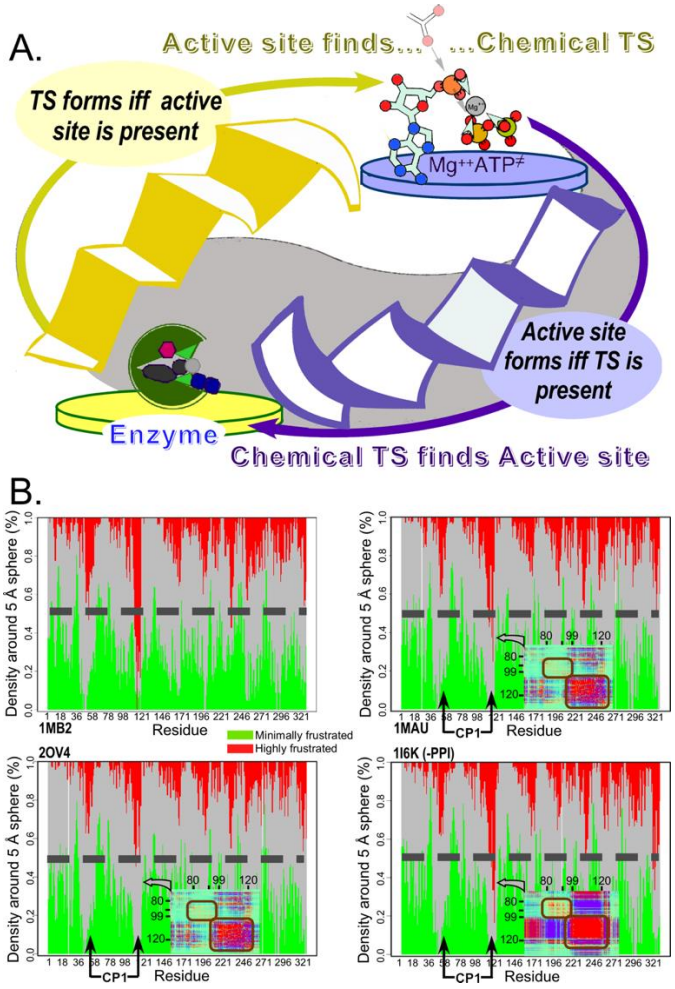

Figure 4. Reciprocally-coupled gating drove the evolution of catalysis. **A.** The paradox of catalysis by molten globular ensembles is that both catalyst and reaction transition state are very rare species whose concentrations are too small to find one another often enough to accelerate the uncatalyzed rate. Yet the catalyzed rates of native and monomeric chorismate mutase are essentially the same, implying that the transition-state complex is as stable for the molten globular catalyst as for the native enzyme. The strange loop representing catalysis exhibits the same formal structure as those in Figs. 1-3, in which both catalysis and folding are governed by bi-conditional if and only if statements for which the antecedent and consequent states are interchanged. **B.** Structure-based bioinformatics evidence suggests that even native TrpRS retains features of such coupling. TrpRS crystal structures of ground-state (1MB2), activated Pre-transition state (1MAU), transition state analog (2OV4), and Products (1I6K) state complexes exhibit a progression in which the regions where the mobile CP1 domain is inserted into the Urzyme maximally relaxes its severe frustration [99, 100] in the transition-state complex. The inserts in the last three frustratograms are vignettes of the covariance heat maps obtained from replica exchange Discrete Molecular Dynamics simulations, and show that correlate motion at the C-terminal boundary between the Urzyme and CP1 intensifies only in the Products state and only after release of product, PPi.

Recent analysis suggests that even modern enzymes undergo something comparable to the dramatic coupling between folding and transition-state binding. As noted in §2, the CP1 insertion in the TrpRS Urzyme moves relative to the anticodon-binding domain. Because CP1 behaves extensively as a rigid body, it turns out that the locus of maximal frustration [100] along the polypeptide chain occurs in two places, where the CP1 inserted into the Urzyme (Fig 4B). Remarkably, the local frustration at those two segments is progressively relaxed as the ground-state monomer in the open conformation (1MB2) binds ligands to form the PreTS state (1MAU), and then the transition state analog complex (2OV4). It re-emerges in the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

Products complex (1I6K). From the definition of frustration [99], this means that the ancestral junction between two different modules of the synthetase becomes most like a molten globule in the transition state complex. From the thermodynamic analysis of Hilvert [97], this means in turn that at that moment the active site is most capable of wrapping tightly around and forming bonds to the transition state configuration of the amino acid, adenosine monophosphate, Mg^{3+}, and PP_i leaving group.

6. Constraints on the emergence of new aaRS•tRNA cognate pairs

Expanding the genetic coding alphabet faced multiple, interrelated challenges (Fig. 5). The first, recognized by Pauling on thermodynamic grounds [101], is that a single binding interaction cannot discriminate sufficiently between amino acids with similar side chains to assure high precision translation. In the most difficult cases, most notably valine, threonine, leucine, and isoleucine, the cognate aaRS must couple a second round of discrimination combined with hydrolytic editing to achieve the necessary quality control. Pauling's observation places a irreducible limitation on the fidelity of codon-dependent protein synthesis accessible via equilibrium binding, ensuring that, absent specialized editing mechanisms, coded proteins would always remain quasispecies-like populations [102]. An experimentally-based estimate of this limitation at the level of aaRS Urymes is given in Fig 4A of [14].

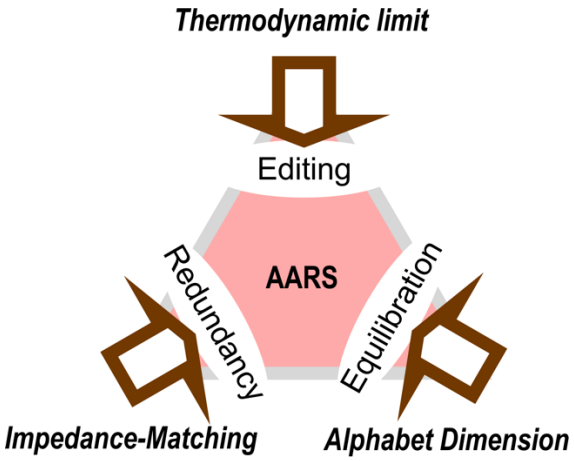

Figure 5. Simultaneous constraints on the precision of aminoacylation by aaRS. The alphabet size and fundamental thermodynamics both place different irreducible limits on aaRS precision. Any change in the dimension of the genetic coding alphabet must be followed by a round of equilibration, as new possibilities are introduced to optimize the catalytic and specificity of the extant aaRSs. Finally, minimizing the dissipation of information similarly constrains optimal aaRS precision according to the frequency of replication and transcription errors in large part because of changes in coding redundancy [14].

Hopfield [103] showed experimentally that high precision in such cases was achieved only by making mistakenly acylated tRNAs energetically costly. That transition converted the dissipation of information by translation errors by the less sophisticated ancestral aaRS into dissipation of free energy by the full-length aaRS (Fig. 6A). The hydrolytic editing entailed a futile cycling of ATP to generate mis-acylated [104] tRNAs or less frequently mis-activated 5'-adenylates [105, 106] that were then hydrolyzed. Mechanisms whereby aaRS gained hydrolytic editing capability were only the latest step in a progressive adaptation of aaRS Urzymes to insertion domains that enabled increased precision even for aaRS specific for amino acids that did not require hydrolytic editing.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

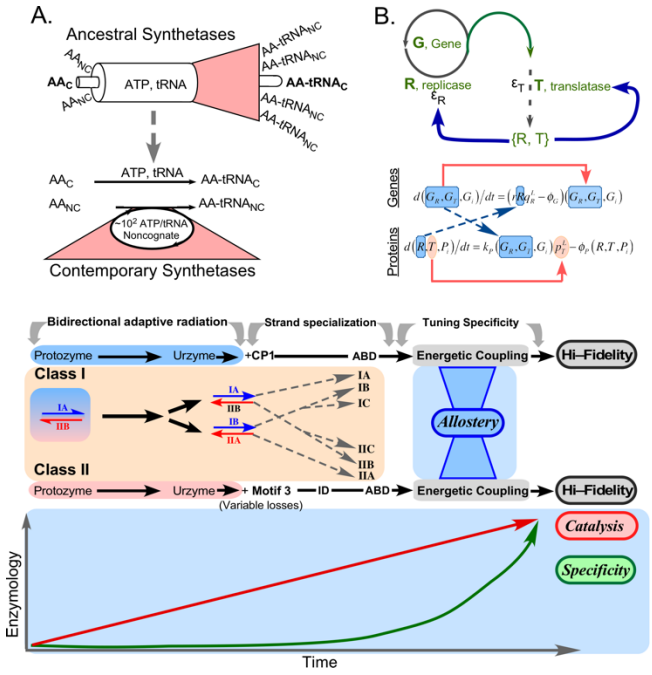

Figure 6. Challenges facing the evolution of aaRS associated with expanding the coding alphabet introducing new aaRS•tRNA cognate pairs. **A.** High precision cannot be achieved using a single association dissociation equilibrium because some amino acids have similar side chains and bind to non-cognate aaRS with dissociation constants nearly 1% of the cognate amino acids. Nature had to couple multiple binding events without allowing dissociation to enable contemporary aaRS to surpass this thermodynamic limitation. As a consequence, dissipation of information in mRNA codescripts characteristic of translational errors by ancestral aaRS (horizontal red fan) was converted into dissipation of chemical free energy in contemporary aaRS (vertical red fan). **B.** Coupling of replication to translation illustrated by the Gene-Replicase-Translatase model system. Rudiments of the GRT system are the replicase and translatase catalysts and their genes (green). Processes necessary to generate the active catalysts, R and T, are replication (circle), translation (dashed line), and folding (blue lines). No distinction is made between duplication and transcription of the respective genes, G, in a world where genetic information is instantiated in RNA. Errors are denoted by ε. Differential equations for the two processes [12] are coupled in both directions via the population variables (dashed arrows) and exhibit autocatalysis (red arrows). **C.** Evolutionary events in aaRS evolution. (Top) The schematic illustrates three distinct stages in the structural and genetic development of aaRS. (Bottom) Experimental data have repeatedly shown that the catalytic proficiency of both classes is a linear function of the number of amino acids, and by implication with time. Specificity, however, likely failed to develop until the allosteric mechanisms could accomplish fine tuning.

The dimension of the genetic codon table imposes similar irreducible limits on the precision of tRNA aminoacylation by aaRS. For all coding alphabets of dimension smaller than 20, isoaccepting tRNAs may be acylated by multiple different amino acids, leading to a similar quasispecies-like distribution of coded proteins. Under these circumstances, aaRS within a single quasispecies will have slightly different specificities, and will randomly incorporate different percentages of related amino acid types.

A third challenge to the emergence of new aaRS•tRNA cognate pairs was that the error frequencies of the ancestral aaRS were constrained by an informational impedance matching to the errors of replication/transcription, arising from the coupling between the two information transfer processes (Fig 6B; [14][14]). We have argued that the choice of pathways for the expansion of the coding table was not arbitrary, but formally resembled the shifting gears of a bicycle derailleur [11], because each reduced coding alphabet restricted errors to some irreducible frequency, owing to limitations imposed by the quasi-species like distribution [102] resulting from the inability to select unique amino acids with high precision.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

Each point of the triangle in Fig. 5 can potentially be joined to the center by an antecedent => consequent relationship filtered by bi-conditionals. Errors made by aaRS can be reduced if and only if they become more thermodynamically costly, but errors can be made thermodynamically more costly if and only if errors are reduced. The number of distinct aaRS•tRNA cognate pairs can be increased if and only if the coding redundancy is reduced, but the coding redundancy can be reduced if and only if the number of distinct aaRS•tRNA cognate pairs increases. The dimension of the coding alphabet can be increased if and only if aaRS precision increases, but aaRS precision can be increased if and only if the coding alphabet dimension increases. We examine the latter example of reciprocally-coupled gating in detail in §7.

7. Reciprocally-coupled gating shaped the growth of the genetic code.

Experimental decomposition of several Class I and II aaRS [86-89, 107, 108] defined three stages in aaRS evolution, which are summarized in Fig. 6C. The initial production both Class I (blue) and II (red) aaRS from bidirectional genes likely terminated before the coding alphabet had grown much beyond four distinguishable, but possibly overlapping amino acid types. The early strand-specialized era may have been initialized by the insertion of CP1 into Class I precursors, which inactivated bidirectional coding, and finished with the addition of anticodon-binding domains (ABD). Plausible partial speciation is indicated by the evolutionary trees in the center, and may have resulted in six different aaRS. Addition of both insertion domains and ABDs enabled the final stage in which allosteric energetic coupling developed between the insertion and ABDs, leading eventually to development of high precision aminoacylation. Increasing the coding alphabet beyond a modest number of coding letters required evolving various kinds of specificity determining mechanisms involving both energetic coupling between insertion and anticodon-binding domains and, eventually, specific editing domains.

The experimental bases for identifying such coupling between aaRS specificity and the size of the coding alphabet include the fact that the apparent limit to the alphabet size of which Urzymes (made from amino acids from the contemporary alphabet) is roughly four letters: each of the Urzymes we have characterized can activate ~five of the twenty amino acids. As with speculations earlier in §4, this question is currently under investigation, as is the question of whether Urzymes can themselves be encoded using only a four-letter alphabet. The second line of experimental evidence stems from studies of the extensive allosteric communication by which TrpRS domain motion contributes to the amino acid specificity of full-length TrpRS [14, 62, 92, 109]. Those studies complement work on amino acid discrimination in TrpRS and other aaRS [110-113] showing the difficulty of changing amino acid recognition just by mutating side chains that directly contact the amino acid itself.

The modular thermodynamic cycle comparing the TrpRS Urzyme with full-length TrpRS and the Urzyme plus either CP1 or the ABD [109] showed that domain motions increase specific discrimination of tryptophan vs closely related tyrosine. Combinatorial mutagenesis of the D1 switch confirmed that coordinated motion of D1 switch residues enhanced the rejection of tyrosine by ~4.4 kcal/mole above that exhibited by the Urzyme itself [62]. Neither intermediate modular construct improved specificity significantly. Exclusive dependence of enhanced aminoacylation [109] and specific side chain recognition by full-length TrpRS on interdomain coupling energies between the two accessory modules argues that independent recruitment of CP1 and the ABD during evolutionary development of Urzymes would have entailed significant losses of fitness. Development of high precision aminoacylation during aaRS evolution from the Urzyme stage to the full-length enzyme thus presents a paradox.

Notably, the editing domains present in subclass IA aaRS (ValRS, LeuRS, IleRS) are all outgrowths of the initial CP1 insertion. Achievement of the sophistication necessary for these functions was highly unlikely without increased numbers of coding letters. A final strange loop thus connects the genetic coding table to the proteome itself via the collective precision of the aaRS and the dimension of the genetic coding alphabet

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

(Fig. 7A; [14, 92]). The idea that proteins evolved in distinct stages has found support in the multiple varieties of polypeptide architectures found in ribosomal protein structures [114]. Such variations must have been encoded into the sequences of mRNAs and genes concurrently with the dawn 0f heredity and natural selection [14, 75, 92, 115]. Although this field has just opened and faces many challenges in order to validate and elaborate supporting details, we can outline how the pieces likely fitted together. Key elements of this strange loop are the evolving precision of ancestral aaRS and the dimension of the coding alphabet, which participate in the reciprocally coupled gating (Fig. 7B) that guided the introduction of new aaRS•tRNA cognate pairs together with other constraints illustrated in Fig. 5 and discussed in §6. The emerging tree of new cognate pairs was shaped by the new possibilities introduced by enlarging the dimension of the coding alphabet.

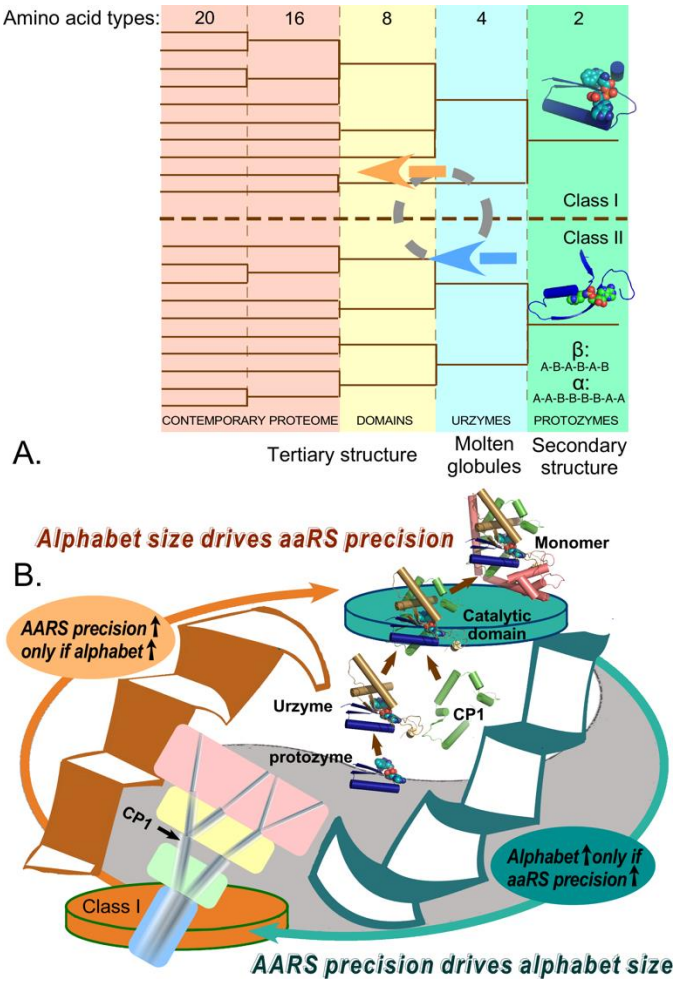

Figure 7. The strange loop connecting the proteome itself to the genetic coding table, via the precision of the aaRS and the coding alphabet size. **A.** Stages in the evolution of the genetic coding alphabet (vertical panels; number of amino acid types) appear to be associated with the evolution of the proteome (bottom), beginning with the introduction of binary patterns enabling secondary structure formation (aaRS protozymes shown as cartoons) through intrinsically disordered molten globules (Urzymes), and finally unique tertiary structures [114]. Trees are represented by thin lines, and major increases in the size of the coding alphabet are represented in differently-colored panels. The circular symbol represents the remodeling of sequences that becomes necessary each time the number of distinct amino acid types increases. That sequence re-shuffling, optimizes in turn the precision of both new and extant aaRS genes, consistent with the new alphabet. **B.** The reciprocally-coupled gating strange loop that drives increases in the alphabet size and aaRS precision.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

8. Conclusions: strange loops tame Eigen's cliff and the paradox of specificity.

We return to the question raised by Dill [1] about "forces" that drive molecular self-organization and, ultimately, produce biology. A central tenet is that the experimental data summarized here furnish sufficient examples to warrant generalizing and arguing by analogy. We also acknowledge the insights of Deacon [116] and Russell and Branscome [69, 70], both of whom recognized the relevance of mutual coupling via reciprocal linkage to the problem of sustaining far from equilibrium states as well as the importance of absence "…use comes from what is not there" [117].

Process control elements regulate events in time. Much effort has been devoted to trying to understand biology n terms of a single process control element, feedback, via autocatalysis and/or hypercycles. In this context, the latter term is reminiscent of introducing increasingly sophisticated "epicycles" to rescue the Ptolemaic model for the solar system. The examples outlined here operate on time scales from sub-millisecond to aeons. They identify a related, but more robust process control element whose novel features appear to surmount the limitations on autocatalysis arising from Eigen's paradox and the paradox of specificity.

The coupling of two XNOR gates together by interchanging the antecedent and consequent logical elements is a recurring formal description for many of biology's most interesting and challenging questions. It is of interest to ask what else these occurrences have in common by collecting the clues we have identified along the way. Fig. 8 is drawn to emphasize that reciprocally-coupled gating functions like a compound logical computer gate. The similarity and differences between Fig. 8 and Fig. 5.8 of [117] should be noted. The latter incorporates reciprocal coupling, but lacks the logical connectives that constrain dissipation. We note below how Fig. 8 could be described as a "teleodynamic" because of how it creates the pretense of purpose.

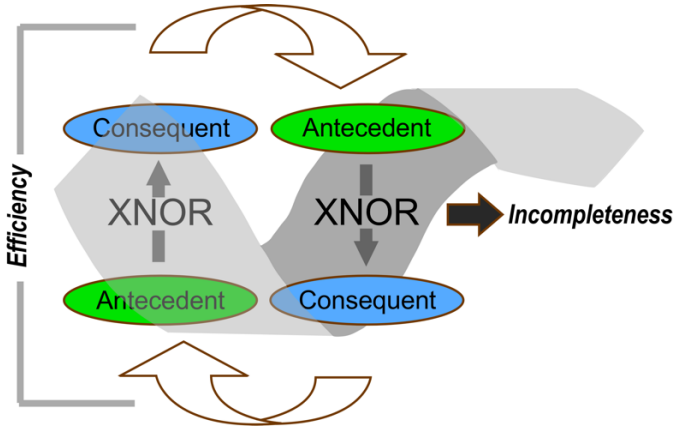

Figure 8. Elements of reciprocally-coupled gating. Two coupled logical XNOR gates are joined head to tail into a single gate, showing the coupling of antecedent and consequent "statements", which assume different meanings in each of the examples described in Figures 2-4 and 7. Three aspects are illustrated: the coupling is brought about by interchanging the antecedent and consequent. This type of coupling constitutes self-reference, hence creates incompleteness. The two XNOR filters together greatly damp unintended variants of the antecedent and consequent, hence reduce dissipation.

The properties of reciprocally coupled gating are two-dimensional, exhibiting different properties—efficiency and incompleteness—in the vertical and horizontal directions of Fig. 8, respectively. We argue that gradients of these properties are functionally analogous to gravitational and chemical potential "forces". Further, the two forces are associated, respectively, with the biological properties of survival and discovery.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

Efficiency and survival. The vertical direction in Fig. 8 corresponds to filtering each stream by the other. Filtering enhances efficiency because the two XNOR gates limit dissipation much as the centripetal force from gravity restricts planetary motion. The strange loop in bioenergetics (Fig. 2) ensures that NTP-dependent free energy transduction processes—in biomechanical work, biosynthesis, signaling, and ATP synthesis—are highly efficient, for example, because the variants of each type of antecedent—conformational state and transition-state binding affinity—exclude those that do not both meet the coupled logical bi-conditional filters.

Efficiency is also a hallmark of the remaining examples, most evidently in Figs. 4 and 7. The efficiency associated with the strange loop in Fig. 3 is subtler, but no less important. We noted in [11] that the feedback loops illustrated in Fig. 3A are more efficient at discovering (and installing) the coding table because they bypass long searches via natural selection. Another way to say this is to recognize that the coding assignments and mRNA sequences of the aaRS genes search much more restricted spaces because they are restricted to sequences that fold and folded proteins that enforce the coding rules. Thus, they are vastly more efficient because of the reciprocally-coupled gating, and are much less vulnerable to the toxic non-functional "parasites" that cripple autocatalytic sets. By this mechanism, the coupled XNOR gates surmount Eigen's paradox/the paradox of specificity.

Incompleteness and discovery. In the horizontal dimension reciprocally-coupled gating achieves explicit self-reference, a property broadly linked to incompleteness because it facilitates construction of paradoxical sentences whose truth can neither be verified nor rejected within the axiomatic system in which they are constructed [16, 118]. The statements associated with each of the examples described in §1(i) – (v) are inherently Gödelian sentences describing puzzles in Biology, each composing a similar "chicken and egg" paradox. Resolution of the paradoxes requires stepping outside the box, so to speak.

Although incompleteness is explicitly defined in Gödelian logic, we [9, 11, 119] have followed others [16, 120-124] in highlighting that natural science may rhyme with logic. Most citations of incompleteness emphasize the negative, disappointing side of incompleteness—that logical systems cannot be complete. We emphasize the "cup half full" interpretation shared by Dyson—"… because of Gödel's theorem, physics is inexhaustible too." [125] and Jaki—"… he (Hawking) made the erroneous claim that Gödel's theorem means the end of physics" [120]. It means exactly the opposite."—that incompleteness implies inexhaustibility.

This argument parallels that of Deacon's citation of Lao-Tse: "…use comes from what is not there" [117]. Completeness implies closure; it is self-limiting. Its opposite behaves like a high chemical potential of yet unformed novelty, hence its capacity as a "teleodynamic" force [126, 127]. Incompleteness equips the strange loop in Fig. 3B with the ability to function as a boot-block, i.e., to discover new functionalities not prefigured in the initial antecedent statements—in this case aaRS genes and coding rules. The strange loop in Fig. 7 behaves similarly, combining efficiency with discovery of new aaRS•tRNA cognate pairs.

The strange loop in Fig. 4 is subtler with respect to its incompleteness. It suggests, however, a plausible model for the evolution of protein catalysts. Directed evolution of native and molten globular forms of dihydrofolate reductase improved catalysis of both forms using non-overlapping sets of mutants, without altering the basic structural differences between them [128]. The authors observed that selected mutations enhanced catalysis indirectly, by strengthening dynamic fluctuations that couple to the reaction coordinate. That conclusion obscures ongoing controversy over the manner in which conformational fluctuations contribute to catalysis [51, 55-57, 129, 130]. The strange loop in Fig. 4A may help reconcile the apparent conflict between the pre-organizational basis of catalysis [51] and recurring evidence from multiple systems suggesting that dynamic networks can contribute to catalysis if transition state complementarity is a transient phenomenon [2].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2021 doi:10.20944/preprints202101.0057.v1

Characterizing the dimensions of Fig. 8 as "forces" leaves many details to fill in. Forces in physics arise from well-defined potentials and explicit gradients. They also are used to define explicit quantities such as work. Our rigorous derivations of the informational Ohm's law in terms of errors and impedance matching [14] suggest that similar formalisms can be devised to bring forces associated with efficiency and incompleteness into better accordance with physics. Even if such developments fail, we believe the metaphorical use of Fig. 8 opens new windows on biology and the origin and behavior of biomolecules.

 doi:10.20944/preprints202101.0057.v1

Acknowledgments: This work reflects extended discussion (CWCjr) over five decades with F. Raymond Salemme, to whose memory this paper is dedicated. The National Institute of General Medical Sciences supported work described in papers cited (NIGMS 78227 and 40906).

Conflict of Interest: Neither author is aware of any conflict of interest.

References

1. Dill KA, Agozzino L. Driving Forces in the Origins of Life. Open Biology. 2020;In Press.
2. Carter CW, Jr., Chandrasekaran SN, Weinreb V, Li L, Williams T. Combining multi-mutant and modular thermodynamic cycles to measure energetic coupling networks in enzyme catalysis Structural Dynamics. 2017;4:032101.
3. Carter CW, Jr. High-Dimensional Mutant and Modular Thermodynamic Cycles, Molecular Switching, and Free Energy Transduction. Annual Review of Biophysics. 2017;46:433-53. doi: 10.1146/annurev-biophys-070816-033811.
4. Chandrasekaran SN, Das J, Dokholyan NV, Carter CW, Jr. A modified PATH algorithm rapidly generates transition states comparable to those found by other well established algorithms. Structural Dynamics. 2016;3:012101. doi: 10.1063/1.4941599. PubMed Central PMCID: PMCPMC4769271.
5. Weinreb V, Li L, Carter CW, Jr. A Master Switch Couples Mg^{2+}-Assisted Catalysis to Domain Motion in B. stearothermophilus Tryptophanyl-tRNA Synthetase. Structure. 2012;20:128-38.
6. Carter CW, Jr. Escapement mechanisms: efficient free energy transduction by reciprocally-coupled gating. Proteins: Structure, Function, and Bioinformatics. 2019;88:710–7. doi: 10.1002/prot.25856.
7. Carter CW, Jr. Coding of Class I and II aminoacyl-tRNA synthetases. Advances in Experimental Medicine and Biology: Protein Reviews. 2017;18:103-48. doi: DOI 10.1007/5584_2017_93.
8. Carter CW, Jr., Wolfenden R. Acceptor-stem and anticodon bases embed amino acid chemistry into tRNA. RNA Biology. 2016;13(2):145–51. doi: 10.1080/15476286.2015.1112488.
9. Carter CW, Jr., Wolfenden R. tRNA Acceptor-Stem and Anticodon Bases Form Independent Codes Related to Protein Folding. Proc Nat Acad Sci USA. 2015;112 (24):7489-94. doi: http://www.pnas.org/cgi/doi/10.1073/pnas.1507569112.
10. Wolfenden R, Lewis CA, Yuan Y, Carter CW, Jr. Temperature dependence of amino acid hydrophobicities. Proc Nat Acad Sci USA. 2015;112 (24):7484-8. doi: 10.1073/pnas.1507565112.
11. Carter CW, Jr, Wills PR. Interdependence, Reflexivity, Fidelity, and Impedance Matching, and the Evolution of Genetic Coding. Molecular Biology and Evolution. 2018;35(2):269-86. doi: 10.1101/139139. PubMed Central PMCID: PMC29077934.
12. Wills PR, Carter CW, Jr. Insuperable problems of an initial genetic code emerging from an RNA World. BioSystems. 2018;164:155-66. doi: 10.1101/140657
13. Sapienza PJ, Li L, Williams T, Lee AL, Carter CW, Jr. An Ancestral Tryptophanyl-tRNA Synthetase Precursor Achieves High Catalytic Rate Enhancement without Ordered Ground-State Tertiary Structures. ACS Chemical Biology. 2016;11:1661−8. doi: 10.1021/acschembio.5b01011.
14. Wills PR, Carter CW, Jr. Impedance matching and the choice between alternative pathways for the origin of genetic coding. International Journal of Molecular Sciences. 2020;21:7392. doi: 10.3390/ijms21197392.
15. Alex Blokhuis, David Lacoste, Philippe Nghe. Universal motifs and the diversity of autocatalytic systems. Proc Nat Acad Sci USA. 2020;117 (October 13, no. 41):25230–6 doi: 10.1073/pnas.2013527117.
16. Hofstadter DR. Gödel, Escher, Bach: an eternal golden braid. New York: Basic Books, Inc; 1979. 777 p.
17. Kauffman SA. Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. Journal of Cybernetics. 1971;1(1):71–96.
18. Eigen M. Selforganization of Matter and the Evolution of Biological Macromolecules. Naturwissenschaften. 1971;58:465-523.
19. Eigen M. Molecular self-organisation and the early stages of evolution. Quart Rev Biophys 1971;4:149-212.
20. Kauffman SA. Autocatalytic Sets of Proteins. J Theor Bio,. 1986; 119:1–24.
21. Eigen M, Schuster P. The Hypercycle: A Principle of Natural Self-Organization Part C: The Realistic Hypercycle. Die Naturwissenschaften. 1978;65:341-69.
22. Eigen M, Schuster P. The Hypercycle: A Principle of Natural Self-Organization Part A: Emergence of the Hypercycle. Naturwissenschaften. 1977;64:541-65.

23. Hordijk W. A History of Autocatalytic Sets: A Tribute to Stuart Kauffman. Biological Theory. 2019;14:224–46. doi: 10.1007/s13752-019-00330-w.
24. Hordijk W. Evolution: Limited and Predictable or Unbounded and Lawless? Biol Theory 2016;11:187–91. doi: 10.1007/s13752-016-0251-5.
25. Sousa FL, Hordijk W, Steel M, Martin WF. Autocatalytic sets in E. coli metabolism. Journal of Systems Chemistry. 2015;6:4. doi: 10.1186/s13322-015-0009-7.
26. Smith JI, Steel M, Hordijk W. Autocatalytic sets in a partitioned biochemical network. Journal of Systems Chemistry. 2014;5:2. doi: 10.1186/1759-2208-5-2.
27. Hordijk W, Wills PR, Steel M. Autocatalytic Sets and Biological Specificity. Bull Math Biol 2014;76:201–24. doi: 10.1007/s11538-013-9916-4.
28. Hordijk W, Steel M, Kauffman S. The Structure of Autocatalytic Sets: Evolvability, Enablement, and Emergence. Acta Biotheoretica. 2012;60:379–92. doi: 10.1007/s10441-012-9165-1.
29. Kauffman S. What Is Life, and Can We Create It? BioScience. 2013;63 (August No. 8):609-10. doi: 10.1525/bio.2013.63.8.2.
30. Lancet D, Zidovetzki R, Markovitch O. Systems protobiology: origin of life in lipid catalytic networks. J R Soc Interface. 2018; 15::20180159. doi: 10.1098/rsif.2018.0159.
31. Orgel LE. Self-organizing biochemical cycles. Proc Nat Acad Sci USA. 2000;97(23):12503–7. doi: 10.1073/pnas.220406697.
32. Stubbs RT, Yadav M, Krishnamurthy R, Springsteen GR. A plausible metal-free ancestral analogue of the Krebs cycle composed entirely of α-ketoacids. Nature Chemistry. 2020;12(| November):1016–22.
33. Sobotta J, Geisberger T, Moosmann C, Scheidler CM, Eisenreich W, Wächtershäuser Gn, et al. A Possible Primordial Acetyleno/Carboxydotrophic Core Metabolism. Life. 2020;10:35. doi: 10.3390/life10040035.
34. Eigen M. Error catastrophe and antiviral strategy. Proc Nat Acad Sci USA. 2002;99(October 15, no. 21):13374–6 doi: 10.1073 pnas.212514799.
35. Szathmàry E. The evolution of relicators. Phil Trans R Soc Lond B. 2000;355:1669-76. doi: 10.1098/rstb.2000.0730.
36. Koonin EV. The Logic of Chance: The Nature and Origin of Biological Evolution. Upper Saddle River, NJ: Pearson Education; FT Press Science; 2011.
37. Turing AM. The Chemical Basis of Morphogenesis. Philosophical Transactions of the Royal Society of London Series B, BiologicalSciences,. 1952;237(641 (Aug. 14)):37-72.
38. Wills PR. Autocatalysis, information, and coding. BioSystems. 2001;50:49-57.
39. McCaskill J, Miller JF, Stepney S, Wills PR. Encoding and representation of information processing in irregular computational matter. 2018. In: Computational Matter [Internet]. Berlin: Springer; [227-43].
40. Wills PR. The generation of meaningful information in molecular systems. Phil Trans R Soc A. 2016;A374:20150016. doi: 10.1098/rsta.20150066.
41. Wills PR, Nieselt K, McCaskill JS. Emergence of Coding and its Specificity as a Physico-Informatic Problem. Orig Life Evol Biosph. 2015;45(June):249-55. doi: 10.1007/s11084-015-9434-5.
42. Wills PR. Reflexivity, Coding and Quantum Biology. BioSystems. 2019;185:104027.
43. Boyer PD. Energy, Life, and ATP. In: Grenthe I, editor. Nobel Lectures, Chemistry 1996-2000. Singapore: The Nobel Foundation; 1997.
44. Boyer PD. The ATP synthase – a splendid molecular machine, Annual Review in Biochem. 1997;66:717-49.
45. Boyer PD. The binding change mechanism for ATP synthase – Some probabilities and possibilities. Biochimica et Biophysica Acta. 1993;1140:215-50.
46. Jencks WP. Reaction mechanisms, catalysis, and movement. Prot Sci. 1994;3:2459-64.
47. Jencks WP. Utilization of Binding Energy and Coupling Rules for Active Transport and Other Coupled Vectorial Processes. Meth Enz. 1989;171:145-64.
48. Hill TL. Some general principles in free energy transduction. Proc Nati Acad Sci USA. 1983;80(May):2922-5.

49. Hill TL, Eisenberg E. Can free energy transduction be localized at some crucial part of the enzymatic cycle? Quarterly Review of Biophysics. 1981;14:463-511.
50. Eisenberg E, Hill TL. A Cross-Bridge Model of Muscle Contraction. Prog Biophys Mol Biol. 1978;33:55-82.
51. Warshel A, Bora RP. Perspective: Defining and quantifying the role of dynamics in enzyme catalysis. The Journal of Chemical Physics. 2016;144:180901. doi: 10.1063/1.4947037.
52. Astumian RD, Mukherjee S, Warshel A. The Physics and Physical Chemistry of Molecular Machines. ChemPhysChem. 2016;17:1719–41. doi: 10.1002/cphc.201600184.
53. Prasad BR, Warshel A. Prechemistry versus preorganization in DNA replication fidelity. PROTEINS, Structure, Function and Bioinformatics. 2011;79:2900-19.
54. Knapp MJ, Rickert K, Klinman JP. Temperature-dependent isotope effects in soybean lipoxygenase-1: Correlating hydrogen tunneling with protein dynamics. J Am Chem Soc 2002;124:3865–74.
55. Villali J, Kern D. Choreographing and enzyme's dance. Current Opinion in Chemical Biology. 2010;14:636-43.
56. Henzler-Wildman KA, Lei M, Vu T, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature. 2007;450(6 December):913-8.
57. Henzler-Wildman KA, Lei M, Vu T, Ott M, Wolf-Watz M, Fenn T, et al. Intrinsic motions along an enzymatic reaction trajectory. Nature. 2007;450(6 December):838-49.
58. Pisliakov AV, Cao J, Kamerlin SCL, Warshel A. Enzyme millisecond conformational dynamics do not catalyze the chemical step. Proc Nat Acad Sci, USA. 2009;106:17359-64.
59. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM. Electrostatic Basis for Enzyme Catalysis. Chemical Reviews. 2006;106:3210-35.
60. Kapustina M, Weinreb V, Li L, Kuhlman B, Carter CW, Jr. A Conformational Transition State Accompanies Tryptophan Activation by *B. stearothermphilus* Tryptophanyl-tRNA Synthetase. Structure. 2007;15:1272-84.
61. Weinreb V, Weinreb G, Chandrasekaran SN, Das J, Dokholyan NV, Carter CW, Jr. Thermodynamic impacts of combinatorial mutagenesis on protein conformational stability: precise, high-throughput measurement by Thermofluor. BioRxiv. 2019;591495. doi: 10.1101/591495.
62. Weinreb V, Li L, Chandrasekaran SN, Koehl P, Delarue M, Carter CW, Jr Enhanced Amino Acid Selection in Fully-Evolved Tryptophanyl-tRNA Synthetase, Relative to its Urzyme, Requires Domain Movement Sensed by the D1 Switch, a Remote, Dynamic Packing Motif J Biol Chem 2014;289:4367-76. doi: 10.1074/jbc.M113.538660.
63. Chandrasekaran SN, Carter CW, Jr. Adding torsional interaction terms to the Anisotropic Network Model improves the PATH performance, enabling detailed comparison with experimental rate data Structural Dynamics. 2017;4:032103.
64. Burbaum J, Schimmel P. Structural Relationships and the Classification of Aminoacyl-tRNA Synthetases. J Biol Chem. 1991;266(26):16965-8.
65. Burbaum JJ, Schimmel P. Assembly of a Class I tRNA Synthetase from Products of an Artificially Split Gene. Biochem. 1991;30(2):319-24.
66. Kapustina M, Carter CW, Jr. Computational Studies of Tryptophanyl-tRNA Synthetase: Activation of ATP by Induced-Fit. J Mol Biol. 2006;362:1159-80.
67. Retailleau P, Weinreb V, Hu M, Carter CW, Jr. Crystal Structure of Tryptophanyl-tRNA Synthetase Complexed with Adenosine-5' tetraphosphate: Evidence for Distributed Use of Catalytic Binding Energy in Amino Acid Activation by Class I Aminoacyl-tRNA Synthetases. Journal of Molecular Biology. 2007;369:108-28.
68. Retailleau P, Huang X, Yin Y, Hu M, Weinreb V, Vachette P, et al. Interconversion of ATP binding and conformational free energies by Tryptophanyl-tRNA Synthetase: structures of ATP bound to open and closed, pre-transition conformations. Journal of Molecular Biology. 2003;325:39-63.
69. Branscomb E, Russell MJ. Frankenstein or a Submarine Alkaline Vent: Who Is Responsible for Abiogenesis? Part 1: What is life–that it might create itself? BioEssays 2018;40:1700179.
70. Branscomb E, Biancalani T, Goldenfeld N, Russell M. Escapement mechanisms and the conversion of disequilibria; the engines of creation. Physics Reports. 2017;677 1–60. doi: http://dx.doi.org/10.1016/j.physrep.2017.02.0010370-1573.

71. Wolfenden R. Experimental Measures of Amino Acid Hydrophobicity and the Topology of Transmembrane and Globular Proteins. Journal of General Physiology. 2007;129(5 May):357–62. doi: 10.1085/jgp.200709743.
72. Gibbs PR, Radzicka A, Wolfenden R. The Anomalous Hydrophilic Character of Proline. J Am Chem Soc. 1991;113:4714-5.
73. Radzicka A, Wolfenden R. Comparing the Polarities of the Amino Acids: Side-Chain Distribution Coefficients between the Vapor Phase, Cyclohexane, 1-0ctanol, and Neutral Aqueous Solution. Biochem. 1988;27(5):1664-70.
74. Wolfenden R, Cullis PM, Southgate CCF. Water, Protein Folding, and the Genetic Code. Science. 1979;206:575-7.
75. Carter CW, Jr., Wills PR. The Roots of Genetic Coding in Aminoacyl-tRNA Synthetase Duality Annual Review of Biochemistry. 2021;89:In Press.
76. Eriani G, Delarue M, Poch O, Gangloff J, Moras D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature. 1990;347(6289):203-6.
77. Cusack S. Eleven down and nine to go. Nat Str Biol. 1995;2:824-31.
78. Cusack S, Berthet-Colominas C, Hartlein M, Nassar N, Leberman R. A second class of synthetase structure revealed by X-ray analysis of Escherichia coli seryl-tRNA synthetase at 2.5 A. Nature. 1990;347(6290):249-55.
79. Ruff M, Krishnaswamy S, Boeglin M, Poterszman A, Mitschler A, Podjarny A, et al. Class II Aminoacyl Transfer RNA Synthetases: Crystal Structure of Yeast Aspartyl-tRNA Synthetase Complexed with $tRNA^{Asp}$. Science. 1991;252(6):1682-9.
80. Carter CW, Jr. Cognition, Mechanism, and Evolutionary Relationships in Aminoacyl-tRNA Synthetases. Ann Rev Biochem. 1993;62:715-48.
81. Carter CW, Jr, Wills PR. Class I and II aminoacyl-tRNA synthetase tRNA groove discrimination created the first synthetase•tRNA cognate pairs and was therefore essential to the origin of genetic coding. IUBMB Life. 2019;71(8, August):1088–98. doi: 10.1002/iub.2094. PubMed Central PMCID: PMC31190358.
82. Carter CW, Jr , Wills PR. Hierarchical groove discrimination by Class I and II aminoacyl-tRNA synthetases reveals a palimpsest of the operational RNA code in the tRNA acceptor-stem bases. Nucleic Acids Research. 2018;46(18):9667–83. doi: 10.1093/nar/gky600.
83. Carter CW, Jr., Wills PR. Experimental Solutions to Problems Defining the Origin of Codon-Directed Protein Synthesis. BioSystems. 2019;1832019(September):103979. doi: 10.1016/j.biosystems.2019.103979. PubMed Central PMCID: PMC31176803.
84. Rodin A, Rodin SN, Carter CW, Jr. On Primordial Sense-Antisense Coding. Journal of Molecular Evolution. 2009;69:555-67.
85. Rodin SN, Ohno S. Two Types of Aminoacyl-tRNA Synthetases Could be Originally Encoded by Complementary Strands of the Same Nucleic Acid. Orig Life Evol Biosph. 1995;25:565-89.
86. Carter CW, Jr. Urzymology: Experimental Access to a Key Transition in the Appearance of Enzymes. J Biol Chem. 2014;289(44):30213–20. doi: 10.1047/jbcR114.576495.
87. Li L, Francklyn C, Carter CW, Jr. Aminoacylating Urzymes Challenge the RNA World Hypothesis. J Biol Chem. 2013;288:26856-63. doi: 10.1074/jbc.M113.496125
88. Li L, Weinreb V, Francklyn C, Carter CW, Jr. Histidyl-tRNA Synthetase Urzymes: Class I and II Aminoacyl-tRNA Synthetase Urzymes have Comparable Catalytic Activities for Cognate Amino Acid Activation. J Biol Chem. 2011;286:10387-95. doi: 10.1074/jbc.M110.198929.
89. Pham Y, Kuhlman B, Butterfoss GL, Hu H, Weinreb V, Carter CW, Jr. Tryptophanyl-tRNA synthetase Urzyme: a model to recapitulate molecular evolution and investigate intramolecular complementation. J Biol Chem. 2010;285:38590-601. doi: 10.1074/jbc.M110.136911
90. Pham Y, Li L, Kim A, Erdogan O, Weinreb V, Butterfoss G, et al. A Minimal TrpRS Catalytic Domain Supports Sense/Antisense Ancestry of Class I and II Aminoacyl-tRNA Synthetases. Mol Cell. 2007;25:851-62.
91. Martinez L, Jimenez-Rodriguez M, Gonzalez-Rivera K, Williams T, Li L, Weinreb V, et al. Functional Class I and II Amino Acid Activating Enzymes Can Be Coded by Opposite Strands of the Same Gene. J Biol Chem. 2015;290(32):19710–25. doi: 10.1074/jbc.M115.642876

92. Carter CW, Jr., Popinga A, Bouckaert R, Wills PR. High-Resolution, Multidimensional Phylogenetic Metrics Identify Class I Aminoacyl-tRNA Synthetase Evolutionary Mosaicity and Inter-modular`Coupling. BioRxiv. 2020;2020/033712.

93. Chandrasekaran SN, Yardimci G, Erdogan O, Roach JM, Carter CW, Jr. Statistical Evaluation of the Rodin-Ohno Hypothesis: Sense/Antisense Coding of Ancestral Class I and II Aminoacyl-tRNA Synthetases. Molecular Biology and Evolution. 2013;30(7):1588-604. doi: 10.1093/molbev/mst070.

94. Goto Y, Katoh T, Suga H. Flexizymes for genetic code reprogramming. Nature Protocols. 2011;6(NO.6):779-90.

95. Pervushin K, Vamvaca K, Vogeli B, Hilvert D. Structure and dynamics of a molten globular enzyme. Nat Struct Mol Biol. 2007;14(December):1202-6.

96. Vamvaca K, Vögeli B, Kast P, Pervushin K, Hilvert D. An enzymatic molten globule: Efficient coupling of folding and catalysis. PNAS 2004;101(August 31, no. 35):12860–4. doi: 10.1073pnas.0404109101.

97. Vamvaca K, Jelesarov I, Hilvert D. Kinetics and Thermodynamics of Ligand Binding to a Molten Globular Enzyme and Its Native Counterpart. J Mol Biol. 2008;382:971–7.

98. Hu H. Wild-type and molten globular chorismate mutase achieve comparable catalytic rates using very different enthalpy/entropy compensations. Science China. 2014;57(No.1):156–64. doi: 10.1007/s11426-013-5021-7.

99. Parra RG, Schafer NP, Radusky LG, Tsai M-Y, Guzovsky AB, Wolynes PG, et al. Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. Nucl Acids Res. 2016;44(Web Server issue):W357.

100. Freibergera MI, Guzovskya AB, Wolynes PG, Parra RG, Ferreiroa DU. Local frustration around enzyme active sites. PNAS. 2019;116(March 5 no. 10):4037–43. doi: 10.1073/pnas.1819859116

101. Pauling L. The probability of errors in the process of synthesis of protein molecules. In: Pauling L, editor. Festschrift Prof Dr Arthur Stoll Basel: Birkhauser; 1957. p. 597-602.

102. Bull JJ, Meyers LA, Lachmann M. Quasispecies Made Simple. PLoS Computational Biology | wwwploscompbiolorg |. 2005;1 (November Issue 6):e61. doi: 10.1371/journal.pcbi.0010061.

103. Hopfield JJ, Yamane, T., Yue, V., and Coutts, S.M. Direct Experimetnal Evidence for Kinetic Proofreading in Amino "Acylation of tRNA Ile". Proc Natl Acad Sci USA. 1976;73:1164-8.

104. Fersht AR, Kaethner MM. Enzyme Hyperspecifity. Rejection of Threonine by the Valyl-tRNA Synthetase by Misacylation and Hydrolytic Editing. Biochem. 1976;15(15):3342-6.

105. Martinis SA, Boniecki MT. The balance between pre- and post-transfer editing in tRNA synthetases. FEBS Letters 2010;584:455–9.

106. Boniecki MT, Vu MT, Betha AK, Martinis SA. CP1-dependent partitioning of pretransfer and posttransfer editing in leucyl-tRNA synthetase. Proc Nat Acad Sci USA 2008;105(49 December 9):19223–8. doi: 10.1073 pnas.0809336105.

107. Carter CW, Jr. What RNA World? Why a Peptide/RNA Partnership Merits Renewed Experimental Attention. Life. 2015;5:294-320. doi: 10.3390/life5010294.

108. Carter CW, Jr., Li L, Weinreb V, Collier M, Gonzales-Rivera K, Jimenez-Rodriguez M, et al. The Rodin-Ohno Hypothesis That Two Enzyme Superfamilies Descended from One Ancestral Gene: An Unlikely Scenario for the Origins of Translation That Will Not Be Dismissed. Biology Direct. 2014;9:11.

109. Li L, Carter CW, Jr. Full Implementation of the Genetic Code by Tryptophanyl-tRNA Synthetase Requires Intermodular Coupling. J Biol Chem. 2013;288(29 November):34736–45. doi: 10.1074/jbc.M113.510958.

110. Perona JJ, Gruic-Sovulj I. Synthetic and Editing Mechanisms of Aminoacyl-tRNA Synthetases. Topics in Current Chemistry 2013;DOI: 10.1007/128_2013_456. doi: 10.1007/128_2013_456.

111. Praetorius-Ibba M, Stange-Thomann N, Kitabatake M, Ali K, Soll I, Carter CW, Jr., et al. Ancient adaptation of the active site of tryptophanyl-tRNA synthetase for tryptophan binding. Biochem. 2000;39(43):13136-43.

112. Bullock TL, Rodríguez-Hernández A, Corigliano EM, Perona JJ. A rationally engineered misacylating aminoacyl-tRNA synthetase. Proc Nat Acad Sci, USA. 2008;105:7428-33.

113. Bullock T, Uter N, Nissan TA, Perona JJ. Amino Acid Discrimination by a class I aminoacyl-tRNA synthetase specified by negative determinants. J Mol Biol. 2003;328:395-408.

114. Kovacs NA, Petrov AS, Lanier KA, Williams LD. Frozen in Time: The History of Proteins. Mol Biol Evol. 2017;34(5):1252–60 doi: 10.1093/molbev/msx086.
115. Carter CWJ. Simultaneous codon usage, the origin of the proteome, and the emergence of de-novo proteins. Current Opinion in Structural Biology. 2021;68:In Press.
116. Deacon TW. Reciprocal Linkage between Self-organizing Processes is Sufficient for Self-reproduction and Evolvability. Biological Theory 2006;1(2):136–49.
117. Deacon TW. Emergence: The Hole at the Wheel's Hub. In: Clayton P, Davies P, editors. The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion. Oxford UK: Oxford University Press; 2008.
118. Ben-Ya'acov U. G•odel's incompleteness theorem and Universal physical theories. Journal of Physics: Conference Series 2019;1391:012067. doi: 10.1088/1742-6596/1391/1/012067.
119. Carter CW, Jr. The Evolution of Genetic Coding. Scientia: Science Diffusion, Ltd Bristol, UK; 2019.
120. Jaki SL. On a Discovery About G•odel's Incompleteness Theorem Paths of Discovery Pontical Academy of Sciences Acta Vatican City 2006 18:49-60.
121. Jaki SL. A Late Awakening to Gödel in Physics 2004 [cited 2020 29/12]. Article]. Available from: http://www.sljaki.com/texts.html.
122. Feferman S. G•odel, Nagel, Minds, and Machines. The Journal of Philosophy. 2009;106:201-19.
123. Hawking S. Gödel and the End of Physics 2002 [cited 2020 29/12/20]. Available from: http://yclept.ucdavis.edu/course/215c.S17/TEX/GodelAndEndOfPhysics.pdf.
124. Dyson F. The World on a String. New York Review of Books. 2004:13 May.
125. Wright PE, Dyson HJ. Linking folding and binding. Curr Op Struct Biol. 2009;19:31–8.
126. Deacon TW. Incomplete Nature. New York: W. W. Norton & Company; 2012.
127. Deacon TW, Cashman T. Steps to a Metaphysics of Incompleteness. Theology and Science. 2016;14(4):401-29. doi: 10. 1080/ 14745700 2016.1231977.
128. Schulenburg C, Stark Y, Künzle M, Hilvert D. Comparative Laboratory Evolution of Ordered and Disordered Enzymes. J Biol Chem. 2015; 290, (15, April 10):9310–20.
129. Kern T, Schanda P, Brutscher B. Sensitivity-enhanced IPAP-SOFAST-HMQC for fast-pulsing 2D NMR with reduced radiofrequency load. Journal of Magnetic Resonance. 2008;190:333–8. doi: 10.1016/j.jmr.2007.11.015.
130. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnefv DM, Wolf-Watz M, Bosco DA, et al. Intrinisic dynamics of an enzyme underlies catalysis. Nature. 2005;438:117-37.