

Article

Not peer-reviewed version

---

# Kappa-Frameshift Background Mutations and Long-Range Correlations of the DNA Base Sequences

---

[Elias Koorambas](#) \*

Posted Date: 17 December 2025

doi: 10.20944/preprints202503.0891.v2

Keywords: informational-entropy; mutations, correlations; DNA base sequences



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.

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.

Article

# Kappa-Frameshift Background Mutations and Long-Range Correlations of the DNA Base Sequences

Elias Koorambas

Computational Applications, Group, Division of Applied Technologies, National Center for Science and Research 'Demokritos', Aghia Paraskevi-Athens, Greece; elias.koor@gmail.com

## Abstract

Following Livadiotis G. and McComas D. J. (2023) [1], we propose a new type of DNA frameshift mutations that occur spontaneously due to information exchange between the DNA sequence of length bases ( $n$ ) and the mutation sequence of length bases ( $m$ ), and respect the kappa-addition symbol  $\oplus\kappa$ . We call these proposed mutations Kappa-Frameshift Background (KFB) mutations. We find entropy defects originate in the interdependence of the information length systems (or their interconnectedness, that is, the state in which systems with a significant number of constituents (information length bases) depend on, or are connected with each) by the proposed KFB-mutation). We also quantify the correlation among DNA information length bases ( $n$ ) and ( $m$ ) due to information exchange. In the presence of entropy defects, the Landauer's bound and minimal metabolic rate for a biological system are modified. We observe that the different  $n$  and  $\kappa$  scales are manifested in the double evolutionary emergence of the proposed biological system through subsystems correlations. For specific values of the kappa parameter we can expect deterministic laws associated with a single biological polymer in the short term before the polymer explores over time all the possible ways it can exist.

**Keywords:** informational-entropy; mutations, correlations; DNA base sequences

## 1. Introduction

Informational entropy analysis is often used in biology, but its meaning should be distinguished from that of thermodynamic entropy in molecular biology and evolution [2–26]. The genetic information in DNA supplies straightforward examples for informational entropy concepts. The characteristic nucleotide sequence of the four bases (Adenine A, Guanine G, Cytosine C, and Thymine T) in an organism's DNA is the same in all cells of the organism, and results in the combination of random mutations and functional selection during biological evolution. There is no preference for a specific sequence;  $4^n$  different DNA sequences are equally possible, and the actual DNA sequence has only a  $4^{-n}$  probability ( $p$ ) (where " $n$ " is the number of bases in the DNA sequence, ranging from one million in bacteria to billions in many animals and plants). Therefore, the space of possible molecular configurations of molecules within an organism is astronomically large. We may assume that the system is "non-ergodic" in the sense that not every part of the configuration space can be explored [3,27–30]. There are several major reasons to expect convergence [3]. First, all evolutionary trajectories occur under certain generic selective constraints, such as the laws of mathematics, physics, and chemistry; these should lead to some universal features [31,32]. For instance, a novel quantum system (quantum dots) recently proposed by the author (2024) [33], due to the system's peculiar quantum state—a Diophantine Markov state, would be trapped in a sequence of self-replicating states and would never reach a classical limit, no matter how big the cube (since gaps between states are always finite). Second, many of these spaces are not explored randomly [34–43]. Finally, evolutionary trajectories happen within a certain system: Darwinian evolution will act on populations of entities. Entities within the system emerge within assembly spaces [44–46]. They are

constructed from components over evolutionary timescales. These shape evolution and shift the way one should think about sequence search conditioned on the past, where the logic of assembly spaces may lead to certain types of universal convergence [3,46].

DNA replication can lead to the introduction of small insertions or deletions. The addition or removal of one or more base pairs  $m$  leads to insertion or deletion mutations, respectively. As shown in [47], the loss or addition of a single nucleotide causes all of the subsequent three-letter codons to be changed [48]. These mutations are called frameshift mutations because the frame of the triplet reading is altered during translation. A frameshift mutation occurs when any number of bases  $m$  are added or deleted, except multiples of three which reestablish the initial reading frame.

Kappa-distributions emerge within the framework of statistical mechanics by maximizing the associated kappa-entropy  $S_\kappa$  under the constraints of the canonical ensemble (e.g., Livadiotis G. and McComas D. J. (2009, 2014, 2023) [1,49,50]). Kappa-distributions are consistent with thermodynamics; yet this fact alone cannot justify the generation and existence of these distributions. Examples of kappa-distributions include super-statistics [51–56], the effect of shock waves [57], turbulence [58–60], the effect of pickup ions [61–63], the pump acceleration mechanism [64], colloidal particles [65], and polytropic behavior [66,67]. For all these phenomena, there is a unique thermodynamic origin. The existence of correlations among the particles of a system does not permit the usage of the classical framework based on the Boltzmann [68]–Gibbs [69] (BG) entropy and Maxwell–Boltzmann (MB)[70] canonical distribution of particle velocity ( $u$ ). The thermodynamic origin of kappa-distributions has recently been connected to the new concept of “entropy defect”, proposed by Livadiotis G. and McComas D. J. (2021) [71,72]. Recently, Livadiotis and McComas (2023) [1] introduced a specific algebra of the addition rule with entropy defect. The addition rule forms a mathematical Abelian group on the set of any measurable physical-quantity (including entropy), called the kappa-addition and refer to by the symbol  $\oplus_\kappa$ . The physical meaning of the kappa index can be understood through particle (subsystems) correlations: the kappa index provides a measure of the loss (for  $\kappa > 0$ ) or gain (for  $\kappa < 0$ ) of entropy  $\Delta S_{system}$  after the mixing of the partial systems A and B to compose the combined system [1].

Following Livadiotis and McComas (2023) [1], we propose a new type of DNA frameshift mutations that occur spontaneously due to information exchange between the DNA sequence of length bases ( $n$ ) and the mutation sequence of length base ( $m$ ), and respect the kappa-addition symbol  $\oplus_\kappa$ . We call these proposed mutations Kappa-Frameshift Background (KFB) Mutations. We find that the origin of an entropy defect is based on the interdependence of the information length systems (or their interconnectedness, that is, the dependence, or connection, of systems with a significant number of constituents (information length bases) on/with each other. The proposed KFB-mutation quantifies the correlations among the information length bases, or constituents of these systems due to information exchanges. In the presence of entropy defects the Landauer’s bound and the minimal metabolic rate for a biological system are modified. We observe that that the different  $n$  and  $\kappa$  scales are manifested in the double appearance of the evolution of the proposed biological system through subsystems correlations. In this evolution, for short time scales and for specific values of the kappa-parameter, we can expect deterministic laws associated with a single biological polymer.

## 2. Kappa-Frameshift Background DNA Mutations

DNA replication can lead to the introduction of small insertions or deletions of base pairs. The addition or removal of one or more base pairs  $m$  results in insertion or deletion mutations, respectively. As shown in [47], the loss or addition of a single nucleotide causes all of the subsequent three-letter codons to be changed [48]. These mutations are called frameshift mutations because they alter the frame of the triplet reading during translation. Spontaneous DNA mutations occur suddenly and their origin is unknown. Such mutations, also called “background mutations”, have been reported in many organisms, such as *Oenothera*, maize, bread molds, microorganisms (bacteria and

viruses), *Drosophila*, mice, humans, etc. (see Ref [48] for details). A frameshift mutation occurs when any number of bases  $m$  are added or deleted, except multiples of three which reestablish the initial reading frame. Mutations of multiples of three nucleotides just add one or several reading frames into the sequence without altering the composition of those reading frames thereafter. It is possible, however, that the frameshift causes early termination of the translation (e.g., when a 3-nucleotide stop codon sequence is introduced). The results of frameshift mutations can, therefore, be very severe if the mutations occur early in the coding sequence. Frameshift mutations can occur at microsatellites (tandem repeats of 1–6 base pairs per repeat unit) and also at non-iterated or short repetitive sequences. Proofreading can remove insertion or deletion mutations in short repetitive and non-iterated sequences with almost the same efficiency as that for point mutations; however, it is much less efficient in removing frameshift mutations in homopolymeric runs [47]. These mutations are removed efficiently by mismatch repair systems. The instability of microsatellite sequences is associated with many diseases, including cancer [48].

Following Livadiotis and McComas (2023) [1], we propose a new type of DNA frameshift mutations that occur spontaneously due to information exchanges between the DNA sequence of base lengths ( $n$ ) and the mutation sequence of base lengths ( $m$ ), and respects the kappa-addition symbol  $\oplus_{\kappa}$ . We call the proposed mutations Kappa-Frameshift Background (KFB) Mutations. KFB mutations will occur when any number of bases  $m$  are added or deleted, to the initial DNA sequence  $n$  with respecting kappa-addition symbol  $\oplus_{\kappa}$  (except multiples of three, which will reestablish the initial reading frame). In the case of the kappa-addition of any number of bases  $m \in \mathbb{N}$  to the initial DNA base sequence  $n \in \mathbb{N}$ , we define the set of kappa-natural numbers  $(\mathbb{N}, \oplus_{\kappa})$  as the set of natural numbers  $\mathbb{N}$  that satisfies kappa addition  $\oplus_{\kappa}$  law then the proposed KFB mutations are given as follows:

$$n \oplus_{\kappa} m = n + m - \frac{nm}{\kappa}, \quad (n, m \text{ no multiples of three, and } \kappa \in \mathbb{N}^*), \quad (1)$$

which brings the usual sum as a particular case  $\oplus_{\infty} \equiv +$ . The relation 1 can be expressed in the convenient product form [1]:

$$1 - \frac{n \oplus_{\kappa} m}{\kappa} = \left(1 - \frac{n}{\kappa}\right) \left(1 - \frac{m}{\kappa}\right), \quad (n, m \text{ no multiples of three, and } \kappa \in \mathbb{N}^*), \quad (2)$$

where  $n$  is the initial length of information (string length) of the DNA or biological polymer that is being copied and  $m$  (no multiples of three) is the length of information (string length) of the mutation added to the initial base  $n$ . The  $\kappa$ -sum is commutative ( $n \oplus_{\kappa} m = m \oplus_{\kappa} n$ ), associative ( $(n \oplus_{\kappa} m) \oplus_{\kappa} z = n \oplus_{\kappa} (m \oplus_{\kappa} z)$ ), but it is not distributive in relation to the usual multiplication ( $\alpha(n \oplus_{\kappa} m) \neq (\alpha n) \oplus_{\kappa} (\alpha m)$ ). The neutral element of the  $\kappa$ -sum is zero,  $n \oplus_{\kappa} 0 = n$ . Similarly, in the case of  $\kappa$ -deletion of any number of bases  $m \in \mathbb{N}$  (no multiples of three) to the initial DNA base sequence  $n \in \mathbb{N}$ , the proposed KFB mutation is given by the following:

$$n \oplus_{\kappa} \bar{m} = \frac{n - m}{\left(1 - \frac{m}{\kappa}\right)}, \quad (m \neq \kappa, n \text{ and no multiples of three, } \kappa \in \mathbb{N}^*). \quad (3)$$

Equation (3) is  $\kappa$ -subtraction, the inverse operation of  $\kappa$ -addition (1): physically, it means that some base(s) are deducted from the initial DNA or biological polymer. The  $\kappa$ -difference obeys  $n \oplus_{\kappa} \bar{m} = \bar{m} \oplus_{\kappa} n$  and  $n \oplus_{\kappa} (m \oplus_{\kappa} z) = (n \oplus_{\kappa} \bar{m}) \oplus_{\kappa} z$ , but  $\alpha(n \oplus_{\kappa} \bar{m}) \neq (\alpha n) \oplus_{\kappa} (\alpha \bar{m})$ . For both  $\kappa$ -addition and  $\kappa$ -subtraction to the initial base  $n$  of any bases  $m$  which are multiples of three, we obtain the KFB changes of the information length of the DNA or biological polymer being copied. Then the

KFB mutation just adds one or several reading frames into the sequence without altering the composition of those reading frames thereafter.

The meaning of the kappa index in the proposed KFB mutation can be understood through the subsystem (subbases) of correlations between  $n$  and  $m$  due to information exchanges within a given length of information (string length) in the DNA or biological polymer being copied. In fact, a simple relation exists between the Pearson correlation coefficient  $R$  and the kappa index  $\kappa$ :

$$R\kappa = d_\kappa / 2, \quad \text{where } \kappa \in \mathbb{R}^* \quad (4)$$

for subbases with  $d_\kappa$ -degrees of freedom ([71–73];[74], Chapter 5).

The largest value of kappa,  $\kappa \rightarrow \infty$ , corresponds to DNA or biological polymer characterized by the absence of any correlations between subbases of  $n, m \in \mathbb{N}$

$$n \oplus_\infty m = n + m, \quad (n, m \text{ no multiples of three}). \quad (5)$$

This condition is the frameshift mutation. The smallest possible kappa value,  $\kappa_{\min} = d_\kappa / 2$ , corresponds to the state furthest from the ordinary frameshift mutation of a DNA or biological polymer, and is characterized by the strongest correlation between  $n$  and  $m$  due to information exchange. The parameter  $\kappa$  provides a measure of the loss ( $\kappa > 0$ ) or gain ( $\kappa < 0$ ) of information after the mixing of the partial bases  $n$  and  $m$  to compose the combined final DNA or biological polymer.

### 3. Information Defects in the DNA or Biological Polymer

Following Ricard Solé et al. (2024) [3], the size of sequence space  $\Omega(n)$  associated with a biological polymer of length  $n \in \mathbb{N}$ , built from a molecular alphabet  $\Sigma = \{A_1, A_2, A_3, \dots, A_N\}$  of size  $|\Sigma| = N$  (where  $N \in \mathbb{N}$ ), is:

$$\Omega(n) = |\Sigma|^n = N^n. \quad (6)$$

The space of possible proteins with a length base  $n$  of 1000 amino acids is  $\Omega = 20^{1000}$  – a space so large that it could not be explored [27]. The space of possible molecular configurations within an organism is astronomically large. The probability  $(p)$  for such a such a protein to exist in nature is  $p(\Omega) = 20^{-1000}$ . We may assume that the biological system the is “non-ergodic” in the sense that not every part of the configuration space can be explored [28–30]. (In statistical physics, an ergodic system is one which, over sufficiently long time-scales, explores all possible microstates that are consistent with its macroscopic properties.) Mathematically, for any physically observable  $Q(q)$ , where  $q$  is the microstate that specifies all of the system’s coordinates ( $q = (q_1, q_2, q_3, \dots)$ ), the long-time average of  $Q(q)$  converges to the ensemble average, so

$$\langle O \rangle = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau O(q(t)) dt. \quad (7)$$

In intuitive terms, an ergodic system explores (over time) all the possible ways it can exist. Given enough time, the system will visit all its available configurations or states [3]. There are several major reasons to expect convergence [3]. Firstly, all evolutionary trajectories occur under certain generic selective constraints, such as the laws of mathematics, physics, and chemistry. These constraints should lead to universal features [31,32]. Secondly, many of these states are not explored randomly. Finally, evolutionary trajectories happen within a certain system: Darwinian evolution acts on populations of entities, which emerge within the system in assembly spaces [44,45, and 46].

The second law of thermodynamics states that a physical process must generate an overall increase  $\Delta S_{total}$  in the entropy of a system ( $\Delta S_{system}$ ) and its environment ( $\Delta S_{env}$ ):

$$\Delta S_{total} = \Delta S_{system} + \Delta S_{env}. \quad (8)$$

Therefore, living systems can only perform processes that reduce entropy internally if, at the same time, they produce an even greater increase in entropy in the environment. Thus, the universal thermodynamic logic of life is that low-entropy input ("resource") is turned into high-entropy output ("waste"), which can be expressed as:

$$\Delta S_{system} < 0, \quad \Delta S_{env} > 0, \quad |\Delta S_{env}| > |\Delta S_{system}|, \quad (9)$$

where

$$\Delta S_{env} = Q / (k_B T), \quad (10)$$

is the entropy production in the environment,  $Q$  is the amount of heat released,  $T$  is the thermal reservoir at temperature  $k_B = 1.38 \times 10^{-23} J / K$  Boltzmann's constant.

This suggests that all life forms face the problem of maintaining low internal entropy in a race against the second law of thermodynamics.

In non-equilibrium thermodynamics [3,75], the rate of entropy production per unit time  $\tau$  can be expressed as an integral:

$$\Delta S_{total} / \tau = \int_V \sigma dV, \quad (11)$$

for a system of volume  $V$ , with the specific entropy production rate,  $\sigma$ , written as follows:

$$\sigma = \sum_{k=1}^n X_k J_k. \quad (12)$$

In Equation (12),  $\{X_1, X_2, X_3, \dots, X_n\}$  indicates a set of "forces" (such as temperature gradients or chemical affinities) and  $\{J_1, J_2, J_3, \dots, J_n\}$  is the set of conjugate fluxes (e.g., heat flows, reaction rates). According to Shannon's formula [2,22-24], the evolution of the base sequences of modern organisms from an unspecified base sequence resulted in a gain of information (a decrease in entropy: negative  $\Delta S_{system}$ ), as follows:

$$\Delta S_{system}(n) = S_f(n) - S_i(n). \quad (13)$$

For the case of writing a specific string from a set of unordered letters ( $n = 0$ ), we have:

$$S_f(n) = \ln(N^n) = n \ln N = 0 \quad (\text{if and only if } n = 0). \quad (14)$$

Then

$$\Delta S_{system}(n) = -S_i(n) = -\ln(N^n) = -n \ln N, \quad (15)$$

$N = |\Sigma|$ , (where  $N \in \mathbb{N}$ ) is the number of unique elements, or letters, in the informational system and  $n$  is the length of information (the string length), that is being copied.

Let us now examine the possible evolutionary events resulting from the combination of the proposed KFB mutations and functional selection during biological evolution in the presence of information exchanges between the length bases. This examination involves the following steps: Firstly, we introduce the kappa-multiplication ( $\boxtimes_\kappa$ ) acting on the spaces of possible proteins  $\Omega(n)$  and  $\Omega(m)$  with a length base of  $n$  and  $m$ , respectively (Equation (6)), and resulting in the kappa-addition  $\oplus_\kappa$  of the bases  $n$  and  $m$  (in the similar way that the  $\kappa$ -addition  $\oplus_\kappa(n, m)$  results with the help of the usual addition  $+(n, m)$  in the Equation (1); namely:

$$\begin{aligned} \Omega(n) \boxtimes_\kappa \Omega(m) &\equiv \Omega(n) \times \Omega(m) \times \Omega(n)^{-m/\kappa} \\ &= \Omega(n+m) \times \Omega(n)^{-m/\kappa} = |\Sigma|^{n+m} \times (|\Sigma|^n)^{-m/\kappa} \\ &= |\Sigma|^{n+m-\frac{nm}{\kappa}} = |\Sigma|^{n \oplus_\kappa m} = \Omega(n \oplus_\kappa m) \quad (n, m \in (\mathbb{N}, \oplus_\kappa) \text{ and } \kappa \in \mathbb{N}^*). \end{aligned} \quad (16\alpha)$$

Equation (16 $\alpha$ ) shows the proposed KFB mutation space of the two possible proteins, with  $\Sigma$  being a molecular alphabet of size  $|\Sigma| = N$ , (where  $N \in \mathbb{N}$ ). For instance, the Equation (16 $\alpha$ ) is the standard morphism from  $(\mathbb{N}, \oplus_{\kappa})$  to  $(\mathbb{N}^*, \boxtimes_{\kappa})$ . The morphism (16 $\alpha$ ) is defined with the help of the morphisms:  $\Omega(n) \times \Omega(m) = \Omega(n+m)$  and  $\Omega(n)^{-m/\kappa}$  from  $(\mathbb{N}, +)$  to  $(\mathbb{N}^*, \times)$ .

We note that the standard morphism (16 $\alpha$ ) brings the standard morphism  $\Omega(n) \times \Omega(m) = \Omega(n+m)$  as a particular case  $\odot_{\infty} \equiv \times$  and  $\oplus_{\infty} \equiv +$  for  $(\kappa \rightarrow \infty)$ . Through reciprocity, the logarithm function is the standard morphisms morphism from  $(\mathbb{N}^*, \boxtimes_{\kappa})$  to  $(\mathbb{N}, \oplus_{\kappa})$ , namely

$$\begin{aligned} \log_{|\Sigma|}(\Omega(n) \boxtimes_{\kappa} \Omega(m)) &\equiv \log_{|\Sigma|}(\Omega(n) \times \Omega(m) \times \Omega(n)^{-m/\kappa}) \\ &= \log_{|\Sigma|}(\Omega(n+m) \times \Omega(n)^{-m/\kappa}) \\ &= \log_{|\Sigma|} \Omega(n+m) + \log_{|\Sigma|} \Omega(n)^{-m/\kappa} = \log_{|\Sigma|} \Omega(n) + \log_{|\Sigma|} \Omega(m) - \frac{m}{\kappa} \log_{|\Sigma|} \Omega(n) \\ &= \log_{|\Sigma|} \Omega(n) + \log_{|\Sigma|} \Omega(m) - \frac{\log_{|\Sigma|} \Omega(n) \log_{|\Sigma|} \Omega(m)}{\kappa} \\ &= \log_{|\Sigma|} \Omega(n) \oplus_{\kappa} \log_{|\Sigma|} \Omega(m) \end{aligned} \quad (16\beta)$$

which brings the standard morphism  $\log_{|\Sigma|}(\Omega(n) \times \Omega(m)) = \log_{|\Sigma|} \Omega(n) + \log_{|\Sigma|} \Omega(m)$  from  $(\mathbb{N}^*, \times)$  to  $(\mathbb{N}, +)$  as a particular case  $\odot_{\infty} \equiv \times$  and  $\oplus_{\infty} \equiv +$  for  $(\kappa \rightarrow \infty)$ . We impose the condition:

$$n \oplus_{\kappa} m = n + m - \frac{nm}{\kappa} \geq 0, \quad (n, m \text{ no multiples of three, and } \kappa \in \mathbb{N}^*). \quad (17)$$

This implies the following restriction to the values of the  $\kappa$ -parameter:

$$\frac{1}{\kappa} \leq \frac{n+m}{nm}, \quad \kappa \in \mathbb{N}^*. \quad (18)$$

Using Equation (4), we obtain the following restriction to the values of the correlation coefficient  $R$ :

$$R = \frac{d_{\kappa}}{2\kappa} \leq \frac{d_{\kappa}(n+m)}{2nm} \quad (19)$$

for subbases with  $d_{\kappa}$ -degrees of freedom ([71-73];[74],Chapter 5). The largest value of kappa,  $\kappa \rightarrow \infty$ , corresponds to DNA or biological polymer without any correlation between the subbases of information length  $n$  and  $m$  given by the Equation (5). The space of the two possible proteins (Equation (16 $\alpha$ )) then becomes:

$$\Omega(n) \otimes_{\infty} \Omega(m) = \Omega(n) \times \Omega(m) \times 1 = |\Sigma|^{n \oplus_{\infty} m} = |\Sigma|^{n+m}, \quad (n, m \in (\mathbb{N}, \oplus_{\kappa}), \kappa \in \mathbb{N}^*). \quad (20)$$

This is the frameshift mutation space of the two possible proteins. We may write Equation (16 $\alpha$ ) as follows:

$$\Omega(n) \otimes_{\kappa} \Omega(m) = \Omega(n) \times \Omega(m) \times \Omega(n, m)^{(correl)} = |\Sigma|^{n \oplus_{\kappa} m}, \quad (n, m \in (\mathbb{N}, \oplus_{\kappa}), \kappa \in \mathbb{N}^*). \quad (21)$$

In Equation (21), the correlation space of the two possible proteins,  $\Omega(n)$  and  $\Omega(m)$  (with a length base  $n$  and  $m$ , respectively) due to information exchanges is given by:

$$\Omega(n, m)^{(correl)} = \Omega(n)^{-m/\kappa} = |\Sigma|^{\frac{-nm}{\kappa}}, \quad (n, m \in (\mathbb{N}, \oplus_{\kappa}), \kappa \in \mathbb{N}^*). \quad (22)$$

The  $\kappa$ -product (16 $\alpha$ ) is commutative  $\Omega(n) \odot_{\kappa} \Omega(m) = \Omega(m) \odot_{\kappa} \Omega(n)$  and associative  $(\Omega(n) \odot_{\kappa} \Omega(m)) \odot_{\kappa} \Omega(z) = \Omega(n) \odot_{\kappa} (\Omega(m) \odot_{\kappa} \Omega(z))$ . The number one is the neutral element of the  $\kappa$ -product  $(\Omega(n) \odot_{\kappa} 1 = \Omega(n))$ , and this permits us to define the inverse multiplicative,  $1 \odot_{\kappa} \Omega(\bar{n})$ , by means of  $\Omega(n) \odot_{\kappa} (1 \odot_{\kappa} \Omega(\bar{n})) \equiv 1$ .

Similarly, the space of the two possible proteins,  $\Omega(n)$  and  $\Omega(m)$  in the case of the deletion of any number of bases  $m$  (except multiples of three) from the initial DNA sequence of base  $n$  is:

$$\Omega(n) \otimes \Omega(\bar{m}) = |\Sigma|^{(n-m)/1-m/\kappa} = |\Sigma|^{n \oplus_{\kappa} \bar{m}} \quad (n, m \in (\square, \oplus_{\kappa}), \kappa \in \square^*) \quad (23)$$

Using Equations (13)–(15), the kappa-gain of information (a decrease in kappa-entropy: negative  $\Delta S_{\kappa\text{-system}}$ ) of the system (DNA or biological polymer) during the proposed KFB mutations is given by:

$$\begin{aligned} \Delta S_{\kappa\text{-system}}(n, m) &= -\ln(|\Sigma|^{n \oplus_{\kappa} m}) = -(n \oplus_{\kappa} m) \ln(|\Sigma|) = -\left(n + m - \frac{nm}{\kappa}\right) \ln(|\Sigma|) \\ &= S(n) + S(m) + \Delta S_{\text{Def}}(n, m) \end{aligned} \quad (24)$$

$\Delta S_{\kappa\text{-system}}(n, m)$  is the entropy of a composite system (DNA or biological polymer) with correlations between  $n$  and  $m$ , where:

$$S(n) = -n \ln(|\Sigma|), \quad (25)$$

$$S(m) = -m \ln(|\Sigma|), \quad (26)$$

$$\Delta S_{\text{Def}}(n, m) = \frac{nm}{\kappa} \ln(|\Sigma|) \quad (27)$$

Equation (27) is the proposed entropy defect for the DNA or biological polymer.

The characteristic non-additive rule of the kappa entropy [1] was first derived on a physical basis through the concept of entropy defect [1]. Then, it was demonstrated that the entropy defect can lead to the generalized, kappa-associated entropy [1]. Here, we explain the origin of entropy defects based on the interdependence of information length systems (or interconnectedness of these systems: the condition whereby systems with a significant number of constituents (information length bases) dependent on, and/or are connected to, each other). The proposed KFB mutation quantifies the correlations among the information length of the bases (DNA or polymer constituents) due to information exchanges. Interdependence induces order and reduces the entropy-information in DNA or biological polymer systems thus producing entropy-information defects. Interdependence is measured through the magnitude  $1/\kappa$  of the entropy-information defect, defend as:

$$\frac{1}{\kappa} = \frac{1}{\Delta S_{\kappa\text{-system}}(n, m)} \frac{\partial \Delta S_{\text{Def}}(n, m)}{\partial \Delta S_{\infty\text{-system}}(n, m)}, \quad (28)$$

where  $\Delta S_{\kappa\text{-system}}(n, m)$  (see Equation (24)) is the entropy-information of the system with correlations among its constituents (e.g., information length bases), while

$$\Delta S_{\infty\text{-system}}(n, m) = -(n + m) \ln(|\Sigma|), \quad \text{for } (\Delta S_{\text{Def}}(n, m) \rightarrow 0 \text{ hence } \kappa \rightarrow \infty) \quad (29)$$

denotes the entropy-information of this system if there was no correlation among the information length bases. The smallest possible kappa value,  $\kappa = d_{\kappa} / 2$ , corresponds to the state furthest from the ordinary frameshift mutation of a DNA or biological polymer, characterized by the highest correlations between bases due to information exchange. The parameter  $\kappa$  provides a measure of the loss ( $\kappa > 0$ ) or gain ( $\kappa < 0$ ) of information after the mixing of the partial bases  $n$  and  $m$  to compose the combined final DNA or biological polymer.

#### 4. Properties of the Information Defect in DNA or Biological Polymers

The information defect determines how the information of the DNA or biological polymer partitions into the entropy information of its constituents. The participation of each constituent's entropy-information in the information defect must be (1) separable, (2) symmetric, and (3) (upper- or lower-) bounded for information loss ( $\kappa > 0$ ) or gain ( $\kappa < 0$ ). Below, we analyze the physical

meaning and mathematical formulation of the three properties that constitute the foundations of information defect:

(1) Separable. The entropy-information defect relates to the constituent's entropies (informations) through a separable function:

$$S_{Def}(n, m) \propto S_{Def}(n, m = const.) \times S_{Def}(n = const., m) \quad (30)$$

(for details see Ref [1])

(2) Symmetric. The information of the total system is symmetric to the  $n$  and  $m$  bases:  $S_{sys}(n, m) = S_{sys}(m, n)$ . The same holds for the information defect:

$$S_{Def}(n, m) = S_{Def}(m, n) \quad (31)$$

(for details see Ref [1])

(3) Bounded. This means that the entropic information has an upper ( $S \leq S_{max}$ ) or lower ( $S_{min} \leq S$ ) for information loss ( $\kappa > 0$ ) or gain ( $\kappa < 0$ ).

We note that the upper limit tends to infinity in the classical thermal equilibrium:  $S \leq S_{max} = \kappa \rightarrow \infty$  or  $\kappa \rightarrow -\infty = S_{min} \leq S$  for information loss ( $\kappa > 0$ ) or gain ( $\kappa < 0$ ) respectively. This indicates that, in the classical case, entropy changes can constantly increase or decrease tending to infinity, and eventually can attain any possible positive or negative values for information loss ( $\kappa > 0$ ) or gain ( $\kappa < 0$ ), respectively [1].

## 5. Kappa Metabolic Rate in the Presence of Information Defects

Landauer's bound [76] gives us the minimal energy required to perform an abstract computation and many string-writing operations, including the copying and transformation of DNA or biological polymer [77]. We may ask if the available energy is sufficient for the string copying and processing of a very small amount of stored information. The fundamental bounds of information can be connected to metabolism via Landauer's bound by calculating the minimal metabolic rate,  $W$ , needed to replicate genetic information for a given genome length and cellular growth rate [3,76,78]. In the presence of entropy defects, Landauer's bound is modified due to the KFB mutations, as follows:

$$Q_{\kappa} \geq k_B T (S_{\kappa-I} - S_{\kappa-F}), \quad (32)$$

where  $Q_{\kappa}$  is the kappa heat released to a bath at temperature  $T$ ,  $k_B$  is Boltzmann's constant,  $S_{\kappa-I}$  and  $S_{\kappa-F}$  are the initial and final kappa system entropy, respectively. This is nothing more than the expression of the second law of thermodynamics (Equations (13–15)), where  $\Delta S_{\kappa-system} = S_{\kappa-I} - S_{\kappa-F}$ .

To write a specific string from a set of unordered letters ( $n = 0$ ), we have  $S_{\kappa-F} = 0$  and

$$S_{\kappa-I}(n, m) = \ln(|\Sigma|^{n \oplus_{\kappa} m}) = S_I(n) + S_I(m) - \Delta S_{I-Def}(n, m) \quad (33)$$

In Equation (33), terms  $S_I(n)$ ,  $S_I(m)$  and  $\Delta S_{I-Def}(n, m)$  are calculated from Equations (25–27),  $|\Sigma| = N$ , the number of unique elements, or letters, in the informational system (Equation(6));  $n$  is the length of information (the string length) that is being copied; and  $m$  is the length of information (the string length) of the KFB mutation. Here,  $n \oplus_{\kappa} m$  is the kappa-composed length of information (the composed string length) after the action of the KFB mutation (Equation (1)). Using Equations (25–27) and (33), the kappa-modified Landauer's bound becomes:

$$Q_{\kappa} \geq Q - \Delta Q_{Def}, \quad (34)$$

where

$$Q = k_B T S_I(n, m) = k_B T (n + m) \ln(|\Sigma|) \quad (35)$$

is the heat released to a bath at temperature  $T$  independent of the kappa index, and

$$\Delta Q_{Def} = k_B T \Delta S_{I-Def}(n, m) = k_B T \frac{nm}{\kappa} (\ln |\Sigma|) \quad (36)$$

is the heat released to a bath at temperature  $T$  dependent of the kappa index and corresponding to the entropy-information defects of the DNA or biological polymer. The term  $nm/\kappa$  in Equation (36) is the correlation between the information lengths  $n$  and  $m$  due to information exchange after the action of the KFB mutation.

We can convert the kappa-modified Landauer's bound into a kappa-modified metabolic rate  $W_\kappa$ , due to the presence of entropy defects by considering how fast the information is copied. Given division time,  $t_{div}$ , of seconds, the kappa-modified minimal metabolic rate of copying is:

$$W_\kappa = W - \Delta W_{Def} = k_B T \frac{(n \oplus_\kappa m) \ln(|\Sigma|)}{t_{div}} = \frac{k_B T \ln(|\Sigma|)}{t_{div}} \left( n + m - \frac{nm}{\kappa} \right), \quad (37)$$

where

$$W = k_B T \frac{(n + m) \ln(|\Sigma|)}{t_{div}} \quad (38)$$

is the metabolic rate independent of the kappa index, and

$$\Delta W_{def} = k_B T \frac{nm \ln(|\Sigma|)}{\kappa t_{div}} \quad (39)$$

is the metabolic rate dependent of the kappa index and corresponding to the entropy-information defects of the DNA or a biological polymer. Again, in limit (29), Equations (34) and (37) become

$$Q_{\infty-system}(n, m) \geq k_B T (n + m) \ln(|\Sigma|), \text{ for } (\Delta Q_{Def}(n, m) \rightarrow 0 \text{ hence } \kappa \rightarrow \infty), \quad (40)$$

$$W_{\infty-system}(n, m) = k_B T \frac{(n + m) \ln(|\Sigma|)}{t_{div}}, \text{ for } (\Delta W_{Def}(n, m) \rightarrow 0 \text{ hence } \kappa \rightarrow \infty). \quad (41)$$

Equation (40) denotes Landauer's bound whereas Equation 41 shows the minimal metabolic rate for a biological system if there are absolutely no correlations among the information length bases.

## 6. Kappa Distribution of DNA Length Bases or Biological Polymers

Starting from the connection of probability distribution  $\{p\}$  with  $(\kappa \rightarrow \infty)$ -entropy, and following Gibb's path, we maximize the BG entropy  $S = -p \ln p$  under the constraint of normalization  $p(\Omega) = 1$ , and derive the Maxwell-Boltzmann (MB) probability distribution  $p(\Omega) = |\Sigma|^{-n}$ . In the latter,  $\Omega(n) = |\Sigma|^n$  and  $N = |\Sigma|$  is the number of letters in the molecular alphabet. These steps can be reversed: the maximization of entropy,  $S = f(p)$ , leads to the distribution,  $f'(p) = -n + const$ ; compared to the MB distribution, we find  $f'(p) = \ln(1/p) + const$ , leading to  $f(p) = -p \ln p$  (taking also into account that in the case of one single possibility, i.e.,  $p = 1$ , we have  $S = 0$ ) – that is, the BG entropy. From this we derive the connection of entropy with  $(\kappa \rightarrow \infty)$ -entropic addition rule: Starting from the BG entropy applied to the two DNA systems,  $n$  and  $m$ , and their composite system,  $n + m$ , and noting the respective distributions  $\{p\}$  and entropies  $S(n)$  and  $S(m)$ , we apply, the entropic equation,  $S = -p \ln p$ , and the property of statistical independence:

$$p(n + m) = p(n)p(m) \text{ or } \ln p(n + m) = \ln p(n) + \ln p(m). \quad (42)$$

The latter is deduced from the exponential distribution function:

$$p(n+m) = 4^{-(n+m)} = 4^{-n} \times 4^{-m} = p(n)p(m), \quad (43) \quad \text{that}$$

maximizes this entropy and the energy summation of the DNA length bases  $n+m$ . Then, we obtain

$$S(n+m) = -p(n) \ln p(n) - p(m) \ln p(m) = S_n(n) + S_m(m). \quad (44)$$

The reverse process starts from the addition rule and seeks the entropic function that obeys this rule, as follows: The addition of two thermodynamically equivalent systems (i.e., systems sharing the same thermodynamic parameters) produces a similar system whose only difference from its constituents is its size. This requires that there be no information exchange and long-range correlation among the DNA bases; otherwise, the order induced by these correlations would have affected the entropy rule. The absence of correlations means that the two systems,  $n$  and  $m$ , are independent – a characteristic of the MB exponential distribution function. The associated entropy of each of these systems is given by the BG formulation.

#### Connection of the probability distribution with ( $K < \infty$ )-entropy

Following Gibb's path, we maximize the kappa entropy

$$S_\kappa(\{p\}) = \ln_\kappa(p(\Omega)) = \kappa \left(1 - \sum p^{1+\frac{1}{\kappa}}(\Omega)\right) = \kappa \left(1 - \sum p_\kappa(\Omega)\right) \quad (45)$$

under the constraint of normalization  $p(\Omega) = 1$  and derive the kappa-probability distribution:

$$p_\kappa(\Omega) = \frac{|\Sigma|^{-n(1+\frac{1}{\kappa})}}{|\Sigma|^{-(1+\frac{1}{\kappa})}} = p^{1+\frac{1}{\kappa}}(\Omega), \quad (46)$$

where  $p(\Omega) = \frac{|\Sigma|^{-n}}{|\Sigma|^{-n}}$ ,  $\Omega(n) = \frac{|\Sigma|^n}{|\Sigma|^n}$ , and

$$\Omega_\kappa(n) = \frac{|\Sigma|^{n(1+\frac{1}{\kappa})}}{|\Sigma|^{(1+\frac{1}{\kappa})}} = \Omega^{1+\frac{1}{\kappa}}(n). \quad (47)$$

The expectation values are determined through the probability  $p^{1+\frac{1}{\kappa}}(n) / \sum p^{1+\frac{1}{\kappa}}(n)$ , called "escort" [4,48]. In Equation (45), given that  $S = 0$  for  $p = 1$ . This is the kappa-related entropy, also named after Havrda, Charvát, Daróczy and Tsallis [79–81]. In Equation (45), we use the  $\kappa$ -deformed exponential function and its inverse, the deformed logarithm function (Equations [47–50]), to get:

$$\exp_\kappa(x) \equiv \left(1 - \frac{x}{\kappa}\right)^{-\kappa}, \quad \ln_\kappa(x) \equiv \kappa \left(1 - x^{-\frac{1}{\kappa}}\right), \quad (48)$$

with the properties of the  $\kappa$ -logarithm and the  $\kappa$ -exponential in a compact form:

$$\exp_\kappa[\ln_\kappa(x)] = \ln_\kappa[\exp_\kappa(x)] = x. \quad (49\alpha)$$

$$\exp_\kappa(x \oplus_\kappa y) = \exp_\kappa(x) \exp_\kappa(y), \quad (49\beta)$$

$$\exp_\kappa(x) \otimes_\kappa \exp_\kappa(y) = \exp_\kappa(x+y), \quad (49\gamma)$$

$$\ln_\kappa(xy) = \ln_\kappa(x) \oplus_\kappa \ln_\kappa(y), \quad (49\delta)$$

$$\ln_\kappa(x \otimes_\kappa y) = \ln_\kappa(x) + \ln_\kappa(y). \quad (49\epsilon)$$

This leads us to the definition of the  $\kappa$ -product ( $\otimes_\kappa$ ) between two real numbers  $x, y$ :

$$x \otimes_\kappa y = \left[x^{1/\kappa} + y^{1/\kappa} - 1\right]_+^{1/\kappa}, \quad (x, y > 0). \quad (49\zeta)$$

The product (49 $\zeta$ ) is commutative ( $x \otimes_\kappa y = y \otimes_\kappa x$ ) and associative ( $(x \otimes_\kappa (y \otimes_\kappa z)) = (x \otimes_\kappa y) \otimes_\kappa z$ ), provided  $x \otimes_\kappa y$  and  $y \otimes_\kappa z$  differ from zero and from infinity. The number one is the neutral element of the  $\kappa$ -product (49 $\zeta$ ) ( $x \otimes_\kappa 1 = x$ ), and this permits us to define the inverse multiplicative ( $\oslash_\kappa$ ),  $1 \oslash_\kappa x$ , by means of  $x \otimes_\kappa (1 \oslash_\kappa x) \equiv 1$  [82].

Using the  $\kappa$ -logarithm and  $\kappa$ -exponential functions the two standard morphisms Equations (16 $\alpha$ - $\beta$ ) have split into four ones. Let us denote the definition set of the generalized exponential function  $\exp_\kappa(x)$  by  $D_\kappa$  [82]. From Equation (49 $\beta$ ), it comes that the generalized exponential function is

morphism from  $(D_\kappa, \oplus_\kappa)$  to  $(\square_{+}^*, +)$ . But from Equation (49 $\gamma$ ), we note that this function is also morphism from  $(D_\kappa, +)$  to  $(\square_{+}^*, \otimes_\kappa)$ . From Equation (49 $\delta$ ), it comes that the generalized logarithm function is morphism from  $(\square_{+}^*, \otimes_\kappa)$  to  $(D_\kappa, +)$ . But from Equation (49 $\epsilon$ ), we note that this function is also morphism from  $(\square_{+}^*, \times)$  to  $(D_\kappa, \otimes_\kappa)$  [82]. Here, we use the kappa index as a subscript to denote the deformed functions, but please note that the q-index could have been used as well (see also: Ref [73] Appendix A for more details). We also recall the known notation (49 $\alpha$ ) of deformed exponential/logarithm functions:

$$\Omega_\infty(n) = \exp_\kappa(S_\kappa(n)) = \exp(S_\infty(n)), \quad (50)$$

$$S_\kappa(n) = \ln_\kappa[\exp(S_\infty)] \Leftrightarrow S_\infty(n) = \ln[\exp_\kappa(S_\kappa(n))], \quad (51)$$

which constitutes the multiplicity  $\Omega_\infty(n)$ . This multiplicity (or statistical weight) is the space of possible molecular configurations of molecules within that biological system. We observe that the multiplicity  $\Omega_\infty(n) = \exp_\kappa(S_\kappa(n))$  becomes a thermodynamic quantity independent of the kappa index. In particular, given the probability distribution  $\{p\}$ , the statistical definition of this entropy, is formulated by:

$$S_\kappa(\{p\}) = \kappa \left( 1 - \sum p^{1+\frac{1}{\kappa}}(\Omega) \right) = \kappa \left( 1 - \sum p^q(\Omega) \right) / (q-1), \quad (52)$$

expressed in terms of the kappa  $\kappa$ -parameter (mostly used by the space-science community, or equivalently, the q-index (mostly used by the community of no-extensive statistical mechanics [74]):

$$q = 1 + \frac{1}{\kappa} \Leftrightarrow \kappa = \frac{1}{q-1}. \quad (53)$$

The BG entropy for  $\kappa \rightarrow \infty$  or  $q \rightarrow 1$  is:

$$S = -\sum p \ln p. \quad (54)$$

#### Connection of ( $K < \infty$ )-entropy with a generalized addition rule

Starting from Equation (45), we calculate the information function

$$1 - \frac{S(n)}{\kappa'} = \sum p(n)^{1+\frac{1}{\kappa'}}, \quad (55)$$

where

$$\kappa' = \kappa \ln |\Sigma|, \quad (56)$$

is the kappa index transformation  $\kappa \rightarrow \tilde{\kappa}$ , and  $|\Sigma| = N$  is the number of letters in the molecular alphabet. Applied again to the two independent systems of the bases  $n$  and  $m$  and their kappa-composite system, base  $n \oplus_\kappa m$ , we find:

$$\sum (p(n \oplus_\kappa m))^{1+\frac{1}{\kappa'}} = \sum (p(n))^{1+\frac{1}{\kappa'}} \sum (p(m))^{1+\frac{1}{\kappa'}}, \quad (57)$$

or

$$1 - \frac{S(n) \oplus_\kappa S(m)}{\kappa'} = \left( 1 - \frac{S(n)}{\kappa'} \right) \left( 1 - \frac{S(m)}{\kappa'} \right). \quad (58)$$

Equations (57) and (58) lead to the standard form of entropy-information defects for a DNA or biological polymer, as in Ref [1]:

$$S(n) \oplus_\kappa S(m) = S(n) + S(m) - \Delta S_{Def}(n, m), \quad (59)$$

where

$$S(n) = -n \ln(|\Sigma|), S(m) = -m \ln(|\Sigma|), \quad (60)$$

and

$$\Delta S_{Def}(n, m) = \frac{S(n)S(m)}{\tilde{\kappa}} = \frac{nm}{\tilde{\kappa}} \ln(|\Sigma|)^2 \quad (61)$$

Then, after the kappa-index transformation  $\kappa \rightarrow \tilde{\kappa}$ :

$$\tilde{\kappa} = \kappa \ln(|\Sigma|) \Leftrightarrow \kappa = \frac{\tilde{\kappa}}{\ln(|\Sigma|)}, \quad (62)$$

in Equation (61), the entropy defects in the standard form (Refs [1]) can take the form of the proposed entropy defects (Equation (27)). Similarly, by applying the kappa-index transformation  $\tilde{\kappa} \rightarrow \kappa$  in Equation (27), the entropy defects take the standard form (Equation (61)), as in Refs [1]. The reverse process can be found by starting from this generalized addition rule and seeking the entropic function that obeys this rule; this has been already shown in several earlier publications (e.g. [1]).

## 7. Discussion

Following Ref [3], the required metabolic rate for information copying alone is calculated from Equation (6), with the key parameters being the number of letters in the molecular alphabet  $N = |\Sigma|$ , the length of the composite genome ( $n \oplus_{\kappa} m$ ) after the kappa frameshift mutation, (with  $n$  and  $m$  being the length of the genome and the mutation length, respectively), and the time to copy the information ( $t_{dev}$ ). For  $N = |\Sigma| = 2^2 = 4$  letters, at a typical division time  $t_{div} = 10^3 s$  and a Typical Genome Length (TGL) of  $n \approx 5 \times 10^6 bases$ , Equation (6) gives an ordinary Typical Bacterial Metabolic Rate (TBM) of  $W_{\infty}^{(TBM)} \approx 10^{-12} J/s$  (Ref[82]). Now, using Equation (37), the kappa-modified TBM  $W_{\kappa}^{(TBM)}$  is calculated in terms of the ordinary TBM and KFB-mutation rate  $W_{KFB}$  as follows:

$$W_{\kappa}^{(TBM)} = W_{\infty}^{(TBM)} \left(1 - \frac{m}{\kappa}\right) + W_{KFB}, \quad (63)$$

where

$$W_{KFB} = k_B T \frac{m}{t_{dev}} \ln(|\Sigma|) \quad (64)$$

The smallest possible kappa value,  $\kappa_{min} = d_{\kappa} / 2$ , ( $d_{\kappa}$ : degrees of freedom ([71,72];[74], Chapter 5) corresponds to the state furthest from the ordinary frameshift mutation of a DNA or biological polymer, characterized by the highest correlation between bases. Using the  $\kappa_{min}$  value, Equation (63) becomes:

$$W_{\kappa}^{(TBM)} = W_{\infty}^{(TBM)} \left(1 - \frac{2m}{d_{\kappa}}\right) + W_{KFB} \quad (65)$$

In the MB probability distribution  $p_{\infty}(\Omega) = |\Sigma|^{-n}$  for a DNA or biological polymer, entropy is zero  $S_{\infty}(\Omega) = 0$  only when a specific information string is written from a set of unordered letters, ( $n = 0$ ). The probability value for such an unordered information string is  $p_{\infty}(\Omega) = 1$ . Furthermore, in the MB statistical mechanics for a specific information string from a set of ordered letters ( $n > 0$ ), there is no preference for a specific sequence: if  $\Omega_{\infty}(n) = 4^n$ , different DNA sequences are equally possible and the actual DNA sequence has only a probability  $p_{\infty}(\Omega) = 4^{-n}$  (where  $n$  is the number

of bases (ordered letters) in the DNA sequence, ranging from one million in bacteria to billions in many animals and plants). For example, the space of possible proteins  $\Omega_{\infty}(n)$  with a base length  $n$  of 1000 amino acids is  $\Omega(1000) = 20^{1000}$  – a space so large that it could never be explored in our universe [27]. The space of possible molecular configurations within an organism is astronomically larger. The probability  $P$  of such a protein (ordered information string) to exist in nature is  $p_{\infty}(\Omega) = 20^{-1000}$ . The requirements of longer time and a larger number of individuals are even more conducive to the observation of deterministic laws associated with the base length  $n$ . The evolutionary fitness of organisms relates to the progressive reduction of the entropy production rate. This, and the Boltzmann's ergodic principle, allow us to consider organisms as evolving chemical entities (Sabater 2006, 2009[2,4]). Accordingly, the time-average state of a single individual over a long period (evolutionary time) coincides with the average state of a sufficiently large number of individuals for shorter time periods [5]. We note that Sabater (2009[5]) worked with an alternative theory with biological time arrows, showing that deterministic behavior of a biological system can arise.

Let us now examine the evolutionary events that could result from the combination of functional selection during biological evolution and information exchanges between the base lengths of biological polymers, represented by the kappa parameter. For DNA or biological polymer with the kappa-probability distribution (Equation (46)) derived from the maximized kappa entropy (Equation (45)) under the constraint of normalized  $p_{\kappa}(n) = 1$ , we observe that the different  $n$  and  $\kappa$  scales in Equations (45) and (46) are manifested in the double evolutionary emergence of the proposed biological polymers through correlations between its bases. For specific values of the kappa parameter, deterministic laws associated with a single biological polymer can be expected for short timescales. Thus, for the kappa-probability distribution, we have the following short-timescale possibilities for an individual:

$$p_{\kappa}(\Omega) = p^{1+\frac{1}{\kappa}}(\Omega) = \begin{cases} 1 & , \kappa = -1, \forall n \in \mathbb{N}^* \\ |0 \leq |\Sigma|^{-2n} < 1 & , \kappa = 1, \forall n \in \mathbb{N}^* \end{cases} \quad (66)$$

Equation (66) measures the loss (for  $\kappa = 1$ ) or gain (for  $\kappa = -1$ ) of information. In Equation (66), for  $p(\Omega) = |\Sigma|^{-n} \neq 1$  and non-zero integer (ordered letters  $n > 0$ ), it is  $p_{\kappa=-1}(\Omega) = 1$ . Therefore, for kappa-entropy, we find the following short-timescale possibilities for a single biological polymer:

$$S_{\kappa}(\Omega) = \kappa \left( 1 - p^{1+\frac{1}{\kappa}}(\Omega) \right) = \begin{cases} 0 & , \kappa = -1, \forall n \in \mathbb{N}^* \\ 0 < 1 - p^2(\Omega) \leq 1 & , \kappa = 1, \forall n \in \mathbb{N}^* \end{cases} \quad (67)$$

In Equation (67), for  $p(\Omega) = |\Sigma|^{-n} \neq 1$ ,  $p_{\kappa=-1}(\Omega) = 1$ , and non-zero integer (ordered letters  $n > 0$ ), it is  $S_{\kappa=-1}(\Omega) = 0$ .

Let us consider the example of the state characterized by  $\kappa = -1$ , that is, the strongest correlation between bases due to information exchange. The space of possible final proteins  $\Omega_{\kappa=-1}^{(F)}(n)$  with a length base  $n$  of 1000 amino acids is:

$$\Omega_{\kappa=-1}^{(F)}(1000) = \Omega_{\kappa=\infty}^{(I)}(1000) p_{\kappa=\infty}^{(I)}(1000) = p_{\kappa=\infty}^{(I)}(1000) \Omega_{\kappa=\infty}^{(I)}(1000) = p_{\kappa=-1}^{(F)}(1000) = 1 \quad (68)$$

The final protein entropy  $S_{\kappa=-1}^{(F)}(1000) = 0 \text{ bit}$ . From this, we obtain entropy changes (gain of information) of the following value:

$$\Delta S_{\kappa\text{-system}}(1000) = -S_{\kappa=\infty}^{(I)}(1000) = -1301,029 \text{ bit} \quad (69)$$

with  $\Delta n = n_F - n_I = 0$  for equal final and initial base length of amino acids ( $n_F = n_I = 1000$ ). We call these proposed changes *iso-order*  $n$  changes. These changes are characterized by the parameters:  $\{(n_I, \kappa), (n_F, \kappa)\} = \{(1000, -\infty), (1000, -1)\}$ . (70)

The space of possible final proteins  $\Omega_{\kappa=-1}^{(F)}(n)$  with a base length  $n$  of 1000 amino acids is 1. An initial space of a magnitude as large as  $\Omega_{\kappa=\infty}^{(I)}(1000) \approx 20^{1000}$  has thus now collapsed to the one final possibility:  $\Omega_{\kappa=-1}^{(F)}(1000) = 1$ . The probability  $P$  for such a protein (ordered information string) final state to exist in nature is now one  $p_{\kappa=-1}^{(F)}(1000) = 1$ .

Let us now consider another example for ( $\kappa = -1$ ) state characterized by the highest correlations between bases due to information exchanges. By writing a specific initial information string from a set of unordered letters for kappa values ( $n = 0, \forall \kappa \in \square^* \setminus \{-1\}$ ) we have:

$$S_{\kappa=-1}(n) = n \left( 1 + \frac{1}{\kappa} \right) \ln(20) = 0 \text{ bit}, \quad (\text{if and only if, } n = 0, \forall \kappa \in \square^* \setminus \{-1\}). \quad (71)$$

We then obtain the entropy changes (information gain):

$$\Delta S_{\text{system}}(n) = S_{\kappa=-1}^{(F)}(1000) - S_{\kappa}^{(I)}(0) = 0 \text{ bit}, \quad (72)$$

where  $S_{\kappa=-1}^{(F)}(1000) = 0 \text{ bit}$  is the final entropy of the “mathematical” protein Equation (72) is the *iso-entropic* process for a given “mathematical” protein, and is adiabatic and reversible [83–86], with equal initial and final probability (that is,  $p_{\kappa}^{(I)}(0) = p_{\kappa=-1}^{(F)}(1000) = 1$ ).

Furthermore, we have the order  $n$  changes:  $\Delta n = n_F - n_I = n_F = 1000$ . It seems that order  $n_F$  emerges spontaneously from the unordered initial information string. This emergence of order  $n_F$  may relate to the quantum self-organization process (see Refs [33,87–90] for details).

Using Equations (37) and (41) the *iso-entropic* changes for the proposed “mathematical” protein yield to the zero kappa-metabolic rate. This implies that, at this optimum state ( $\kappa = -1$ ), to protect itself from damage it cannot endure, the “mathematical” protein must not waste energy (produce entropy) in reactions. We call this state of the “mathematical” protein the kappa-functional optimum state. This state is given by Equation (72). In nature, it may be possible for a real protein to approximate the kappa-functional optimum state with increasing correlation between its parts:

$$\{(n_I, \kappa), (n_F, \kappa)\} \xrightarrow{\text{approx}} \{(0, \kappa), (1000, -1)\}. \quad (73)$$

Due to the strongest correlation between its parts, such a protein would need to waste much less energy (produce less entropy) in reactions and would be less susceptible to damage than the classical information theory predicts.

## 8. Conclusion

Following Livadiotis G. and McComas D. J., (2023) [1], we propose a new type of DNA frameshift mutations: Kappa-Frameshift Background (KFB) mutations. KFB mutations occur spontaneously due to information exchange between the DNA sequence of length base ( $n$ ) and the mutation sequence of length base ( $m$ ), and respect the kappa-addition symbol  $\oplus_{\kappa}$ . KFB mutations occur when any number of  $m$  bases are added or deleted to the initial DNA sequence of  $n$  bases. This repeats the kappa-addition symbol  $\oplus_{\kappa}$ , except in the case of multiples of three, which reestablish the initial reading frame. In the proposed KFB mutation, the kappa index can be understood through the

subsystem (subbase) of information length correlation between  $n$  and  $m$  due to information exchange within the given length of DNA or biological polymer (the string length) that is being copied.

We find that entropy defects originate in the interdependence of information length systems (or their interconnectedness: the state at which systems with a significant number of constituents (information length bases) depend on, or are connected with, each other) by the proposed KFB mutations. In general, interdependence induces order in the systems involved (DNA or biological polymers) and reduces their entropy-information, thus producing entropy-information defect.

In the presence of entropy defects, the Landauer's bound and minimal metabolic rate for a biological system is modified. We observe that that the different  $n$  and  $\kappa$  scales in Equations (45) and (46) are manifested in the double evolutionary emergence of proposed biological polymers due to subsystems correlations. In this process, for specific values of the kappa parameter, we can expect deterministic laws associated with a single biological polymer for short timescales. We find that the possible evolutionary events resulting from the combined effects of functional selection during biological evolution and information exchange between the length bases of the biological polymer are represented by the kappa parameter. We also identify *iso-order*  $n$  changes characterized by the parameters (70). Using a proposed "mathematical" protein, we find that the *iso-entropic* changes this protein yield to the zero kappa-metabolic rate. This implies that, at the optimum state ( $\kappa = -1$ ), to protect itself from damage it cannot endure, the "mathematical" protein must not waste energy (produce entropy) in reactions. We call this state of the "mathematical" protein the kappa-functional optimum state (Equation (72)). In nature, with increasing correlation between its parts, a real protein may approximate the kappa-functional optimum state. Due to the strongest correlation between its parts, such a protein would waste much less energy (produce less entropy) in reactions, and would be less susceptible to damage, than the classical information theory predicts.

**Data Availability Statement:** Data will be made available on reasonable request. [Author's comment: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.]

**Code Availability Statement:** This manuscript has no associated code/software. [Author's comment: Code/Software sharing not applicable to this article as no code/software was generated or analyzed during the current study.]

## Declarations

**Competing Interests:** The authors declare no competing interests.

## References

1. George Livadiotis and David J. McComas 2023 EPL 144 21001.
2. B.Sabater, Entropy Perspectives of Molecular and Evolutionary Biology. Int. J. Mol. Sci. (2022), **23**, 4098. <https://doi.org/10.3390/ijms23084098>.
3. Solé, R. Kempes, C.P. Corominas-Murtra, B. De Domenico, M. Kolchinsky, A. Lachmann, M. Libby, E. Saavedra, S. Smith, E. Wolpert, D. Fundamental Constraints to the Logic of Living Systems. Interface Focus (2024), **14**, doi:10.1098/rsfs.2024.0010.
4. B.Sabater, Are organisms committed to lower their rates of entropy production? Possible relevance to evolution of the Prigogine theorem and the ergodic hypothesis. Biosystems (2006) **83**: 10–17.
5. B.Sabater, Time arrows and determinism in Biology, Biological Theory **4**(2) (2009), 174–182.
6. Martyushev, L.M. Entropy and entropy production: Old Misconceptions and new breakthroughs. Entropy (2013), **15**, 1152–1170.
7. Baez, J.C. Blake, S.; Pollard, B.S. Relative Entropy in Biological Systems. Entropy (2016), **18**, 46.
8. Roach, T.N.F. Use and abuse of entropy in Biology: A case for Caliber. Entropy (2020), **22**, 1335.

9. Demetrius, L. Directionality principles in thermodynamics and evolution. *Proc. Natl. Acad. Sci. USA* (1997), 94, 3491–3498.
10. Skene, K.R. Life's a gas: A thermodynamic theory of biological evolution. *Entropy* (2015), 17, 5522–5548.
11. Skene, K.R. Thermodynamics, ecology and evolutionary biology: A bridge over troubled water or common ground? *Acta Oecol.* (2017), 85, 116–125.
12. Doyle, S.R.; Carusela, F. Guala, S. Momo, F. A null model for testing thermodynamic optimization in ecological systems. (2019), arXiv:1111.2019v1.
13. Dewar, R.C. Maximum entropy production and plant optimization theories. *Phil. Trans. R. Soc. B* (2010), 365, 1429–1435.
14. Kondepudi, D.K. De Bari, B. Dixon, J.A. Dissipative structures, organisms and evolution. *Entropy* (2020), 22, 1305.
15. Dobovišek, A. Županovič, P. Brumen, M. Bonačič-Lošič, Z. Kuič, D. Juretic, D. Enzyme kinetics and the maximum entropy production principle. *Biophys. Chem.* (2011), 154, 49–55.
16. Gaiseanu, F. What is life: An informational model of the living structures. *Biochem. Mol. Biol.* (2020), 5, 18–28.
17. Pulselli, R.M. Simoncini, E. Tiezzi, E. Self-organization in dissipative structures: A thermodynamic theory for the emergence of prebiotic cells and their epigenetic evolution. *Biosystems* (2009), 96, 237–241.
18. Annala, A. Kuismanen, E. Natural hierarchy emerges from energy dispersal. *Biosystems* (2009), 95, 227–233.
19. Hui, D. Liao-Fu, L. Lin Hao, L. Entropy production rate changes in lysogeny/lysis switch regulation of bacteriophage Lambda. *Commun. Theor. Phys.* (2011), 55, 371.
20. Trevors, J.T. Saier, M.H., Jr. Thermodynamic perspectives on genetic instructions, the laws of biology and diseased states. *Comptes Rendus Biol.* (2011), 334, 1–5.
21. Tretiakov, K.V. Szleifer, I. Grzybowski, B.A. The rate of energy dissipation determines probabilities of non-equilibrium assemblies. *Angew. Chem. Int. Ed. Engl.* (2013), 52, 10304–10308.
22. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Technol. J.* (1948), 27, 623–656.
23. Adami, C. Ofria, C. Collier, T.C. Evolution of biological complexity. *Proc. Natl. Acad. Sci. USA* (2000), 97, 4463–4468.
24. Vingaa, S. Almeida, J.S. Rényi continuous entropy of DNA sequences. *J. Theor. Biol.* (2004), 231, 377–388.
25. Demetrius, L. Thermodynamics and evolution. *J. Theor. Biol.* (2000), 206, 1–16.
26. Rakoczy, R. Kordas, M. Story, G. Konopacki, M. The characterization of the residence time distribution in a magnetic mixer by means of the information entropy. *Chem. Engin. Sci.* (2014), 105, 191–197.
27. Dryden, D.T. Thomson, A.R. White, J.H. How much of protein sequence space has been explored by life on Earth? *Journal of The Royal Society Interface* (2008), 5, 953–956.
28. Kauffman. *Investigations*; Oxford University Press, (2000).
29. Kauffman, S.A.; Roli, A. A third transition in science? *Interface Focus* (2023), 13, 20220063.
30. Kauffman, S.A. Prolegomenon to patterns in evolution. *Biosystems* (2014), 123, 3–8.
31. Kempes, C.P. Koehl, M. West, G.B. The scales that limit: the physical boundaries of evolution. *Frontiers in Ecology and Evolution* (2019), 7, 242.
32. Kempes, C.P. Krakauer, D.C. The multiple paths to multiple life. *Journal of molecular evolution* (2021), 89, 415–426.
33. Koorambas, E. Particle in a Markov Cube by the Non-Classical Information Entropy Space. *Int J Theor Phys* 63, 81 (2024). <https://doi.org/10.1007/s10773-024-05609-8>.
34. Aguirre, J. Catalán, P. Cuesta, J.A. Manrubia, S. On the networked architecture of genotype spaces and its critical effects on molecular evolution. *Open biology* (2018), 8, 180069.
35. Manrubia, S. The simple emergence of complex molecular function. *Philosophical Transactions of the Royal Society A* (2022), 380, 20200422.
36. Alberch, P. From genes to phenotype: dynamical systems and evolvability. *Genetica* (1991), 84, 5–11.
37. Gavrillets, S. Gravner, J. Percolation on the fitness hypercube and the evolution of reproductive isolation. *Journal of theoretical biology* (1997), 184, 51–64.
38. Fontana, W. Modelling 'evo-devo' with RNA. *BioEssays* (2002), 24, 1164–1177.

39. Schultes, E.A. Bartel, D.P. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* (2000), 289, 448–452.
40. Li, H. Helling, R.Tang, C. Wingreen, N. Emergence of preferred structures in a simple model of protein folding. *Science* (1996), 273, 666–669.
41. Li, H. Tang, C. Wingreen, N. Are protein folds atypical? *Proc. Nat. Acad. Sci. USA* (1998), 95, 4987–4990.
42. Denton, M. The Protein Folds as Platonic Forms: New Support for the Pre-Darwinian Conception of Evolution by Natural Law. *J. Theor. Biol.* (2002), 219, 325–342.
43. Banavar, J.R. Maritan, A. Colloquium: Geometrical approach to protein folding: a tube picture. *Reviews of Modern Physics* (2003), 75, 23.
44. Marshall, S.M. Mathis, C. Carrick, E. Keenan, G. Cooper, G.J. Graham, H. Craven, M. Gromski, P.S. Moore, D.G. Walker, S.I. others. Identifying molecules as biosignatures with assembly theory and mass spectrometry. *Nature communications* (2021), 12, 3033
45. Ahnert, S.E. Johnston, I.G. Fink, T.M. Doye, J.P. Louis, A.A. Self-assembly, modularity, and physical complexity. *Physical Review E* (2010), 82, 026117.
46. Sharma, A. Czégel, D. Lachmann, M.; Kempes, C.P. Walker, S.I. Cronin, L. Assembly theory explains and quantifies selection and evolution. *Nature* (2023), 622, 321–328.
47. N.V. Bhagavan, *Medical Biochemistry, United States* (2002), ISBN978-0-12-095440-7, New York (2002), Elsevier Inc.
48. Chang-Hui Shen, *Diagnostic Molecular Biology, United States* (2023), ISBN 978-0-323-91788-9 New York (2023) Elsevier Inc. doi: <https://doi.org/10.1016/C2020-0-01054-6>.
49. Livadiotis G. and McComas D. J., *Entropy*, 23 (2021) 1683.
50. Livadiotis G., *Entropy*, 16 (2014) 4290.
51. Beck, C. & Cohen, E. G. D. Superstatistics. *Phys. A* 322, 267–275 (2003).
52. Schwadron, N. A. et al. Superposition of stochastic processes and the resulting particle distributions. *Astrophys. J.* 713, 1386–1392 (2010).
53. Hanel, R., Turner, S. & Gell-Mann, M. Generalized entropies and the transformation group of superstatistics. *Proc. Natl. Acad. Sci. U.S.A.* (2012) 108, 6.
54. Livadiotis, G., Assas, L., Dennis, B., Elaydi, S. & Kwessi, E. Kappa function as a unifying framework for discrete population modelling. *Nat. Res. Mod.* 29, 130–144 (2016).
55. Livadiotis, G. Rankine-Hugoniot shock conditions for space and astrophysical plasmas described by Kappa distributions. *Astrophys. J.* 886, 3 (2019).
56. Gravanis, E., Akylas, E. & Livadiotis, G. Physical meaning of temperature in superstatistics. *Europhys. Lett.* 130, 30005 (2020).
57. Zank, G. P. et al. Particle acceleration at perpendicular shock waves: Model and observations. *J. Geophys. Res.* 111, A06108 (2006).
58. Bian, N., Emslie, G. A., Stackhouse, D. J. & Kontar, E. P. Te formation of a kappa-distribution accelerated electron populations in solar fares. *Astrophys. J.* 796, 142 (2014).
59. Yoon, P. H. Electron kappa distribution and steady-state Langmuir turbulence. *Plasma Phys.* 19, 052301 (2012).
60. Yoon, P. H. Electron kappa distribution and quasi-thermal noise. *J. Geophys. Res.* 119, 7074 (2014).
61. Livadiotis, G. & McComas, D. J. Exploring transitions of space plasmas out of equilibrium. *Astrophys. J.* 714, 971–987 (2010).
62. Livadiotis, G. & McComas, D. J. Influence of pickup ions on space plasma distributions. *Astrophys. J.* 738, 64 (2011).
63. Livadiotis, G., McComas D. J., Transport equation of kappa distributions in the heliosphere. *Astrophys. J.*, to appear (2023).
64. Fisk, L. A. & Gloeckler, G. The case for a common spectrum of particles accelerated in the heliosphere: Observations and theory. *J. Geophys. Res.* 119, 8733–8749 (2014).
65. Peterson, J., Dixit, P. D. & Dill, K. A. A maximum entropy framework for non-exponential distributions. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20380–20385 (2013).

66. Livadiotis, G., Desai, M. I. & Wilson, L. B. III. Generation of kappa distributions in solar wind at 1 AU. *Astrophys. J.* 853, 142 (2018).
67. Livadiotis, G. On the origin of polytropic behavior in space and astrophysical plasmas. *Astrophys. J.* 874, 10 (2019).390–6394 (2011).
68. Boltzmann, L. Über die Mechanische Bedeutung des Zweiten Hauptsatzes der Wärmetheorie. *Wiener Berichte* 53, 195–220 (1866).
69. Gibbs, J. W. *Elementary principles in statistical mechanics* (Scribner's sons, 1902).
70. Maxwell, J. C. Illustrations of the dynamical theory of gases, on the motions and collisions of perfectly elastic spheres. *Philos. Mag.* 19, 19–32 (1860).
71. Livadiotis, G.; McComas, D.J. Invariant kappa distribution in space plasmas out of equilibrium. *Astrophys. J.* (2011), 741, 88.
72. Livadiotis, G. Kappa and q indices: Dependence on the degrees of freedom. *Entropy* (2015), 17, 2062–2081.
73. Livadiotis, G. *Kappa Distribution: Theory & Applications in Plasmas*, 1st ed.; Elsevier: Amsterdam, The Netherlands; London, UK; Atlanta, GA, USA, (2017).
74. Livadiotis, G.; McComas, D.J. Understanding kappa distributions: A toolbox for space science and astrophysics. *Space Sci. Rev.* (2013), 75, 183–214.
75. Glansdorff, P.; Prigogine, I. On a general evolution criterion in macroscopic physics. *Physica* (1964), 30, 351–374.
76. Landauer, R. Irreversibility and heat generation in the computing process. *IBM journal of research and development* (1961), 5, 183–191.
77. Kolchinsky, A. Corominas-Murtra, B. Decomposing information into copying versus transformation. *Journal of the Royal Society Interface* (2020), 17, 20190623.
78. Kempes, C.P. Wolpert, D. Cohen, Z Pérez-Mercader, J. The thermodynamic efficiency of computations made in cells across the range of life. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* (2017), 375, 20160343.
79. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* 52, 479–487 (1988).
80. Havrda, J. & Charvát, F. Quantification method of classification processes concept of structural  $\alpha$ -entropy. *Kybernetika* 3, 30–35 (1967).
81. Daróczy, Z. Generalized information functions. *Inf. Control* 16, 36–51 (1970).
82. L. Nivanen, A. Le Méhauté, Q.A. Wang, *Reports on Mathematical Physics*, 52, 437 (2003).
83. Sears, W., Salinger L. (1986), *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics*, 3rd edition, Addison-Wesley.
84. Sandler, S. I. *An Introduction Applied to Statistical Thermodynamics*; John Wiley & Sons, NJ, (2011).
85. McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, CA, (2000).
86. Lucia, U. (2008). "Probability, ergodicity, irreversibility and dynamical systems". *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 464: 1089–1104.
87. Popovic M., Entropy change of open thermodynamic systems in self-organizing processes, *Thermal Science*, vol 18, Nr 4, pp 1425-1432, (2014).
88. P.W. Anderson, (1972), More is Different, *SCIENCE*, 177, 4047.
89. Bohm, D, A Suggested Interpretation of the Quantum Theory in Terms of "Hidden Variables" I. *Phys. Rev.* 85(2), 166–179 (1952).
90. Bohm, D, A Suggested Interpretation of the Quantum Theory in Terms of "Hidden Variables", II. *Phys. Rev.* 85(2), 180–193 (1952).

**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.