

Revisiting Crick's Dogma and the Impossibility of Reverse Translation

Jan Charles Biro^{1,2*}

¹Department of Biotechnology, Karolinska Institute, Stockholm, Sweden

²Homulus Foundation, USA

Abstract

The origin, validity and role of the so-called "Central Dogma" of Molecular Biology are reviewed. This survey leads to the conclusion that after the pioneering days of the new discipline, during which the Dogma provided a preliminary orientation in the newly discovered world of nucleic acid and protein sequences, its value probably expired. It a function for over 50 years but its subsequent rise to almost unchallengeable status is no longer compatible with the fundamental spirit of any natural science. Therefore I suggest that Crick's Dogma be reconsidered and reallocated

Background

Crick [1] (1916-2004) formulated his Dogma in October 1958, five years after he had proposed the structure of dsDNA in 1953. The correctness of his dsDNA structure was not yet universally accepted and many colleagues expressed concerns about how the helical DNA unfolds and how the inverted bases become accessible. However, these concerns were not taken too seriously. The consensus was that the key to the "Secret of Life" had been found and it would open all the remaining doors. Crick had obviously founded a new scientific field.

Nucleic acids were first identified in 1868 by Miescher [2] (1844-1895) as the fourth broad category of principal substances distinctive of living organisms (fats, sugars, proteins and nucleic acids). Until the early 1950s, almost all biologists believed that the hereditary message, the gene, consisted of protein. The elementary chemistry of nucleic acids (the sugar + phosphate + base composition) had been established quickly and DNA was thought to be constructed in the simplest way imaginable, with the nucleotides following one another in fixed order in repeated sets of four. This extremely elementary picture was called the *tetranucleotide hypothesis* and was propounded by Levene and La Forge [3] (born Fishel Aaronovich Levin, 1869 - 1940), an organic chemist with an excellent reputation.

The belief that DNA could only be some sort of structural stiffening, since the genetic material must be protein, was held with dogmatic tenacity. Levene's tetranucleotide hypothesis, formulated around 1910, required that DNA consist of equal amounts of adenine, guanine, cytosine, and thymine. Before the later work of Chargaff et al. [4], it was widely speculated that "tetranucleotides" were organized in DNA molecules in a way that could not carry genetic information.

Rigorous proof that the gene is DNA and not protein appeared in 1944, when Avery et al. [5] (1877-1955) discovered that inheritable transformations occur when a strain of bacteria was mixed with DNA extracted from a different strain.

The first observations connecting nucleic acids to protein synthesis were made in 1939, when Caspersson [6] (1910-1997), a Swedish cytologist and geneticist, published jointly with Brachet (1909-1998) the finding that cells making proteins are rich in ribonucleic acids, RNA, implying that RNA is required to make proteins. A remarkable structural connection between nucleic acids and proteins was also known from Astbury's early X-ray studies, which showed that the spacing between nucleotides along the DNA column is the same as the distance between amino acid residues along an extended polypeptide, suggesting a stereochemical correlation of deep significance [7].

The principle that *DNA makes RNA makes protein* was first put into print in 1947 (at least 10 years before Crick) by two bacteriologists, Boivin and Vendrely [8]; or rather, by the anonymous editor who compressed their paper in *Experientia* into an English-language summary, less ambiguous than the text.

The "*tetranucleotide hypothesis*" was challenged by Chargaff et al. [4] (1905-2002) in 1951. He discovered that in natural DNA the number of guanine units equals the number of cytosine units and the number of adenine units equals the number of thymine units. This strongly hinted at the base pair makeup of DNA, although Chargaff failed to make this connection himself. The second of Chargaff's rules is that the composition of DNA varies from one species to another, in particular in the relative amounts of A+T and G + C bases. Such evidence of molecular diversity, which had not been anticipated, made DNA a more credible candidate for the genetic material than protein.

In 1953 the experimental studies and intellectual efforts of Miescher [2], Avery et al. [5], Chargaff et al. [4], Wilkins et al. [9] and Franklin and Gosling [10] were successfully combined into a simple helical model of B-DNA by Watson and Crick [11]. This was an important synthesis of the available research data and created novel scientific information of key importance. But it was still only information, which had to be completed and integrated with information from other sources (genetics, biochemistry, histology, bacteriology) if it was to serve as the knowledge-base for the new interdisciplinary science of Molecular Biology. The years between 1953 and 1961 became a period of intensive but rather fruitless speculation - the plateau before the next huge intellectual steps, which were marked by the discovery of the *Genetic Code* by Nirenberg and Matthaei [12] and by the recognition of additional elements (mRNA, rRNA) and formulation of the rules of translation by Jacob and Monod [13]. In 1956, Gamow introduced the

***Corresponding author:** Jan Charles Biro, Homulus Foundation, 612 S Flower, #1229, Los Angeles, CA 90017, USA, Tel: +1-213-627-6134; E-mail: Jan.biro@att.net

Received January 08, 2014; **Accepted** February 25, 2014; **Published** March 05, 2014

Citation: Biro JC (2014) Revisiting Crick's Dogma and the Impossibility of Reverse Translation. J Theor Comput Sci 1: 110. doi: [10.4172/2376-130X.1000110](https://doi.org/10.4172/2376-130X.1000110)

Copyright: © 2014 Biro JC. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

first model, the *overlapping codon - or diamond - model of translation*, which was followed rather quickly by Crick's "*comma-free model*" [14].

The Pursuit and Statement of the Dogma

Given the personality of Crick - which was described by colleagues [15] and by the expert historian of that time, Judson [16] - it is not difficult to imagine that the decade between his discovery of the double-helical model of DNA and its official recognition in 1964 was a difficult period for him. He instinctively knew he had made a significant discovery, but progress and recognition were delayed. Not only delayed; they were also blocked by harsh criticisms that pointed out the weaknesses of the DNA model, namely that its inter-twisted strands made it difficult to understand how unfolding and exposure of the inverted bases might take place [17,18].

The *Central Dogma* of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that information cannot be transferred back from a protein to nucleic acids or to another protein. In other words, '*once information gets into protein, it can't flow back to nucleic acid*'.

Its complement, the *Sequence Hypothesis*, is often conflated with the Central Dogma. According to the *Sequence Hypothesis*, DNA is transcribed to RNA, and RNA is translated into protein. "In its simplest form it (the *Sequence Hypothesis*) assumes that the specificity of a piece of nucleic acid is expressed solely by the sequence of its bases, and that this sequence is a (simple) code for the amino acid sequence of a particular protein. This hypothesis appears to be rather widely held. Its virtue is that it unites several remarkable pairs of generalizations:

1. The central biochemical importance of proteins and the dominating role of genes, and in particular of their nucleic acid [sequences],
2. The linearity of protein molecules (considered covalently) and the genetic linearity within the functional gene,
3. The simplicity of the composition of protein molecules and the simplicity of the nucleic acids."

This description was further amplified in the article and, in discussing how a protein folds up into its three-dimensional structure, Crick suggested that "*the folding is simply a function of the order of the amino acids*" in the protein. This proposal was first presented in 1958 [1], but it was widely misunderstood and the critics, who called it a "considerable over-simplification", forced Crick to provide further explanation and a softer formulation [19]. In this second version, Crick gives a clear explanation:

"Because it was abundantly clear by that time that a protein had a well-defined three dimensional structure, and that its activity depended crucially on this structure, it was a necessary to put the folding-up process on one side, and postulate that, by and large, the polypeptide chain folded itself up. This temporarily reduced the central problem from a three dimensional one to a one dimensional one". ... "The principal problem could then be stated as the formulation of the general rules for information transfer residue-by-residue [flow of sequence information and not flow of matter] from one polymer with a defined alphabet to another".

Crick showed that information transfer processes could be divided roughly into three groups. For the first group (class I), some direct or indirect evidence seemed to exist. This is indicated by the solid arrows in Figure 1.

Crick's Dogma, year 1958

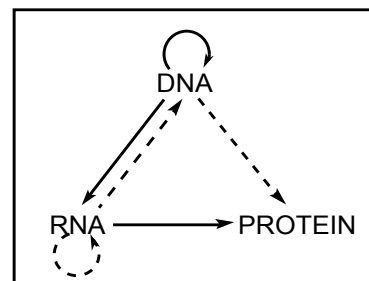


Figure 1: A tentative of the present day [1958]. Solid arrows show general transfers; dotted arrows show special transfers. Again, the absent

Next (class II) there were two transfers (shown in Figure 1 as dotted arrows) for which there was neither any experimental evidence nor any strong theoretical requirement. The third group (class III) comprised three transfers, arrows for which have been omitted from Figure 1.

The general opinion at the time was that class I almost certainly existed, class II was probably rare or absent, and class III was very unlikely to occur. It is very revealing to read how this fundamental classification was made:

"... there were good general reasons against all the three possible transfers in class III. In brief, it was most unlikely, for stereochemical reasons, that protein-protein transfer could be done in the simple way that DNA-DNA transfer was envisaged. The transfer protein-RNA (and the analogous protein-DNA) would have required (back) translation, that is, the transfer from one alphabet to a structurally quite different one. It was realized that forward translation involved very complex machinery. Moreover, it seemed unlikely on general grounds that this machinery could easily work backwards. The only reasonable alternative was that the cell had evolved an entirely separate set of complicated machinery for back translation, and of this there was no trace, and no reason to believe that it might be needed. I decided, therefore, to play safe, and to state as the basic assumption of the new molecular biology the non-existence of transfers of class III. Because these were all the possible transfers from protein, the central dogma could be stated in the form "*once (sequential) information has passed into protein it cannot get out again*". About class II, I decided to remain discreetly silent".

Crick had been forced several times to explain (and soften) his dogmatic statement of 1958. He had to recognize that his words had caused a lot of trouble (he called it misunderstanding). The phrase "*Central Dogma*" itself produced irritation in the scientific literature: it stank of authority. Crick practically retreated by calling his idea a Dogma, admitting that he used the term "Dogma" to mean "unfounded, religious belief", and he had recognized his mistake only after Monod explained it to him many years later. This rather naïve excuse is cited and accepted even by Judson [16]. I must honestly challenge this explanation. We have access to Crick's original draft of "Ideas on protein synthesis" (Oct. 1956) [1], which starts with a statement about "The Doctrine of the Triad" in the first line, followed by the familiar "The Central Dogma" in the second line. The use of these two words ("dogma" and "doctrine") is unlikely to have been a mistake. ('Dogma' is a Greek word for 'unproven opinion'. It has nothing inherently to do with religion, but it came to be derogatively associated with certain religious persuasions in Western Europe during the Middle Ages. In

the context of Crick's proposal as seen through 21st century eyes, we could regard it as an idea for which there is no reasonable evidence.)

What's wrong with the Dogma?

We know today that Crick's so-called "*discrete silence*" about class II information transfers was fully justified, because RNA>DNA transfer does occur (now well known as reverse transcription): RNA can replicate to RNA [20] and there are direct DNA templates for protein synthesis [21,22]. Only information transfer from proteins to nucleic acids remains in the "forbidden fruit" category. But is *reverse translation* really impossible? What are the arguments for its non-existence?

1. It hasn't been found yet - but that is not an acceptable argument for impossibility. Absence of evidence is not evidence of absence.

2. Forward translation is a very complex process; therefore it must be difficult to reverse. That is not a scientific argument either.

3. Proteins have a well-defined 3D structure that is known to be essential for their function, but transfer of 1D information to 3D information (or the reverse) is very complex. Therefore (?) "...it was a necessary to put the folding-up process on one side, and postulate that, by and large, the polypeptide chain folded itself up. This temporarily reduced the central problem from a three dimensional one to a one dimensional one." - Yes! And at the same time to postulate that protein > nucleic acid information transfer is impossible for exactly the same reason.

4. A very strong argument against the possibility of reverse translation is that Nirenberg's Genetic Code is redundant, so translation is an information-losing process and reconstruction of lost information is, of course, not possible. This argument was not known until 1964 and is usually not mentioned as support for Crick's Dogma of 1958.

5. There is supposedly no stereochemical connection between codons and the amino acids they encode; such a connection should be expected for any form of secondary or higher structural information transfer. This statement is another dogma of Crick, known as the "frozen accident" hypothesis [23]. It has been carefully analyzed, and rejected, by many scientists including Woese [24].

There is other, more general and indirect reasons to question Crick's Dogma (as well as any other dogmatic generalization in response to scientific or theoretical questions of central importance):

6. Protein synthesis from a nucleic acid template is a series of enzyme reactions, and enzyme reactions are reversible (at least in theory). Therefore it is to be expected that even the enzyme reactions involved in translation should be (at least in theory) reversible. As a consequence, translation as whole should be reversible, even if the process is very complex and complicated.

7. All natural material processes (and there are no others) follow the second law of thermodynamics, which entails the successive breakdown of order and replacement by disorder and heat. The only known natural process that might appear to be an exception is Life, which periodically shows signs of negative entropy, creation of order from disorder. Translation, as we know today, is a strongly information-losing process (proteins contain only 1/3 of the sequence information originally presents in the corresponding nucleic acids). Crick's Dogma justifies this information loss without proving it.

8. Mathematics, an extremely abstract, non-experimental science, makes a very clear distinction between proof and conjecture. Crick's Dogma is neither proof nor conjecture; it is a personal instinctive

feeling that became an unchallengeable statement for some.

How did the Dogma survive 50+ years and achieve such a high status? For one thing, it is not completely wrong: direct evidence for the existence of reverse translation has never been found. For another, lack of rejection, year after year, created positive feedback and strengthened the Dogma, which finally became so strong that its mere existence effectively prohibited tests of it in any "serious" laboratory, or publication of any doubt about its general validity in a "respected" scientific journal.

There are additional strong *indirect reasons for keeping Crick's Dogma alive*:

1. It is clearly consistent with neo-Darwinian evolutionary theory [25], which states that all changes in a species result from spontaneous mutations (changes in nucleic acid sequence) that survive because they provide some biological / reproductive advantage. Changes occur in the order nature > DNA > protein > function > survival and not the reverse. Permitting a protein > DNA order of information transfer could open the gates to non-Darwinist ideas about evolution.

2. The possible existence of a protein > DNA order of information transfer could revitalize Lamarckism [26], which held that acquired characteristics might become inherited. We all know that Lamarck's ideas were thoroughly tested in practice by Lysenko [27], causing widespread death from hunger in his country.

3. Proteins have long been regarded as the carriers of important biological (including genetic) information (since before 1853), and many scientists still feel it necessary to state this every day, completely forgetting that inherited information is not the only biologically important type of information.

4. The whole world of recent protein biotechnology is based on Crick's proposal that "*by and large, the polypeptide chain folds itself up*", an entirely groundless assumption in 1958. However, this statement acquired a basis somewhat later, in 1961, when Anfinsen et al. and Anfinsen [28,29] showed that ribonuclease could be refolded after denaturation and its enzyme activity was preserved. This suggested (and seemed to confirm) that all the information required for a protein to adopt its final conformation is encoded in its primary structure.

Objections to the Dogma

There are three categories of objection against the Dogma: Formal, Conceptual and Experimental.

1. There is a formal mistake in the presentation of the Dogma. Information transfer between macromolecules has two different, readily distinguished forms. The *first* category is *physical transfer (or transformation) of information* from one sequence to another, as in transcription and translation. The second category is *the recognition type of information transfer*, such as specific binding (recognition) between complementary nucleic acid sequences. Specific interactions between proteins and nucleic acids (for example transcription factor binding to promoters or enhancers, or restriction enzyme binding to cut-sites) also belong to this category, as do specific protein-protein interactions (such as receptor-ligand and antigen-antibody interactions). Crick's Dogma is obviously about the first type of information transfer but this is easily misunderstood: in the case of nucleic acids, the first type of information transfer automatically results in the second. Much criticism was directed against the Dogma because there was disagreement about the existence of recognition-type information transfer between proteins or between proteins and nucleic acids [30]. Figure 2 clarifies this situation.

2. Some parts of the Dogma have become scientifically obsolete because of direct experimental evidence: RNA>DNA, RNA>RNA and DNA>protein types of direct information transfer do exist [20-22].

3. One crucial statement of the Dogma, the prohibition of any kind of direct information transfer from proteins, is still neither confirmed nor refuted by direct laboratory experiments. (Information here means the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein).

However, indirect laboratory observations, theoretical studies and bioinformatics / system-biological concepts clearly indicate that this statement is conceptually wrong. The essence of the statement is that (a) proteins fold spontaneously (amino acid sequences contain all the information necessary for folding - no additional information is required); (b) transfer of folding (3D) information from proteins to nucleic acid sequences (1D) is not possible because they are too markedly different (!) from each other.

In addition, there are the following points.

- The limitation of the 'correct and spontaneous protein folding' conjecture has become more and more obvious: in addition to primary sequence information, folding information exists and it is often necessary for correct folding. The entire literature on *prions* [31] and *chaperons* [32] is about this "additional" 3D protein folding information and its biological necessity.

- A large and increasing number of publications suggest a connection between wobble bases in mRNA and the secondary structure of the folded protein [33,34].

- We constructed a Common Periodic Table of Codons and Nucleic Acids, which provides strong evidence that the physicochemical properties of amino acids are coded, primarily by the central codon letter. This is strong support for the co-evolution of codons and the encoded amino acids (as opposed to Crick's so-called "frozen accident"

hypothesis, which states that the codon-amino acid connection is merely accidental) [35].

- We have shown that amino acid residues preferentially interact with their codon-like sequences in the highly specific and unique restriction-enzyme / restriction-site interactions [36]. This is also indirect evidence for the possibility and the real existence of stereochemically-specific interactions between codons and their encoded amino acids.

- We have shown by extensive statistical analysis of real, known X-ray protein structures in PDB that co-locating (interacting) amino acids (3D) are preferentially encoded in partially complementary codons (i.e. the 1st and 3rd codon bases are complementary in reverse orientation, while the 2nd codon residue may be, but is not necessarily, complementary) [37]. This observation is the basis of the *Proteomic Code* and explains how specific (3D) protein information is encoded in a 1D nucleic acid sequence.

- We have shown by statistical analysis of 113 species-specific Codon Usage Tables that wobble bases are not randomly chosen, but their frequencies are highly predictable from the rest [38]. This observation further confirms the importance (and non-randomness) of wobble bases. In other words, the genetic code (64 codons) is redundant in respect of encoding the 20 amino acids, but the excess codon information is used to store and encode specific protein configurations and protein-protein interactions.

- Comparison of mRNA folding plots with the corresponding protein structure plots indicated that the energetically most favorable configurations of mRNAs are similar to the specific structures of their encoded proteins. This observation, together with the concept of the *Proteomic Code*, led us to formulate the hypothesis of nucleic acid-assisted protein folding [39].

Many of the aforementioned observations are of fairly recent date and were made possible by the "bioinformatics data boom", which was started by the Human Genome Project. None of these observations directly or indirectly indicates the existence of transcription-like transfer of biological information from proteins to nucleic acids or to other proteins. However, it has become clear that there is no conceptual hindrance to such kinds of information transfer. Storage and transfer of 3D structural information is possible in nucleic acids in addition to the storage and transfer of amino acid sequence coding information described by Nirenberg.

Prions and Protein-Mediated Epigenetic Inheritance

We use the term "reverse translation" to denote a biochemical process in which the molecular information in a protein (P, structure, function) is transferred into nucleic acids (N), and translation of the nucleic acid product of the reverse translation will reproduce the original protein (P1>reverse translation>N; N>translation>P2; P1=P2).

However, there is an alternative interpretation of "reverse translation", when the phrase is used to denote *any* kind of information transfer from proteins to nucleic acids (or phenotypic information into genotypic information). In this case the information in the nucleic acids corresponds to that in the proteins, but the nucleic acids and proteins, the carriers of that information, are not connected to each other by rules such as the Genetic Code. Prions provide a possible example of such "phenotype-to-genotype transformation".

Prions [31], proteinaceous and infectious pathogens, cause a group of invariably fatal neurodegenerative diseases: bovine spongiform encephalopathy (BSE) in cattle; scrapie in sheep; Creutzfeldt-Jakob

Biological Information Flow (Transformation & Recognition) between Nucleic Acids and Proteins

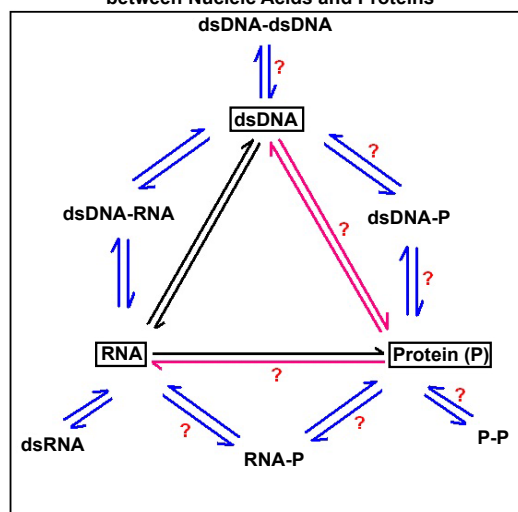


Figure 2: Biological information (transformation and recognition) between nucleic acids and proteins. Transcription and translation are indicated by black arrows, while the red arrows indicate the theoretical (but not observed) possibility of reverse translation. Information transfer through macromolecular interactions is indicated by blue arrows. There are many unanswered questions.

disease (CJD), Gerstmann-Straussler-Scheinker syndrome (GSS), fatal familial insomnia (FFI) and kuru [40] in humans; and chronic wasting disease in some types of deer. The prion gene, PrNP [41], is normally present in the genomes of many species and is expressed predominantly in the nervous system in mammals. The product is the prion protein PrP; the human version is 253 amino acids long (GenBank® accession no. AAD46098). The normal form, called cellular PrP (PrPC), has a novel physiological function [42]. It can be converted into a modified protein (PrPSc) through a post-translational process. PrPSc has a strong tendency to aggregate into amyloid-like material, which is biologically undegradable. Intracellular excess of PrPSc does not reduce the production of PrPC, which continues and sustains the conversion of PrPC into PrPSc. The main catalyst of the PrPC→PrPSc conversion is PrPSc itself.

Two hypotheses might serve to explain this reaction: PrPSc might be a protein template for a self-perpetuating conformation change of PrPC (template assistance model); alternatively, prion replication might be a process similar to crystallisation in that prions are propagated in a chain reaction [43] and PrPSc aggregates by 'nucleation' on a pre-existing 'seed' of PrP (nucleated polymerisation model) [44,45]. It is important to note that these nucleation and catalyzed-conversion hypotheses are not mutually exclusive. Whatever the action of PrPSc, this molecule represents a molecular phenotype that is normally represented in the genome no differently from normal PrPC. The PrPSc phenotype cannot be inherited because it kills cells merely by its presence.

However, this protein-induced self-perpetuating mechanism becomes very interesting in situations where the structural variant (Pv, corresponding to PrPSc) of a naturally occurring protein (Pn, corresponding to PrPC) provides benefits for the cells that carry it. These cells survive and – as the cytoplasmic proteins in “germ” or “mother” cells are equally segregated between “daughter” cells – Pv therefore persists into future cell generations without having genotypic representation in the nucleic acids. The Pv phenotype becomes inherited by the simple mechanical segregation of cytoplasmic proteins, without the involvement of a specific genotype or the complex expression of this particular phenotype via transcription and translation. This entirely protein-based virtual inheritance might go on for an unlimited number of generations (Figure 3).

(It is necessary to emphasize here that proteins are always involved in the division and multiplication of cells, with the sole exception of the “host-based” reproduction of viruses and phages. Even if nucleic acids are regarded as the sole carriers of genetic (heritable) information, they are always accompanied by proteins (from the “mother” cell) during cell division and reproduction).

There are numerous examples of this protein-based inheritance from yeast prions [46]. It is hypothesized that such protein-mediated epigenetic inheritance buys time (“general look-ahead effect” [47]) for the organism to survive until some genetic mutation renders the temporary advantage permanent. If and when that happens, the phenotypic information is formally transferred (reverse-translated?) to the genome.

Protein-mediated epigenetic inheritance might explain several early observations in which acquired characteristics seemingly became inherited, “seemingly” because they were not necessarily represented in the genotype [48,49]. However, it might be difficult for some scientists to see this as an example of reverse translation, or reverse transfer of biological information from proteins to nucleic acids.

Prions & the Protein-Mediated Epigenetic Inheritance

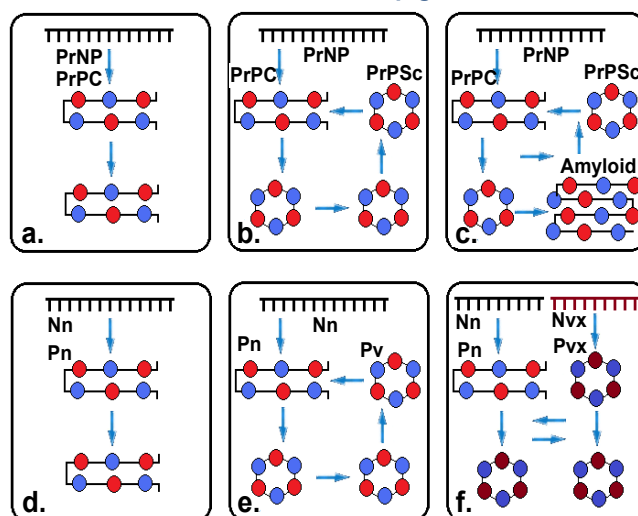


Figure 3: Prions and protein-mediated epigenetic inheritance. Prion protein (PrPC) is normally present in nerve cells and encoded by the PrNP gene (a). An abnormally-folded variant (PrPSc) appears in the cells under pathological conditions. PrPSc acts as chaperone and refolds even normal PrPC molecules into PrPSc (b). It also aggregates into amyloid, which forms undegradable deposits, degenerates and kills the cells (c). However there are examples where a differently folded variant (Pv) of a naturally occurring protein (Pn, encoded by the Nn gene) is for cells (d, e) and the Pn>Pv conversion provides survival advantages to the host cells; Pv is transferred not “genetically” by a separate gene, but “epigenetically” by the protein itself. It is theorized that, after several generations, mutation and natural selection might provide a genetically inherited variant (Pvx) of Pv (encoded by the Pvx gene), which replaces the original Pv and its functions (f).

Recent Status of Reverse Translation

Historically, speculations have been published that poly-amino acid reverse translation (PAA-RT) may have existed in nature in prebiotic evolution and could exist undiscovered in nature today [50-54]. There are several patents relating to this notion. Their common feature is an attempt to arrange codon-amino acid complexes along a template peptide sequence, polymerize the codon parts, and use the product as the nucleic acid template for synthesis of DNA and of proteins that are similar or identical to the template protein [55,56].

One product under development is called Peplica™ [57]. In Peplica™, translation is made to function in reverse (“reverse translation”). Thus, the protein code (encoded in its unique amino acid sequence) is used to make DNA or RNA genes. After the reverse translation step, the resulting nucleic acid is amplified by a conventional amplification method such as PCR. Using the amplification product, the identity of the protein can be determined and, if desired, more of it can be produced. In principle, Peplica™ could be used to detect as little as one copy of a protein molecule.

The differences between genome and proteome are of course well recognized, as is the importance of epigenetic modifications of proteins. The example of prions raises a very intriguing question: are there proteins in the cytoplasm that pass from generation to generation (are inherited) but are not present or expressed in the current genome (SIC! i.e. from DNA or RNA templates)? It might be necessary in future to consider sequencing proteomes, just as genomes have been sequenced recently. The development of bioinformatics, sequence databases and computational tools also offer exceptionally effective approaches to

reviewing and further developing classical statements in molecular biology [58-62].

Storage and Transfer of Folding Information Using Synonymous Codons

Nirenberg's Genetic Code is a clear and general explanation of the transfer of primary structure information from nucleic acids to proteins. However, this code is ~3-fold redundant, which means that ~2/3 of the genetic information remains on the "nucleic acid side" and is never transferred to proteins. Simultaneously, there is a shortage of molecular information on the "protein side". Larger proteins cannot fold correctly without the assistance of other proteins, called chaperons. Most recently-identified chaperons are themselves proteins, which are also produced by equally information-losing translations, so they may also need chaperones for their own folding. Thus, the information deficit in proteins potentially leads to a process of infinite regression.

The concept of the *Proteomic Code* gives a relatively simple explanation of the storage and transfer of folding information in nucleic acids using the excess (redundant) amino acid coding information. The *Proteomic Code* means that co-locating amino acids are preferentially encoded by partially complementary codons [63] (Figure 4).

Similarly, separated amino acids are preferentially encoded by non-complementary synonymous codons. The *Proteomic Code* indicates the action of a statistical rule: there is a tendency for amino acid co-locations in protein structures to be reflected in codon co-locations in the corresponding mRNA structures.

By extension, this means that the peptide structure is represented, at least to some degree, in mRNA structure (Figure 5).

Codon redundancy is mainly represented by the wobble bases. These wobble bases make it possible for most of the 400 potential amino acid pairs to be represented by codon pairs, which may optionally be complementary at the 1st and 3rd codon positions in reverse orientation. In this way, a wide variety of mRNA structures can be achieved without affecting the translational meaning of the codons. Thus, folding information in addition to the sequence coding information may be stored in nucleic acids. Transfer of structural information to proteins requires that a temporary contact between mRNA and the newly-translated peptide persists during peptide folding, conveniently on the surface of

Coding of Folding by Synonymous Codon Usage

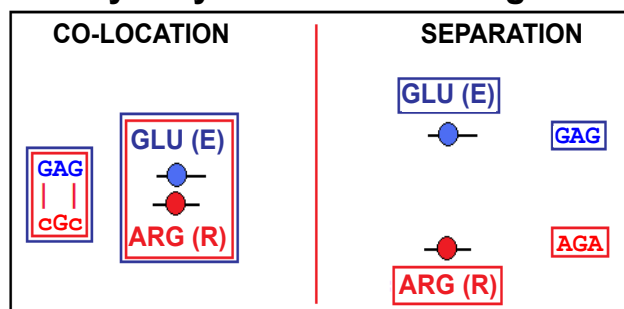


Figure 4: Coding of folding by synonymous codon usage. Two amino acids may be coded by partially complementary or non-complementary codons. This structural code (the *Proteomic Code*) determines whether the coded amino acids will preferentially co-locate or separate in a translated protein.

Coding of Folding by Synonymous Codon Usage

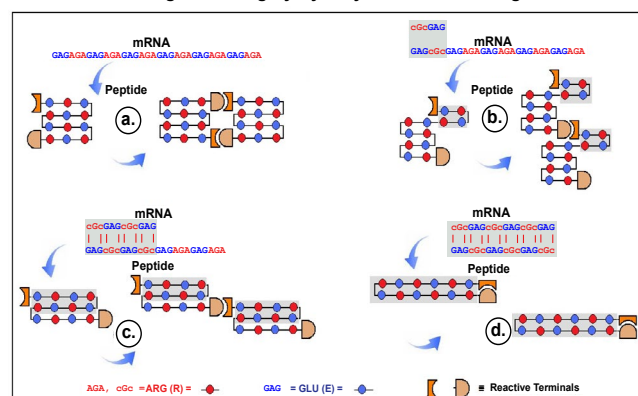


Figure 5: Effect of synonymous codons on the folding (structure) of mRNA and coded peptides. A peptide consists of 6R (positively charged, red) and 6E (negatively charged, blue) amino acid residues. It contains reactive termini that interact with each other. This peptide has many equally possible and favored tertiary structures (tertiary structure) and several copies might interact with each other (quaternary structure), for example a compact, globular, that forms dimers (a). However its mRNA may contain structural information simply by replacing AGA with its synonymous codon (cGc). This structural information can be transferred into the peptide during translation and different 3D structures and interactions (b, c). A hairpin-like structure, for example, "shortcuts" the reactive termini, resulting in only monomer formation (d). Structurally coded parts of sequences are shaded (grey).

rRNA. This contact may involve tRNA, though it does not necessarily do so. We have shown that some amino acids collocate with their codon-like sequences in specific protein-nucleic acid interactions [36].

We suggest the extension and completion of Nirenberg's Genetic Code by the *Proteomic Code* [63].

Consequences of the Dogma, Conclusion and Suggestion

- Crick's Dogma was a helpful summary of knowledge in the pioneering years of molecular biology. However, this field of science has developed significantly during the past half century and the Dogma is no longer a correct summary of our core knowledge in it.
- The Dogma became elevated to an almost unchallengeable status, making it inaccessible to serious scientific criticism or revision. This kind of ideological protection of an outmoded idea is not compatible with a modern (21st century) scientific spirit.
- The Dogma symbiotically promotes and protects other now-obsolete concepts such as Anfinsen's thermodynamic principle, and has become a serious obstacle to developing some really central aspects of molecular biology.
- The Dogma has serious "side effects", never intended by Crick and his fellows: it permits and authenticates the idea of irreversible and information-losing biological processes. These might exist, but it is conceptually wrong to accept them without compelling proof.
- The Dogma biases the whole of recent research in molecular biology. Our pathological focus on proteins permits and favors the rejection of 98% of our genetic material as "junk DNA" [64], canalizing huge intellectual and economic resources into examining minor genetic variations such as Single Nucleotide Polymorphisms, SNPs [65].

Since the elaboration of the central dogma of molecular biology, our understanding of cell function and genome action has benefited

from many radical discoveries. These discoveries contradict atomistic pre-DNA ideas of genome organization and violate the central dogma at multiple points. In place of the earlier mechanistic understanding of genomics, molecular biology has led us to an informatic perspective on the role of the genome [66].

It is my opinion that the characterization of Crick's proposal as a significant historical biological idea, instead of a 'Central Dogma', would be more appropriate nowadays. I would propose such a change, without denigrating its past contribution to science.

Competing Interests

The author declares he has no competing interests.

Author's Contributions

The whole of this manuscript was conceived, written and revised by JCB.

Acknowledgment

This work was supported by grants from the Homulus Foundation, Los Angeles, CA, USA. JCB has been a "scientist in the US National Interest" since 2006. JCB wishes to thank the trust and support of this great Nation. The author also wishes to thank and he very much appreciates the continuous attention and advice of George L G Miklos PhD (Director of Secure Genetics, Sydney, Australia). The pioneering works and views of Prof. Carl Woese were very useful and indeed essential for parts of our work. We respectfully recognize and acknowledge this.

References

- Crick FHC (1958) On protein synthesis. Symp Soc Exp Biol XII: 139-163.
- Miescher F (1871) Ueber die chemische Zusammensetzung der Eiterzellen. Med-Chem Unters 4: 441-460.
- Levene PA, La Forge FB (1915) On Chondrosamine. Proc Natl Acad Sci U S A 1: 190-191.
- Chargaff E, Lipshitz R, Green C, Hodes me (1951) The composition of the deoxyribonucleic acid of salmon sperm. J Biol Chem 192: 223-230.
- Avery OT, Macleod CM, McCarty M (1944) studies on the chemical nature of the substance inducing transformation of pneumococcal types : induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. J Exp Med 79: 137-158.
- Caspersson (1950) Cell growth and cell function, a cytochemical study W. W. Norton & Company, Newyork.
- Astbury WT (1947) X-ray studies of nucleic acids. Symp Soc Exp Biol : 66-76.
- Boivin A, Vendrely R (1947) Experientia 3: 32-34.
- Wilkins MH, Stokes AR, Wilson HR (1953) Molecular structure of deoxypentose nucleic acids. Nature 171: 738-740.
- Franklin RE, Gosling RG (1953) Molecular structure of sodium thymonucleate. Nature 171: 740-741.
- Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171: 737-738.
- Nirenberg MW, Matthaei JH (1961) The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci U S A 47: 1588-1602.
- Jacob F, Monod J (1964) Biochemical and Genetic Mechanisms of Regulation in the bacterial cell. Bull Soc Chim Biol (Paris) 46: 1499-1532.
- Hayes B (1998) The invention of the genetic code. American Scientist. 86: 814.
- Watson JD (1980) The Double Helix: A personal account of the discovery of the structure of DNA, New York: Norton.
- Judson HF (1996) The Eighth Day of Creation: Makers of the Revolution in Biology, Cold Spring Harbor Laboratory Press, New York.
- Biro JC (2003) Speculations about alternative DNA structures. Med Hypotheses 61: 86-97.
- Biro JC (2009) Principia Biro-Informatica: Creative Ideas in Molecular Biology & Bioinformatics. Los Angeles, CA, Homulus Foundation.
- Crick F (1970) Central dogma of molecular biology. Nature 227: 561-563.
- Ahlquist P (2002) RNA-dependent RNA polymerases, viruses, and RNA silencing. Science 296: 1270-1273.
- McCarthy BJ, Holland JJ (1965) Denatured DNA as a direct template for in vitro protein synthesis. Proc Natl Acad Sci U S A 54: 880-886.
- Uzawa T, Yamagishi A, Oshima T (2002) Polypeptide synthesis directed by DNA as a messenger in cell-free polypeptide synthesis by extreme thermophiles, *Thermus thermophilus* HB27 and *Sulfolobus tokodaii* strain 7. J Biochem 131: 849-853.
- Crick FH (1968) The origin of the genetic code. J Mol Biol 38: 367-379.
- Woese CR (1967) In the Genetic Code: The Molecular Basis for Gene Expression. Harper & Row, New York.
- Theissen G (2009) Saltational evolution: hopeful monsters are here to stay. Theory Biosci 128: 43-51.
- Lamarck JB (1809) Philosophie zoologique ou exposition des considérations relatives à l'histoire naturelle des animaux ("Zoological Philosophy: Exposition with Regard to the Natural History of Animals").
- Lysenko TD (1948) The Science of Biology Today. New York: International Publishers.
- CB, Haber E, Sela M, White FH Jr (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci U S A 47: 1309-1314.
- CB (1973) Principles that govern the folding of protein chains. Science 181: 223-230.
- Werner E (2005) Genome semantics, in silico multicellular systems and the Central Dogma. FEBS Lett 579: 1779-1782.
- Prusiner SB (1998) Prions. Proc Natl Acad Sci U S A 95: 13363-13383.
- Macario AJ, Conway de Macario E (2007) Molecular chaperones: multiple functions, pathologies, and potential applications. Front Biosci 12: 2588-2600.
- Xie T, Ding D (1998) The relationship between synonymous codon usage and protein structure. FEBS Lett 434: 93-96.
- Mukhopadhyay P, Basak S, Ghosh TC (2007) Synonymous codon usage in different protein secondary structural classes of human genes: implication for increased non-randomness of GC3 rich genes towards protein stability. J Biosci 32: 947-963.
- Biro JC, Benyó B, Sansom C, Szilávecz Á, Fördös G, et al. (2003) A common periodic table of codons and amino acids. Biochem Biophys Res Commun 306: 408-415.
- Biro JC, Biro JM (2004) Frequent occurrence of recognition site-like sequences in the restriction endonucleases. BMC Bioinformatics 5: 30.
- Biro JC (2007) The Proteomic Code: a molecular recognition code for proteins. Theor Biol Med Model 4: 45.
- Biro JC (2008) Does codon bias have an evolutionary origin? Theor Biol Med Model 5: 16.
- Biro JC (2005) Nucleic acid chaperons: a theory of an RNA-assisted protein folding. Theor Biol Med Model 2: 35.
- Gajdusek DC: Unconventional viruses and the origin and disappearance of kuru. In Nobel lectures, physiology or medicine 1971-1980. Edited by Lindsten J. Singapore: W
- Lee IY, Westaway D, Smit AF, Wang K, Seto J, et al. (1998) Complete genomic sequence and analysis of the prion protein gene region from three mammalian species. Genome Res 8: 1022-1037.
- Martins VR, Mercadante AF, Cabral AL, Freitas AR, Castro RM (2001) Insights into the physiological function of cellular prion protein. Braz J Med Biol Res 34: 585-595.
- Aranda-Anzaldo A (1992) Possible cell-free prion replication. Med Hypotheses 38: 249-251.
- Singh N, Gu Y, Bose S, Kalepu S, Mishra RS, et al. (2002) Prion peptide 106-126 as a model for prion replication and neurotoxicity. Front Biosci 7: a60-71.

45. Tompa P, Tusnady GE, Friedrich P, Simon I (2002) The role of dimerization in prion replication. *Biophys J* 82: 1711-1718.
46. Koonin EV (2012) Does the central dogma still stand? *Biol Direct* 7: 27.
47. Whitehead DJ, Wilke CO, Vernazobres D, Bornberg-Bauer E (2008) The look-ahead effect of phenotypic mutations. *Biol Direct* 3: 18.
48. Waddington CH: Canalisation of development and the inheritance of acquired characters. *Nature* 1942, 150:563.
49. Waddington CH: Genetic assimilation of an acquired character. *Evolution* 1953, 7:118-126.
50. Biro J (1983) General theory of cell regulation. *Med Hypotheses* 12: 203-226.
51. Craig R (1981) The theoretical possibility of reverse translation of proteins into genes. *J Theor Biol* 88: 757-760.
52. Cook ND (1977) The case for reverse translation. *J Theor Biol* 64: 113-135.
53. Mekler LB (1967) Mechanism of biological memory. *Nature* 215: 481-484.
54. Nashimoto M (2001) The RNA/protein symmetry hypothesis: experimental support for reverse translation of primitive proteins. *J Theor Biol* 209: 181-187.
55. Gold L, Tuerk C, Pribnow D, Drew Smith J (1991) Cell-free synthesis and isolation of novel genes and polypeptides. US Patent: US5658754.
56. Martin MT: Methods and compositions for reverse translation. US Patent: US 7,169,894,B2, Filed Nov. 4. 2002.
57. Peplica TM : Mirari Biosciences, Inc. 9610 Medical Center Drive, Suite 240, Rockville, MD 20850
58. Bussard AE (2005) A revolution? The prion anomaly may challenge the central dogma of molecular biology. *EMBO Rep* 6: 691-694.
59. Li M, Wang IX, Li Y, Bruzel A, Richards AL, et al. (2011) Widespread RNA and DNA sequence differences in the human transcriptome. *Science* 333: 53-58.
60. Lin W, Piskol R, Tan MH, Li JB: Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". *Science* 2012, 335(6074):1302. author reply 1302.
61. Pickrell JK, Gilad Y, Pritchard JK: Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". *Science* 2012, 335(6074):1302. Author reply 1302.
62. Kleinman CL, Majewski J: Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". *Science* 2012, 335(6074):1302. author reply 1302.
63. Biro JC (2008) Discovery of proteomic code with mRNA assisted protein folding. *Int J Mol Sci* 9: 2424-2446.
64. Ohno S (1972) So much "junk" DNA in our genome. *Brookhaven Symp Biol* 23: 366-370.
65. Thorisson GA, Stein LD (2003) The SNP Consortium website: past, present and future. *Nucleic Acids Res* 31: 124-127.
66. Shapiro JA (2009) Revisiting the central dogma in the 21st century. *Ann N Y Acad Sci* 1178: 6-28.