Pontificia Universidad JAVERIANA Bogotá

Univ. Sci. 28(1): 1–21, 2023
doi: 10.11144/Javeriana.SC281.npor

ORIGINAL ARTICLE

New Perspectives on Reverse Translation: Brief History and Updates

Adhityo Wicaksono*[1], Viol Dhea Kharisma[2], Arli Aditya Parikesit[3]**

Edited by
Juan Carlos Salcedo-Reyes
salcedo.juan@javeriana.edu.co

1. Division of Biotechnology, Generasi Biologi (Genbinesia) Foundation, Jl. Swadaya Barat no. 4, Gresik Regency 61171, Indonesia

2. Division of Molecular Biology and Genetics, Generasi Biologi (Genbinesia) Foundation, Jl. Swadaya Barat no. 4, Gresik Regency 61171, Indonesia

3. Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences, Jakarta, Indonesia

*adhityo.wicaksono@gmail.com
**arli.parikesit@i3l.ac.id

Received: 11-02-2022
Accepted: 21-10-2022
Published online: 15-01-2023

Citation: Wicaksono A, Kharisma VD, Parikesit AA. New Perspectives on Reverse Translation: Brief History and Updates, *Universitas Scientiarum*, 28(1): 1–20, 2023.
doi: 10.11144/Javeriana.SC281.npor

Funding: n.a.

Electronic supplementary material: n.a.

Abstract

Since the 1950s, reverse translation has been an enigmatic part of Crick's central dogma of molecular biology. It might be described as the possibility to back-translate information from proteins to nucleic acids (or codons). A few studies have attempted to theorize and/or conduct *in vitro* experiments to test the likelihood of reverse translation, with ideas often involving the creation of peptide recognition sites that bridge the peptide and the codon. However, due to many constraints including an asymmetrical informational transfer, the stability of protein-peptide bonds, the structural non-uniformity of protein R-groups, and the informational loss in post-translational protein modifications, this concept requires follow-up studies. On the other hand, current bioinformatic tools that rely on computational programs and biological databases represent a growing branch of biology. Bioinformatics-based reverse translation can utilize codon usage tables to predict codons from their peptide counterparts. In addition, the development of machine learning tools may allow for the exploration of biological reverse translation *in vitro*. Thus, while *in vivo* reverse translation appears to be nearly impossible (due to biological complexity), related biological and bioinformatics studies might be useful to understand better the central dogma's informational transfer and to develop more complex biological machinery.

Keywords: amino acids; central dogma; genetic transfer; polypeptide; RNA

1. Introduction

Francis Crick [1, 2] proposed molecular biology's central dogma to explain the flow of basic chemical information from DNA to protein (*i.e.*, DNA replication, transcription into RNA as a mediator molecule, and translation of RNA into proteins, the polymeric building blocks of life). This process is also known as "the general transfer" (**Figure 1**, green arrows). For transcription, DNA-dependent RNA polymerase will transcribe (*i.e.*, copy) DNA's genetic information onto messenger RNA (mRNA). For translation to take place, a complex ensemble of ribosomes and transfer RNA (tRNA) will read the mRNA (*i.e.*, the transcribed DNA copies) attaching coded amino acids to form proteins. The dogma was extended with "special transfers" after the discovery of the RNA-dependent DNA polymerase or reverse transcriptase in viruses, which back-transcribes RNA into DNA [3, 4], and RNA-dependent RNA polymerase (RdRp) which transcribes an RNA strand into its complementary RNA strand [5] (Figure 1, blue arrows). Lastly, Crick's original central dogma also referred to "the unknown transfers" of protein which involves an informational reversal into the corresponding nucleic acids, and the replication of protein molecules to amplify a set of similarly composed protein molecules; both processes remain a mystery (Figure 1, red arrows).

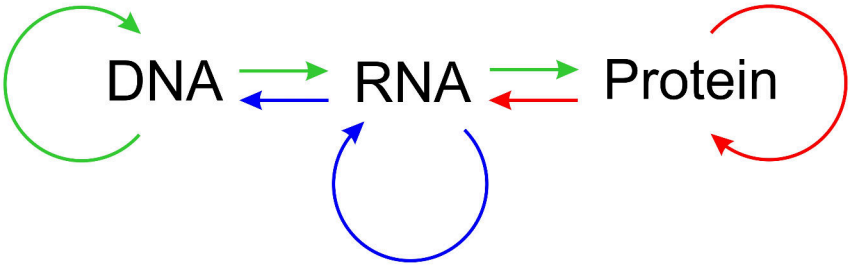

Figure 1. Schematic of molecular biology's central dogma with general (green arrows), special (blue arrows), and unknown (red arrows) transfer processes.

Koonin [6, 7] stated that an information reversal from protein to nucleic acid may be possible via certain mechanisms (*e.g.*, prions) but that a reversal of the genetic code is impossible. Furthermore, this information back flow itself is a violation of the known central dogma. Assuming that reverse translation is possible, we will refer to information reversal as "the unknown transfers" (Figure 1, red arrows), a possible extended part of the dogma. We will cover two reverse translation perspectives, (i) as an actual biological process (either natural or engineered) and (ii) as a bioinformatics subject. Furthermore, this review addresses three main questions: Is it possible for a polypeptide sequence to be translated back into a nucleic acid sequence? If so, will it be significantly applied to many present-day molecular biology studies? and, if not, what are its implications? The discussion includes transfer-related findings from ancient earth biochemistry to present-day prion-mediated epigenetics, comparing previous reverse translation studies and hypotheses, and reviewing reverse translation limitations. Lastly, the bioinformatics approach as a tool of transfer between protein to nucleic acid sequence data is reviewed.

2. The concept of reverse translation in ancient life and information flow from protein to nucleic acids

RNA possesses both genetic and catalytic properties [8]. Gilbert highlighted this decisive RNA dual condition among biochemical molecules within the "RNA world hypothesis" [9, 10]. For viruses, both DNA and RNA serve as genetic blueprint molecules, with RNA also serving as catalytic molecule (*e.g.*, mRNA carries genetic information from the DNA to be translated by tRNA, or RNA enzymes that are known as ribozymes). However, objections to the RNA world hypothesis claim that unstable and easily degradable RNA molecules are too complex to be synthesized prebiotically and, more importantly, that RNA catalytic properties are limited and are not present in all RNA types (only in long RNA sequences) [11]. DNA, being more stable, displaced RNA in the RNA world hypothesis, which became the "DNA world hypothesis", highlighting DNA's role as the molecule that carries the information encoding proteins. This change in perspective followed the description of ribosome formation and was triggered by the discovery of viral reverse transcriptases. In addition to viral reverse transcriptases, two enzymes (ribonucleotide reductases and thymidylate synthases) can convert ribonucleotides (rNTPs) into deoxyribonucleotides (dNTPs) [12]. This conversion from RNA to DNA not only increases nucleotide stability but allows for convenient restoration of cytosine (C) upon deamination to uracil (U) (not possible in RNA), which otherwise may lead to mutation [12] — allowing DNA to thrive and dominate over RNA as the genetic molecule of life, except within the realm of viruses.

Moreover, it was suggested that in ancient time there was a two-way information transfer between proteins and nucleic acids, originating what we now understand as the genetic code [13]. The mechanism, however, remains unclear.

Proteins are not known as information templates. Yet, as illustrated by the "unknown transfer" pieces of the central dogma, one may discuss whether protein information can be transformed back into that of nucleic acid, or whether proteins are capable of replicating other proteins. Some instances imply protein self-replicability. For example, the 32-residue α-helical peptide of yeast transcription factor GCN4 leucine-zipper domain can catalyze itself (by the acceleration of thioester-promoted amide-bond condensation of 15- and 17-residue fragments), serving as an auto-replication template in a dilute aqueous solution [14]. This self-replicating ability shows that peptide sequences can store information to make copies of their own molecules. Prions (viral proteins) constitute another example of self-replicating proteins. The accumulation of abnormally conformed and infectious PrP^{Sc} proteins in the central neural system, PrP^{Sc} being the "scrapie isoform" of original cellular prion proteins (PrP^{C}), can cause transmissible spongiform encephalopathies (TSE) [15]. After protein translation, mutant PrP^{Sc} proteins serve as templates that bind to normal PrP^{C} proteins. By turning α-helices into dominantly β-sheets, whereby changing protein stability, the prions catalyze the formation and aggregation of anomalous isoforms [16]. It was, however, later discovered that prions only use their own structures to modify the conformation of other proteins into similar prion proteins and that the prion protein itself is synthesized from a PRNP gene (https://www.ncbi.nlm.nih.gov/gene/5621) that is stored as nucleotide sequences [17].

Another study revealed the phenomenon of "prion-mediated inheritance" where a yeast prion-forming protein, namely the Sup35 translation-termination factor, serves as an epigenetic modulator (called [PSI+]) that allows for diversification of protein translation [18]. With a normal Sup35 protein, a regular protein is translated from mRNA once the ribosome reaches the stop codon. In contrast, if the Sup35 protein is in prion aggregate (amyloid) form, a protein variant will be translated by reading through the stop codon [18, 19]. This protein expression alteration shows that, despite being gene-encoded, a prion can direct its prion-independent inheritance to "prion-mediated inheritance" via the epigenetic influence of the prion [6]. It thus appears as if a certain backflow of information from protein to DNA is possible, *i.e.*, describing "the unknown transfer" in the central dogma and providing support for Lamarckian inheritance (phenotype-based inheritance) [6].

While nucleic acids possess simplistic "one-dimensional" information that could be referred to as "the digital code", proteins with their more complex information and structural variants might be referred to as "the analog code" [7]. The natural flow of information is from a simple model to a more complex one. The irreversible nature of the translational process serves as an exclusion principle where, if reverse translation were possible, it would entail reversal from complexity to a simpler version (*i.e.*, devolving life processes). The information transfer from RNA to amino acids is very complex and involves the pairing of amino acids to tRNA (via aminoacyl-tRNA synthetase), the tRNA-related detection of corresponding mRNA (via anticodon-codon recognition), and the complex enzymatic process in the ribosome which "sews" tRNA-carried amino acids together. This intricate machinery thus allows for simpler RNA nucleotides to become more complex and highly versatile protein structures. As mentioned earlier, prions may provide a small opportunity for protein-to-nucleic acid communication. However, this phenomenon does not support the concept of full physical information reversal from proteins to nucleic acids. Instead, these phenotypic changes may occasionally lead to genetic mutations [6]. To better understand the

possibilities and limitations of protein-based information transfer, whether in ancient or current times, some researchers developed ideas and experiments over the last few decades to explore "what if" reverse translation were possible.

3. Prior studies and hypotheses

In the past, some hypothetical theories and preliminary studies supported the idea of reverse translation by suggesting the existence of an alternative system to the existing ribosomal one. Mekler in 1967 [20] suggested the idea of tRNA anticodons being reverted into mRNA codons (**Figure 2**). This could be achieved if the specific tRNAs were bound to amino acids which, in turn, were docked onto complementary antigen determinant sites.

However, the conversion of tRNA anticodons to mRNA codons was not fully explained by Mekler's hypothesis [20]. It was only postulated that a polymerase might be involved in aligning the anticodon arm of tRNA and transferring the anticodon information back to mRNA codons [20, 21]. To date, no studies have filled the gap of missing information. During normal translation, the anticodon loading of tRNA (via aminoacyl-tRNA synthetase) is achieved with extremely high accuracy before recognition of mRNA codons in the ribosome-mRNA complex takes place [22]. Yet, for Mekler's hypothetical process, the process behind the rebinding of individual peptides in the polypeptide or protein sequences back to tRNA was not fully described. Without these details, the nature of this hypothetical process remains a mystery.

In 2001, Nashimoto [13] suggested a protocol using a reverse translational RNA (rtRNA) consisting of a hammerhead self-cleaving ribozyme [23] (sequence: 5'-GGGAACUCCCUCUGAUGAGCAGGACCGUGCACUCUUCGAACAUUUCGACAUGAGACACGGAUCCUGCGAAACACCGUGUC-3'). The rtRNA had an arginine codon (peptide binding site) and an arginine

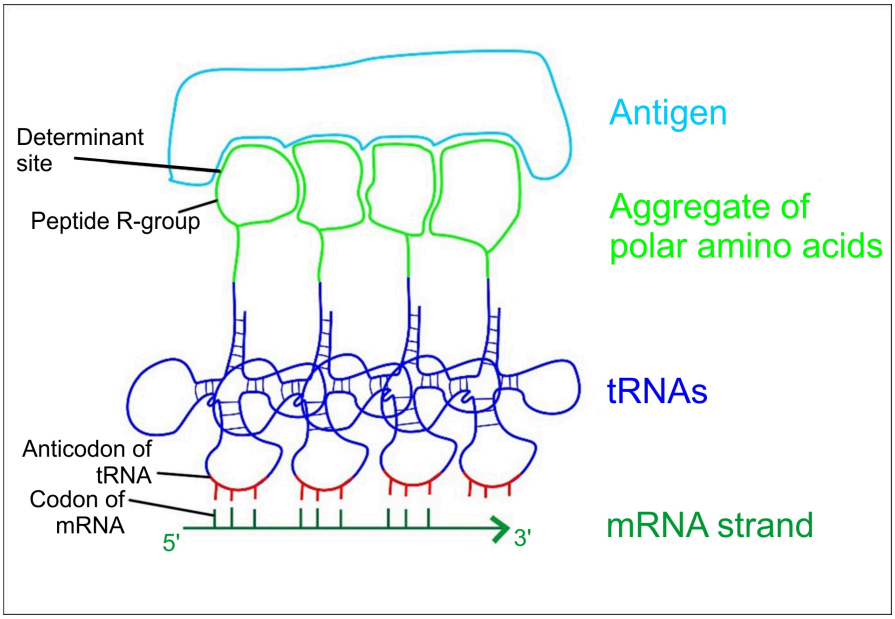

Figure 2. The concept of a reverse translation method where the determinant site of an antigen binds to the R-groups of polar amino acids. Each amino acid is bound to a specific tRNA sequence (having its own anticodon). As the tRNA anticodon sites line up, they could serve as a template to synthesize a new mRNA strand (carrying their series of codons that encode amino acids). Concept by Mekler [20], illustration inspired by and modified from Cook [21].

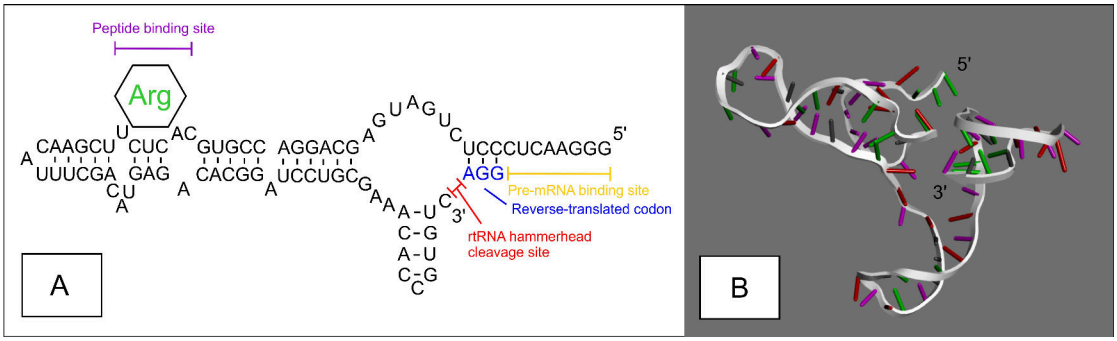

Figure 3. A reverse translation RNA (rtRNA) hammerhead ribozyme that can serve as a catalytic biomolecule for the information sequence reversal from a primitive protein polypeptide back into an mRNA molecule. (A) The rtRNA consists of an amino acid/peptide binding site sequence (in this case: arginine (purple)), a hammerhead ribozyme self-cleavage site (red) connected to the arginine (AGG) codon (blue), and a pre-mRNA binding site (yellow). Illustration inspired by and modified from Nashimoto [13]. (B) Predicted three-dimensional rtRNA (viewed using the BIOVIA Discovery Studio 2021). Tertiary structure predicted using the iFoldRNA v2 web server [27].

(AGG) RNA binding sequence [24] (**Figure 3**). In this case, the amino acid-binding RNA had an arginine-binding sequence, which was synthesized by exposing pure arginine to a randomized pool of RNA sequences to reveal regions within the latter that would strongly bind to this specific amino acid [25]. This protocol is now recognized as the systematic evolution of ligands by exponential enrichment (SELEX) method and uses DNA or RNA libraries to bind desired target molecules [26]. Nashimoto predicted that the rtRNA and pre-mRNA (containing two stop codons) would dock into the primitive protein (**Figure 4**).

SELEX procedures use as many as 10^{14} to 10^{15} DNA sequences or RNA molecules that consist of 40–80 random bases attached with PCR primers on both sides. Several thousand sequences in the library will form the correct secondary structure that will bind with high affinity to the target molecules [28]. These high-affinity RNA-amino acid complexes are then selected with their nucleotide sequences for each amino acid. For example, the SELEX procedure allowed for the identification of tryptophan-binding RNA [24, 29]. Moreover, information on S-adenosylmethionine riboswitches [30, 31, 32], aptamer domain for lysine [33, 34], and aptamer domain for citrulline and arginine [35] have been discovered, making it easier to apply this SELEX protocol to other amino acids in the future. Unfortunately, the unstable nature of RNA structures remains challenging [28].

The combination of a machine learning program with molecular docking and docking score analyzer programs could potentially shorten the *in vitro* SELEX procedures by providing predictive results. Specifically, *in silico* machine learning methods can be built to generate a randomized RNA sequence library, before docking it with 20 proteinogenic amino acids and revealing the best specific RNA sequence with the highest binding affinity for each amino acid. So far, a program such as AnnapuRNA [36] can be used to determine a docking score between RNA and a small molecule (to validate, once the RNA and the small molecules have been previously docked by another molecular docking program such as Autodock Vina [37]).

Nashimoto [13] also suggested that transferring 20 amino acids back into 20 codons (instead of the full 64 codons) would work for the simulation of early Earth conditions (where codons were less abundant) and may greatly reduce the number of reverse translational tests. However, it is still unclear whether the amino binding site in the rtRNA would recognize the respective amino acids during reverse translation, especially as amino acids are not free ligands but rather integrated into

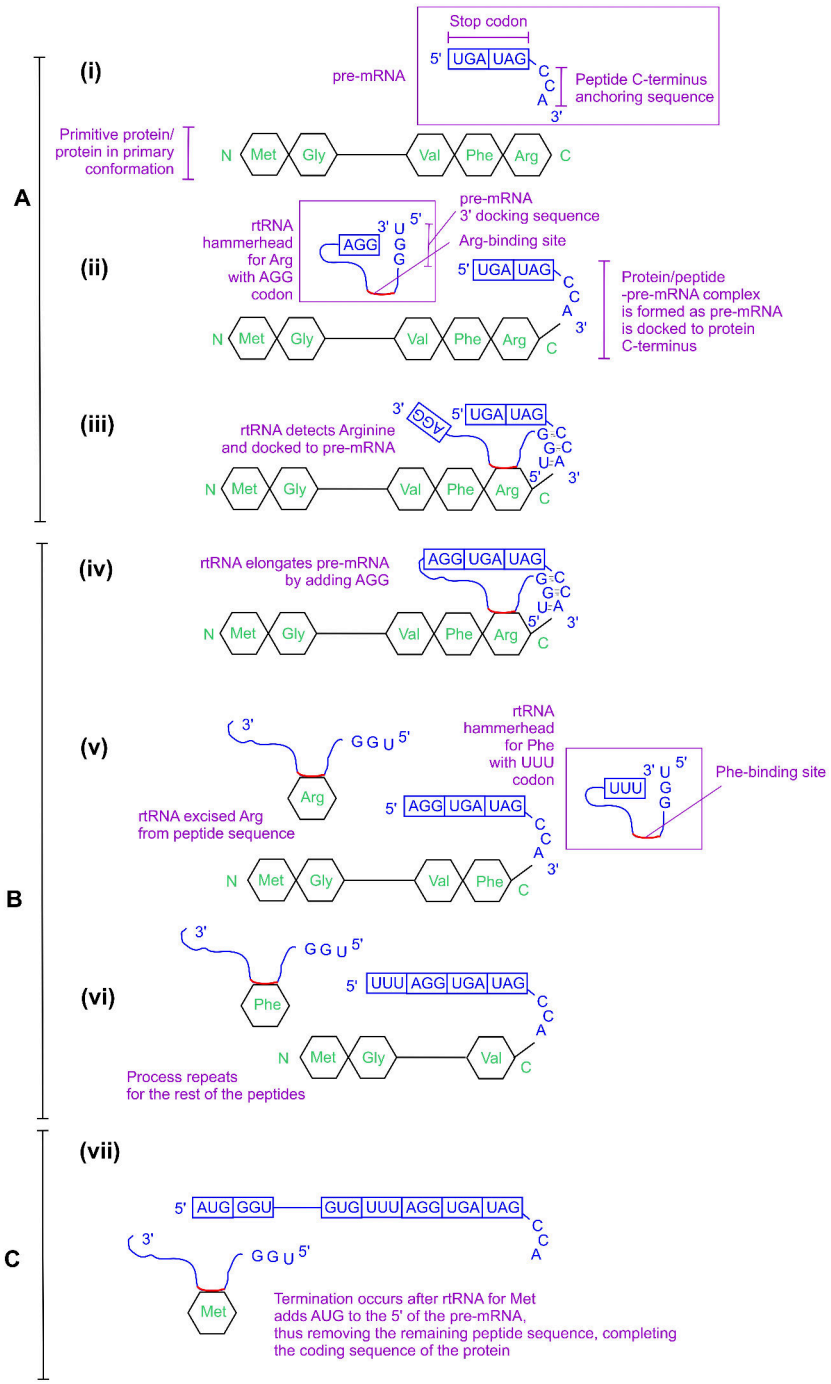

Figure 4. Speculative RNA-mediated reverse translation on a primitive protein or simple protein in its primary structure. (A) The process starts with the initiation step, where the protein polypeptide strand and pre-mRNA are forming a complex (i-ii) and the first rtRNA recognizes the first amino acid/peptide on the C-terminus by docking on it (iii). (B) Elongation begins when, according to the amino-peptide detected by the rtRNA, a codon is added to the pre-mRNA (iv) followed by excision of the first C-terminus peptide (v). This process is repeated for the next peptide in the protein sequence (vi). (C) Termination occurs after rtRNA detected the last amino acid, methionine (Met), in the polypeptide strand, and the start codon (AUG) is added to the pre-mRNA (vii), thus completing the mRNA coding sequence. Illustration inspired by and modified from Nashimoto [13].

a long polypeptide chain. To potentially counter this problem, a modified SELEX procedure that detects a binding site for amino acids in both the C and N terminus of an oligo/polypeptide can be suggested. However, unlike the individual amino acid SELEX procedure, this would lead to the recognition of only partial amino acid structures within an oligo/polypeptide sequence (instead of the entire structure as for free amino acids). In turn, this will affect the binding force between the amino acid molecules and the RNA nucleotides in the binding site. Information reversal poses another problem, which will be discussed in section 4.1.

The newest suggestion in reverse translation has been described by Martin [38] in a patent claim (US 7,169,894 B2). The patent refers to critical steps not described in Nashimoto's paper on reverse translation [13], some of which are re-addressed in this review: (i) The mechanism of RNA synthesis through amino acid binding, (ii) the mechanism to "read" the peptide/protein in sequence (related to rtRNA binding of an amino acid in sequence), and (iii) the unfolding of folded proteins for the reverse translation process. The reverse translation element (RTE), a key element in the reverse translation process, was described in the patent as polypeptide degradation (via Edman or other chemical degradation) to acquire a single terminal (C or N terminal) amino acid derivative. The RTE is described as a macromolecule, macromolecular assemblies, or particles (could be molecularly imprinted polymers or antibody-based structure). Reverse translation starts via exposure of the cleaved amino acid derivative to the RTE mixture and various single codons, followed by RTE binding to the amino acid derivative and encoding codon derivative, and lastly, polynucleotide synthesis [38]. The problem with this idea is that details regarding the RTE chemical component are unclear (or perhaps not fully described yet). The patent only mentions that RTE can form a suspension and that it can non-covalently and specifically bind to one or more copies of the cleaved amino acid derivative as well as its corresponding codon.

Physical barriers can also limit the overall informational transfer process between amino acid sequences, in proteins, and RNA or DNA nucleotides. These barriers include the asymmetrical transfer size between 20 amino acids and their 64 codons, the robustness of amino acid peptide bonds in the protein compared to the phosphodiester bonds in the nucleic acids, and the unevenness of peptide R-groups compared to the uniformity of the four nucleobases of nucleic acids (rendering it the R-groups hard to be detected). Finally, post-translational modifications of proteins may reduce the information that had been present in originally translated proteins.

4. Limitations

4.1. Data transfer asymmetry

As mentioned earlier, Nashimoto's rtRNA SELEX process entails a genetic code limitation [13]. There are 64 natural combinations of nucleotides that create codon triplets which can be translated into 20 proteinogenic amino acids and three stop codons. For reverse translation to occur, those 20 amino acids should be returned into 64 codons. However, in Nashimoto's reverse translation [13], the three stop codons are used as "primers" for reverse translation (Figure 4) and the 20 amino acids will be directly translated into the 20 codons. Therefore, a way to translate the rest of the 41 codons into RNA remains necessary. According to Nashimoto (personal communication), simulating early RNA-protein symmetry with reverse translation will result in the transferring of 20 amino acids into 20 codons. To make a full reverse translation to run, a reversion into 64 codons with consideration of tRNA variation (see codon table concept in section 5) reflecting different protein foldings [39] could be achieved with several tRNA even if they carry the same

amino acids. However, translational folding associated with tRNA abundance poses a problem. Hypothetically, reverse translation might thus determine which codons should fit into the mRNA sequence by detecting certain folds in the protein structure.

4.2. Peptide bond stability

The first problem of protein transfer is the existence of stable peptide bonds between amino acids in the polypeptide chain (compared to the phosphodiester bonds in nucleic acids). Although the C–O bond in an ester and the C–N bond in an amide entail 358 kJ mol^{-1} and 305 kJ mol^{-1} of energy, respectively [40], these bonds differ in their favor to hydrolysis. During hydrolysis, a phosphodiester bond releases 5.3 kcal mol^{-1} (22.175 kJ mol^{-1}) of free energy (Gibb's energy, or ΔG) [41], whereas a peptide bond releases c.a. 2 kcal mol^{-1} to 4 kcal mol^{-1} (8.368 kJ mol^{-1} to 16.736 kJ mol^{-1}) of ΔG [42]. In that sense, the peptide bond releases less energy which indicates that the bond is closer to equilibrium and is, therefore, more stable.

In a biological system, it is likely that a catalyst (like proteolytic enzymes) could be a ribozyme. Although no proteolytic ribozyme has been discovered or created thus far, it would hypothetically be able to break a polypeptide into amino acids during reverse translation. It could potentially work as a separated set of enzymes or be integrated into the reverse-translation system. For example, mimicking the two-sided functionality of the reverse transcriptase HIV-1 enzyme, one part (the polymerase) could revert RNA into DNA while another part (the nuclease) could break the RNA apart [43]. Thus, in creating a big protein scaffold, one could potentially revert protein to RNA while a protease hydrolyzes the peptide bonds of the polypeptide. Yet, once again, it would be highly difficult to recognize and transfer varied peptide R-groups back into nucleotide codons (*i.e.*, aside from protein cleaving and amino acid-to-codon reversal, this process would be similar to Martin's RTE machinery [38]).

4.3. Variation in protein R-groups, peptide pairing, and structural recognition

Proteins are comprised of various amino acids (*i.e.*, combination of 20 proteinogenic amino acids with 20 different R-group compositions). Some amino acids are formed with simple molecules (*e.g.*, a single hydrogen in glycine) while others are more complex in their structure (*e.g.*, ring structures in phenylalanine, tryptophan, and tyrosine) [44]. In comparison, there are only two groups of nucleobases in nucleotides: those with a single-ringed pyrimidine (*i.e.*, thymine (T), U, and C) and those with a double-ringed purine (*i.e.*, adenine (A) and guanine (G)) [45]. These monomeric variation differences make the probing of molecules required for informational transfer difficult. In DNA and RNA, purine and pyrimidine bases are paired. As such, one base could be used as the probe for the other. During DNA replication, mRNA transcription (where template DNA nucleotides are copied into their mRNA nucleotide pairs), and the translation of mRNA codons with tRNA anticodons. Other paired examples include the reverse transcription of RNA templates into cDNA [46] and the replication of RNA using RdRP (where one RNA strand serves as a template which results in another RNA strand) [47]. However, as no amino acids are paired, this principle does not apply to protein or polypeptide sequences. The intermolecular bonds between amino acids (found in secondary and tertiary protein structures) can be formed by hydrogen bonds, van der Waals interactions, hydrophobic interactions, salt/ionic bridges, and disulfide bridges [48].

Structural recognition problems will lead to difficulty in the building of peptide-recognition components such as Nashimoto's rtRNA peptide-binding nucleotide sequence [13], Mekler's antigenic surface [20], and Martin's RTE surface [38]. During normal translation, aminoacyl-

tRNA synthetase adds peptides to the 3' end of tRNA and this pairing is performed with high accuracy. The latter is supported by the structurally maintained positioning of the tRNA acceptor end and the anticodon loop (and sometimes the variable loop), the corresponding amino acid, and the ATP inside the aminoacyl-tRNA synthetase binding pocket domains [49]. Another factor is the optimal ratio between aminoacyl-tRNA synthetase and tRNA concentrations inside a living cell [50]. Furthermore, aminoacyl-tRNA synthetase possesses proofreading capabilities whereby mismatched amino acids can be hydrolyzed either before or after their transfer to tRNA [51]. In a cocktail of RNA nucleotides and proteins or peptides (whether single or as oligopeptides) inside a living cell, a perfect balance in concentration, a rate of structural recognition by an enzymatic or RNA nucleotide surfaces, and an error proofreading mechanism must all co-occur. Therefore, for reverse translation to be biologically viable, equally complex mechanisms will have to exist or will have to be incorporated.

4.4. Post-translational protein modifications

Following the ribosomal translational process, a subset of proteins will undergo certain modifications. Polypeptides will be folded and reshaped into their secondary and tertiary forms (with some tertiary structures later merging to form a quaternary structure). The folding of proteins is heavily affected by their internal peptide sequences (peptide hydrophobic effects are mainly responsible) but is also supported by electrostatic interactions (*e.g.*, salt bridging) to maintain structural flexibility, stability, and functionality [48]. Since proteins are linear and long unbranched polymers, their folding process is reversible [52]. A Nobel-winning experiment by Anfinsen [53] revealed that protein ribonuclease can be unfolded (when subjected to 8 M urea and β-mercaptoethanol) and later refolded into a new isomer structure (via random cross-linking). Another experiment by Yuan et al. [54] allowed for the "unboiling" of an egg (*i.e.*, protein reversion to its native form) by applying shear stress using a vortex fluid device. Since the folded protein structure presents a problem in reverse translation (as previously discussed), these methods of protein denaturation or unfolding could potentially be used to obtain a linear protein template for reverse translation. Unfortunately, another significant hurdle toward successful reverse translation exists in proteolysis of the protein sequence.

Partial proteolysis (*i.e.*, reducing the protein size with new C- and N-termini) is one process of protein maturation. As a result, the protein can be either activated, inactivated, has its function altered, or excised with growth factors to allow for a wide variety of biological regulations [55]. Such excision by proteolysis makes it difficult to predict protein function by only reading the nucleic acid transcript (mRNA) or by studying the newly translated polypeptide. The informational loss by proteolysis thus reduces the information available for reverse translation. Even if the latter could be achieved, the resulting RNA would only be partially complete compared to the original transcript. This implies that *in vivo* (if not *in vitro*) reverse translation would greatly reduce the complexity of biological life and revert evolutionary progress. However, to make reverse translation workable as an *in vitro* process, all the limitations described in this section would still require much study and experimental work.

Regarding reverse translation, one hopeful approach is the perception of protein as a digital code and not an analog code as previously suggested by Koonin [7]. While the latter code would refer to a protein in terms of its structures (*i.e.*, both sequences and more complex structures that direct its function), the digital code refers only to its peptide sequence. Long sequenced functional RNA molecules (just like proteins) often form secondary, tertiary, and even quaternary structures while nucleic acids comprise sequences made up of four nucleobases in triplet combinations or codons (*i.e.*, the templates of protein information). Proteins as digital codes are thus peptide sequences

with information that can be useful in omics and structural bioinformatics. In bioinformatics, a forward-reverse informational flow supports the mapping of sequences between nucleic acids and proteins. When applied across different organisms, this creates a useful tool for phylogenetic analyses. This forward-reverse informational flow can also be incorporated into omics databases or be used for structural predictions in computational biology.

5. Reverse translation as a bioinformatics tool

Bioinformatics is an emerging branch of biology that utilizes computational tools to simulate biological processes by using information from biological databases (*i.e.*, extracted *de novo* from sampled living organisms, or collected from previous studies and stored in an online database). Previously, Edmann's degradation could be used for peptide identification (using a sequencing machine) [56]. Today, protein sequencing can be performed using tandem mass spectrometry [57]. Some software programs are also able to convert peptide codes into nucleic acid sequences. Thus, in a scenario where a protein sequence is known, but not that of the RNA or DNA, reverse translation could be useful to determine gene or transcript sequences. For instance, a paleogenomics study using a preserved mammoth specimen taken from permafrost in Russia compared the mammoth genome with that of modern biota [58]. Since protein fragments are often found in much older fossil samples, *e.g.*, with a highly preserved Early Jurassic era (195 million years old) collagen sample from a sauropodomorph dinosaur, *Lufengosaurus* [59], reverse translation of the peptide fragment to a nucleic acid sequence could potentially be incorporated in comparative paleomics studies (including paleoproteomics, paleogenomics, and even paleometagenomics). Current advances in sequencing and identification technologies hold great potential (especially since these advanced methods are more accurate in determining sequences). However, further development and integration of machine learning algorithms with online databases could also support computational-based reverse translation.

Older bioinformatics software such as DNASTAR EditSeq (previously LASERGENE EditSeq [60]) reverse translated proteins and predicted codons, although the reversal algorithm was undescribed. Today, many include tools for reverse translation. For example, the Sequence Manipulation Suite (https://www.bioinformatics.org/sms2/rev_trans.html) will convert a protein sequence into a DNA sequence and the user can even enter their desired codon table (in GCG format) as a template upon sequence submission. Alternatively, a default codon table from the coding sequence of *Escherichia coli* (available on GenBank) and the Kazusa's Codon Usage Database (https://www.kazusa.or.jp/codon) [61] is employed. The latter is a composed list of each codon's total quantity in the selected organism genome, the number of codons per 1,000 nucleotides, and the fraction of codons. After input and submission of the protein peptide sequence (in FASTA format), the website will generate two DNA sequences (also in FASTA format) based on the most likely codons and the consensus codons (while also listing base probabilities throughout the sequences). A similar open web service is provided by the European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI), namely EMBOSS Backtranseq (which also uses selected parameters from its own codon usage table; https://www.ebi.ac.uk/Tools/st/emboss_backtranseq) and EMBOSS Backtranambig for the back translation of specifically ambiguous nucleotide sequences (using selected parameters of the updated codon table of NCBI's Taxonomy Browser; https://www.ebi.ac.uk/Tools/st/emboss_backtranambig). Other software such as Qiagen's CLC Main Workbench also supports reverse translation features with codon parameters such as random codons, only the most frequent codon, codons based on the frequency distribution of biological organisms (also based on Kazusa's Codon Usage Database, or using custom codon frequency tables). All these bioinformatics reverse translation tools share the same principle:

the use of codon-usage tables with listed frequencies of codons. Then, based on probabilities, these tools will convert a peptide sequence to a DNA sequence. Previous studies that used reverse translation tools include those on the development of non-replicating mRNA and self-amplifying mRNA vaccine candidates where the epitope can be combined or manipulated before being back-translated to DNA sequences for production against SARS-CoV-2 and HPV [62, 63, 64].

Other than vaccine development, or the similar modification and production of short peptide sequences, bioinformatics could be used for the development of nucleic acid and protein engineering. For example, methods combining directed evolution with omics databases [65] and protein structure databases [66] have been described. Applications may include the development of precision medicines [65, 67], protein engineering [66], and nucleic acid aptamers [68] (although these applications are independent of reverse translation). As described in section 3, machine learning may be used to simulate and predict RNA sequences that will bind to specific amino acids with the highest binding affinity (*i.e.*, using SELEX, but *in silico* version), and later to create the supporting organic environment for the *in vitro* testing for reverse translation itself. Overall, more research is required to close the scientific gaps around reverse translation possibilities, and future attempts to recreate or innovate reverse translation might also unlock new knowledge.

6. Conclusions

Reverse translation, the informational flow between proteins and nucleic acids, remains a mysterious phenomenon. This off-limits concept in biology bridges the less complex nucleic acids code (*i.e.*, the digital code) and the complex peptide code (*i.e.*, the analog code) [7]. However, future *in vitro* studies will have to complete Nashimoto's rtRNA peptide-binding nucleotide sequence [13], Mekler's antigenic surface [20], and Martin's RTE surface [38]. The resulting studies will also likely give rise to equally important side products (*e.g.*, to understand how ancient Earth model of RNA-peptide interaction system, to make a *de novo* RNA-peptide molecular interaction machinery, and for *in silico* modeling) and could support the development of ribozymes and the engineering of aptamers or peptide-nucleic acid recognition molecules. In addition, reverse translation bioinformatics tools could be improved for better and more accurate sequence prediction. Furthermore, these tools could be useful for protein engineering (providing a more direct approach than that of existing databases). In nature, the protein itself can be an intriguing machine. By learning more about unique proteins like prions, we can better understand how some proteins can epigenetically affect the synthesis of other proteins. Ultimately, however, reverse translation is mainly useful in the form of bioinformatics (especially for studies that involve short peptide sequences). Based on the information thus far, not much can be done for *in vitro* reverse translation in the future, let alone *in vivo* reverse translation. Nevertheless, as the fields of molecular biology and bioinformatics advance, it is likely that "reverse translation" might gain further focus or be redefined. Similar to how program-based tools on proteins, nucleic acids, or bioinformatics are currently useful in answering old problems from the ancient RNA world, the development of new molecular machinery and tools for biomolecular engineering in the future might extend biology's central dogma in new ways.

7. Acknowledgements

The authors thank Prof. Dr. Masayuki Nashimoto of Niigata University of Pharmacy and Applied Life Sciences, Japan, for the early insights and discussions. Our thanks also go to Andamar Pradipta, M.A. from the i3L Community Language Center for his proofreading of this manuscript early stage.

8. Conflict of interest

All authors declare having no conflict of interest to disclose.

References

[1] Crick F. On Protein Synthesis. *Symposia of the Society for Experimental Biology,* 12: 139–163, 1958.

[2] Crick F. Central dogma of molecular biology. *Nature*, 227(5258): 561–563, 1970.

doi: 10.1038/227561a0

[3] Baltimore D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, 226(5252): 1209–1211, 1970.

doi: 10.1038/2261209a0

[4] Temin HM, Mizutani S. RNA-dependent DNA polymerase in virions of *Rous sarcoma virus*. *Nature*, 226(5252): 1211–1213, 1970.

doi: 10.1038/2261211a0

[5] Baltimore D, Eggers HJ, Franklin RM, Tamm I. Poliovirus-induced RNA polymerase and the effects of virus-specific inhibitors on its production. *Proceedings of National Academy of Science USA*, 49(6): 843–849, 1963.

doi: 10.1073/pnas.49.6.843

[6] Koonin EV. Does the central dogma still stand? *Biology Direct*, 7(1): 27, 2012.

doi: 10.1186/1745-6150-7-27

[7] Koonin EV. Why the Central Dogma: on the nature of the great exclusion principle. *Biology Direct*, 10(1): 52, 2015.

doi: 10.1186/s13062-015-0084-3

[8] Rich A. On the problems of evolution and biochemical information transfer. In: Kasha M., Pullman B. (eds.). *Horizons in Biochemistry.* Academic Press; New York, NY, USA: 1962. pp. 103–126

[9] Lehman N. The RNA World: 4,000,000,050 years old. *Life*, 5(4): 1583–1586, 2015.

doi: 10.3390/life5041583

[10] Gilbert W. Origin of life: The RNA world. *Nature*, 319(6055): 618, 1986.

doi: 10.1038/319618a0

[11] Bernhardt HS. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). *Biology Direct*, 7(1): 23, 2012.

doi: 10.1186/1745-6150-7-23

[12] Forterre P. The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochimie*, 87(9-10): 793–803, 2005.

doi: 10.1016/j.biochi.2005.03.015

[13] Nashimoto M. The RNA/protein symmetry hypothesis: experimental support for reverse translation of primitive proteins. *Journal of Theoretical Biology*, 209(2): 181–187, 2001.

doi: 10.1006/jtbi.2000.2253

[14] Lee DH, Granja JR, Martinez JA, Severin K, Ghadiri MR. A self-replicating peptide. *Nature*, 382(6591): 525–528, 1996.

doi: 10.1038/382525a0

[15] Beekes M, McBride PA. The spread of prions through the body in naturally acquired transmissible spongiform encephalopathies. The FEBS Journal, 274(3): 588–605, 2007.

doi: 10.1111/j.1742-4658.2007.05631.x

[16] Chiti F, Dobson CM. Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, 75: 333–366, 2006.

doi: 10.1146/annurev.biochem.75.101304.123901

[17] Ridley RM. What Would Thomas Henry Huxley Have Made of Prion Diseases? In: Baker HF (ed.). *Molecular Pathology of the Prions*; Humana Press Inc., NY, 2001; Vol. 59, pp.1-16.

[18] Serio TR, Cashikar AG, Kowal AS, Sawicki GJ, Lindquist SL. Self-perpetuating changes in Sup35 protein conformation as a mechanism of heredity in yeast. *Biochemical Society Symposia*, 68: 35–43, 2001.

doi: 10.1042/bss0680035

[19] Serio TR, Lindquist SL. [PSI+]: an epigenetic modulator of translation termination efficiency. *Annual Review of Cell and Development Biology*, 15: 661–703, 1999.

doi: 10.1146/annurev.cellbio.15.1.661

[20] Mekler LB. Mechanism of biological memory. *Nature*, 215(5100): 481–484, 1967.

doi: 10.1038/215481a0

[21] Cook ND. The case for reverse translation. *Journal of Theoretical Biology*, 64: 113–135, 1977.

doi: 10.1016/0022-5193(77)90116-3

[22] Tawfik DS, Gruic-Sovulj I. How evolution shapes enzyme selectivity–lessons from aminoacyl-tRNA synthetases and other amino acid utilizing enzymes. *The FEBS Journal*, 287(7): 1284–1305, 2020.

doi: 10.1111/febs.15199

[23] Symons RH. Small catalytic RNAs. *Annual Review of Biochemistry*, 61: 641–671, 1992.

doi: 10.1146/annurev.bi.61.070192.003233

[24] Connell GJ, Illangesekare M, Yarus M. Three small ribooligonucleotides with specific arginine sites. *Biochemistry*, 32(21): 5497–5502, 1993.

doi: 10.1021/bi00072a002

[25] Ellington AD, Szostak JW. *In vitro* selection of RNA molecules that bind specific ligands. *Nature*, 346(6287): 818–822, 1990.

doi: 10.1038/346818a0

[26] Tuerk C, Gold L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science*, 249(4968): 505–510, 1990.

doi: 10.1126/science/2200121

[27] Krokhotin A, Houlihan K, Dokholyan NV. iFoldRNA v2: folding RNA with constraints. *Bioinformatics*, 31(17): 2891–2893, 2015.

doi: 10.1093/bioinformatics/btv221

[28] Bowser MT. SELEX: Just another separation? *Analyst*, 130(2): 128–130, 2005.

doi: 10.1039/B412492H

[29] Famulok M, Szostak JW. Stereospecific recognition of tryptophan agarose by *in vitro* selected RNA. *Journal of American Chemical Society*, 114(10): 3990–3991, 1992.

doi: 10.1021/ja00036a065

[30] Gilbert SD, Rambo SP, Van Tyne D, Batey RT. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nature Structural & Molecular Biology*, 15(2): 177–182, 2008.

doi: 10.1038/nsmb.1371

[31] Lu C, Smith, A.M., Fuchs, R.T., Ding, F., Rajashankar, K., Henkin, T.M., Ke, A. Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nature Structural & Molecular Biology*, 15(10): 1076–1083, 2008.

doi: 10.1038/nsmb.1494

[32] Montange RK, Batey RT. Structure of the S-adenomethionine riboswitch regulatory mRNA element. *Nature*, 441(7097): 1172–1175, 2006.

doi: 10.1038/nature04819

[33] Garst AD, Heroux A, Rambo RP, Batey RT. Crystal structure of the lysine riboswitch regulatory mRNA element. *Journal of Biological Chem*istry, 283(33): 22347–22351, 2008.

doi: 10.1074/jbc.C800120200

[34] Serganov A, Huang L, Patel DJ. Structural insight into amino acid binding and gene control by a lysine riboswitch. *Nature*, 455(7217): 1263–1267, 2008.

doi: 10.1038/nature07326

[35] Yang Y, Kochoyan M, Burgstaller P, Westhof E, Famulok F. Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science*, 272(5266): 1343–1346, 1996.

doi: 10.1126/science.272.5266.1343

[36] Stefaniak F, Bujnicki JM. AnnapuRNA: A scoring function for predicting RNA-small molecule binding poses. *PloS Computational Biology*, 17(2): p.e1008309, 2021.

doi: 10.1371/journal.pcbi.1008309

[37] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2): 455–461, 2010.

doi: 10.1002/jcc.21334

[38] Martin MT. Methods and compositions for reverse translation. United State Patent, no: US 7,169,894 B2 (Jan 30, 2007).

[39] Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature Structural & Molecular Biology*, 16(3): 274–280, 2009.

doi: 10.1038/nsmb.1554

[40] Cottrell TL. The strengths of chemical bonds (2^{nd} ed). Butterworths, London 1958.

[41] Dickson KS., Burns CM, Richardson JP. Determination of the free-energy change for repair of a DNA phosphodiester bond. *Journal of Biological Chem*istry, 275(21): 15828–15831, 2000.

doi: 10.1074/jbc.M910044199

[42] Martin RB. Free energies and equilibria of peptide bond hydrolysis and formation. *Biopolymers*, 45(5): 351–353, 1998.

doi: 10.1002

[43] Tu X., Das K, Han Q, Bauman JD, Clark Jr, AD, Hou, X, Frenkel YV, Gaffney BL, Jones RA, Boyer PL, Hughes SH. Structural basis of HIV-1 resistance to AZT by excision. *Nature Structural & Molecular Biology*, 17(10): 1202–1209, 2010.

doi: 10.1038/nsmb.1908

[44] Alhazmi HA. Mobility shift-affinity capillary electrophoresis for investigation of protein-metal ion interactions: aspects of method development, validation and high throughput screening. Doctorate Thesis, Faculty of Life Sciences, Technische Universität Braunschweig, Germany. p.2.

[45] Bakış Y, Otu HH, Sezerman OU. Inferring Phylogenies from Physico Chemical Properties of DNA. *American Journal of Bioinformatics Research*, 2(1): 1–6, 2012.

doi: 10.5923/j.bioinformatics.20120201.01

[46] Andréola ML, Parissi V, Litvak S. 2013. DNA Polymerases: Reverse Transcriptase Integrase, and Retrovirus Replication. In: Lennarz WJ, Lane MD (eds.). *Encyclopedia of Biological Chemistry (2nd edition)*. Academic Press, Massachusetts, USA. pp.101–107, 2013.

doi: 10.1016/B978-0-12-378630-2.00258-9

[47] Ahlquist P. RNA-dependent RNA polymerases, viruses, and RNA silencing. *Science*, 296(5571): 1270–1273, 2002.

doi: 10.1126/science.1069132

[48] Kumar S, Nussinov R. Close-range electrostatic interactions in proteins. *ChemBioChem*, 3(7): 604–617, 2002.

doi: 10.1002/1439-7633(20020703)

[49] McClain WH. Rules that Govern tRNA Identity in Protein Synthesis. *Journal of Molecular Biology*, 234(2): 257–280.

doi: 10.1006/jmbi.1993.1582

[50] Swanson R, Hoben P, Sumner-Smith M, Uemura H, Watson L, Söll D. Accuracy of in Vivo Aminoacylation Requires Proper Balance of tRNA and Aminoacyl-tRNA Synthetase. *Science*, 242(4885): 1548–1551.

doi: 10.1126/science.3144042

[51] Shen C-H. *Diagnostic Molecular Biology*, Academic Press, Massachusetts, USA. pp.87–116, 2019.

doi: 10.1016/B978-0-12-802823-0.00004-3

[52] Rose GD, Fleming PJ, Banavar JR, Maritan A. A backbone-based theory of protein folding. *Proceedings of the National Academy of Sciences of the United States of America* 103(45): 16623–16633.

doi: 10.1073/pnas.0606843103

[53] Anfinsen CB. Principles that Govern the Folding of Protein Chains. *Science*, 181(4096): 223–229.

doi: 10.1126/science.181.4096.223

[54] Yuan TZ, Ormonde CF, Kudlacek ST, Kunche S, Smith JN, Brown WA, Pugliese KM, Olsen TJ, Iftikhar M, Raston CL, Weiss GA. Shear-Stress-Mediated Refolding of Proteins from Aggregates and Inclusion Bodies. *ChemBioChem*, 16(3): 393–396, 2015.

doi: 10.1002/cbic.201402427

[55] Rogers LD, Overall CM. Proteolytic Post-translational Modification of Proteins: Proteomic Tools and Methodology. *Molecular & Cellular Proteomics*, 12(12): 3532–3542.

doi: 10.1074/mcp.M113.031310

[56] Edman P, Begg G. A protein sequenator. *European Journal of Biochemistry*, 1: 80–91, 1967.

doi: 10.1007/978-3-662-25813-2_14

[57] Medzihradszky KF, Chalkley RJ. Lessons in de novo peptide sequencing by tandem mass spectrometry. *Mass Spectrometry Reviews*, 34(1): 43–63, 2015.

doi: 10.1002/mas.21406

[58] Poinar HN, Schwarz C, Qi J, Shapiro B, MacPhee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, Rampp M. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science*, 311(5759): 392–394, 2006.

doi: 10.1126/science.1123360

[59] Lee YC, Chiang CC, Huang PY, Chung CY, Huang TD, Wang CC, Chen CI, Chang RS, Liao CH, Reisz RR 2017. Evidence of preserved collagen in an Early Jurassic sauropodomorph dinosaur revealed by synchrotron FTIR microspectroscopy. *Nature Communications*, 8: 14220.

doi: 10.1038/ncomms14220

[60] Arnold C., Clewley JP. From ABI Sequence Data to LASERGENE's EDITSEQ. In: Swindell SR. (eds) *Sequence Data Analysis Guidebook. Methods In Molecular Medicine™*, Springer, Totowa, 70: 65–74, 1997.

doi: 10.1385/0-89603-358-9:65

[61] Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from the international DNA sequence databases. *Nucleic Acids Research*, 28(1): 334–334, 1998.

doi: 10.1093/nar/26.1.334

[62] Bhattacharya M, Sharma AR, Ghosh P, Patra P, Patra BC, Lee SS, Chakraborty C. Bioengineering of novel non-replicating mRNA (NRM) and self-amplifying mRNA (SAM) vaccine candidates against SARS-CoV-2 using immunoinformatics approach. *Molecular Biotechnology*, 64(5): 510–525, 2022.

doi: 10.1007/s12033-021-00432-6

[63] Hosseini NG, Tebianian M, Farhadi A, Khani, AH, Rahimi A, Mortazavi M, Hosseini SY, Taghizadeh M, Rezaei M, Mahdavi M. *In silico* analysis of L1/L2 sequences of human papillomaviruses: implication for universal vaccine design. *Viral Immunology*, 30(3): 210–223, 2017.

doi: 10.1089/vim.2016.0142

[64] Yazdani Z, Rafiei A, Yazdani M, Valadan R. 2020. Design an efficient multi-epitope peptide vaccine candidate against SARS-CoV-2: an *in silico* analysis. *Infection and Drug Resistance*, 13: 3007–3022, 2020.

doi: 10.2147/IDR.S264573

[65] Li R, Li L, Xu Y, Yang J. 2022. Machine learning meets omics: applications and perspectives. *Briefings in Bioinformatics*, 23(1): bbab460, 2022.

doi: 10.1093/bib/bbab460

[66] Verma R, Schwaneberg U, Roccatano D. Computer-aided protein directed evolution: a review of web servers, databases and other computational tools for protein engineering. *Computational and Structural Biotechnology Journal*, 2(3): e201209008, 2012.

doi: 10.5936/csbj.201209008

[67] Lheureux S, Braunstein M, Oza AM. Epithelial ovarian cancer: evolution of management in the era of precision medicine. *CA: A Cancer Journal for Clinicians*, 69(4): 280–304, 2019.

doi: 10.3322/caac.21559

[68] Kinghorn AB, Fraser LA, Liang S, Shiu SCC, Tanner JA. 2017. Aptamer bioinformatics. *International Journal of Molecular Sciences*, 18(12): 2516.

doi: 10.3390/ijms18122516

Nuevas perspectivas sobre traducción inversa: breve historia y estado actual

Resumen: Desde los años 50s del siglo XX, la traducción inversa ha sido una parte enigmática del dogma central de la biología molecular de Crick. Podría describirse como la posibilidad de retro-traducir información de proteínas a ácidos nucleicos (o codones). Unos pocos estudios han intentado teorizar y/o conducir experimentos *in vitro* para comprobar la probabilidad de la traducción inversa, con ideas que frecuentemente involucran la creación de sitios de reconocimiento de péptidos que actúan como puentes entre el péptido y el codón. Sin embargo, debido a muchas restricciones, incluida una transferencia asimétrica de información, la estabilidad de los enlaces proteína-péptido, la no-uniformidad estructural de los grupos R de las proteínas y la pérdida de información en las modificaciones postraduccionales de las proteínas, este concepto requiere estudios posteriores. De otro lado, las herramientas bioinformáticas actuales que se basan en programas computacionales y en bases de datos biológicas representan una rama en crecimiento de la biología. La traducción inversa basada en bioinformática puede utilizar tablas de uso de codones para predecir codones a partir de sus contrapartes peptídicas. Además, el desarrollo de herramientas de aprendizaje automático podría permitir la exploración de la traducción inversa biológica *in vitro*. Así, mientras la traducción inversa *in vivo* parecer ser casi imposible (debido a su complejidad biológica), estudios biológicos y bioinformáticos relacionados pueden ser útiles para entender mejor el dogma central de transferencia de información y para desarrollar maquinaria biológica más compleja.

Palabras Clave: aminoácidos; dogma central; transferencia genética; polipéptidos; ARN

Novas Perspectivas sobre Tradução Reversa: Breve História e Atualizações

Resumo: Desde a década de 1950, a tradução reversa tem sido uma parte enigmática do dogma central da biologia molecular proposto por Crick. Esta pode ser descrita como a possibilidade de retro traduzir a informação das proteínas para os ácidos nucleicos (ou códons). Alguns estudos tentaram teorizar e/ou conduzir experimentos *in vitro* para testar a probabilidade de tradução reversa, com ideias baseadas na criação de locais de reconhecimento de peptídeos para ligar o peptídeo com o códon. No entanto, devido a muitas restrições, incluindo a transferência informacional assimétrica, a estabilidade das ligações proteína-peptídeo, a não uniformidade estrutural dos grupos R das proteínas e a perda informacional durante as modificações proteicas pós-translacionais, esse conceito requer de mais estudos. Por outro lado, as ferramentas bioinformáticas atuais que dependem de programas computacionais e bancos de dados biológicos representam um ramo crescente da biologia. A tradução reversa baseada em bioinformática pode utilizar tabelas de uso de códons para predizer os códons a partir das suas contrapartes peptídicas. Além disso, o desenvolvimento de ferramentas de aprendizado de máquina pode permitir a exploração da tradução reversa biológica *in vitro*. Assim, enquanto a tradução reversa *in vivo* parece ser quase impossível (devido à complexidade biológica), estudos biológicos e bioinformáticos podem ser úteis para entender melhor a transferência informacional do dogma central e desenvolver máquinas biológicas mais complexas.

Palavras-chave: aminoácidos; Dogma central; Transferência genética; Polipeptídeo; RNA

Adhityo Wicaksono Ph. D, MRSB is the Head of Biotechnology Division at Genbinesia Foundation, Indonesia. He graduated on his doctorate in the laboratory of natural material technology of the Faculty of Science and Engineering, Åbo Akademi University, Finland. Currently working as a postdoctoral researcher in Chulalongkorn University, Thailand on plant gene expression study via bioinformatics and wet lab. He interested in multiomics study and molecular biology of nucleic acids.

ORCID: 0000-0003-3298-3577

Viol Dhea Kharisma M.Sc has completed a Bachelor of Biology, then continued his studies and obtained a Master of Biology at Universitas Brawijaya, Indonesia. He is a researcher on the SARS-CoV-2 vaccine design by in silico approach and revealing of antiviral agent from natural compound. He is also the Research Scientist of Computational Virology Research Unit, Division of Molecular Biology and Genetics, Genbinesia Foundation, Indonesia.

ORCID: 0000-0001-9060-0429

Arli Aditya Parikesit Assoc. Prof. Dr.rer.nat. is the Faculty Member of Bioinformatics Department at i3L. He finished his bachelor's and master's studies in the Chemistry Department, Faculty of Mathematics and Sciences, University of Indonesia. He was awarded the DAAD fellow to conduct doctorate research at Faculty of Informatics and Mathematics, University of Leipzig, Germany. Dr. Arli is also an expert on immunoinformatics, in silico drug design, and in silico transcriptomics.

ORCID: 0000-0001-8716-3926