

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Cumulative Specificity: A Universal Mechanism for the Initiation of Protein Synthesis

---

Tokumasa Nakamoto and Ferenc J. Kezdy

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/51179>

---

## 1. Introduction

The main problem in the initiation of protein synthesis is the determination of how the ribosome recognizes and binds to the initiation site (IS) of the mRNA. There are currently three major hypotheses that address this problem, all differently. The Shine-Dalgarno (SD) hypothesis for prokaryotes proposed in 1974, postulates that the IS is selected by base pairing of a segment of the 3' end of the 16S rRNA of the small ribosomal subunit and a complementary segment in the leader sequence of the IS [1]. The scanning hypothesis for eukaryotes proposed in 1978, postulates that the 40S ribosome initiation complex recognizes and binds to the 5' end of the mRNA and scans the IS until it finds the initiator codon [2]. The cumulative specificity (CS) hypothesis has its origin in a 1966 proposal that provided an essentially unique accessibility of the IS proposal for prokaryotes [3], but with a number of recent modifications, it evolved into its current form in 2007 [4]. The CS mechanism postulates that the IS of the mRNA is selected by incremental ribosomal binding of the IS, the ribosomal binding subsites interacting with their respective IS subsites, one or a few subsites at a time.

An important aspect of any good hypothesis is its ability to stimulate research. The long tenure of the SD and the scanning hypotheses as the bases for numerous researches attest to the credibility and appeal of the two hypotheses. However, with manifold increase of knowledge in the field, evaluation of the current hypotheses is timely, especially the SD and the scanning hypotheses, which were proposed so long ago. Repeating, the major hypotheses are the SD proposal for prokaryotes, the scanning mechanism for eukaryotes, and the CS mechanism, for both prokaryotes and eukaryotes. On another postulated mechanism, the internal ribosome entry site (IRES) for eukaryotes [5], there is some question whether it and the scanning mechanism are indeed distinct and unconnected [6] so the former will not be considered here as a major mechanism.

The model IS of the *E. coli* mRNA, generated by computer analysis from 68 non-identical IS sequences, consists of 46-48 nucleotides with preferred bases (recognition elements) in given positions, but without a specific base in any given position except for the initiator codon [7]. This means that the IS surface is extensive, nonrigid, complex and the IS is a non-unique sequence. The IS of the eukaryotic mRNA has characteristics similar to those obtained for *E. coli*, although not documented as convincingly as in *E. coli* [8]. The extensiveness, nonrigidity, and complexity of the IS of the mRNA would make its binding to the ribosome — that is, the perfect meshing together of the large and complex surface of the IS and that of the ribosome — unlikely to occur in a single collision. Rather, the binding of the IS of the mRNA by the ribosome should occur by one or severable sub-segments at a time as proposed by the CS model reaction, which was originally developed to account for the specificity of substrate binding to the HIV Type 1 and 2 proteases [9]. The CS reaction was then proposed as a paradigm for the initiation of protein synthesis [4]. The main difference between the two above binding reactions is that in one case, an enzyme binds its substrate, and in the other, a ribosome binds a template mRNA.

The specificity sites of the protease substrates consist of a sequence of 6-8 amino acids, and the recognition signals of these substrate sites also consist of preferred recognition elements — viz. amino acids — in given positions [9]. Thus these active sites are also extended, nonrigid and complex and have non-unique sequences of amino acids and are similar in general characteristics to those described above for the IS of mRNA. The CS model reaction assumes that the protease binds the specificity site of the substrate, by a sub-segment or several sub-segments at a time. In the initial collision of the enzyme and substrate, the enzyme first binds to one or more subsites of the substrate. This process is then followed in the protease and protein synthesis reactions by sequential zipper-like juxtapositions and bindings of the appropriate, remaining subsites of the enzyme, or the ribosome, and the substrate (or the mRNA), until completion of the enzyme-substrate (ribosome-mRNA) binding.

A very important feature of the CS mechanism is that it is able to recognize as a signal, not a rigid, immovable molecular structure, but a structure variable within certain limits, namely a sequence of preferred molecular elements in given positions. The consequence of such a heterogeneous, multi-pronged binding is that the ensuing chemical transformation, or the binding of two large surfaces, does not occur with a single reaction rate, which would reflect an all-or-none specificity. Rather, the rate — and hence the specificity — spans over a wide range, favorable interactions at each subsite contributing incrementally to the overall recognition of the substrate active site or of the IS of the mRNA.

Further insights on the initiation mechanism of protein synthesis can be obtained by considering the implications on the general mechanisms of protein synthesis of the evolutionary process and logic [6,10,11]. An important implication of this process and logic is that the general mechanisms of protein synthesis, including the initiation mechanism, are expected to be universal in all domains of life. The universality of the mechanisms of protein synthesis follows from the conservation of the complex protein synthesizing apparatus, exhibiting very similar basic components, consisting of ribosome, mRNA, tRNAs (including an initiator methionyl-tRNA), aminoacyl-tRNA synthetases, a universal genetic code, and

numerous other proteins and factors. This review will examine whether the various current hypotheses consider these implications of evolution, along with the characteristics of the IS and other logical points. Such an examination will help to evaluate the comprehensiveness and reasonableness of the initiation mechanisms advanced by the various hypotheses.

## 2. Body

### 2.1. Basis for evaluating the hypotheses for the initiation of protein synthesis

The evolutionary logic, previously presented for evaluating the various hypotheses proposed for the mechanism of initiation of protein synthesis [6,10,11], should perhaps rather be thought of as an ensemble of logical considerations gleaned from general knowledge, compiled here to contribute to the evaluation of current hypotheses. We hope that these intuitive and axiomatic considerations do provide a firm foundation for the assertions and conclusions made in this chapter.

As mentioned above, evolutionary logic leads one to conclude that with the conservation of the numerous constituent elements of the complex protein synthetic apparatus, the underlying mechanisms of the synthetic process should also be conserved in all domains of life and are universal. As reviewed later, this conservation also extends in the domains of prokaryotes and eukaryotes to the similarity of mRNA initiation signals and to the ability of the ribosomes to recognize these signals [10,11]. Ribosomes can recognize the mRNA from a different domain of life only if the initiation signals are identical or very closely related and if ribosomes have also conserved their ability to recognize these signals. Taken together, these facts can indicate a conserved initiation mechanism.

The mechanism of initiation of protein synthesis must have evolved gradually and without abrupt changes, from prokaryotes to eukaryotes. Therefore, any hypothesis postulating that this evolution is accompanied by profound changes in the mechanistic characteristics of the ribosomes is rather improbable. This will have bearing in the evaluation of the scanning hypothesis for eukaryotes.

When a hypothesis postulates an exclusive and primary biological pathway — such as the pathway for the initiation of protein synthesis — then the pathway's unique initiation signals and the other essential components of the reaction, must all be obligatory. Moreover, the proposed pathway must be in accord with all experimental observations, such as the characteristics of the initiation sites. These considerations will be central in the evaluation of the SD hypothesis and, to a lesser degree, the other hypotheses.

Evolution is a very efficient process and is not likely to tolerate the conservation of unused information. For instance, the characteristics of the model *E. coli* initiation site suggests the presence of 46-48 nucleotides with signal character. Thus, any comprehensive initiation mechanism should have an initiation signal that encompasses the entire nucleotide sequence shown to have signal character, which includes essentially the entire IS, even the amino acid coding region. The existence and translation of non-SD and leaderless mRNAs indicate that any hypothesis that requires the recognition of only the

leader sequence cannot be a complete hypothesis. A comprehensive initiation mechanism must, therefore, be able to account for the initiation of translation of canonical, as well as, of leaderless mRNAs.

Lastly, as mentioned, the hypotheses will be evaluated on the basis of whether the characteristics of the IS and the dictum of a universal initiation mechanism were considered in formulating each respective initiation mechanism.

## 2.2. Characteristics of the Initiation Site

The characteristics of the initiation site (IS) were determined by computer analysis of 68 non-identical *E. coli* ribosome binding site sequences [7]. The analysis generated a model IS containing 46-48 nucleotides that assigns only preferred bases rather than an absolute requirement for a specific base in any given position. The only exception to the preceding is the absolute requirement of the initiator codon in a given position in every IS.

The model IS reveals important characteristics of the IS. Part of the model sequence is complementary to the 3' of the 16S rRNA, that is, its base frequency profile reflects the SD sequence. Nonetheless, there is no unique initiation sequence or even a unique SD sequence in the IS, only ensembles of preferred bases. The prokaryotic ISs thus constitute a large multiplicity of loosely related nucleotide sequences, and the protein synthesizing system, which recognizes all of them, has broad substrate specificity [4]. Nearly half of the model site covers the amino acid coding region. In other words, it is the entire IS — about one-half of it consisting of the leader region and the other half of the amino acid coding region — that serves as the ribosome recognition or initiation signal. The authors suggested that the finding that the amino acid coding region has recognition features, or signal character, might explain how leaderless mRNAs are recognized by the ribosomes.

The model *E. coli* IS is a composite of three sequences derived by analysis of the 68 *E. coli* ribosome binding site sequences using the initiator codon, the SD leader sequence and the CU at the very 3' end of the IS as the common reference in aligning the IS sequences for analysis. When the initiator codon was the common reference, the model sequence derived showed considerable ambiguity in the area where the SD sequence was expected to be located. When the SD sequence was used as the common reference and the leader region was analyzed, the AUG start codon had a variable locus of 6-9 positions in the 3' direction from the SD sequence. Examination of the sequences showed that CU seemed to be the 3' end of the model sequence. As with the SD sequence and its variable distance from the AUG codon, analysis found a completely random behavior with the CU as the common reference. A realignment of the sequences and subsequent analysis, however, showed a contiguous and almost identical model sequence between the AUG codon and the assumed CUCG end signal. The best model sequences fragments obtained from the analyses using the three common references described above were then combined to form the composite model *E. coli* initiation site.

### 2.3. The conservation of prokaryotic initiation signals and prokaryotic initiation mechanism in eukaryotes

Before discussing specific conservations, it helps to clarify the conclusion that all underlying mechanisms of the protein synthetic process should be conserved in all domains of life because of the conservation of the numerous constituent elements of the complex protein synthetic apparatus. Does this mean that the entire underlying protein synthetic mechanisms are universally conserved? We shall review below that, in addition to conservation of the physical apparatus and the underlying general mechanisms, other aspects of the process, like initiation mechanism features, such as prokaryotic initiation signals with the underlying prokaryotic initiation mechanism, are conserved in eukaryotes. This could also imply that the mechanisms of peptide bond formation, peptide chain elongation and protein chain termination could in all domains of life be identical or very similar to each other.

It may be at present convenient then, when concluding that the initiation mechanism is conserved in eukaryotes, to assume tacitly that the rest of the mechanisms of the protein synthetic pathway are conserved as well. After all, initiation signals alone do not comprise all of the initiation mechanism, and some additional ribosome functions or mechanism are needed to complete the process. In any case, this tacit assumption does not invalidate the experimental observations from which the conclusion of conservation was deduced, nor does it invalidate the deductive process.

The conservation of prokaryotic initiation signals and the prokaryotic initiation mechanism in eukaryotes deduced from experiments using heterologous systems, in which eukaryotic mRNAs were translated in *E. coli* cell-free systems and prokaryotic mRNAs were translated in eukaryotic cell-free systems, has been reviewed [6,10,11]. The experiments shed light especially on the mechanism of initiation of eukaryotic protein synthesis and provided support for a common or a universal initiation mechanism shared by prokaryotes and eukaryotes. The studies are reviewed, again, with emphasis on the significance of the conservation of initiation signals and the initiation mechanism.

Polypeptides were synthesized in an *E. coli* cell-free system with poliovirus as messenger whose tryptic digests were found to correspond to tryptic digests of authentic poliovirus proteins [12]. The tobacco mosaic virus (TMV) RNA in an *E. coli* cell-free system directed the synthesis of several discrete polypeptides, including one similar to TMV coat protein by criteria of polyacrylamide electrophoresis and peptide mapping [13]. Avian myeloblastosis viral RNA was translated by *E. coli* ribosomes to yield a protein that was antigenically identical to the group-specific antigen 4 of the virus [14]. The preceding experiments clearly show that *E. coli*, or prokaryotic, ribosomes can recognize eukaryotic viral initiation signals and translate the eukaryotic viral mRNAs. This indicates that the eukaryotic viral mRNAs contain evolutionarily conserved prokaryotic initiation signals since it is unlikely that the two sets of mRNAs would have dissimilar initiation signals and that the same ribosomes would recognize both sets of signals.



Studies were also performed using prokaryotic mRNAs in a eukaryotic cell-free system. Bacteriophage Q $\beta$  RNA, a polycistronic prokaryotic messenger, was translated in extracts of Krebs II mouse ascites cell-free system [15]. Viral coat protein was identified as the primary product by co-migration on polyacrylamide gel with authentic coat protein, and by mapping of tryptic digests. A specific mRNA for a structural lipoprotein of *E. coli* was translated in a wheat germ cell-free system [16]. Eukaryotic ribosomes can thus faithfully translate prokaryotic initiation signals and initiate translation. The translation of prokaryotic mRNAs by eukaryotic ribosomes also means that the initiation signals in prokaryotes and eukaryotes are identical or very similar, and that the eukaryotic ribosomes translate prokaryotic mRNAs by initiating translation with a prokaryotic-like mechanism, or with an evolutionarily conserved prokaryotic mechanism.

Experiments with heterologous systems that were extremely important in establishing the universality of the initiation signals involved the translation of capped prokaryotic mRNAs by eukaryotic ribosomes. In the experiments,  $\lambda$  phage 8S cro mRNA and other  $\lambda$  transcripts that still retained their prokaryotic ISs, when capped in vitro, were found to be translated in a wheat germ cell-free system as efficiently as — or even more efficiently than — naturally capped eukaryotic mRNAs [17,18]. The efficient translation of capped prokaryotic mRNAs by a eukaryotic cell-free system means that prokaryotic initiation signals are equivalent to, or as effective as, the initiation signals of naturally capped eukaryotic mRNAs. Additionally, in light of the cross heterologous translations reviewed above, the preceding can be construed as evidence that prokaryotic initiation signals with the underlying prokaryotic initiation mechanism are conserved in eukaryotes. Thus, the heterologous experiments indicate the conservation of a universal initiation mechanism, or at least, the conservation of a common initiation mechanism in the domains of the prokaryotes and the eukaryotes.

## 2.4. Ribosomal initiation complex

The ribosome is the predominant constituent of the complex protein synthesizing apparatus. In all domains of life it is composed of two subunits, one small and one large. It was assumed for some time that the subunits of the ribosome functioned in protein synthesis combined as a 70 S ribosome unit. In 1967, however, it was demonstrated that the prokaryotic 70S ribosome operates in a cycle when participating in protein synthesis. The 70S ribosome dissociates into subunits when initiating synthesis, re-associates into the 70S ribosome at the completion of initiation and during polypeptide chain elongation, and again dissociates into subunits when synthesis is complete [19]. Phage f2 RNA as well as poly AUG (1:1:1) in random sequence was shown to stimulate binding of fMet-tRNA to the 30S ribosomal subunits, but not to the 70S ribosome. The presence of the 50S subunit inhibited the binding of fMet-tRNA to the 30S subunit with phage f2 RNA and the poly AGU. It was proposed that the first step in protein synthesis is the formation of a complex consisting of the 30S subunit, mRNA, and fMet-tRNA. Later studies established that the binding reaction required the participation of GTP and 3 specific initiation protein factors [20].

Although the 70S ribosome does not bind fMet-tRNA with f2RNA or polyAGU, it does bind fMet-tRNA with the triplet AUG. This indicates that the mRNA binding site of the 70S ribosome can accommodate a small single triplet and not the larger f2 RNA or polyAGU containing an AUG codon. These observations support the view that the 30S subunit, being only about a third the size of the 70S ribosome, can access the IS more effectively. Another obvious possibility for the lower effectiveness of the 70S ribosome to bind mRNA may be due in part to the shielding by the 50S subunit in the 70S ribosome of the mRNA binding site. A major reason for the difficulty of ribosome binding the IS may be that RNA interacts intramolecularly so extensively that a ribosome binding to a nucleotide sequence of about 47 nucleotides long, which is the length of the model *E. coli* mRNA initiation site, is no easy feat. A synthetic RNA polymer containing 4 bases in equal proportions in random sequence has been found to have about 50% of its bases paired [21]. The effectiveness of the small ribosomal subunit to access the IS is probably the primary advantage of the ribosome cycle. Besides, the 70S ribosome is effective and essential for polypeptide chain elongation.

Initiation of eukaryotic protein synthesis follows the same pattern as in prokaryotes with the small ribosome subunit, the 40S ribosome, forming an initiation complex with eukaryotic initiator Met-tRNA that, however, is not formylated. The initiation of the eukaryotic protein synthesizing process requires at least 5 initiation factors and is more complex [22].

## **2.5. A Cumulative Specificity (CS) reaction model for proteases from human immunodeficiency virus (HIV) Types 1 and 2**

The substrate specificities of the aspartyl proteases from HIV Types 1 and 2 were not readily discernible. The scissile bond of the substrate chain is surrounded by at least 3-4 amino acids on each side without any specific sequence or even a specific amino acid in any given position. To gather some evidence for the basis of the specificity of these enzymes, the frequencies of amino acid distribution at each position surrounding the cleavage site of the of HIV 1 protease substrate were statistically analyzed in 40 substrates with known amino acid sequences [9]. The analysis revealed that certain amino acid residues had quite higher than normal frequencies at three subsites in addition to the positions of the amino acids directly involved in the cleavage, but there was no absolute requirement for any given specific amino acid in any of these positions. Thus, each subsite appeared to have a marked specificity toward some residues, and also a marked negative specificity, since some amino acid residues did not occur at all at those subsites. Inferring that the characteristics of the frequencies of the particular amino acids in given positions are the result of actual molecular interactions between the substrate and the active site of the enzyme, a mechanistic model was proposed to account for the broad specificity of the HIV proteases.

The model postulates that the positioning of the cleavage site with respect to the catalytic groups of the enzyme is the result of the cooperative interaction between the amino acid residues of the substrate subsites and corresponding subsites surrounding the active site of the enzyme. Each of these mutually independent interactions contributes incrementally to the optimization of the positioning of the cleavage site with respect to the catalytic groups



and for this reason the model is called the cumulative specificity mechanism. According to this model, the reaction begins with the collision of the enzyme and the substrate, resulting in binding at one of the subsites, presumably the most accessible one and the one with the most favorable interactions. Once the substrate peptide chain is immobilized, this initial binding is then rapidly followed by the independent and sequential interactions with the other substrate subsites. The sequential sub-segment interaction mechanism is most likely, since the peptide chain of the substrate is not rigid and the binding of the enzyme and the subsites could not occur in a single collision.

In summary, the mechanistic model postulates that the catalytically productive positioning of the scissile bond results from the cumulative effect of independent interactions between each substrate side-chain and its respective enzyme subsite. According to this view, none of these individual interactions is absolutely essential as long as the peptide is properly anchored at a sufficient number of adjacent subsites. These subsites should be independent, since; there is no discernible cross-correlation between any pair of amino acids occupying any two subsites. Finally, the concept of negative specificity accounts for the fact that the presence of unfavorable interactions at certain subsites can actually prevent any peptide chain from being a substrate for the enzyme.

The "cumulative specificity model" provides a rational interpretation of the puzzling multiplicity of natural substrates for HIV proteases. The essential features of the model are that no subsite has absolute specificity and that a combination of several mediocre interactions is at least equivalent to the combination of a few strong interactions, as far as the catalytic, as opposed to the binding specificity is concerned. The broad specificity of the HIV- 1 protease appears to follow from its ability to bind productively substrates in which interactions with only a few of the amino acid residues in the subsites need be optimized, that is, the amino acids need to have sufficiently high frequencies.

The analysis, extended to 22 peptide segments cleaved by the HIV 2 protease, delineated marked differences in specificity from that of the HIV 1 enzyme. Since the HIV 1 and 2 proteases are very similar enzymes, both having extended substrate active sites and recognition signals of the sites appear to be preferred amino acids in given positions, it was concluded that the cumulative specificity model was also the mechanism for the HIV 2 protease, as well.

### **3. Current hypotheses for initiation mechanisms**

#### **3.1. Shine-dalgarno hypothesis for prokaryotes**

Until recently no controversy existed about the initiation mechanism in prokaryotic protein synthesis. Most of the attention has been focused on the Shine-Dalgarno (SD) hypothesis [1]. It postulates that the initiation site or signal is selected by the base pairing of a nucleotide sequence preceding the initiation codon, currently known as the Shine-Dalgarno (SD) sequence, and a complementary nucleotide sequence at the 3' end of the 30S ribosome's 16S rRNA (the anti-SD sequence). The proposal that a unique nucleotide sequence base pairing

was the basis for IS selection was so attractive that the SD mechanism quickly became accepted even without rigorous proof, and even in the presence of conflicting evidences [23].

The initiation signal of the SD hypothesis, the SD sequence, is composed of about 8 nucleotides [7]. Thus, a sequence of about 8 nucleotides, which is less than 20% of the approximately 47 nucleotides of the *E. coli* model IS with signal character, raises the evolutionary problem of conservation of about 80% of unused information of initiation signals. Further studies showed that the nucleotide spacing between the SD sequence and the initiator codon can vary as much as 6 to 9 nucleotides [7,24]. All the above raise doubts on whether the SD interaction alone can effectively direct the in-reading-frame binding of the ribosomes to the mRNA.

As discussed before, the essentials of a proposed mechanism for a central biological reaction should be obligatory. In the case of the SD mechanism, the SD sequence, the anti-SD sequence of the 16S rRNA, and the base pairing of the SD sequence with the anti-SD sequence of the ribosome (the SD interaction) should all be obligatory for the selection of the initiation site. In fact, they turn out to be not absolutely necessary for the initiation reaction: there exists non-SD and leaderless mRNAs in the cell that do not contain any SD sequence and yet they are translated efficiently. Additionally, the 30S ribosomal subunits, which were reconstituted with 16S rRNA from which the anti-SD segment was deleted, functioned effectively in initiation [25].

The participation of the SD sequence in the initiation reaction, however, has been convincingly demonstrated. SD sequences were isolated base paired to the anti-SD segments of the 30S ribosomes from a reaction mixture in which ribosomes were incubated with mRNA and fMet-tRNA [26]. Although the SD interaction does participate in the initiation reaction, if present in the IS, it does not appear to be absolutely required, or obligatory. Thus, there is a conflict between the hypothesis and reality because the hypothesis postulates that the SD interaction is the only pathway for initiation, that is, it is obligatory. A resolution can be provided for this problem by the CS mechanism if it is accepted as the initiation mechanism for prokaryotes, which will be discussed briefly now and then later, more in depth.

According to the CS initiation mechanism of protein synthesis the binding of the ribosome to the IS occurs one or a few subsites (sub-segments) at a time, but, except for the initiator codon containing subsite, none of the subsite interactions is absolutely essential as long as the IS is anchored to the ribosome by a sufficient number of adjacent subsite interactions. This would be the case with the SD interaction if it is considered as one of the multiple subsite interactions of the CS mechanism, it would not always be essential. In other words, initiation of proteins synthesis is not solely dependent on the SD interaction and it is not obligatory.

### 3.2. Scanning hypothesis for eukaryotes

Since there were no obvious initiation recognition signals in eukaryotic mRNAs, it was proposed that the ribosomes do not outright recognize the IS, but only the 5' end of the

mRNA [2]. According to this proposal, the process begins with the 40S ribosome-Met-tRNA complex, which first recognizes and binds to the 5' end of the mRNA [2], found to consist of a 7-methyl guanosine and referred to as cap [27]. The ribosomal complex then scans the initiation site for the first AUG codon, which was found subsequently not always to be the initiator codon so modifications were made to account for the findings [28]. Later studies have shown that initiation factors (eIF-4F) facilitated the binding of the 40S ribosomal subunit to the mRNA [29]. When the initiator codon is located, the 60S ribosome joins the 40S complex to form the 80S ribosome complex, an aminoacyl-tRNA is bound next and the first peptide bond is formed.

Numerous exceptions to the mechanism have been observed where eukaryotic ribosomes do bind directly to internal sites of the mRNA and initiate synthesis. The observations were construed as evidence for another, separate pathway, referred to as the cap-independent or IRES (internal ribosome entry sites) –mediated translation [5]. According to advocates of the IRES mechanism, a complex IRES RNA structure of the initiation site somehow promotes the correct binding of the 40S ribosomes to internal sites. If as concluded earlier, however, that the prokaryotic initiation signals, along with the prokaryotic initiation mechanism, are conserved in the eukaryotic ribosomal system is true, then there is a simple and alternative explanation for the IRES observations than a separate IRES synthetic pathway.

Before turning to the alternative explanation for the IRES observations, the weaknesses of the proposals of the scanning and the IRES reactions will be addressed. The proposed two mechanistically quite different reactions immediately pose a difficult problem: the same ribosome cannot perform the two very different functions, even with the aid of auxiliary proteins. This, implausibly, would require the evolution of another species of ribosomes. A related major weakness is the absence of a gradual evolutionary change in the scanning proposal: there would not be a gradual evolutionary change in the case of a prokaryotic ribosome that recognizes the IS, binds directly to it, changes to a eukaryotic ribosome that recognizes only the end of the mRNA, and then scans the IS for the initiator codon. Such a process would involve too great a change in the mechanistic characteristics of the ribosome. These shortcomings of the scanning mechanism suggested the proposal of the modified CS initiation mechanism for eukaryotes, which will be reviewed below.

The above complications may be resolved if the modified CS mechanism for the initiation of eukaryotic protein synthesis is indeed adopted as the mechanism for eukaryotes. This mechanism postulates its evolution from the prokaryotic mechanism [6,10,11]. According to this proposal, the initiation of eukaryotic protein synthesis involves two steps. The first step now needs to be restated in less specific, but broader terms. Instead of specifying that an initiation factor complex (eIF-4F) binds the cap of the mRNA, it is now proposed that evolved initiation factors, including eIF-4F, bind to, or interact with the mRNA [30]. Thus, the revised version of step 1 is: the evolved initiation factors, including eIF-4F, first bind to, or interact with, the mRNA. These bindings or interactions are assumed to make the IS eminently accessible, and enhance step 2, which is the initiation of translation by the conserved prokaryotic CS mechanism. As mentioned before, in the CS mechanism the ribosome binds directly to the IS of the mRNA, without ribosomal scanning of the IS. In

summary, this modification of the prokaryotic CS initiation mechanism only adds the participation of evolved initiation factors to the function of the basic prokaryotic CS mechanism, and thus renders the eukaryotic model compatible with the proposal of a universal initiation mechanism, basically identical in all domains of life.

The alternative explanation for the IRES observations can now be conveniently discussed. From the viewpoint of the modified CS initiation mechanism, the existence of the scanning mechanism is only theoretical. There is no experimental evidence that proves that this mechanism actually operates in eukaryotic protein synthesis. Furthermore, the existence of the IRES pathway as a separate protein synthesizing pathway is questioned. The IRES pathway is viewed only as the *in vitro* expression of the activity of step 2 of the modified CS initiation mechanism without the expression of step 1. The IRES pathway is, in other words, only the *in vitro* expression of the conserved prokaryotic initiation mechanism in the eukaryotic ribosomes. This, then, is the alternative explanation for the IRES observations.

### 3.3. Cumulative specificity hypothesis for prokaryotes and eukaryotes

As mentioned earlier, the CS initiation mechanism of protein synthesis has its origin in a 1966 proposal that provided an essentially unique accessibility of the IS mechanism for prokaryotes [3]. It postulated that the initiator codon is selected by virtue of its unique accessibility. All non-initiator internal methionine codons were assumed to be sequestered and inaccessible to the ribosomes by secondary structure. This assumption was based on the observation that synthetic polynucleotides containing all four bases in equal proportions in random sequence, failed to act as mRNA in a cell-free protein synthesizing system, which was interpreted as that all AUG codons in the synthetic polynucleotide were inaccessible to ribosomes because of secondary structure (unpublished experiment). An experiment much later, however, indicated the need for an extension of the proposal. In that study, a non-SD model mRNAs was prepared for kinetic measurements [31]. The mRNAs—created to minimize secondary structures—had an accessible AUG but no other obvious recognition signal, and yet, they were able to direct the initiation of polypeptide synthesis, apparently in full agreement with our proposed mechanism. However, a second, still unrestrained and thus supposedly accessible AUG, failed to act as an initiator. This seems to indicate that the ribosomes were somehow able to discriminate the IS negatively by rejecting certain bases surrounding the initiator codon.

The unique accessibility proposal was thus modified to postulate that a site containing a non-initiator methionine codon can be made functionally inaccessible by sequestration with secondary structure or by other unfavorable local interactions, i.e., through steric hindrance, hydrophilic/hydrophobic mismatch, or by electrostatic repulsion, which all contribute negative specificity. The proposal was renamed as the unique accessibility hypothesis. However, when more was learned about the novel features of the CS model reaction for HIV proteases [9], the model reaction was incorporated into the unique accessibility proposal. The incorporation of CS reaction replaced the discrimination of ISs by the negative specificity of the unique accessibility hypothesis with the positive recognition of the IS by

cumulative specificity. The modified reaction was renamed as the CS hypothesis for the initiation of protein synthesis [4].

The CS initiation mechanism of protein synthesis incorporates the key features of the CS model reaction of HIV proteases and postulates that recognition of the initiation signal is the result of interactions of one or a few subsites (sub-segments), at a time, between the ribosome and the IS of the mRNA. Thus, the selection of the IS occurs through cooperativity and cumulative specificity of subsite interactions that allow a reaction to occur even if not all subsites are occupied [4,32]. This enables many subsites of the IS that share only some of the structural elements to be accepted as ligands by the ribosomal subsites, and hence, the broad substrate specificity of the protein synthesizing system follows.

According to the CS model for the HIV protease reaction, none of the individual subsite interactions of the substrate active site with the respective subsite of the enzyme is absolutely essential as long as the peptide is properly anchored at a sufficient number of adjacent subsites. The same rule applies to the CS initiation of protein synthesis that none of the subsite interactions is absolutely essential except that the interaction of the initiator codon subsite of the IS and the ribosomal subsite with the accessible Met-tRNA anti-AUG codon is absolutely required. In the case of the initiation of protein synthesis, there must also be a sufficient number of adjacent subsite interactions to anchor the IS properly to the ribosome.

Another important feature of CS initiation mechanism is the role played by the secondary structure of the mRNA. It keeps the ISs accessible to the ribosomes and it also reduces the accessibility of non-initiator methionine codons by sequestering them and thus favoring the recognition of the initiator codon. The multiple roles played by the secondary structure were demonstrated in a study in which the secondary structure of bacteriophage f2 RNA was disrupted by treatment with formaldehyde. The treated RNA was shown to yield three new, non-viral polypeptides when the RNA was translated in an *E. coli* cell-free protein synthesizing system [33]. Three new non-initiator methionine codons were evidently selected as initiators because filtering by secondary structure was eliminated, which showed that the specificity of the ribosome alone was not sufficient to eliminate all of the non-initiator methionine codons.

#### *Cumulative specificity in-reading-frame binding of ribosomes to mRNA*

The proposed CS initiation mechanism [6,10, 11] for the in-reading-frame binding of the mRNA by the ribosomes will now be reviewed, critically evaluated as were the SD and scanning hypotheses, and then revised with an admission of mea culpa by one of the authors (T.N.). This revision of the in-reading-frame binding aspect of the basic CS initiation mechanism must not be confused with the previously described modified CS initiation mechanism, which only added the participation of evolved initiation factors to the function of the basic prokaryotic CS mechanism for the initiation of eukaryotic protein synthesis.

The initiation reaction was postulated to begin with a relatively strong interaction of the small ribosomal subunit initiation complex with an accessible subsite of the IS that contains



the initiator AUG codon. The base pairing of the initiator AUG codon and the anti-AUG codon of the ribosome bound initiator Met-96tRNA, along with the strong binding of the entire sub-segment, secures the mRNA onto the ribosomal complex in reading frame. This first interaction is stabilized by subsite interactions that reach out in both directions of the AUG subsite in zipper-like fashion, until the ribosomal complex is firmly bound to the mRNA. Initiation can then occur when the first designated aminoacyl-tRNA is bound.

According to the above proposal, the ribosome binds to the mRNA in reading frame at the initiator codon containing subsite in a first interaction with the mRNA. This remarkable feat is possible only because of a probably invalid assumption of the proposal, namely that the initiator codon of the IS of about 50 nucleotides is always accessible for binding to the ribosomal initiation complex. As mentioned earlier, for the 30S ribosomal subunit, as an example, to bind, meshed perfectly, to a nucleotide sequence of about 47 nucleotides long, which is the length of the model *E. coli* IS, is no easy feat. This is because RNA interacts intra-molecularly so extensively by base pairing. A synthetic RNA polymer containing 4 bases in equal proportions in random sequence has been found to have about 50% of its bases paired [21]. Therefore, the mRNA would not likely to have its initiator codon located in the middle of the IS of 47 nucleotides more accessible to the ribosome than any other IS subsites, even in natural mRNAs. Prokaryotic viral mRNAs have been shown to have 60-70% of their nucleotides involved in base pairing [34]. However, one cannot rule out the possibility that secondary structure increased with a precise arrangement could enhance the accessibility of the initiator codon subsite. So the possibility of an intrinsically accessible initiator codon subsite does exist.

The major problem in binding of the mRNA IS by the ribosome is that the IS, which is about 50 nucleotides long, must minimize not only its base interactions within the IS, but also its intra-molecular interactions with nearby adjoining regions, and even with distant regions of the mRNA. Local interactions may be minimized by appropriate evolutionary base selections, that is, by controlling the primary sequence of the IS. There are studies indicating that regions around and at initiation sites are low in secondary structure. One group of researchers determined the secondary structure of the region by computer analysis of the nucleotide sequence of the intracistronic initiation sites of *infB* mRNA for various bacterial species [35]. They found that the mRNA has an open structure around the initiation site. A second group computationally folded human and mouse mRNA sequences on sets of transcripts, and found that the initiation site is characterized by a relaxed secondary structure [36].

In any case, a significant portion of the IS, nonetheless, must be accessible to the ribosome, for otherwise, one would not have a functioning mRNA. The problem, then is surmised to be, that despite the accessibility of the IS to the ribosome, the IS may not be fully extended with the initiator codon subsite readily accessible. Therefore, some mechanism is needed to stretch or to extend the IS on the ribosome so that the initiator codon subsite is accessible. As will be suggested by the revised basic CS initiation mechanism below, this can be done by

anchoring the IS onto the ribosome at a sufficient number of adjacent subsite interactions. Such a mechanism essentially stretches the IS to make the initiator codon subsite accessible to the appropriate ribosomal binding subsite.

The original proposal, that is, the prokaryotic CS initiation mechanism, will now be revised and recapitulated. The revised proposal, which assumes that all subsites of the IS, are more or less, equally accessible, postulates that the first subsite interaction will occur, more or less randomly, between one or more of all of the various subsites of the ribosome and its or their respective IS subsites. If the first interaction happens to be between the ribosomal subsite with the anti-AUG codon of the initiator Met-tRNA and the initiator AUG codon containing IS subsite, then the remaining subsites will interact in zipper-like fashion to stabilize the first interaction. In this manner an in-reading-frame binding of the mRNA by the ribosome will be completed. In all other first subsite interactions, the critical interaction which involves the anti-AUG codon of the initiator Met-tRNA and the initiator codon IS subsite will occur only after the IS is anchored onto the ribosome by a sufficient number of adjacent IS subsites and the initiator codon subsite is made accessible. In other words, the critical subsite interaction of the ribosomal subsite with the anti-AUG codon of the initiator Met-tRNA and the initiator codon IS subsite, occurs towards the end of the subsite interactions for all first non-initiator codon subsite interactions.

If, however, the IS initiator codon subsite is somehow intrinsically accessible because of the organized secondary structures of the mRNA, then all pathways will be a single pathway as originally proposed, i.e., all first ribosomal subsite bindings will be of the initiator codon IS subsite, and the remaining subsites will interact in zipper-like fashion to stabilize the first interaction. The in-reading-frame binding of the mRNA by the ribosome will all be completed in this manner. Another way in which the preceding single pathway may predominate under conditions in which the initiator codon subsite is not intrinsically accessible and the ribosome interacts with it only randomly, is the favorability or strength of the first interaction as described below.

Evidence for the strength of the critical subsite interaction where the initiator codon containing IS subsite is bound to the ribosomal subsite with the anti-AUG codon of the initiator Met-tRNA is the strength of the specific base pairing of the initiator AUG codon and the anti-AUG codon of the initiator Met-tRNA. Further evidence for the strength of this interaction was provided by studies in which nucleotides around the AUG initiation codon were replaced. Protein synthesis was decreased by as much as 95% upon replacement of three nucleotides adjoining the initiator AUG codon on the 5' side [37]. Similarly, the replacement of three nucleotides just next to the AUG codon on the 3' side decreased translation by more than 65% [38]. These observations underscore the importance of the AUG segment in initiation, and suggest a strong interaction of the ribosome with this particular IS subsite.

The process of the binding of ribosomes to leaderless mRNA is as described above for canonical mRNAs, except that the binding of eIF4F may not occur and subsite interactions are only between the ribosomal binding subsites and the initiator codon subsite and the amino acid coding region subsites of the IS. The final step in the initiation of translation of

canonical and leaderless mRNAs is the binding of the aminoacyl-tRNA directed by the codon following the initiator codon, and the formation of the first peptide bond.

Base pairing is probably not the only means of molecular recognition of the nucleotides of the IS by the ribosome in the interactions, since the ribosomal binding site is composed of RNA and proteins. In the interaction of the ribosomal subsites containing the anti-SD sequence, base pairing is the predominant means of nucleotide recognition when the SD sequence is present in the mRNA. The Shine-Dalgarno base-pairing interaction may be considered as just one of the multiple independent interactions of the CS initiation mechanism. Recognition in other ribosomal binding sub-sites may also involve steric fit, steric hindrance, hydrophilic or hydrophobic match or mismatch, and electrostatic attraction or repulsion. In other words, the recognition is also the product of both positive and negative specificities.

The revised CS hypothesis for the initiation of protein synthesis assumes that the productive positioning of the initiator codon of the IS on the ribosome results from the cumulative effect of independent interactions between each base and its respective subsite on the ribosome. According to this view, except for the subsite containing the initiator codon, none of these individual interactions is absolutely essential as long as the IS is properly anchored at a sufficient number of adjacent subsites. It is assumed that, to make the initiator codon region accessible, enough adjacent subsite interactions are needed to anchor the IS to the ribosome, which extends or stretches out the IS. The extension or stretching of the IS exposes the initiator codon region, making the initiator codon accessible. This allows the in-reading-frame binding of the ribosome to the mRNA to be completed by the interaction of the ribosomal subsite with the accessible anti-codon of the initiator Met-tRNA and the IS subsite containing the initiator codon AUG. In this manner the base-pairing of the AUG and the anti-AUG codon of the ribosome bound initiator Met-tRNA can occur.

Thus, it follows that recognition of the IS does occur by recognition of a number of individual subsites. As mentioned several times, it is most unlikely that collision of the ribosome binding site with the IS of the mRNA would be a single step in which the ribosome binding site and the IS would already be perfectly oriented to achieve an optimal fitting of all subsites. Rather, the initial collision probably results in the binding at just one or a few of the subsites, each contributing incrementally to anchoring the IS to the ribosome. Then if these interactions are favorable, the rest of the subsites will be filled in cooperatively, in a zipper-like fashion. The ultimate result is, however, the binding of the whole IS as a block, and the exact positioning on the ribosome of the subsite containing the initiator codon. In that ultimate ribosome-mRNA complex the global strength of binding and the precise positioning of the initiator codon subsite of the IS at the reaction site still depends on the sum of the contributions of the individual subsites, i.e., the cooperativity of the subsites.

#### 4. Conclusion

It is important to understand that this evaluation of current hypotheses have the advantage of hindsight provided by knowledge of the initiation of protein synthesis, not available

when the SD and the scanning hypotheses were proposed. Critical evaluations are thus made while appreciating the great value of the two older hypotheses in stimulating research.

Returning to the conclusion of this review, we have essentially taken the view that there are three keys to unlocking the secrets of the initiation of protein synthesis. The first two keys provide insights into the nature of the initiation mechanism, and the third key is the initiation mechanism that is compatible with those insights.

The first key consists of the implications of the characteristics of the IS. The *E. coli* model IS consists of a nucleotide sequence of about 47 bases, with preferred bases in given positions, but no particular base in any given position, except for the initiator codon located in a specific position in every IS of the mRNA. An important feature of the model IS is that all, or nearly all of its nucleotides presumably have signal character, which includes the leader and amino acid coding regions. This means that the initiation signal of a comprehensive mechanism must include the leader, as well as, the amino acid coding regions. The length of the model IS indicates that the IS is extensive, nonrigid, and complex, and that the ribosome is unlikely to bind the IS of the mRNA, meshed perfectly, in a single collision with the mRNA. The unlikelihood of the binding of the mRNA by the ribosome in a single collision predicts that the ribosome would bind the IS of the mRNA via a sub-segment or a few sub-segments, at a time. The ribosomal binding of the IS of the mRNA thus must happen between the ribosomal subsites and the respective IS subsites, one or a few at a time. The initiation mechanism must also account for the recognition of initiation signals consisting of about 47 nucleotides, at least in *E. coli*, with preferred bases in given positions.

The second key consists of the implications of evolutionary evidence and logic that point to a universal initiation mechanism in all domains of life. This implies that a common or universal initiation mechanism should be constantly favored, even when one is faced with an appealing mechanism for a particular phylum, and when one examines any data, always being alert for indications of conservation of initiation signals or mechanisms. As reviewed earlier in this chapter, when such diligence was maintained in reviewing the studies of heterologous systems in which ribosomes or mRNAs of different phyla were interchanged, the conclusion was reached that prokaryotic initiation signals as well as the underlying prokaryotic initiation mechanism are conserved in eukaryotes. The conservation of the initiation signals and the initiation mechanism has been interpreted as evidence of the existence of a universal initiation mechanism.

The third key consists of the initiation mechanism for protein synthesis that is most compatible with the insights provided for a mechanism by keys 1 and 2. Unfortunately the SD and scanning hypotheses were proposed before publication of the study revealing the IS characteristics of *E. coli*, although many base sequences of IS were known. That evolutionary evidence and logic supported a universal initiation mechanism was not unknown, but the proposal of an initiation mechanism for eukaryotes vastly different for that of prokaryotes indicated a lack of conviction in a universal mechanism, or at least, in the evidence for it. Thus, key 1 was not in existence

at the time of proposal of the two hypotheses, and there was not much faith in key 2. For these reasons, each of the SD and the scanning hypotheses has only considered its own particular facet of the mechanism and ignored most of the insights provided by keys 1 and 2.

For example, despite the observation that the signal character of the IS is divided about half in the leader region and the other half in the amino acid coding region, the SD hypothesis postulates an initiation signal located exclusively in the leader region, less than 20% of the nucleotides with signal character, while the scanning hypothesis postulates a single nucleotide at the 5' end of the mRNA as a recognition signal, not even included in the IS. For this reason, the two hypotheses cannot account for the initiation of translation of leaderless mRNAs. Furthermore, the two hypotheses also do not acknowledge the problem of the need of ribosomes to bind to an extensive, nonrigid and complex IS, sub-segment by sub-segment, nor do they address the possibility of a universal initiation mechanism. The SD and the scanning hypotheses, in other words, essentially ignored most of the insights provided by keys 1 and 2.

Recapitulating, in the view described above, the SD as well as the scanning hypotheses hardly account for the insights of keys 1 and 2, i.e., the implications of the characteristics of the IS of the mRNA and of the dictum of universality of the initiation mechanism. The CS hypothesis, on the other hand, is more compatible with the insights provided by keys 1 and 2. The CS mechanism for initiation of prokaryotic synthesis is essentially the mechanism formulated for the HIV proteases with minor changes to adapt it to the initiation of protein synthesis. This mechanism postulates a cooperative and cumulative, sub-segment by sub-segment recognition binding of the IS. As initiation signals, the CS mechanism recognizes nucleotide sequences with preferred bases in given positions, that is, the entire IS with signal character.

Finally, the CS initiation mechanism for prokaryotes modified for eukaryotic protein synthesis appears to be in accord with all experimental observations. The proposal postulates an evolutionary link of the initiation mechanism of eukaryotic protein synthesis to that of the prokaryotes. It assumes that evolved eukaryotic initiation factors interact with the mRNA and make the IS eminently accessible. This dramatically enhances the ribosomal binding to the IS and greatly increases the rate of initiation of protein synthesis by the conserved CS mechanism in the eukaryotic ribosomes. This modification keeps the eukaryotic initiation mechanism basically identical to the prokaryotic mechanism, and therefore, one may conclude that the two mechanisms are essentially identical as well as universal. The modified CS mechanism is compatible with the conclusion that the prokaryotic initiation signals and the prokaryotic initiation mechanism are conserved in eukaryotes.

## Author details

Tokumasa Nakamoto and Ferenc J. Kezdy

*Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, USA*



## Acknowledgement

The authors offer their profound thanks to Dr. Herbert Friedmann, a colleague at The University of Chicago, for his conscientious and excellent editing in the preparation of this manuscript.

## 5. References

- [1] Shine J, Dalgarno L (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosome binding sites, *Proc. Natl. Acad. Sci. USA*, 71: 1342-1346.
- [2] Kozak M (1978) How do eukaryotic ribosomes select initiation in regions in messenger RNA?, *Cell*, 15: 1109-1123.
- [3] Kolakofsky D, Nakamoto T (1966) The initiation of viral protein synthesis in *E. coli*. Extracts, *Proc. Natl. Acad. Sci. U.S.A.* 56: 1786-1793.
- [4] Nakamoto T (2007) The initiation of eukaryotic and prokaryotic protein synthesis: A selective accessibility and multisubstrate enzyme system, *Gene*, 403: 1-5.
- [5] Hellen CUT, Sarnow P (2001) Internal ribosome entry sites in eukaryotic mRNA molecules, *Genes Dev.*, 15: 1593-1612.
- [6] Nakamoto T (2009) Evolution and the universality of the mechanism of initiation of protein synthesis, *Gene*, 432: 1-6.
- [7] Scherer GFE, Walkinshaw MD, Arnott S, Morre SDJ (1980) The ribosome binding sites recognized by *E. coli* ribosomes have regions with signal character in both leader and protein coding segment, *Nucl. Acid Res.*, 8: 3895-3907.
- [8] Kozak M (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs, *Nucl. Acid Res.*, 12: 857-872.
- [9] Poorman RA, Tomasseli AG, Henrikson RL, Kézdy FJ (1991) A cumulative specificity model for proteases from human immunodeficiency virus types 1 and 2, inferred from statistical analysis of an extended substrate database, *J. Biol. Chem.* 266: 14554-14561.
- [10] Nakamoto T (2010) Mechanisms of the Initiation of Protein Synthesis: In Reading Frame Binding of Ribosomes to mRNA, *Mol. Biol. Rep.* DOI 10.1007/s11033-010-0176-1
- [11] Nakamoto T (2011) Mechanisms of the Initiation of Protein Synthesis: In Reading Frame Binding of Ribosomes to mRNA, *Mol. Biol. Rep.* 38:847-855
- [12] Rekosh DM, Lodish HF, Baltimore D (1970) Protein synthesis in *Escherichia coli* extracts programmed by poliovirus RNA, *J. Mol. Biol.* 54: 327-340.
- [13] Glover JF, Wilson MA (1982) Efficient translation of the coat protein cistron of tobacco mosaic virus in a cell-free system from *Escherichia coli*, *Eur. J. Biochem.*, 122: 485-492.
- [14] Siegert W, Konings RNH, Bauer H, Hofschneider PH (1972) Translation of avian myeloblastosis virus RNA in a cell-free Lysate of *Escherichia coli*, *Proc. Natl. Acad. Sci. USA*, 69: 888-891.
- [15] Morrison TG, Lodish HF (1973) Translation of bacteriophage Q $\beta$  RNA by cytoplasmic extracts of mammalian cells, *Proc. Natl. Acad. Sci. USA*, 70: 315-319.

- [16] Wang S, Marcu KB, Inouye M (1976) Translation of a specific mRNA from *Escherichia coli* in a eukaryotic cell-free system, *Biochem. Biophys. Res. Comm.*, 68: 1194-1200.
- [17] Paterson BM, Rosenberg M (1979) Efficient translation of prokaryotic mRNAs in a eukaryotic cell-free system requires addition of a cap structure, *Nature*, 279: 692-696.
- [18] Rosenberg M, Paterson BM (1979) Efficient cap-dependent translation of polycistronic prokaryotic mRNAs is restricted to the first gene in the operon, *Nature* 279: 696-701.
- [19] Nomura M, Lowry CV (1967) Phage f2 RNA=Directed Binding of Formylmethionyl-tRNA to Ribosomes and the Role of 30S ribosomal Subunits in Initiation of Protein Synthesis, *Proc. Natl. Acad. Sci. USA* 58: 946-953.
- [20] Gualerzi CO, Pon CL (1990) Initiation of mRNA Translation in Prokaryotes. *Biochem* 29: 5881-5889.
- [21] Gralla J, Delisi C (1974) mRNA is Expected to Form Stable Secondary Structures. *Nature* 248: 330-332.
- [22] Merrick WC (2004) Cap-independent and cap-dependent translation in eukaryotic systems, *Gene* 332: 1-11.
- [23] Nakamoto T (2006) A unified view of the initiation of protein synthesis, *Biochem. Biophys. Res. Commun.* 341: 675-678.
- [24] Steitz J.A. (1979) Genetic signals and nucleotide sequences in messenger RNA, in *Biological Regulation and Development* (Goldberger, R. ed.) pp 349-399, Plenum Publ. Co., New York.
- [25] Melancon P, Leclerc D, Destroismaisons N, Brakier-Gingras L (1990) The anti-Shine-Dalgarno region in *Escherichia coli* 16S ribosomal RNA is not essential for the correct selection of translational starts, *Biochem.*, 29: 3402-3407.
- [26] Steitz JA, Jakes K (1975) How ribosomes select initiator regions in mRNA: base-pair formation between the 30 terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* 72: 4734-4738.
- [27] Shatkin J (1976) Capping of eukaryotic mRNAs, *Cell* 9: 645-653.
- [28] Kozak M (1995) Adherence to the first-AUG rule when a second AUG codon follows closely upon the first, *Proc. Natl. Acad. Sci. USA*, 92: 2662-2666.
- [29] R.D. Abramson, T.E. Dever, T.G. Lawson, B.K. Ray, R.E. Thach, W.E. Merrick, The ATP-dependent interaction of eukaryotic initiation factors with mRNA, *J. Biol. Chem.*, 262 (1987) 3826-3832.
- [30] Anthony DD, Merrick WC (1991) Eukaryotic initiation factor (eIF)-4F: Implications for a role in internal initiation of translation, *J. Biol. Chem.*, 266: 10218-10226.
- [31] Calogero RA, Pon CL, Canonaco A, Gualerzi CO (1988) Selection of the mRNA translational initiation region by *Escherichia coli* ribosomes, *Proc. Natl. Acad. Sci. USA*, 85: 6427-6431.
- [32] D.E. Koshland Jr DE, Hamadani K (2002) Proteomics and models for enzyme cooperativity, *J. Biol. Chem.*, 27: 46841-46844
- [33] Lodish, HF (1970) Secondary structure of bacteriophage f2 ribonucleic acid and the initiation of in vitro protein biosynthesis. *J. Mol. Biol.* 50: 689-702.
- [34] Ricard B, Salser W (1976) Optical measurements reveal base-pairing in T4-Specific mRNAs, *Biochim. Biophys. Acta* 425: 196-201.

- [35] Laursen BS, de A. Steffensen SA, Hedegaard J, Moreno JMP, Mortensen KK, Sperling-Petersen HU (2002) Structural requirements of the mRNA for intracistronic translation initiation of the enterobacterial *infB* gene, *Genes to Cells*, 7: 901-910.
- [36] Shabalina SA, Ogurtsov AY, Nikolay A, Spiridonov NA (2006) A periodic pattern of mRNA secondary structure created by the genetic code, *Nucl. Acid Res.*, 34: 2428-2437.
- [37] Hui A, Hayflick J, Dinkelspiel K, de Boer HA (1984) Mutagenesis of the three bases preceding the start codon of the galactosidase mRNA and its effect on translation in *Escherichia coli*, *EMBO J.*, 3: 623-629.
- [38] Looman AC, Bodlaender J, Comstock J, Eaton D, Jhurani P, de Boer HA, van Knippenberg PH (1987) Influence of the codon following the AUG initiation codon on the expression of a modified *lac Z* gene in *Escherichia coli*, *EMBO J.*, 6: 2489-2492.