Proteome-wide analysis reveals clues of complementary interactions between mRNAs and their cognate proteins as the physicochemical foundation of the genetic code

Anton A Polyansky, Mario Hlevnjak, and Bojan Zagrovic*

Department of Structural and Computational Biology; Max F. Perutz Laboratories; University of Vienna; Vienna, Austria

*Correspondence to: Bojan Zagrovic;
Email: bojan.zagrovic@univie.ac.at

Despite more than 50 years of effort, the origin of the genetic code remains enigmatic. Among different theories, the stereochemical hypothesis suggests that the code evolved as a consequence of direct interactions between amino acids and appropriate bases. If indeed true, such physicochemical foundation of the mRNA/protein relationship could also potentially lead to novel principles of protein–mRNA interactions in general. Inspired by this promise, we have recently explored the connection between the physicochemical properties of mRNAs and their cognate proteins at the proteome level. Using experimentally and computationally derived measures of solubility of amino acids in aqueous solutions of pyrimidine analogs together with knowledge-based interaction preferences of amino acids for different nucleobases, we have revealed a statistically significant matching between the composition of mRNA coding sequences and the base-binding preferences of their cognate protein sequences. Our findings provide strong support for the stereochemical hypothesis of genetic code's origin and suggest the possibility of direct complementary interactions between mRNAs and cognate proteins even in present-day cells.

Introduction

The discoveries of the genetic code,[1] mRNA,[2] and protein synthetic machinery,[3] together with the structure of DNA,[4] provided a powerful foundation for explaining information passage from genes to proteins in atomistic and molecular terms. However, our understanding of the relationships between these major cellular biopolymers still lacks a strong evolutionary perspective. In particular, the physicochemical driving forces, which have led to the establishment of the fundamental organization of information transfer in the cell, remain to be elucidated.

A particularly pertinent problem in this regard is the interconnection between mRNA and proteins. Whereas the two are obviously related by the universal genetic code, the principal building block of all life, its origin still remains one of the most important, yet unanswered foundational questions in biology.[5] What is more, the mRNA–protein relationship extends significantly beyond just the coding context, especially if one considers the wide cellular interaction network contributing to the regulation of protein expression. For example, recent studies of protein–mRNA interactomes in eukaryotic cells have revealed a surprising fact that a number of proteins without recognizable RNA binding domains, including several transcription factors and metabolic enzymes, nevertheless display mRNA binding ability.[6-9] Such interactions also include direct binding to cognate mRNAs, which has over the years been detected for a diverse set of proteins.[6,10-15] These findings open up a possibility that protein–mRNA interactions could be a more common phenomenon in the cell than previously thought, and invite formulation of novel, fundamental physicochemical principles behind such interactions.
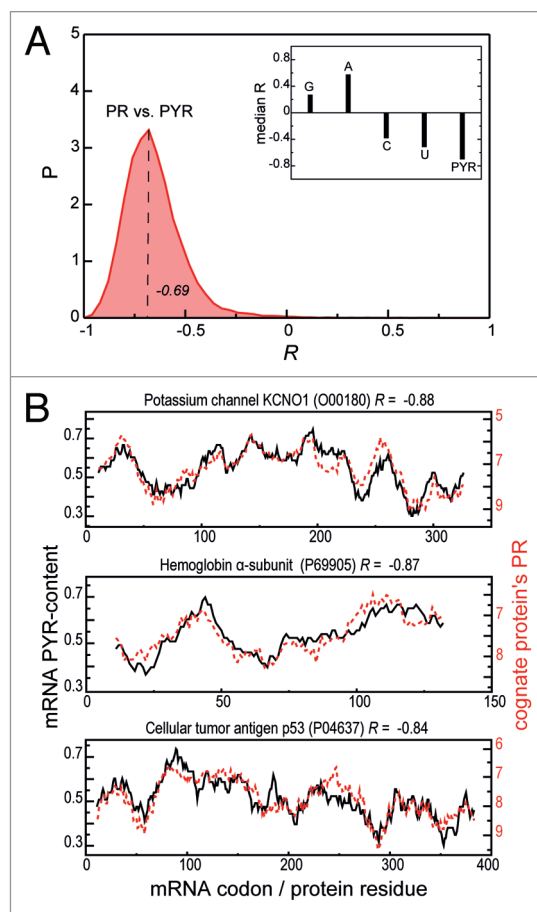
Figure 1. Matching of mRNA coding-sequence pyrimidine (PYR) profiles and their cognate protein sequence polar-requirement (PR) profiles. (**A**) Distribution of Pearson correlation coefficients (R, x-axis) between window-averaged PYR-content profiles of individual mRNA-coding sequences and window-averaged PR sequence profiles of their cognate proteins for the human proteome (window size is 21 amino acids/codons). P (y-axis) corresponds to bin-size-normalized probability density. Inset: the median Rs for the human proteome obtained using protein PR sequence profiles and different nucleobase density mRNA profiles. (**B**) Comparison between mRNA PYR-content and their cognate protein PR profiles for three exemplary human proteins. Experimental PR scale[18] has been used for all comparisons presented here. Note that due to the definition of the PR scale, negative correlations indicate positive matching between the given content of mRNAs and the affinity for PYR analogs of their cognate proteins and vice versa.

In two recent studies, we have presented evidence of intrinsic potential of mRNAs and their cognate proteins to interact, which could simultaneously also explain fixation of the relationship between these two biopolymers in the modern genetic code.[16,17] More specifically, we have compared nucleotide content of naturally occurring mRNA coding sequences with the propensity of cognate protein sequences to interact with different nitrogenous bases. This was performed using both experimentally[18] and computationally[19] derived polar requirement (PR) scales capturing the solubility of amino acids in aqueous solutions of pyrimidine

analogs, and knowledge-based interaction preferences between amino acids and RNA bases derived by analyzing binding interfaces in the known 3D-structures of protein–RNA complexes. Already in the 1960s and 1970s, Carl Woese and co-workers have shown that depending on type, amino acids exhibit different, clearly defined preferences for interacting with pyrimidine analogs (the PR scale).[18,20] This was then used to support the stereochemical hypothesis for the origin of the genetic code, the idea that the specific pairing between individual codons and their cognate amino acids stems from direct binding preferences of the two for

each other.[5,21-25] At the same time, the connection between the PR of individual amino acids and pyrimidine (PYR) content of their codons had never been quantitatively explored, and the same is true for amino-acid propensities to interact with other nucleobases. Even more so, the relationship between the nucleobase density of mRNA-coding regions and the base-binding propensities of their cognate protein sequences at the whole proteome level remained fully unexplored.

Matching Between mRNA Composition and Nucleobase Binding Propensities of their Cognate Proteins

We have studied this relationship for complete proteomes of 15 different organisms, five from each domain of life.[16] First, we have shown that the average PYR content of mRNA coding sequences exhibits an extremely strong inverse correlation with the average PR of their cognate protein sequences over complete proteomes. For example, the Pearson correlation coefficient R between the two variables approaches -0.9 for the human proteome. Since the PR scale assigns low values to amino acids with high propensity to interact with PYR analogs and vice versa, this means that the PYR content of mRNA coding sequences is directly proportional to the average affinity of their cognate protein sequences for PYR-like compounds. Second, we have evaluated the level of matching between window-averaged PYR profiles of individual mRNA coding sequences and PR profiles of their cognate protein sequences and observed that the distributions of the thus-obtained Rs for all 15 species exhibit great similarity, with the average R typically in the vicinity of -0.7. To illustrate this finding, we present here the results of such a comparison for the human proteome using the original experimental PR scale derived by Woese and co-workers[18] (**Fig. 1A**). The distribution of Pearson Rs obtained by comparing individual mRNA PYR profiles and their cognate protein PR profiles is prominently shifted toward strong negative correlations with the median value of -0.69. Importantly, the PYR density profiles of mRNA display stronger correlations

with PR profiles of their cognate protein sequences than any other type of mRNA compositional profiles (**Fig. 1A**, inset). As shown in our original study,[16] these correlations further improve by several percentage points if one uses a computationally derived PR scale, albeit with no impact on any qualitative conclusions.

What do these correlations mean at the level of individual mRNA/protein pairs? In **Figure 1B** we present PYR mRNA density profiles aligned with PR profiles of their cognate proteins for three select cases where the two display strong correlation—a membrane protein (potassium channel KCNQ1), a cytosolic globular protein (hemoglobin α-subunit), and a significantly unstructured nuclear/cytosolic protein (tumor suppressor p53). While these three well-known proteins were purposefully chosen to illustrate a high level of matching, one should emphasize that even the median level of correlation (-0.69) corresponds to profiles with strong matching, as discussed in more detail in our original work. As one can see in these examples, the mRNA PYR densities quantitatively mirror sequence profiles of cognate proteins capturing their affinity for PYR analogs. In other words, PYR-rich regions in mRNAs directly correspond to the regions in their cognate proteins with high affinity for PYR analogs and vice versa. What is more, the universal genetic code appears to be highly optimized with respect to maximizing this matching: for more than half of the proteomes tested, the native genetic code outperforms each and every one of the 10^6 randomized codes tested.[16]

These compelling results notwithstanding, the PR scale still leaves several important questions unanswered. First, while this scale captures amino-acid affinities for pyrimidine-like compounds, it remains silent about the potentially equally important purine (PUR) affinities. Second, substituted pyridines, as used in deriving the PR scales, are only proxies for the real biologically relevant nucleobases. How does the picture change if one looks at amino-acid propensities to bind uracil, cytosine, guanine, or adenine? To address these questions, we have derived interaction preferences between different amino acids and RNA nucleobases by focusing on binding interfaces in the

known 3D structures of protein–RNA complexes (approximately 300 high-resolution X-ray and NMR structures, including five ribosomal structures).[17] In the process, we have isolated sequence-specific protein–RNA contacts where each amino-acid side chain is surrounded by more than one RNA base, while ignoring non-specific interactions defined solely by protein or RNA backbones. The scales of amino acid/nucleobase interaction preferences were derived by employing standard distance-independent contact potential formalism[26-30] and allowed us to explore the connections between composition of mRNA codons and the preferences of amino acids to interact with different nucleobases at protein/RNA interfaces (see also the section below). Moreover, this approach also allowed us to examine analogous matching for sequence profiles of complete mRNA coding sequences and their cognate proteins. Strikingly, we have observed a strong correlation between preferences of amino acids to interact with guanine (G-preference) and the average PUR-content of their respective codons (R = -0.84).[17] In other words, amino acids, which are predominantly encoded by PUR bases, display a strong tendency to co-localize with G at protein–RNA interfaces and vice versa. Consequently, we have also found that sequence profiles of protein G-preferences strongly match PUR density in cognate mRNAs (median R = -0.80 for the human proteome, **Fig. 2A**) and that this effect is stronger than anything one sees with other possible amino-acid/nucleobase interaction preferences (**Fig. 2A**, inset). A similar, but somewhat weaker signal was observed when it comes to mRNA PYR or PUR profiles and sequence preferences of their cognate proteins to interact with pyrimidine (PYR-preference) or purine bases (PUR-preference), respectively.[17] Most notably, for adenine (A-preferences) this correlation is significant, but reverse (median R = 0.53 for the human proteome, **Fig. 2A**): protein sequence regions encoded by PUR-rich mRNA sequence stretches prefer not to interact with adenines and vice versa, and the same is true for A-rich mRNA sequences.[17]

Using the same mRNA/protein pairs as in **Figure 1**, we demonstrate the

close matching between mRNA PUR density profiles and knowledge-based G-preference profiles of their cognate protein sequences (**Fig. 2B**), which for these three proteins actually exceeds the matching seen for mRNA PYR density profiles and their cognate proteins' PR profiles. Taken together, the results of both of our studies clearly demonstrate a striking level of mirroring between PYR- or PUR-density profiles of mRNA coding sequences and base-binding-preference profiles of their cognate proteins, independently of how the respective amino-acid preferences are obtained (e.g., an experimental or computational PR scale, and a number of knowledge-based scales). These observations have allowed us to propose a novel and potentially wide-reaching hypothesis that mRNA-coding regions may be physicochemically complementary to the respective cognate protein regions and bind, especially if the complementary segments are available for interaction such as in the case of unstructured mRNA and protein stretches. In fact, we would like to propose that such direct complementary binding interactions may be a key element underlying the whole mRNA/protein relationship when it comes to both its evolutionary development as well as present-day biology.[16,17]

Correlations in the Genetic Code Reveal its Physicochemical Nature

Knowledge-based amino-acid interaction preference scales allowed us also to test the physicochemical relationship between amino acids and the composition of their codons. This is especially important in the stereochemical perspective of the origin of the genetic code.[5,21-25] For this purpose, we have calculated correlations between the average codon composition in the human proteome and the respective amino-acid interaction preferences. We divide amino acids into two equally sized subsets: (1) amino acids, which can be obtained abiotically in similar conditions as in the classic Miller-Urey experiments,[31] and (2) amino acids whose synthesis requires more complicated biochemical pathways.[5] According to the coevolution theory,[32,33] amino acids in the first subset are
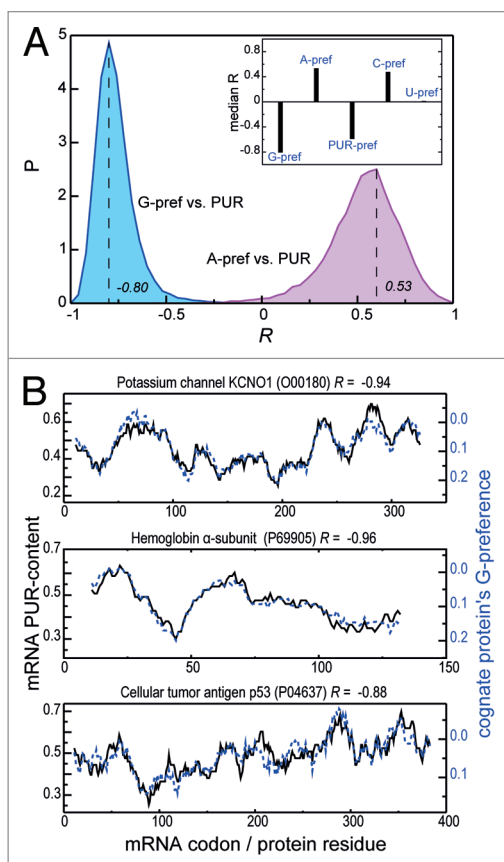
Figure 2. Matching of mRNA coding-sequence purine (PUR) density profiles and protein sequence knowledge-based interaction preference profiles. (**A**) Distribution of Pearson correlation coefficients (*R*, x-axis) between mRNA PUR profiles and G- (blue) or A-preference (magenta) sequence profiles of human proteins. *P* (y-axis) corresponds to bin-size-normalized probability density. *Inset*: the median *Rs* for the human proteome obtained using PUR mRNA profiles and different knowledge-based interaction preference scales for cognate protein sequences. (**B**) Comparison between mRNA PUR density profiles and their cognate protein G-preference profiles for the same three exemplary human proteins as given in **Figure 1B**. Note that due to the definition of knowledge-based scales, negative correlations indicate positive matching between the given content of mRNAs and the binding preferences of their cognate proteins and vice versa.

evolutionarily "old," in a sense that they were utilized during code evolution from prebiotic synthesis, while the second type of amino acids entered the code by means of biosynthesis from the "old" amino acids or were introduced into proteins through post-translational modifications. Interestingly, we observe relatively strong correlations between G-binding preferences and G-content of codons for the first type of amino acids (**Fig. 3A**), and similarly so for their C-binding preferences and C-content of codons. On the other hand, these correlations are weak for the "new" amino acids and also decrease when it comes to all 20 amino acids (**Fig. 3C**). In contrast, strong anti-correlation in the case of A is observed for the "new" amino

acids (**Fig. 3B**) and it remains so over all 20 (**Fig. 3C**). In other words, the G- and C-content of codons for biosynthetically more primitive amino acids from the first set is strongly related to their tendency to preferentially co-localize with G and C at protein–RNA interfaces, respectively. Conversely, the A-content of biosynthetically more complex amino acids is inversely related to their A-binding preferences at protein-RNA interfaces.

Interestingly, the codons of the "old" amino acids are enriched in G and C as compared with the "new" amino acids (the average G/C content of 0.58 vs. 0.41, respectively, *P* value = 0.097, Welch two sample t-test), which was also noticed before.[34] Summarizing these facts, we can

speculate that a stereochemical connection could have played a role in shaping the genetic code in the early stages by providing a mapping between prebiotic amino acids and G/C-rich codons, whereas engagement of new amino acids required more of A and U to be included in the codons. The latter could have been driven by other factors, such as evolution of metabolic pathways[32,33] or optimization of code stability,[35,36] which also further reduced the direct correspondence between codon content and amino-acid affinities. Our findings may also explain in part some of the difficulties in trying to prove the direct stereochemical nature of the genetic code[23-25]: for example, our results suggest that some amino acids (especially evolutionarily more recent ones whose anticodons are A-rich, such as phenylalanine) may also exhibit appreciable affinity for their anticodons. This suggestion agrees with the findings of Johnson and Wang who have detected significant co-localization of a number of evolutionarily more recent amino acids and their cognate anticodons in ribosomal structures.[25] What is more, following similar previous suggestions,[5] our findings underline the fact that only a combination of different hypotheses could give a complete view of the origin of the genetic code. At the same time, our results give strong support to the possibility of direct templating of proteins from mRNA in the era before the development of ribosomal decoding and code's fixation in that era.[22,37] This scenario, first put forth by Carl Woese, suggests that amino acids in ancient systems associated with mRNA directly in the course of translation, following their intrinsic physicochemical propensities. Relating this to our findings, the ancient physicochemical background of the genetic code appears to still allow G- or PYR-affinity of amino acids in present-day protein sequences to closely follow PUR/PYR density profiles, respectively, of their cognate mRNAs. Finally, direct binding of mRNAs to proteins they code for is also consistent with the proposal that the principal function of ancient proteins during the transition from the RNA world was indeed structural stabilization of RNA molecules.[37] In line with this, our results suggest that strong binding complementarity may
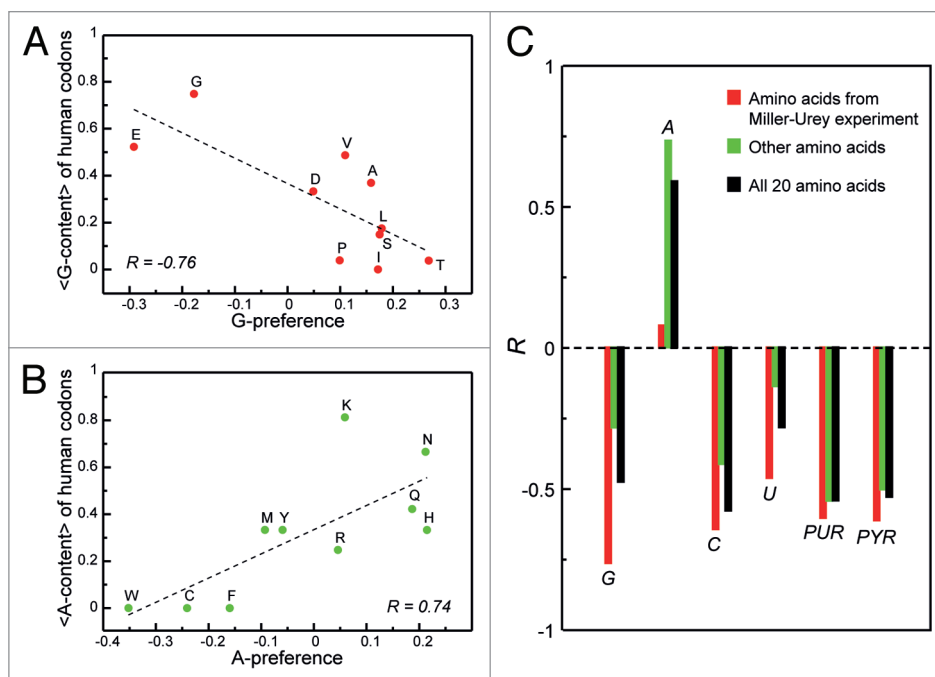
Figure 3. Connection between knowledge-based nucleobase-binding preferences of amino acids and the base content of their cognate codons. (**A**) Correlation between G-interaction preferences of amino acids from Miller-Urey experiment[31] and the average G-content of their codons in mRNAs of the entire human proteome. (**B**) Correlation between A-interaction preferences of other amino acids and the average A-content of their codons. (**C**) Correlation coefficients *(R)* between different amino-acid interaction preferences and the respective compositions of their codons. Amino acids are grouped into three subsets, which are colored according to the legend. Note that due to the definition of knowledge-based scales, negative correlations indicate positive matching between the given content of codons and the binding preferences of their cognate amino acids and vice versa.

exist predominantly at the level of longer mRNA and polypeptide stretches rather than individual amino acids and codons.

Potential Roles of Cognate Interactions in Present-Day Biology

Arguably the strongest experimentally testable prediction of our complementary binding model is that a large fraction of proteins will directly bind to their cognate mRNAs in an in-frame manner, especially if both molecules are structurally destabilized.[16,17] Which processes in present-day biological systems could depend on such binding? Given that one observes complementary matching mainly at the level of primary sequences of the two biopolymers, one should first consider all the contexts in which both polymers are significantly unstructured. However, we do not exclude the possibility of direct cognate binding according to the same physicochemical principles even if both molecules are largely structured. Namely, it remains to be explored how putative binding hot

spots at the level of primary sequence profiles map onto folded mRNA and protein secondary and tertiary structures. It is possible that the binding potential, suggested by primary-sequence profiles, remains present even if both molecules adopt well-defined folds. What is more, the fact that adenines behave qualitatively differently from guanines, as discussed above, points at additional complexities that still need to be fully understood.

In principle, all scenarios where mRNAs and/or proteins exist as partially extended polymers might be affected by cognate interactions of the kind described above, such as, for example, during splicing of pre-mature mRNAs, protein secretion, or translocation. However, probably the most obvious process where binding of mRNAs to cognate proteins could have a regulatory role is translation. Here, the newly synthesized proteins would repress the translation of their own message by directly or indirectly competing with the binding between ribosome and mRNA and forming a negative feedback loop. From thymidylate synthase

to dihydrofolate reductase to different ribosomal proteins, there is a number of known examples where precisely such regulation takes place.[6,10-15,38] It will be interesting to study the parallels as well as discrepancies between our proposal and the specific binding mechanisms, which have been proposed in some of these cases. For example, cognate binding in this context is known to occur not only in mRNA coding sequences, but also in their untranslated regions (UTRs). An important frontier in this regard will be to study potential profile complementarity even between non-coding RNA segments and different protein profiles. Finally, it will be interesting to explore whether cognate binding-based translational regulation may be a more widespread phenomenon than typically considered, as suggested by our model.

Another context in which cognate interactions may also play a natural role is viral assembly. After all, a loaded viral capsid can be thought of as one of the simplest entities in which genetic message resides in close proximity of its own product.

Recently, Stockley and co-workers have used single-molecule fluorescence correlation spectroscopy to show that specific and direct binding of capsid proteins to single-stranded RNA of bacteriophage MS2 and satellite tobacco necrosis virus plays a key role in RNA compaction and capsid assembly and packaging.[39,40] It will be interesting to examine to what extent the binding mechanisms seen in these systems agree with our proposal. What is more, analogous principles may in fact also be applied to single-stranded DNA viruses.

Finally, the mRNA–protein complementarity hypothesis may also be intimately related to the structure and function of ribonucleoprotein complexes and RNA–protein granules. A number of recent studies have reported and characterized instances of phase separation in the cytoplasm[41] and nucleoplasm,[42] a process analogous to lipid-raft formation in the membrane, which in aqueous compartments results in the formation of liquid droplets. These droplets define specific, non-membrane-bound compartments typically rich in proteins and RNA (such as nucleoli, germline P granules, P-bodies, stress granules, Cajal bodies, etc.),[43] and are in many cases known to be the sites of mRNA storage, processing, and decay. Moreover, it has also been suggested that both multi-valency (i.e., many potential binding sites) and presence of low complexity (i.e., disordered) regions in proteins may be important factors in the formation of such phase-separated compartments.[43] It is tempting to relate these observations to the mRNA–protein complementarity hypothesis, whereby transient cognate mRNA-protein interactions would provide the necessary driving force for such compartmentalization. Twenty years ago, Kyrpides and Ouzounis suggested that cognate protein–mRNA binding interactions may represent an ancient mechanism through which mRNA stability is auto-regulated.[10] Although their model of such binding lacked specific mechanistic details, they saw it as an extension of the stereochemical hypothesis at the level of complete polymers, a view now corroborated by our microscopically detailed model. Importantly, in their theoretical scheme, untranslated or excess mRNAs pair up with their cognate proteins and

are subsequently degraded. One may speculate that the above compartments, such as P-bodies or stress granules, may be the actual sites in which such interactions take place. In our picture, being disordered should facilitate interactions between mRNAs and cognate proteins, but this might also promote degradation. What is more, such destabilization might actually be a consequence of different stresses (e.g., heat shock), which also fits well with the role of P-bodies and stress granules in general.

To sum up, our recent findings concerning the close inter-dependence between compositional profiles of mRNAs and their cognate proteins provide a novel framework for analyzing interactions between proteins and not just mRNAs, but rather nucleic acids in general. Namely, all of the findings and principles presented here could also be generalized to other types of RNA molecules in present-day systems, including long non-coding RNAs, as well as DNA. However, it should be emphasized that cognate interactions of the kind discussed here could have been functionally relevant only in ancient cells and it could very well be that in the course of evolution they were replaced by more specific and more efficient mechanisms. Alternatively, as we would like to suggest, it is possible that they actually became a foundation on top of which these more specific mechanisms operate. We hope our results and suggestions will stimulate the field to explore these exciting possibilities.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Crick FHC. The origin of the genetic code. J Mol Biol 1968; 38:367-79; PMID:4887876; http://dx.doi.org/10.1016/0022-2836(68)90392-6

2. Brenner S, Jacob F, Meselson M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature 1961; 190:576-81; PMID:20446365; http://dx.doi.org/10.1038/190576a0

3. Palade GE. A small particulate component of the cytoplasm. J Biophys Biochem Cytol 1955; 1:59-68; PMID:14381428; http://dx.doi.org/10.1083/jcb.1.1.59

4. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953; 171:737-8; PMID:13054692; http://dx.doi.org/10.1038/171737a0

5. Koonin EV, Novozhilov AS. Origin and evolution of the genetic code: the universal enigma. IUBMB Life 2009; 61:99-111; PMID:19117371; http://dx.doi.org/10.1002/iub.146

6. Anders G, Mackowiak SD, Jens M, Maaskola J, Kuntzagk A, Rajewsky N, Landthaler M, Dieterich C. doRiNA: a database of RNA interactions in post-transcriptional regulation. Nucleic Acids Res 2012; 40(Database issue):D180-6; PMID:22086949; http://dx.doi.org/10.1093/nar/gkr1007

7. Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell 2012; 46:674-90; PMID:22681889; http://dx.doi.org/10.1016/j.molcel.2012.05.021

8. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 2012; 149:1393-406; PMID:22658674; http://dx.doi.org/10.1016/j.cell.2012.04.031

9. König J, Zarnack K, Luscombe NM, Ule J. Protein-RNA interactions: new genomic technologies and perspectives. Nat Rev Genet 2011; 13:77-83; PMID:22251872; http://dx.doi.org/10.1038/nrg3141

10. Kyrpides NC, Ouzounis CA. Mechanisms of specificity in mRNA degradation: autoregulation and cognate interactions. J Theor Biol 1993; 163:373-92; PMID:8246508; http://dx.doi.org/10.1006/jtbi.1993.1126

11. Ouzounis CA, Kyrpides NC. Reverse interpretation: a hypothetical selection mechanism for adaptive mutagenesis based on autoregulated mRNA stability. J Theor Biol 1994; 167:373-9; PMID:8207952; http://dx.doi.org/10.1006/jtbi.1994.1076

12. Chu E, Copur SM, Ju J, Chen TM, Khleif S, Voeller DM, Mizunuma N, Patel M, Maley GF, Maley F, et al. Thymidylate synthase protein and p53 mRNA form an in vivo ribonucleoprotein complex. Mol Cell Biol 1999; 19:1582-94; PMID:9891091

13. Tai N, Schmitz JC, Liu J, Lin X, Bailly M, Chen TM, Chu E. Translational autoregulation of thymidylate synthase and dihydrofolate reductase. Front Biosci 2004; 9:2521-6; PMID:15353304; http://dx.doi.org/10.2741/1413

14. Schüttpelz M, Schöning JC, Doose S, Neuweiler H, Peters E, Staiger D, Sauer M. Changes in conformational dynamics of mRNA upon AtGRP7 binding studied by fluorescence correlation spectroscopy. J Am Chem Soc 2008; 130:9507-13; PMID:18576621; http://dx.doi.org/10.1021/ja801994z

15. Zhao X, Liu M, Wu N, Ding L, Liu H, Lin X. Recovery of recombinant zebrafish p53 protein from inclusion bodies and its binding activity to p53 mRNA in vitro. Protein Expr Purif 2010; 72:262-6; PMID:20363330; http://dx.doi.org/10.1016/j.pep.2010.03.028

16. Hlevnjak M, Polyansky AA, Zagrovic B. Sequence signatures of direct complementarity between mRNAs and cognate proteins on multiple levels. Nucleic Acids Res 2012; 40:8874-82; PMID:22844092; http://dx.doi.org/10.1093/nar/gks679

17. Polyansky AA, Zagrovic B. Evidence of direct complementary interactions between messenger RNAs and their cognate proteins. Nucleic Acids Res 2013; In press: http://dx.doi.org/10.1093/nar/gkt618; PMID:23868089

18. Woese CR. Evolution of the genetic code. Naturwissenschaften 1973; 60:447-59; PMID:4588588; http://dx.doi.org/10.1007/BF00592854

19. Mathew DC, Luthey-Schulten Z. On the physical basis of the amino acid polar requirement. J Mol Evol 2008; 66:519-28; PMID:18443736; http://dx.doi.org/10.1007/s00239-008-9073-9

20. Woese CR, Dugre DH, Saxinger WC, Dugre SA. The molecular basis for the genetic code. Proc Natl Acad Sci USA 1966; 55:966-74; PMID:5219702; http://dx.doi.org/10.1073/pnas.55.4.966

21. Woese CR. Order in the genetic code. Proc Natl Acad Sci USA 1965; 54:71-5; PMID:5216368; http://dx.doi.org/10.1073/pnas.54.1.71

22. Woese CR. On the evolution of the genetic code. Proc Natl Acad Sci USA 1965; 54:1546-52; PMID:5218910; http://dx.doi.org/10.1073/pnas.54.6.1546

23. Yarus M. Amino acids as RNA ligands: a direct-RNA-template theory for the code's origin. J Mol Evol 1998; 47:109-17; PMID:9664701; http://dx.doi.org/10.1007/PL00006357

24. Yarus M, Widmann JJ, Knight R. RNA-amino acid binding: a stereochemical era for the genetic code. J Mol Evol 2009; 69:406-29; PMID:19795157; http://dx.doi.org/10.1007/s00239-009-9270-1

25. Johnson DB, Wang L. Imprints of the genetic code in the ribosome. Proc Natl Acad Sci USA 2010; 107:8298-303; PMID:20385807; http://dx.doi.org/10.1073/pnas.1000704107

26. Miyazawa S, Jernigan RL. Estimation of effective interresidue contact energies from protein crystal-structures - quasi-chemical approximation. Macromolecules 1985; 18:534-52; http://dx.doi.org/10.1021/ma00145a039

27. Donald JE, Chen WW, Shakhnovich EI. Energetics of protein-DNA interactions. Nucleic Acids Res 2007; 35:1039-47; PMID:17259221; http://dx.doi.org/10.1093/nar/gkl1103

28. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. RNA 2009; 15:189-99; PMID:19144906; http://dx.doi.org/10.1261/rna.1270809

29. Perez-Cano L, Solernou A, Pons C, Fernandez-Recio J. Structural prediction of protein-RNA interaction by computational docking with propensity-based statistical potentials. Pacific Symposium on Biocomputing 2010; 15:269-80.

30. Tuszynska I, Bujnicki JM. DARS-RNP and QUASI-RNP: new statistical potentials for protein-RNA docking. BMC Bioinformatics 2011; 12:348; PMID:21851628; http://dx.doi.org/10.1186/1471-2105-12-348

31. Miller SL, Urey HC, Oró J. Origin of organic compounds on the primitive earth and in meteorites. J Mol Evol 1976; 9:59-72; PMID:1018332; http://dx.doi.org/10.1007/BF01796123

32. Wong JT. A co-evolution theory of the genetic code. Proc Natl Acad Sci USA 1975; 72:1909-12; PMID:1057181; http://dx.doi.org/10.1073/pnas.72.5.1909

33. Wong JT. Coevolution theory of the genetic code at age thirty. Bioessays 2005; 27:416-25; PMID:15770677; http://dx.doi.org/10.1002/bies.20208

34. Trifonov EN, Kirzhner A, Kirzhner VM, Berezovsky IN. Distinct stages of protein evolution as suggested by protein sequence analysis. J Mol Evol 2001; 53:394-401; PMID:11675599; http://dx.doi.org/10.1007/s002390010229

35. Haig D, Hurst LD. A quantitative measure of error minimization in the genetic code. J Mol Evol 1991; 33:412-7; PMID:1960738; http://dx.doi.org/10.1007/BF02103132

36. Freeland SJ, Knight RD, Landweber LF, Hurst LD. Early fixation of an optimal genetic code. Mol Biol Evol 2000; 17:511-8; PMID:10742043; http://dx.doi.org/10.1093/oxfordjournals.molbev.a026331

37. Noller HF. Evolution of protein synthesis from an RNA world. Cold Spring Harb Perspect Biol 2012; 4:a003681; PMID:20610545; http://dx.doi.org/10.1101/cshperspect.a003681

38. Babitzke P, Baker CS, Romeo T. Regulation of translation initiation by RNA binding proteins. Annu Rev Microbiol 2009; 63:27-44; PMID:19385727; http://dx.doi.org/10.1146/annurev.micro.091208.073514

39. Borodavka A, Tuma R, Stockley PG. Evidence that viral RNAs have evolved for efficient, two-stage packaging. Proc Natl Acad Sci USA 2012; 109:15769-74; PMID:23019360; http://dx.doi.org/10.1073/pnas.1204357109

40. Borodavka A, Tuma R, Stockley PG. A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. RNA Biol 2013; 10:481-9; PMID:23422316; http://dx.doi.org/10.4161/rna.23838

41. Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, Jülicher F, Hyman AA. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science 2009; 324:1729-32; PMID:19460965; http://dx.doi.org/10.1126/science.1172046

42. Brangwynne CP, Mitchison TJ, Hyman AA. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. Proc Natl Acad Sci USA 2011; 108:4334-9; PMID:21368180; http://dx.doi.org/10.1073/pnas.1017150108

43. Weber SC, Brangwynne CP. Getting RNA and protein in phase. Cell 2012; 149:1188-91; PMID:22682242; http://dx.doi.org/10.1016/j.cell.2012.05.022