THE PROTEIN SOCIETY

Organization of the multiaminoacyl-tRNA synthetase complex and the cotranslational protein folding

Igor N. Berezovsky,[1,2]* Zejun Zheng,[1] Atsushi Kurotani,[3] Alexander A. Tokmakov,[4] and Igor V. Kurochkin[1]*

[1]Bioinformatics Institute (BII), Agency for Science, Technology and Research (a*STAR), Singapore 138671
[2]Department of Biological Sciences (DBS), National University of Singapore (NUS), Singapore 117579
[3]RIKEN Center for Sustainable Resource Science (CSRS), Yokohama, Kanagawa 230-0045, Japan
[4]Research Center for Environmental Genomics, Kobe University, Kobe, Japan

Received 14 April 2015; Accepted 24 June 2015
DOI: 10.1002/pro.2735
Published online 30 June 2015 proteinscience.org

Abstract: Aminoacyl-tRNA synthetases (ARSs) play an essential role in the protein synthesis by catalyzing an attachment of their cognate amino acids to tRNAs. Unlike their prokaryotic counterparts, ARSs in higher eukaryotes form a multiaminoacyl-tRNA synthetase complex (MARS), consisting of the subset of ARS polypeptides and three auxiliary proteins. The intriguing feature of MARS complex is the presence of only nine out of twenty ARSs, specific for Arg, Asp, Gln, Glu, Ile, Leu, Lys, Met, and Pro, regardless of the organism, cell, or tissue types. Although existence of MARSs complex in higher eukaryotes has been already known for more than four decades, its functional significance remains elusive. We found that seven of the nine corresponding amino acids (Arg, Gln, Glu, Ile, Leu, Lys, and Met) together with Ala form a predictor of the protein α-helicity. Remarkably, all amino acids (besides Ala) in the predictor have the highest possible number of side-chain rotamers. Therefore, compositional bias of a typical α-helix can contribute to the helix's stability by increasing the entropy of the folded state. It also appears that position-specific α-helical propensity, specifically periodic alternation of charged and hydrophobic residues in the helices, may well be provided by the structural organization of the complex. Considering characteristics of MARS complex from the perspective of the α-helicity, we hypothesize that specific composition and structure of the complex represents a functional mechanism for coordination of translation with the fast and correct folding of amphiphilic α-helices.

Keywords: multiaminoacyl-tRNA synthetase complex (MARS); predictor of α-helicity; cotranslational folding; side-chain rotamers; amphiphilic α-helices

Introduction

Aminoacyl-tRNA synthetases (ARSs) are key enzymes that catalyze an attachment of amino acids to their cognate tRNAs, thereby determining how the genetic code is translated into proteins. The ARSs are divided into two classes, I and II, based on the structural similarity of their active site domains. Unlike prokaryotic counterparts, ARSs in higher eukaryotes form a multiaminoacyl-tRNA synthetase complex (MARS) involving a subset of ARS polypeptides and three auxiliary proteins, p43, p38, and p18.[1] Disintegration of the complex in mice by truncation of p38 leads to lethality within two days of animal birth, pointing to

Abbreviations: ARSs, aminoacyl-tRNA synthetases; GAIT, γ-interferon activated inhibitor of translation; MARS, multiaminoacyl-tRNA synthetase complex; MDR1, multidrug resistance 1

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Igor N. Berezovsky, Bioinformatics Institute (BII), Agency for Science, Technology and Research (a*STAR), Singapore 138671; Department of Biological Sciences (DBS), National University of Singapore (NUS), Singapore 117579.
E-mail: igorb@bii.a-star.edu.sg (or) Igor V. Kurochkin, Bioinformatics Institute (BII), Agency for Science, Technology and Research (a*STAR), Singapore 138671. E-mail: igork@bii.a-star.edu.sg

its essential physiological role.[2] Subcellular localization of MARS complex indicates that it is well-positioned for the delivery of aminoacyl-tRNAs to the translation machinery. Several studies detected ARSs complex in either transient or stable association with polysomes.[3,4] Further, it was hypothesized that aminoacyl-tRNAs produced by MARS complex are directly transferred to the elongation factor followed by the direct transfer to the ribosome without dissociating into the cytosol.[5] Though existence of MARS complex in higher eukaryotic cells has been already known for more than four decades,[6] it is still unclear why assembly of its components is critical for maintaining the normal physiology. For example, association into MARS complex does not affect catalytic properties of individual synthetases.[7] Kyriacou and Deutscher[8] provided an evidence that MARSs complex can be essential for normal protein synthesis and cell growth because it facilitates channeling of aminoacyl-tRNAs to the ribosome. These authors have started from the observation that ArgRS exists in two forms in mammalian cells — an integral component of MARSs complex and in a free form. Arginyl-tRNA synthesized by the ArgRS involved into MARS complex was shown to be far more preferred substrate for translation than arginyl-tRNA made by the free form of ArgRS.[8] While these experiments showed a possible supportive role of MARS complex in translation, they did not explain the most intriguing feature of the complex. It is still an unanswered question why the only nine ARSs out of twenty (specific for Arg, Asp, Gln, Glu, Ile, Leu, Lys, Met, and Pro) are present in MARS complex regardless of the organism, cell, or tissue types.[1] The reason for inclusion of this peculiar set of ARSs into the complex is not clear since its members apparently do not share any common features. Indeed, MARS complex contains both classes of ARSs (class I and class II), and no correlations were found between ARSs presence in the complex and the size, expression profiles, or chromosomal locations of the corresponding genes.[9] The complex-forming and noncomplex-forming ARSs are not different in their preference for amino acids with particular physical-chemical properties, for example polar, hydrophobic or aromatic side-chain. Wolfson and Knight proposed that there is a link between the occurrences of ARSs in the complexes and the size of their substrate amino acids.[10] They observed that amino acids with large or small accessible surface area have ARSs that are excluded from the complex. It was hypothesized, therefore, that binding of aminoacyl-tRNA to eEF-1A for medium-size amino acids is weaker than for large and small amino acids. Therefore, structural organization of MARS complex may facilitate delivery of produced aminoacyl-tRNAs to eEF-1A in case of amino acids of an average size. However, experimental data supporting this assumption was not obtained.[10] Eswarappa and Fox pointed out that all amino acids (except Leu) corresponding to ARSs found in the complex are derivatives of the two citric acid cycle intermediates, α-ketoglutatrate and oxaloacetate.[11] They proposed that citric acid cycle has influenced the evolution of MARS complex formation. Exposure of cells in primitive multicellular organisms to hypoxia reduces a flux through the citric acid cycle. Therefore, an assembly into MARS complex may become advantageous for more efficient utilization of the limited amino acid pool and, as a consequence, for increase of the translation efficiency. This hypothesis does not explain, however, why MARS complex is maintained in higher organisms including mammals in which the availability of certain amino acids does not depend on the intermediary metabolism.[11]

Recently obtained structure of the ARSs complex revealed that it has elongated and multiarmed shape with the accessibility for all the components.[12] Copurification approach coupled with the gene knockdown of each of three nonsynthetase components of MARSs complex, p18, p38, and p43 revealed that the complex consists of two subcomplexes[13]: subcomplex I contains MetRS, IleRS, LeuRS, and two dimers of GluProRSs; subcomplex II includes Gln- and two ArgRSs. These domains are connected by the dimer of p38 proteins, which also binds the dimers of ArgRSs and LysRSs. Association of LysRS or AspRS to these complexes requires p38.[13] Crosslinking experiments proposed a different three domain model[14]: domain I contains MetRS, GlnRS, and the dimer of AspRS, which form one "arm" of the complex; domains II contains the dimers of ArgRSs and LysRSs that form second "arm"; and domain III contains IleRS, LeuRS, and the dimer of ProGluRSs in the base of the complex. The 3D working model of MARS complex obtained by the electron microscopy[15] corroborates grouping and spatial arrangement of synthetase components in MARS complex. Although different in details, all models unravel the sectioning of MARS complex into substructures that include ARSs for either hydrophobic or polar amino acids.

It is well established that protein can fold cotranslationally while it is still attached to the ribosome.[16] The rate of translation on the ribosomes can affect the protein folding pathways, shifting modes of the folding process in the nucleation–condensation mechanism between its extreme regimes described by the hydrophobic collapse model and the framework model.[17] Mechanism-wise, local discontinuous translation can temporally separate translation of the polypeptide chain's segments, actively coordinating their cotranslational folding.[18] Several mechanisms have presumably evolved to coordinate translational kinetics with the folding of a nascent polypeptide chain emerging from the ribosome. They

include ordering of high- and low-efficiency codons along mRNA, changing tRNA concentrations, secondary structure of the 5′-UTR, presence of regulatory sequence motifs in the vicinity of the translation start site, and so forth.[19] While replacement of rare codons by the frequent synonymous ones may enhance translation of the protein, it may result in the reduced folding efficiency.[20] A synonymous substitution in the human multidrug resistance 1 (MDR1) gene affects the timing of cotranslational folding of encoded P-glycoporotein, leading to a slight alternation in the native tertiary structure of the protein and, as a result, altered substrate specificity.[21] Protein folding on the ribosome can be aided by controlled rates of the polypeptide chain elongation coordinated by the biased codon usage. Amino acids in α-helices are typically encoded by frequent codons, while β-strands, turns, loops, and domain linkers encoded by the codons with lower occurrences.[22] Nascent polypeptides emerging from the ribosome and not yet folded can have hydrophobic patches and other structural features transiently exposed, serving as degradation signals.[23] They can also be involved into nonspecific interactions that result in the protein aggregation.[24]

Based on the analysis of crystallized proteins[25] and theoretical estimates of the flexibility of polypeptide backbone, it has been theoretically inferred[26] and experimentally corroborated[27] that closure of long loops is a crucial event in the folding of natural proteins.[28] It reduces the conformational space in orders of magnitude, providing a formation of the hydrophobic core[26] and directing an arrangement of already formed loops into the final fold/domain structure.[29] Different pathways of the protein folding process can be unified under the simple nucleation–condensation mechanism,[17] which considers general collapse/condensation of the polypeptide chain around a diffuse nucleus. In the nucleation–condensation mechanism the interplay of the secondary and tertiary interactions is crucial for reaching the native state of a protein. Specifically, the secondary structure formation is an important part of the folding process, determining the type of the folding pathway in relation to two abovementioned extremes of the nucleation–condensation mechanism, hydrophobic collapse model and framework model.[17]

Secondary structure elements, α-helices and β–hairpins (minimal independently stable β-structure) show significant difference in their rates of formation with the α-helices yielding folding times 30 times faster than β–hairpins.[30,31] Lower stability and stronger context-dependent propensity of the β–structure[32] is reflected in a typically longer folding times as well as in the demand on large initiation free energy and large minimal size of a stable β–sheet.[33–35] The α-helices, on the contrary, are formed locally in the absence of long-range interactions, folding, thus, fast and remaining quasi-independent and individually stable.[30,31,36] It has been even hypothesized in the beginning of the protein studies era that protein globule can be "originally" formed as a highly helical intermediate globule with a subsequent transition of α-helices into different types of secondary structure".[37] Though model of the "α-helical globule" as an initial folded state did not survive, the role of α-helices as fast folders, local elements of stability, and quasi-independent units of folding deserves specific consideration. Propensities of individual amino acids to form particular secondary structure are context-dependent,[32] and nonlocal interactions can determine the secondary structure of peptide sequences of substantial length.[38] Considering secondary structure propensity in different fold types, the α-helical propensities are similar in all folds for exposed and buried residues.[39] The β–sheet propensities calculated for exposed residues show that they determine amino acid composition in β–structures and govern their folding.[39]

Specific repertoire of aminoacyl-tRNA synthetases in the eukaryotic MARS complex prompted us to hypothesize that the very organization of MARS complex can be advantageous for the fast and efficient folding of the α-helical parts of proteins, contributing to the overall folding process and to prevention of degradation or aggregation of the nascent polypeptide chain in the otherwise crowded and potentially hostile environment of the cell.[16,23,24] We found here that there is a remarkably strong association between the occurrence of ARSs in the complex and presence of their amino acid substrates in the α-helical portions of polypeptide chains. Moreover, we found indications that the very structure of MARS complex can apparently provide position-specific propensities of amino acids in amphiphilic α-helices. We propose, therefore, that specific composition and structural organization of MARS complex underlies a functional mechanism for the coordination of translation with the fast cotranslational folding of α-helices.[40]

Results

We explore here a possible role of MARS complex in the cotranslational protein folding, specifically, in maintaining efficient synthesis of α-helices. We show that the composition of MARS complex can play an important role in supporting the α-helical propensity of a protein, and that structural organization of MARS complex can apparently be a provider of the position-specific amino acid propensity in α-helices.

The percentage of residues in the protein involved into α-helices is reflected in the amino acid composition of this protein. Even though α-helical propensity of individual amino acids has been studied for many years,[41–43] the very predictor of α-helicity, that is, the most optimal set of amino

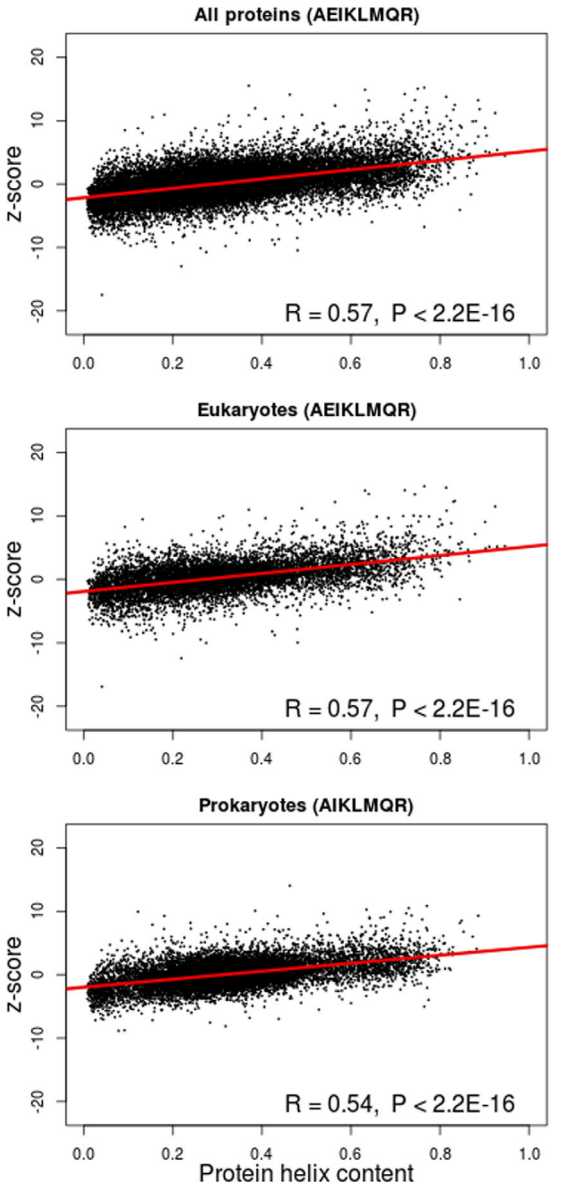

Figure 1. The *Z*-score predictor of protein α-helicity. The predictor is a correlation of the *Z*-scored amino acid composition of a protein, with the fraction of its structure involved into α-helices. Top chart shows predictor derived on the complete database (combination of amino acids is AEIKLMQR, correlation coefficient $R = 0.57$), the middle and bottom charts contain predictors for eukaryotes (AEIKLMQR, $R = 0.57$), and prokaryotes (AIKLMQR, $R = 0.54$), respectively (see also Supplementary Fig. S1).

acids beneficial for α-helical structures, has never been derived. We use here a simple and efficient way for determining such a set of amino acids in form of the *Z*-scored predictor of α-helicity, calculating a correlation of the *Z*-scored amino acid composition of a protein with the fraction of its structure involved into α-helices (see Materials and Methods for explanations). The optimal from the point of view of α-helicity combination, AEIKLMQR (Fig. 1 and Supporting Information Table S1), includes seven of the nine residues, which have corresponding aminoacyl-tRNAsynthetases in MARS complex (all except Ala). The predictor includes three hydrophobic (Ile, Leu, and Met), three charged (Glu, Arg, and Lys), and two polar (Ala and Gln) amino acids. Such well-balanced composition apparently shows that the overall demand on the degree of compactness and proper ratio between the hydrophobic core and hydrophilic surface[44] persists down to the elements of the secondary structure, in particular α-helices. The predictor of α-helicity is very robust, which is reflected in sets of top ten predictors obtained for groups of prokaryotic and eukaryotic proteins, and for sets of α-helices with intervals of sizes 5–40 and 5–70 amino acids, respectively (Supporting Information Table S1). Though AlaRS is not part of the complex, Ala is present in the predictor having the highest α-helical propensity among all 20 amino acids. Remarkably, all amino acids (besides Ala) in the predictor have the highest possible number of side-chain rotamers: Arg and Lys have 81 rotamers each, Gln–36, Glu–27, Ile/Leu–9 each, and Met–27. Thus, compositional bias of a typical α-helix can additionally contribute to the helix's stability by increasing the entropy of a folded state. Two synthetase components of MARS complex, AspRS and ProRS, correspond to amino acids, which are not in the predictor. In case of Asp, small entropic contribution of the side-chains can be an explanation for its absence in the predictor of α-helicity. Indeed, from the point of view of the helix's stability six side-chain rotamers of Asp do not contribute to the entropy of the folded state as much as side-chains of other charged residues favorable for the helix formation. At the same time, there can be other reasons for the presence of AspRS and ProRS in MARS complex, even though Asp and Pro are not crucial or even disadvantageous (as in the case of Pro) for the helix formation. For example, both biochemical analysis[13] and genetic dissection of protein–protein interactions[45] revealed a strong functional link between AspRS and other parts of MARS complex. Proline is involved into capping of α-helices,[46–48] and Asp can serve as a signal of the helix's start in the (−1) position and can substitute Glu in the P[ED][ED] starting pattern of the α-helix (Figs. 2 and 3).

Amino acid composition of α-helices corroborates a preference for the amino acids of predictor (Fig. 2), and it holds for separately considered proteins of eukaryotes and prokaryotes (Supporting Information Fig. S2). Compositional characteristics are slightly stronger pronounced in eukaryotes than in prokaryotes (Supporting Information Fig. S2). Noteworthy, short helices yield access of Phe, Tyr, Val, and Trp in all type of proteins. Presence of three aromatic residues (F, Y, and W) in the short helices points to their potential role in the protein–protein and protein–nucleic acid recognition and interactions, and involvement into the protein function.[49–51] Preliminary

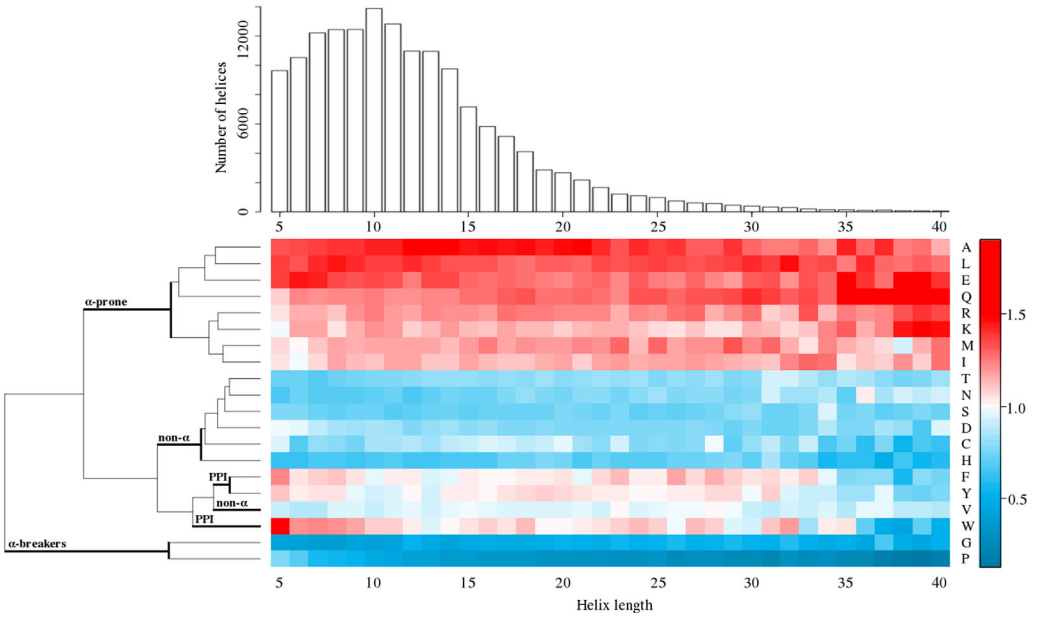

Figure 2. Amino acid compositional bias of α-helices depending on the helix size. Distribution of helices' sizes is mostly located in the interval 5–40 residues, with the maximum at 10 residues. Compositional bias clearly separates group of amino acids in the predictor from others, showing that predictor's amino acids are preferred in helices of all sizes (see also Supplementary Fig. S2). Groups of amino acids with high/low α-helical propensity (α-prone/non-α), of potentially involved into recognition, function, and protein–protein interactions aromatic residues (PPI), as well as Gly and Pro (α-brakers) are bold-marked.

inspection of the secondary structure annotation in PDB was suggestive of considering two sets of α-helices with the sizes 5–40 and 5–70 amino acids, respectively. The former group chiefly represents helices of globular proteins, the latter represents helices of the coiled-coil domains and those of motor, structural, and other nonglobular proteins. Distribution of the helix sizes shows that majority of α-helices in proteins lies in the interval between five and forty residues (total 154879), with only 556 helices having the size between 40 and 70 residues. Though we obtained relatively poor statistics on long helices, it does not affect the predictor (for the comparison predictor for the helices of lengths 5–70 residues is given in the Supporting Information Table S1). Therefore, one can cautiously consider the amino acid composition of long helices. Supporting Information Figure S2(c–e) shows an enrichment of long helices with Gln, Glu, Lys, and Arg along with a decrease in the amounts of Ala, Ile, and Met. All these effects are slightly stronger in eukaryotes, the same as in the case with the predictor. Taking into account that extra long helices are presumably elements of the coiled-coil domains present in transcription factors, motor protein, and so forth, overrepresentation of charged/polar residues is rather expected in such structures.

Helices of different sizes show preference for the predictor's amino acids in all positions, with Leu and Ala being the most frequent ones followed by the Lys and Arg (Fig. 3). The only exception is the first three positions, in which the combination ProGluGlu is systematically preferred as the starting motif [sometimes Glu can be replaced by Asp in the positions (2) and (3)]. It is supplemented by the few more specific biases: in the upstream position (−1) with preference for one of Asp, Asn, Gly, Ser, or Thr; and in downstream positions (+1) with preference for Gly and (+2) with excess of Gly or Pro. The starting ([DNGST]P[ED][ED]) and ending (G[GP]) so-called capping[46–48] signals are strongest for the helix length up to 35-40 residues, apparently showing that this size is the maximal characteristic size of the typical α-helix. Another visible characteristic yield by the helices' logos (Fig. 3) is manifestation of the 3.6-residue periodicity archetypal for α-helices, which is most pronounced in the helices with sizes up to twenty amino acid residues. To rigorously analyze this periodicity and to establish if there are some cellular mechanisms involved in maintaining α-helical patterns, we performed analysis of the position-dependent α-helical propensities. Specifically, we sought for a potential role of the organization of MARS complex in the formation of sequences and structures of α-helices. Figure 4 shows position-specific amino acid propensities for α-helices of sizes 10, 15, and 20 residues. It is easy to see that after the typical P[ED][ED] starting pattern there is a periodic alternation of two polar and two hydrophobic amino acids along the length of the helix. The effect is weakening closer to the C-termini of α-helix, and it is becoming noisy in longer helices (Supporting Information Fig. S3). The observed periodic pattern of the alternation of two hydrophobic

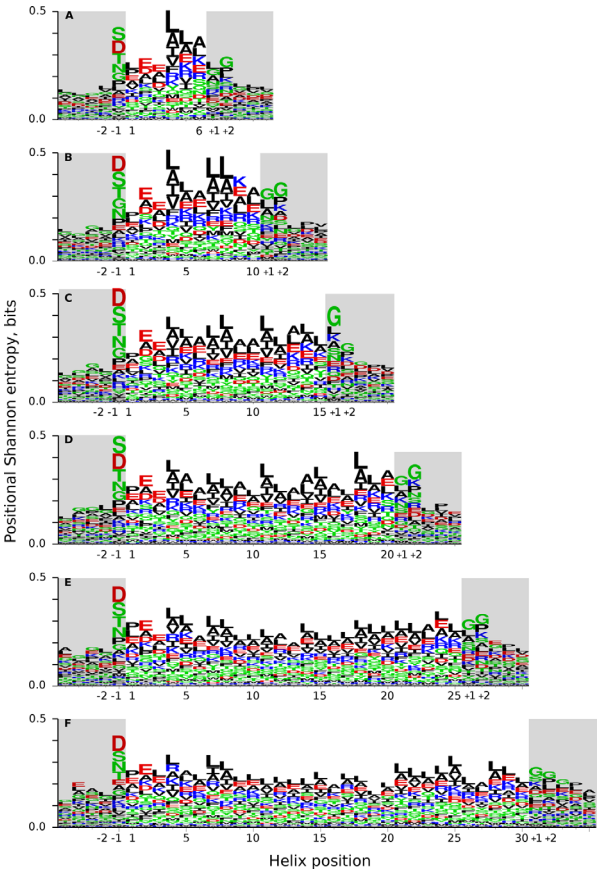

Figure 3. Profiles of the helix signatures in form of logos. A: Helix length is six amino acid residues. B: Helix length – 10 residues. C: Helix length – 15 residues. D: Helix length – 20 residues. E: Helix length – 25 residues. F: Helix length – 30 residues. Height of the letter reflects the percentage of the corresponding residue in the position. Compositional bias toward predictor is clear. Specifically, hydrophobic Ala, Ile, and Leu (black letters) are the most frequent in most of the helical positions, followed by the negatively charged Glu (red) and positively charged Arg and Lys (blue), along with starting signal (position (−1): [DSTGN]) and N-terminal (positions (1): [P], (2): [ED], and (3): [ED]) and C-terminal (position (+1): [G] and (+2): [GP]) capping. Helix periodicity (3.6 residues) is also well detectable up to the helix size of 25 amino acid residues.

and two polar amino acids is typical for amphiphilic α-helices that comprise majority of α-helices.[52] Given an alternation of hydrophobic and polar residues, composition of MARS complex (presence of the Arg-, Asp-, Gln-, Glu-, Ile-, Leu-, Lys-, Met-, and ProRSs) can play a particular role in guiding the folding of α-helices. Recent literature on the structural organization of MARS complex is suggestive of the specific arrangement/grouping of aminoacyl-tRNA synthetases of polar and hydrophobic residues.[13–15,45] Specifically, separation of corresponding ARSs into different "arm domains"[14] hints to the role of MARS complex in maintaining synthesis of amphiphilic α-helices. Persistently high occurrence of Ala in all positions of the helix (also reflected in the presence of Ala in the predictor of α-helicity) provides an additional indirect support for this hypothesis. Indeed, since Ala is over-represented in all positions of α-helices, there is no need for some mechanism that would secure an alternation of Ala with other amino acids in the growing helix. Therefore AlaRS is not a part of MARS complex,[13–15,45] while composition and spatial organization of corresponding ARSs in the complex can provide the alternation of hydrophobic and charged residues in the α-helices.[52]

Discussion

It has been shown that in addition to their role in translation,[53] amynoacyl-tRNA synthetases also work in noncanonical functions, such as regulation of gene expression, angiogenesis, and cellular signaling.[54,55] Mutations in ARSs can affect either translational or nontranslational functions, leading to different diseases.[56] Therefore a comprehensive view of the evolutionary and structural aspects of MARS complex points to the specific role of this complex in organization of the cellular apparatus in eukaryotes: ARSs can switch between translational functions when they are parts of MARS and nontranslational functions after dissociation from the complex and interactions with other partners.[57] In this work, we asked a question about the potential role of MARS complex in translational apparatus of multicellular eukaryotes. It appeared that one of the feasible answers lies in the role of MARS complex in supporting fast and efficient cotranslational folding of nascent α-helices. It has long been known that helices in globular proteins tend to be amphiphilic.[58–61] The noncompact architecture of MARS complex supposedly enables concurrent binding and release of all bulky tRNA substrates,[12] allowing thus simultaneous delivery of aminoacyl-tRNAs that correspond to alternating hydrophobic and polar amino acids. As a result, the α-helices can be rapidly folded, ensuring shortened exposure of the nascent polypeptide chain to surrounding environment and preventing its degradation or aggregation. To coordinate the synthesis of amphiphilic α-helix, MARS complex should optimally alternate between two states: one posed for the catalysis and release of aminoacyl-tRNAs for hydrophilic amino acids, and another one for the catalysis and release of aminoacyl-tRNAs for hydrophobic residues. Various studies established that MARS complex consists of several structurally stable and spatially separated subcomplexes.[13,15,45] The covalent labeling experiments revealed that ARSs in the complex are only partially accessible to the tRNA substrate molecules having on an average four bound tRNAs per complex.[15] This could be explained by only partial access of tRNAs to all ARSs in the complex or by the full access of tRNAs

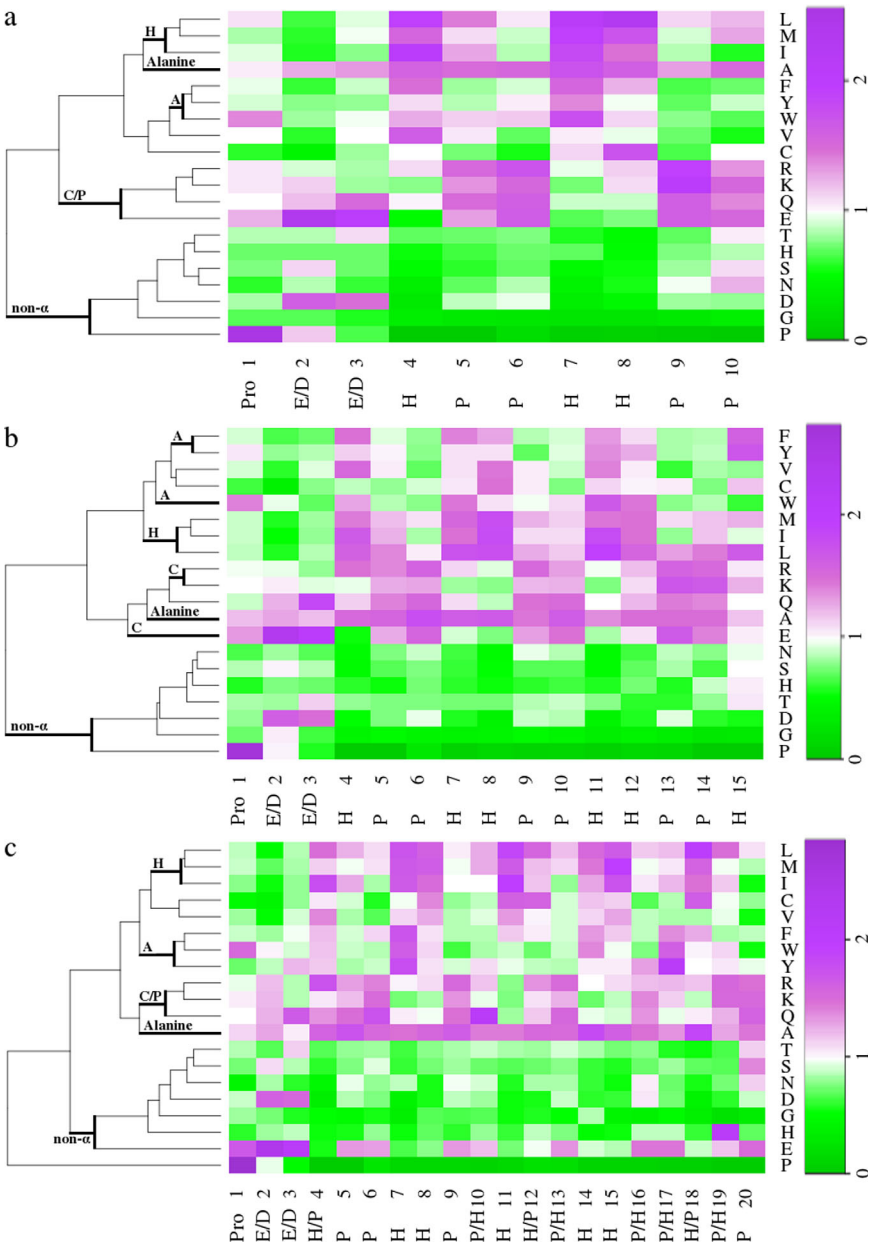

Figure 4. Position-dependent amino acid compositional biases for selected helix sizes. A: Helix length is 10 amino acid residues; (B) helix length – 15 residues; (C) helix length – 20 residues. Major groups of amino acids obtained in the clustering are bold-marked: Alanine; aromatic (A); charged (C); hydrophobic (H); polar (P); and non- α.

to about a half of ARSs in the complex at any given moment of time.

In his original work, Cyrus Levinthal proposed that "protein folding is speeded and guided by the rapid formation of local interactions which then determine the further folding of the peptide; this suggests local amino acid sequences which form stable interactions and serve as nucleation points in the folding process".[62] Levinthal's simple estimate on the number of accessible conformations (10^{300}) for the 150-residue protein served as a motivation for the hypothesis on cotranslational folding. Ironically, his clear statement on what we call nowadays cotranslational folding was turned into so-called "Levinthal paradox", which emphasizes on the astronomical number of conformations to be sampled en route to the native state. This so-called paradox was used during the last three-four decades as a justification for introducing scholastic models of the protein folding[63] largely irrelevant, however, to the real processes in a living cell.[18,64] Recent advances in the understanding of the protein structure had, finally, directed attention of researchers back to the proper formulation of the task: consideration of the cotranslational folding in a crowded cellular environment.[24,65] Here, we tackled only one aspect of this fascinating problem. We hypothesize that MARS complex is tailored to the cotranslational folding, specifically, cotranslational formation of the α-

helices. We propose here that composition and structural organization of MARS complex provides an additional layer of control over the rate of the cotranslational protein folding. Specifically, it represents a functional mechanism for the coordination of translation with the fast cotranslational folding of α-helices.[40] We found several indications of the link between MARS complex's organization and its potential role in the folding of amphiphilic α-helices along with nontranslational ones. First, predictor of the α-helicity consists of eight amino acids, Ala, Gln, Glu, Ile, Leu, Lys, Met, and Arg. For the comparison, synthetase components of MARS complex include seven aminoacyl-tRNA synthetases corresponding to all amino acids in the predictor besides Ala. Second, predictor of the α-helicity includes amino acids (all except Ala) with the highest possible number of the side-chain rotamers, which may contribute to the helix stability by increasing entropy of the folded state.[66] Third, alternation of the hydrophobic and polar residues typical for amphiphilic α-helices (Fig. 4 and Supporting Information Fig. S3) is well correlated with the grouping[13,14,45] and mutual availability[15] of synthetase components of MARS complex into domains/subcomplexes obtained in experiments.[13–15,45] Fourth, AspRS and ProRS synthetase components of MARS complex may play a specific role in the formation of α-helix. On the one hand, the absence of Asp and Pro in the predictor of α-helicity suggests that AspRS and ProRS are apparently not critical for the cotranslational elongation of the α-helices. For example, the AspRS and ProRS are not members of the complex in the nematode *C. elegans*.[67] On the other hand, the [DNGST]P[ED][ED] pattern in positions (−1), (1), (2), and (3) of the helix (Figs. 3 and 4) is suggestive that MARS complex starts a working cycle with GluProRSs initiated by the signal in positions (−1). Noteworthy, of all ARSs of higher eukaryotes only the genes of ProRS and GluRS are fused into one gene, encoding a single polypeptide GluProRSs.[68] Both AspRS and ProRS are involved into the N-terminal capping, and the pattern G[GP] in positions (+1) and (+2) shows that ProRS also works in the C-terminal capping. The noncanonical function of GluProRS in controlling inflammatory response has also been discovered.[69] Stimulation of human monocytes with IFN-γ triggers release of GluProRS from MARS complex, followed by its binding to three other proteins with formation of the heterotetrameric γ-interferon activated inhibitor of translation (GAIT) complex.[69] GAIT complex binds 3'-UTR of selected set of mRNAs involved in inflammation and represses their translation.[70] Interestingly, the release of GluProRS occurs 2 h after IFN-γ treatment, while the functional GAIT complex that represses translation is formed 12–14 hr later. Existence of the time-lag between these events is a puzzling phenomenon, calling for the explanation. To hypothesize, one can imagine, for example, that release of GluProRS from MARS complex might facilitate preferential translation of proteins enriched in β-structural elements. Strong functional involvement of AspRS is supported by the experimental data on copurification coupled with the gene knockdown of each of three nonsynthetase components[13] and by the genetic dissection of the protein–protein interactions in MARS complex.[45] Fifth, absence of the AlaRS in MARS complex indirectly supports a hypothesis that composition and spatial organization of the complex is important for providing the alternation of hydrophobic and polar residues in amphiphilic α-helices.[52] AlaRS is not included into MARS complex, because there is simply no demand on MARS complex's control over the placement of Ala in growing helix since Ala is omnipresent in all positions of the helices. Additionally, Ala is also one of the most abundant amino acids in the organismal tissues and fluids, such as blood plasma[71] and spinal cerebral fluid,[72] possibly also performing essential nontranslational functions.[54]

To conclude, combination of hypothesized involvement of MARS complex into the cotranslational folding of amphipathic α-helices and its possible role in nontranslational functions show that MARS complex can be a worthy target for the future experimental efforts and computational/theoretical modeling. In relation to translational function, MARS complex underwent long evolution that changed its composition and structure from the very simple to complex and culminated in the most advanced MARS of multicellular eukaryotes.[57] Existence of pronounced α-helical propensity already in prokaryotic proteins may well indicate that demand on the efficient folding of α-helices was apparently a driving force from the very beginning of the evolution of MARS complex. To this end, it would be important to experimentally explore how the architecture of MARS complex allows the spatiotemporal release of aminoacyl-tRNAs for the translation of different portions of amphiphilic α-helix. This task can be solved by the high-resolution analysis of the conformational changes of MARS complex caused by the adding of the cognate tRNAs of corresponding amino acids. Independent structural analysis of MARS complex with bound tRNAs of particular groups of amino acids (hydrophobic or polar) will allow one to compare structures of the active complex in different states, showing differences between them and illuminating relevant functional dynamics of MARS complex. Theoretical model of the conformational changes in switching between the states of initiation, active subcomplexes corresponding to hydrophobic and polar component of the helices, and the helix capping would also be of a great interest. Given the size and complexity of MARS complex's oligomeric structure, one should expect an interesting interplay between different levels of structural hierarchy and the allosteric mechanisms of regulation at work.

Materials and Methods

We used set of proteins with 50% redundancy downloaded from the PDB (release: 30 September 2014), which includes 12,981 prokaryotic and 9678 eukaryotic proteins. For selecting and analyzing the α-helical regions, we used a secondary structure annotation in PDB. We considered two sets of α-helices with sizes 5–40 (helices in globular proteins) and 5–70 amino acids (helices in the coiled-coil domains, motor, structural, and other nonglobular proteins). The number of α-helices with the lengths 5–40 and 5–70 residues are 154,879 and 155,435, respectively. We aligned α-helices by the N-end.

Earlier works showed that the compositional predictor, that is the fraction of amino acids that correlates the best with a certain characteristic, is a simple way to highlight the set of amino acids beneficial for selected characteristic.[73] Here we estimate an effect of different combinations of amino acids on the α-helicity of a protein, i.e. fraction of amino acids in the protein that are involved in α-helices. We define and derive *Z*-scored predictor of protein's α-helicity, that is combination of amino acids revealing the highest correlation (we use Pearson correlation coefficient) of the protein's *Z*-scored amino acid composition with the fraction of its structure involved into α-helices. To equalize the individual changes in amino acid frequencies and to make the effects of different amino acids comparable we standardize the frequencies of amino acids j for each protein m:

$$Z_m = \sum_j \frac{f_{mj} - \langle f_j \rangle}{\sigma_j}.$$

The f_{mj} is a frequency of the amino acid j in a protein m, $\langle f_j \rangle$ is an average frequency of amino acid j in the database, and σ_j is a standard deviation of frequencies of the amino acid j in the database. Thus, the "*Z*-score" predictor of the α-helicity correctly takes into account different variance of individual amino acids. We test all meaningful combinations of amino acids (in total $2^{19}-1$), selecting the one with the best correlation. Seven out of eight amino acid comprising the predictor of α-helicity have corresponding amynoacyl-tRNA synthetases in MARS complex. It is a very low probability ($\sim 10^{-5}$, given the equal probability for all combinations) to find specific combination of seven out of 20 amino acids (total number of combinations is 77,520).

We have calculated position-specific amino acids propensities,[74] based on the earlier introduced formalism used for the studies periodicity in α-helices.[52]. The position-specific propensity (P_{ij}) is calculated as:

$$P_{ij} = \frac{f_{ij}}{f_i} = \frac{n_{ij}/\sum_i n_{ij}}{N_i/\sum_i N_i},$$

where n_{ij} and f_{ij} are the number and fraction of an amino acid of type i in helical position j, N_i and f_i are the number and fraction of an amino acid of type i in the entire database of α-helices.

The hierarchical clustering (Figs. 2 and 4; Supporting Information Figs. S2 and S3) was performed in the R package (http://www.r-project.org/) using complete linkage method with Euclidean distance.

References

1. Park SG, Ewalt KL, Kim S (2005) Functional expansion of aminoacyl-tRNA synthetases and their interacting factors: new perspectives on housekeepers. Trends Biochem Sci 30:569–574.
2. Kim JY, Kang YS, Lee JW, Kim HJ, Ahn YH, Park H, Ko YG, Kim S (2002) p38 is essential for the assembly and stability of macromolecular tRNA synthetase complex: implications for its physiological significance. Proc Natl Acad Sci USA 99:7912–7916.
3. Kaminska M, Havrylenko S, Decottignies P, Le Marechal P, Negrutskii B, Mirande M (2009) Dynamic organization of aminoacyl-tRNA synthetase complexes in the cytoplasm of human cells. J Biol Chem 284: 13746–13754.
4. David A, Netzer N, Strader MB, Das SR, Chen CY, Gibbs J, Pierre P, Bennink JR, Yewdell JW (2011) RNA binding targets aminoacyl-tRNA synthetases to translating ribosomes. J Biol Chem 286:20688–20700.
5. Sivaram P, Deutscher MP (1990) Existence of two forms of rat liver arginyl-tRNA synthetase suggests channeling of aminoacyl-tRNA for protein synthesis. Proc Natl Acad Sci USA 87:3665–3669.
6. Bandyopadhyay AK, Deutscher MP (1971) Complex of aminoacyl-transfer RNA synthetases. J Mol Biol 60: 113–122.
7. Mirande M, Cirakoglu B, Waller JP (1983) Seven mammalian aminoacyl-tRNA synthetases associated within the same complex are functionally independent. Eur J Biochem 131:163–170.
8. Kyriacou SV, Deutscher MP (2008) An important role for the multienzyme aminoacyl-tRNA synthetase complex in mammalian translation and cell growth. Mol Cell 29:419–427.
9. Lee SW, Cho BH, Park SG, Kim S (2004) Aminoacyl-tRNA synthetase complexes: beyond translation. J Cell Sci 117:3725–3734.
10. Wolfson A, Knight R (2005) Occurrence of the aminoacyl-tRNA synthetases in high-molecular weight complexes correlates with the size of substrate amino acids. FEBS Lett 579:3467–3472.
11. Eswarappa SM, Fox PL (2013) Citric acid cycle and the origin of MARS. Trends Biochem Sci 38:222–228.
12. Dias J, Renault L, Perez J, Mirande M (2013) Small-angle X-ray solution scattering study of the multi-aminoacyl-tRNA synthetase complex reveals an elongated and multi-armed particle. J Biol Chem 288: 23979–23989.
13. Kaminska M, Havrylenko S, Decottignies P, Gillet S, Le Marechal P, Negrutskii B, Mirande M (2009) Dissection of the structural organization of the aminoacyl-tRNA synthetase complex. J Biol Chem 284:6053–6060.
14. Norcum MT, Warrington JA (1998) Structural analysis of the multienzyme aminoacyl-tRNA synthetase complex: a three-domain model based on reversible chemical crosslinking. Protein Sci 7:79–87.
15. Wolfe CL, Warrington JA, Treadwell L, Norcum MT (2005) A three-dimensional working model of the multienzyme complex of aminoacyl-tRNA synthetases

based on electron microscopic placements of tRNA and proteins. J Biol Chem 280:38870–38878.
16. Choi SI, Kwon S, Son A, Jeong H, Kim KH, Seong BL (2013) Protein folding in vivo revisited. Curr Prot Peptide Sci 14:721–733.
17. Daggett V, Fersht AR (2003) Is there a unifying mechanism for protein folding? Trends Biochem Sci 28:18–25.
18. Zhang G, Hubalewska M, Ignatova Z (2009) Transient ribosomal attenuation coordinates protein synthesis and cotranslational folding. Nature Struct Mol Biol 16:274–280.
19. Gingold H, Pilpel Y (2011) Determinants of translation efficiency and accuracy. Mol Systems Biol 7:481
20. Komar AA, Lesnik T, Reiss C (1999) Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. FEBS Lett 462:387–391.
21. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science 315:525–528.
22. Thanaraj TA, Argos P (1996) Protein secondary structural types are differentially coded on messenger RNA. Protein Sci 5:1973–1983.
23. Turner GC, Varshavsky A (2000) Detecting and measuring cotranslational protein degradation in vivo. Science 289:2117–2120.
24. Gershenson A, Gierasch LM (2011) Protein folding in the cell: challenges and progress. Curr Opin Struct Biol 21:32–41.
25. Berezovsky IN, Grosberg AY, Trifonov EN (2000) Closed loops of nearly standard size: common basic element of protein structure. FEBS Lett 466:283–286.
26. Berezovsky IN, Trifonov EN (2001) Van der Waals locks: loop-n-lock structure of globular proteins. J Mol Biol 307:1419–1426.
27. Ittah V, Haas E (1995) Nonlocal interactions stabilize long range loops in the initial folding intermediates of reduced bovine pancreatic trypsin inhibitor. Biochemistry 34:4493–4506.
28. Berezovsky IN, Trifonov EN (2002) Loop fold structure of proteins: resolution of Levinthas paradox. J Biomol Struct Dynam 20:5–6.
29. Berezovsky IN, Kirzhner VM, Kirzhner A, Trifonov EN (2001) Protein folding: looping from hydrophobic nuclei. Proteins 45:346–350.
30. Williams S, Causgrove TP, Gilmanshin R, Fang KS, Callender RH, Woodruff WH, Dyer RB (1996) Fast events in protein folding: helix melting and formation in a small peptide. Biochemistry 35:691–697.
31. Gilmanshin R, Williams S, Callender RH, Woodruff WH, Dyer RB (1997) Fast events in protein folding: relaxation dynamics of secondary and tertiary structure in native apomyoglobin. Proc Natl Acad Sci USA 94:3709–3713.
32. Minor DL, Jr, Kim PS (1994) Context is a major determinant of beta-sheet propensity. Nature 371:264–267.
33. Hartman R, Schwaner RC, Hermans J Jr. (1974) Beta poly(l-lysine): a model system for biological self-assembly. J Mol Biol 90:415–429.
34. Munoz V, Henry ER, Hofrichter J, Eaton WA (1998) A statistical mechanical model for beta-hairpin kinetics. Proc Natl Acad Sci USA 95:5872–5879.
35. Munoz V, Thompson PA, Hofrichter J, Eaton WA (1997) Folding dynamics and mechanism of beta-hairpin formation. Nature 390:196–199.
36. Schwarz G (1965) On the kinetics of the helix-coil transition of polypeptides in solution. J Mol Biol 11:64–77.
37. Lim VI (1978) Polypeptide chain folding through a highly helical intermediate as a general principle of globular protein structure formation. FEBS Lett 89:10–14.
38. Minor DL Jr., Kim PS (1996) Context-dependent secondary structure formation of a designed protein sequence. Nature 380:730–734.
39. Fujiwara K, Toda H, Ikeguchi M (2012) Dependence of alpha-helical and beta-sheet amino acid propensities on the overall protein fold type. BMC Struct Biol 12:18.
40. Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding 'speed limit'. Curr Opin Struct Biol 14:76–88.
41. Chou PY, Fasman GD (1974) Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. Biochemistry 13:211–222.
42. Eisenhaber F, Persson B, Argos P (1995) Protein structure prediction: recognition of primary, secondary, and tertiary structural features from amino acid sequence. Crit Rev Biochem Mol Biol 30:1–94.
43. Blaber M, Zhang XJ, Matthews BW (1993) Structural basis of amino acid alpha helix propensity. Science 260: 1637–1640.
44. Goncearenco A, Berezovsky IN (2014) The fundamental tradeoff in genomes and proteomes of prokaryotes established by the genetic code, codon entropy, and physics of nucleic acids and proteins. Biol Direct 9:29
45. Rho SB, Kim MJ, Lee JS, Seol W, Motegi H, Kim S, Shiba K (1999) Genetic dissection of protein-protein interactions in multi-tRNA synthetase complex. Proc Natl Acad Sci USA 96:4488–4493.
46. Doig AJ, Baldwin RL (1995) N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. Protein Sci 4:1325–1336.
47. Aurora R, Rose GD (1998) Helix capping. Protein Sci 7: 21–38.
48. Richardson JS, Richardson DC (1988) Amino acid preferences for specific locations at the ends of alpha helices. Science 240:1648–1652.
49. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. J Mol Biol 280:1–9.
50. Ma B, Elkayam T, Wolfson H, Nussinov R (2003) Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. Proc Natl Acad Sci USA 100:5772–5777.
51. Moreira IS, Martins JM, Ramos RM, Fernandes PA, Ramos MJ (2013) Understanding the importance of the aromatic amino-acid residues as hot-spots. Biochim Biophys Acta 1834:404–414.
52. Engel DE, DeGrado WF (2004) Amino acid propensities are position-dependent throughout the length of alpha-helices. J Mol Biol 337:1195–1205.
53. Ibba M, Soll D (2000) Aminoacyl-tRNA synthesis. Ann Rev Biochem 69:617–650.
54. Guo M, Schimmel P (2013) Essential nontranslational functions of tRNA synthetases. Nature Chem Biol 9: 145–153.
55. Hausmann CD, Ibba M (2008) Aminoacyl-tRNA synthetase complexes: molecular multitasking revealed. FEMS Microbiol Rev 32:705–721.
56. Yao P, Fox PL (2013) Aminoacyl-tRNA synthetases in medicine and disease. EMBO Mol Med 5:332–343.
57. Havrylenko S, Mirande M (2015) Aminoacyl-tRNA synthetase complexes in evolution. Intl J Mol Sci 16:6571–6594.
58. Wuilmart C, Urbain J (1984) Alpha secondary structures generate weak but recurrent periodicity in proteins. Eur J Biochem 139:35–40.
59. Kanehisa MI, Tsong TY (1980) Local hydrophobicity stabilizes secondary structures in proteins. Biopolymers 19:1617–1628.
60. Horne DS (1988) Prediction of protein helix content from an autocorrelation analysis of sequence hydrophobicities. Biopolymers 27:451–477.

61. Reyes VE, Phillips L, Humphreys RE, Lew RA (1989) Prediction of protein helices with a derivative of the strip-of-helix hydrophobicity algorithm. J Biol Chem 264:12854–12858.
62. Levinthal C (1969) How to fold graciously. In: Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House. Editors: J.T.P. DeBrunner and E. Munck Monticello, IL: University of Illinois Press, pp 22–24.
63. Garbuzynskiy SO, Ivankov DN, Bogatyreva NS, Finkelstein AV (2013) Golden triangle for folding rates of globular proteins. Proc Natl Acad Sci USA 110:147–150.
64. Zhang G, Ignatova Z (2011) Folding at the birth of the nascent chain: coordinating translation with co-translational folding. Curr Opin Struct Biol 21:25–31.
65. Ellis RJ, Hartl FU (1999) Principles of protein folding in the cellular environment. Curr Opin Struct Biol 9: 102–110.
66. Berezovsky IN, Chen WW, Choi PJ, Shakhnovich EI (2005) Entropic stabilization of proteins and its proteomic consequences. PLoS Comput Biol 1:e47
67. Havrylenko S, Legouis R, Negrutskii B, Mirande M (2011) *Caenorhabditis elegans* evolves a new architecture for the multi-aminoacyl-tRNA synthetase complex. J Biol Chem 286:28476–28487.
68. Cerini C, Kerjan P, Astier M, Gratecos D, Mirande M, Semeriva M (1991) A component of the multisynthetase complex is a multifunctional aminoacyl-tRNA synthetase. Embo J 10:4267–4277.
69. Sampath P, Mazumder B, Seshadri V, Gerber CA, Chavatte L, Kinter M, Ting SM, Dignam JD, Kim S, Driscoll DM, Fox PL (2004) Noncanonical function of glutamyl-prolyl-tRNA synthetase: gene-specific silencing of translation. Cell 119:195–208.
70. Mukhopadhyay R, Jia J, Arif A, Ray PS, Fox PL (2009) The GAIT system: a gatekeeper of inflammatory gene expression. Trends Biochem Sci 34:324–331.
71. Lacey JM, Wilmore DW (1990) Is glutamine a conditionally essential amino acid? Nutr Rev 48:297–309.
72. D'Aniello A, Fisher G, Migliaccio N, Cammisa G, D'Aniello E, Spinelli P (2005) Amino acids and transaminases activity in ventricular CSF and in brain of normal and Alzheimer patients. Neurosci Lett 388: 49–53.
73. Ma BG, Goncearenco A, Berezovsky IN (2010) Thermophilic adaptation of protein complexes inferred from proteomic homology modeling. Structure 18:819–828.
74. Kumar S, Bansal M (1998) Dissecting alpha-helices: position-specific analysis of alpha-helices in globular proteins. Proteins 31:460–476.