Role Of Long-Range Interactions And Ribosomal Electrostatics In Protein Folding

By

Daria V. Fedyukina

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 01/11/13

The dissertation is approved by the following members of the Final Oral Committee:
        Silvia Cavagnero, Professor, Chemistry
        Judith N. Burstyn, Professor, Chemistry
        James C. Weisshaar, Professor, Chemistry
        Thomas M. Record, Jr., Professor, Biochemistry/Chemistry
        John L. Markley, Professor, Biochemistry

Acknowledgements

I would like to express my gratitude to my advisor, Prof. Silvia Cavagnero, for her guidance, patience, and encouragement in my challenging research projects throughout the entire graduate school duration. I greatly appreciate her enthusiasm and persistence in research, no matter how frustrating the results of this research may be at times.

I am acknowledging the current and former members of the Cavagnero group for their support and making my tenure pleasant. Special thank you to Sarah Weinreis, Sneha Kumar, Ashok Sekhar, Arelys Rosado, and Margarita Santiago.

The main strength of the University of Wisconsin-Madison is its collaborative environment across all departments and schools. This is how UW stands out for me and this is why I am proud to become a UW alumnus. Therefore, I would like to specially acknowledge people who helped me selflessly: Newsha Ardalani and Sharad Punuganti (Computer Science); Prof. Julie Mitchell (Math Department); Dr. Milo Westler, Dr. Marco Tonelli, Dr. Mark Anderson, Dr. Rita Hannah, Dr. Arash Bahrami, and Dr. Fariba Assadi-Porter (Biochemistry, NMRFAM); Christopher Kirchgasler (ESL); Jeremy Gordon and Prof. Sean Fain (Med. Physics); Prof. Frank Korosec (Radiology); Daniel Olszewski (Weinert Center); Harry Webne-Behrman (Office of HR Development) and Darin Harris (Office of Quality Improvement).

I am especially grateful to my teachers and instructors who were not only the source of knowledge, but also a source of inspiration: Dr. Charles Fry, Dr. Monika Ivancic, Prof. Edwin Sibert, Prof. Hans Reich and many others.

As a Project and Teaching Assistant, I learned lessons of patience and politeness from Dr. Jeanne Hamers, Prof. Ned Sibert, Dr. Cheri Barta, Dr. Charlie Fry, and Dr. Monika Ivancic, and Prof. Fleming Crim. Thank you for shaping me into a strong leader.

The strength of Chemistry Department is not only in its superior education and research, but also, even more importantly, in its caring staff. I would like to thank Dr. Matt Sanders, Rosana Perez-Ellmann, Betty Harwood, Teresa Knudson, Kat Myhre, and many others without whom my life in the USA would not be perfectly comfortable.

I thank my dissertation committee members, Prof. Silvia Cavagnero, Prof. Judith Burstyn, Prof. John Markley, Prof. Thomas Record, and Prof. James Weisshaar, for dedication their time to read and comment on the material in this dissertation.

I would like to acknowledge Dr. Patricia Kokotailo and Mr. Larry Deroo for letting me live in their house and for creating a perfect atmosphere for writing this dissertation, as well as for celebrating every milestone.

Finally, I would like to specially thank my wise parents and husband who are at the core of my happiness both personally and professionally.

Dedicated to my mom and dad.

Table of Contents

List of Figures

Figure 2-10. Estimated helical populations in the N-terminal region of acid-unfolded (a) (1-77)apoMb, (b) (1-119)apoMb and (c) (1-153)apoMb expressed as percent helicities for the A, B and C regions of the sequence. Percent helicities were estimated from the assigned secondary chemical shifts, averaged over the pertinent portion of the sequence, divided by the experimental chemical shift expected for a 100% helix. The secondary chemical shifts corresponding to a 100% helix were taken to be 2.8 and 2.1 ppm for C$^\alpha$ and C', respectively [23]. Estimated percent helicities are shown for data derived from C$^\alpha$ (light blue) and C' (blue) nuclei. Averages over the values derived from C$^\alpha$ and C'

(IDPs, on the left) from independently folded proteins (on the right). The regions on the left, right, and in-between the dashed lines host IDPs, folded, and partially ordered proteins, respectively. Blue squares and red circles denote proteins with positive and negative mean net charges per residue, respectively. Calculations were performed based on the UniProtKB sequence information (see Materials and Methods). 129

Figure 4-3. Exit tunnel view of the electrostatic potential mapped on the 50S molecular surface of *E. coli* (left) and *H. marismortui* (right). Electrostatic potentials are obtained upon solution of the Linearized Poisson-Boltzmann Equation (LPBE) at 150 mM KCl with a solute dielectric of 2.0 and solvent dielectric of 78.0 by using 3D structures from 2AW4 [30] and 2QA4 [31] Protein Data Bank (PDB) entries for *E. coli* and *H. marismortui*, respectively. A blue color indicates regions with positive potential (> +3 kT/e) values and a red color indicates regions with negative potential (< -3 kT/e) values, where kT – a unit of energy equal to 4.11×10^{-21} Joules at room temperature (k – Boltzmann constant, T – temperature in Kelvin) and e – electric charge in Coulombs. 133

Figure 4-4. Schematic representation of the charge segregation concept. The negative charges enclosed in circles denote charges due to rRNA phosphate groups. 134

Figure 4-5. Change in amino acid composition upon transition from non-halophilic to halophilic species. Black bars: ribosomal proteins. Pattern bars: to some cellular proteins [22]. No data are available for asparagine (N) and glutamine (Q) in cellular proteins. The data were calculated as described in Materials and Methods using amino acid sequences of ribosomal proteins from seven non-halophilic and three halophilic organisms. 136

List of Tables

Role Of Long-Range Interactions And Ribosomal Electrostatics In Protein Folding

Daria V. Fedyukina

Under the supervision of Professor Silvia Cavagnero

in the Department of Chemistry, University of Wisconsin-Madison

Abstract

Protein folding is a ukary-chemical process by which polypeptide chains attain their native 3D structures. Understanding of the mechanisms of protein folding informs the search for a cure to protein misfolding diseases, e.g., Alzheimer's, Parkinson's, and prion-related malfunctions. The need for these cures is unquestionable: Alzheimer's, for instance, is predicted to affect 1 out of 85 people by 2050 and is considered the most costly disease, accounting for ca. 18% of the entire Medicare costs in the US including $21B of direct and $33B of indirect medical costs.

For many decades protein folding was studied in the test tube on isolated protein samples where many physiologically relevant components and conditions were ignored for simplicity. This dissertation attempts to understand how one of such components, the ribosome, modifies the protein folding process. In bacteria, the ribosome is a protein-synthesizing machinery, a 2.4-Mda macromolecular complex consisting of ~ 55 proteins and 3 large RNAs. The process of protein synthesis is known as translation and is highly

directional, i.e., each protein is synthesized from the N to C terminus. The folding of a growing protein chain starts on the ribosome and is a ubiquitous process in all organisms.

In order to understand the details of protein folding on the ribosome, this dissertation tackles two main questions: 1) *how the interactions between N- and C-termini affect the secondary structure of the N-terminus,* and 2) *how to study the protein folding in the presence of the ribosome.* Question 1 is addressed in Chapter 2, while Question 2 is discussed more broadly in all remaining chapters.

Chapter 1 provides a brief overview of the protein folding on the ribosome based on a published review article.

In *Chapter 2*, the C terminus of sperm whale apomyoglobin was removed to analyze its influence on the formation of secondary structure. This chapter proved the need to incorporated the directionality of the translation into the protein folding studies. We determined which long-range interactions between the N- and C- termini are inducing secondary structure formation in unfolded apomyoglobin. This finding informs apomyoglobin folding on the ribosome because the ribosome buries ~35-40 C-terminal residues in its tunnel preventing the interaction between N and C termini.

In *Chapter 3,* I discuss the inclusion of the ribosome into the study of the protein folding at amino-acid resolution. Hydrogen/Deuterium Exchange (HDX) on the ribosome bound nascent proteins (RNCs) was proposed as a methodology for the NMR detection of RNCs folding. I explored the appropriate conditions for HDX on the ribosome and concluded that two features must be taken into account. First, the treatment of RNC with a strong chaotrope is needed after the quenching of HDX. Second, the *in vivo* production of the SecM-stalled RNC is required to attain satisfactory yields for future spectroscopic

and spectrometric experiments. This chapter points out the complications of the HDX and SecM-stalling procedures and suggests modifications to future directions for protein folding on the ribosome studies.

Chapter 4 explores the electrostatic properties of the ribosomal surface and ribosomal proteins. The ribosomes and 50S ribosomal proteins from ten different organisms were analyzed with respect to their mean net charge, hydrophobicity, electrostatic potential, and accessible surface area. I found that ribosomal proteins exhibit a high degree of charge segregation. This charge segregation is more pronounced in the ribosomal proteins from halophilic organisms. I quantify the solvent-exposed fraction of the negatively and positively charged surface of the ribosomal proteins within the ribosome. My findings suggest that in addition to charge segregation, ribosomal proteins from halophiles tend to expose more negatively charged surface to the outer surface of the ribosome. Investigations of hydrophobicity and mean net charge per residue of ribosomal proteins showed that the ribosomal proteins from halophilies tend to be more negatively charged and less hydrophobic than the ribosomal proteins from non-halophilic organisms. Ribosomes from halophilic organisms appeared to have no fully ordered proteins.

Chapter 5 provides a brief description of future directions of the research discussed in this dissertation. The lessons learnt in the previous chapters inform these directions. First, I propose a new direction for the protein folding on the ribosome research (Chapter 3). Then, I suggest to continue the ribosomal electrostatic research (Chapter 4) by defining and quantifying the degree of charge segregation of ribosomal proteins.

Chapter 1   Introduction

Protein folding on the ribosome

The protein folding is a   ukary-chemical process by which a protein chain attains its correct 3D structure. Our review "Protein Folding At the Exit Tunnel" covers four main topics pertaining to this dissertation work:

1) introduction to protein folding *in vitro* and its limitations;

2) protein folding on the ribosome;

3) overview of current methods used to produce ribosome-bound nascent proteins for protein folding studies;

4) overview of physical and biological techniques used to unravel protein folding mechanisms on the ribosome.

This review concludes that there is an urgent need to transition from classical *in vitro* protein folding studies on a simple protein solution to the studies that incorporate other physiologically relevant factors such as the ribosome, chaperons, and molecular crowding. The critical role of the ribosome is highlighted throughout the review.

Protein Folding At The Exit Tunnel

This chapter is derived from a review article published by

Daria V. Fedyukina and Silvia Cavagnero

in

Annual Reviews Biophysics **2011**, 40, 337-359.

1.1 Abstract

Over five decades of research have yielded a large body of information on how purified proteins attain their native state when refolded in the test tube, starting from a chemically or thermally denatured state. Nevertheless, we still know little about how proteins fold and unfold in their natural biological habitat: the living cell. Indeed, a variety of cellular components, including molecular chaperones, the ribosome, and crowding of the intracellular medium, modulate folding mechanisms in physiologically relevant environments. This review focuses on the current state of knowledge in protein folding in the cell with emphasis on the early stage of a protein's life, as the nascent polypeptide traverses and emerges from the ribosomal tunnel. Given the vectorial nature of ribosome-assisted translation, the transient degree of chain elongation becomes a relevant variable expected to affect nascent protein foldability, aggregation propensity and extent of interaction with chaperones and the ribosome.

1.2   Principles Of *In Vitro* Protein Folding

Since Christian Anfinsen's pioneering article on the relation between protein sequence and structure in 1954 [1] and his formulation of the thermodynamic hypothesis of protein folding in 1962 [35], thousands of articles have been written on how proteins travel through energy landscapes and reach their native state. The large majority of this body of work considers the *in vitro* refolding mechanisms of pure proteins, starting from a thermally or chemically denatured state diluted into a buffer at physiologically relevant pH. Most experimental and computational studies have so far been carried out on small single-domain proteins. Multidomain proteins are still largely unexplored and have started to receive attention only recently [4].

Over five decades of research on the mechanisms of protein folding *in vitro* have revealed that there is a wide variability in the way different proteins fold in the test tube [9, 15, 77]. Nonetheless, a few important trends of general significance have emerged. The main concepts are worth a summary here because they can be considered the basis for understanding fundamental aspects and mechanistic differences once proteins are allowed to fold and unfold in the complex cellular environment.

First, protein folding does not proceed via a random search [51], and protein energy landscapes are highly funneled [8, 99]. The above facts greatly contribute to optimize the efficiency of the conformational search to reach the native state. As a result, a variety of parallel paths are typically present as proteins fold, each generally comprising the formation of numerous transiently populated species, i.e., kinetic intermediates (some experimentally undetectable) separated by energy barriers in the case of rugged landscapes, or progressively evolving conformations undergoing barrierless diffusion

toward the native state. The latter scenario typically applies only to very small (<60 residues) proteins. In experimental studies, single-exponential kinetics is often observed. It is important to keep in mind that single-exponential folding is fully compatible with the concept of parallel folding pathways, and it does not necessarily imply a truly two-state folding, which is rarely observed. Indeed, multiple unfolded or partially folded conformations often interconvert faster than the rate-determining steps; hence they do not give rise to distinct kinetic phases [22]. In addition, computer simulations suggest that kinetic intermediates are usually present, yet they may be poorly populated and therefore experimentally undetectable [16]. Several proteins fold via experimentally detectable folding intermediates, which in some cases are en route to the native state.

Second, individual elements of secondary structure may form very fast [18], as in the case of α-helices (typically <1 μs), but are usually not stable in the absence of long-range tertiary contacts. Therefore, protein folding is generally not a rigorously hierarchical process, and it is extremely rare that high populations of secondary structure (e.g., helices) fold first, followed by collapse and tertiary structure formation. This idea is schematically illustrated in Figure 1-1 as the class of paths denoted by dashed gray lines, comprising type 1, 2, and 5 species. Studies on isolated polypeptides representing portions of primary structure of entire proteins show that individual helices and sheets are usually unstructured in the absence of surrounding tertiary contacts [23]. Investigations on the early stages of protein folding showed that only small populations of secondary structure are detectable before chain collapse (exceptions are some members of the engrailed homeodomain family and protein A [15]). Furthermore, protein variants containing destabilized versions of highly intrinsically helical regions of the chain are

folding competent [12].

Third, the timescale for protein chain collapse is highly variable (nanoseconds to seconds) and sequence dependent [77]. Collapse may (a) occur after most of the secondary structure is formed, as rarely observed experimentally (gray path in Figure 1-1); (b) be concurrent with most secondary structure formation, as seen in a number of apparently two-state folders (blue path in Figure 1-1) giving rise to relatively slow collapse with topology-dependent rates; (c) be concurrent with some secondary structure formation followed by slower acquisition of additional secondary structure, as in proteins with detectable folding intermediates such as apomyoglobin (apoMb) [42] (blue path in Figure 1-1), or (d) precede most secondary structure formation (black path in Figure 1-1).
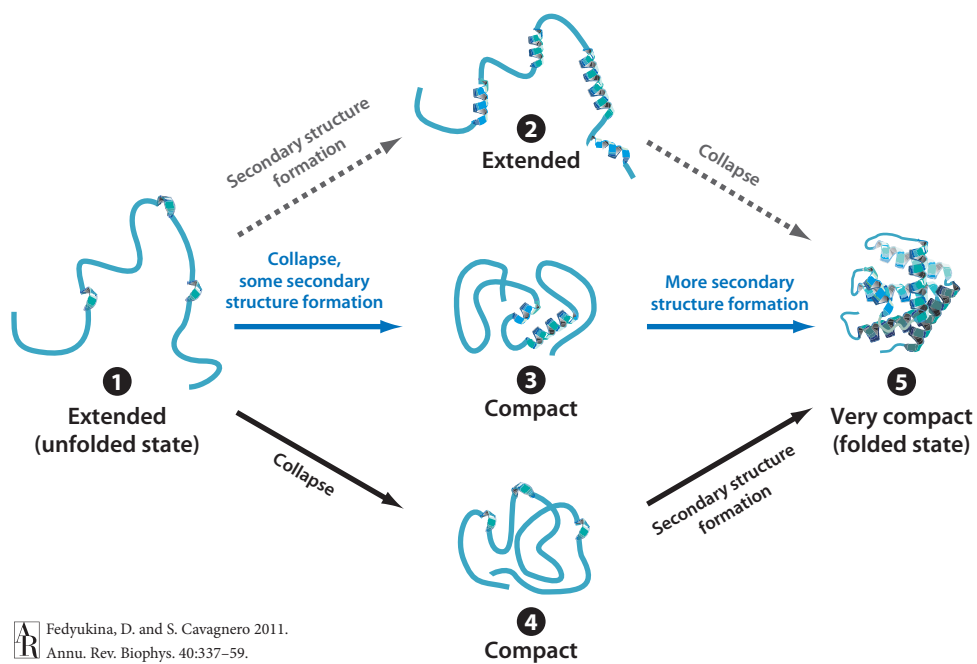
Figure 1-1. In vitro protein folding mechanisms. Different mechanisms are denoted by dashed gray (rarely observed), dark blue, and black arrows. The experiments leading to the formulation of these models are typically performed in purified protein solutions and involve the refolding of unfolded states generated chemically or by temperature jumps. Note that the species other than the unfolded and folded states (denoted 2, 3, and 4) may be either intermediates, transition states, or transient species populated along diffusive downhill routes.

The sequence determinants for the above options are not entirely clear yet and represent an outstanding challenge in *in vitro* protein folding. On the other hand, there are two apparent emerging trends. Collapse is slower when it occurs concomitantly with secondary structure formation, pointing to the kinetic difficulties in assembling secondary and tertiary structure together. In addition, secondary structure formation starting from a collapsed intermediate is also typically slow, pointing to the kinetic challenges in sampling conformational space from collapsed species (especially in large proteins). The above is true even if these species have significant internal dynamics, for instance, in the case of molten globules, and may bear a solvated nonpolar core.

Fourth, the starting species of *in vitro* folding experiments, the so-called unfolded state, is sometimes far from lacking a structure; therefore, it is not truly unfolded [60, 80]. Only expanded highly dynamic unfolded state ensembles follow the three criteria outlined above. Unfolded states bearing significant secondary structure and/or compaction are clearly posed to apply biases to the conformational search, sometimes making it more efficient. The presence of secondary/tertiary structure in proteins under strongly denaturing conditions is particularly interesting in the context of this review, given that the unfolded state populated under physiologically relevant conditions sometimes behaves differently from a self-avoiding Gaussian chain [71].

Fifth, a significant fraction (~40% in Eukarya) of the proteins expressed in the cell is actually natively unfolded [102]. Representatives of this class are known as intrinsically disordered proteins (IDPs) and lack a well-defined independent structure at physiologically relevant pH and ionic strength. IDPs often fold upon interaction with their biological counterparts: their folding mechanisms, still poorly explored, are beyond

the scope of this work.

1.3   Protein Folding In the Cell

1.3.1   *In Vitro* and *In Vivo* Protein Folding

Based on the results of pioneering nuclear magnetic resonance (NMR) experiments in live cells, the native structure of medium-size proteins in the intracellular environment is believed to be similar to the one populated *in vitro* in buffered solution. However, folding mechanisms in the cell are bound to be different from *in vitro* folding (Figure 1-2) due to the presence of a different unfolded state (see below); molecular chaperones; the ribosome; a highly crowded medium (200–300 mg*ml^{-1} total protein concentration); cofactors such as heme, NADH, and others; intracellular processes such as posttranslational modifications; and quality control processes such as protein degradation. In addition, some proteins are also subject to translocation into and out of different cell compartments, secretion, and co-translational insertion into membranes. The latter processes are neglected in this review, which focuses on the folding of cytosolic soluble proteins.
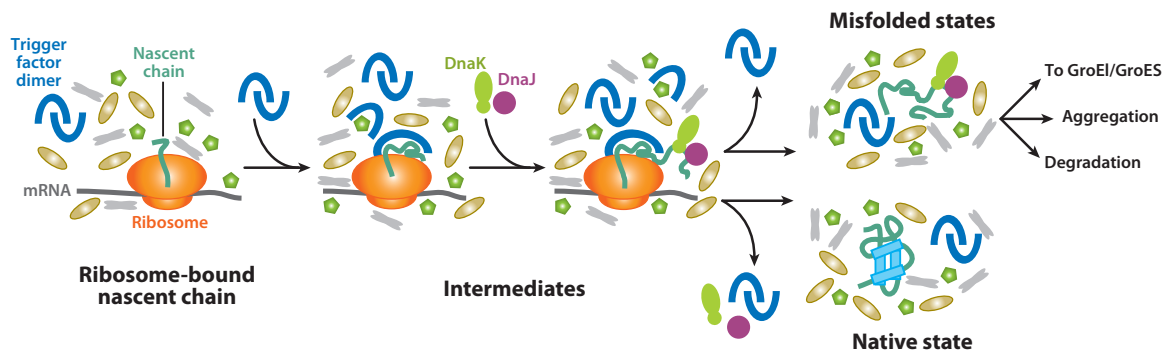
Figure 1-2. Key aspects of cotranslational protein folding in the crowded milieu of the cellular cytosol.

1.3.2   The Unfolded State

Protein folding and unfolding in the cell can occur either during or after protein biosynthesis, i.e., co- or post-translationally. In both cases, the nature of the unfolded state is poorly understood, yet likely profoundly different from the non-physiological unfolded state ensemble of *in vitro* experiments in denaturants. For instance, in the case of full-length proteins in aqueous media at pH 7, the unfolded state is more compact than in the presence of denaturants [101]. The effect of molecular crowding on the unfolded state under native conditions has yet to be studied in depth.

1.3.3   Molecular Chaperones

Molecular chaperones are key components of the cellular environment in bacteria, eukarya, and archaea. Their identity and roles have been reviewed elsewhere [3, 17, 28, 30, 37, 38, 55, 56]. Chaperones assist protein folding in the cell by preventing protein misfolding and aggregation and possibly also promoting folding [58, 82, 86]. Interestingly, many chaperones in bacteria have overlapping specificities and their roles are often overlapping, except for the bacterial GroEL/ES, the lack of which is lethal to the cell [32].

1.4   Folding On the Ribosome: What Is Special About It?

1.4.1   Incomplete Protein Chains from Single-Domain Proteins Do not Generally Assume a Native-Like Conformation

The presence of ribosome-bound incomplete protein chains is one of the unique features of co-translational events. In 1967, i.e., soon after the discovery that the biosynthesis of most proteins is catalyzed by the ribosome and proceeds vectorially from

the N terminus to the C terminus, Phillips [65] formulated the hypothesis that the N-terminal portion of nascent proteins may start folding during translation. Two years later, Taniuchi & Anfinsen [81] responded by showing that co-translational folding is unlikely for small- and medium-size single-domain proteins because individual purified N-terminal fragments of staphylococcal nuclease (Snase) of increasing length do not achieve any stable fold until their length closely approaches that of the complete protein. Since then, additional experimental model studies on Snase showed that the C-terminally truncated protein can indeed become compact yet partially disordered with only some of its secondary structure if very few residues are removed from its C terminus [29].

This finding suggests that the thermodynamic driving force for native-like tertiary structure formation develops during the very latest stages of chain elongation. Analogous studies on chymotrypsin inhibitor 2 and barnase are in agreement with the above ideas [61]. Computational studies based on the burial of nonpolar surface as a function of chain elongation further support this concept [49, 50].

1.4.2   Incomplete Protein Chains can be Prone to Aggregation

Chain elongation model studies on purified model polypeptides from the medium-size (17 kDa) all-α-helical protein sperm whale apoMb [13] provide additional support to the idea that the native fold can be achieved only at lengths close to that of the complete primary structure. In addition, this study shows that incomplete N-terminal chains (from 36 to 119 residues of the 153-residue full-length protein), rich in nonpolar residues, exhibit a strong tendency to aggregate and form nonnative β-strands. The above misfolding/aggregation progressively decreases in magnitude as chain length approaches the full-length protein. This model system study highlights a unique feature of incomplete

protein chains bearing a high nonpolar content: their tendency to aggregate. Aggregation of incomplete nascent chains is not tolerable in the cellular environment. The ability of the ribosome to keep chains maximally segregated during translation was demonstrated by a recent cryo-electron tomography study of *Escherichia coli* polysomes [7]. Individual ribosome components of the polysome adopt a staggered or pseudohelical mutual arrangement, with nascent chains maximally spaced and pointing toward the cytosol. This 3D arrangement is naturally poised to minimize self-association of nascent proteins. Another investigation showed that the ribosome's ability to keep chains segregated prevents the aggregation of incomplete chains of the tailspike protein from the *Salmonella* phage P22 even in the absence of the co-translationally active trigger factor (TF) chaperone [25]. As soon as ribosome release of the tailspike nascent chains is induced, the incomplete-length chains undergo self-association. The intrinsic ability of the ribosome to prevent the aggregation of rhodanese and lysozyme was also shown [33].

1.4.3 The Ribosome and Its Exit Tunnel Provide a Unique Environment for Nascent Chain Conformational Sampling

The tethering of all nascent polypeptides to the ribosome leads to expanding the function of this amazing machine from that of an mRNA decoding center and catalyst for peptide bond formation to an obligatory scaffold, and possibly interaction counterpart, during the co-translational conformational sampling of nascent polypeptides and proteins. The last few years have witnessed enormous progress in the elucidation of the archaeal and bacterial ribosome structure, structure-function relations, and assembly [66, 70, 79, 88, 94]. The structures of the 50S large subunit and the entire ribosome solved at high resolution by X-ray crystallography [2, 36, 74, 91] provide ideal support to all studies of

protein folding in and out of the exit tunnel. Figure 1-3a shows the high-resolution 3D structure of the *E. coli* ribosome, including the small and large subunits and the ribosomal proteins. Figure 1-3b,c provide schematics of a section of both the bacterial and archaeal ribosomes, respectively, highlighting the ribosomal exit tunnel and the proteins that directly face the tunnel's interior (L4, L22, L24) or are in close proximity to the tunnel (L23 and L29 in bacteria, and L23, L29, and L39e in archaea). Nascent proteins traverse the tunnel from the ribosome active site (i.e., the peptidyl transferase center, which houses the nascent protein C terminus) up until the tunnel's exit [31, 85]. The tunnel is not completely straight and has a bend. Its length spans 80 to 100 Å, depending on where the exit side end of the tunnel is defined (Figure 1-3d) [85].

file kindly provided by N.R. Voss and P.B. Moore) [85]. Abbreviations: PTC, peptidyl

transferase center; TF, trigger factor.

1.4.4   Kinetic Considerations on Co-translational Protein Folding

The best way to study folding at the exit tunnel is undoubtedly to watch the development of nascent protein structure and dynamics concurrently with translation. As shown in the next section, following up on this opportunity is especially desirable for large proteins, given that their translation rates approach intrinsic folding rates [62] and that it is likely that codon usage and ribosomal pausing are poised to affect the actual mechanism of folding.

Indeed, studying the co-translational folding of fairly small single-domain proteins would also be extremely useful to verify that translation rates are slower than conformational sampling on the ribosome. However, to the best of our knowledge no such studies have been performed, although there are excellent prospects for progress in this area in the near future.

Co-translational protein folding studies need to preserve the natural translation rates (so that they can be compared with folding rates) and are therefore best performed *in vivo*. However, working in an *in vivo* environment is challenging due to (a) the difficulties in selectively detecting folding in the complex cellular environment, and (b) the inability to synchronize translation given the stochastic nature of the process. Biological approaches pioneered by A. Helenius and F.U. Hartl have solved challenge *a* by monitoring protein activity co-translationally, and challenge *b* by pulse-chase experiments often performed in bulk spheroplasts.

1.5   What A Difference Translation Makes

1.5.1   Biosynthesis Rates Affect the Extent of Cotranslational Folding in Multidomain Proteins and Can Be Ad Hoc Modulated

Protein synthesis proceeds at variable rates in different environments and organisms (see Reference 93 and Table 1 in Reference 13). For instance, translation rates are faster *in vivo* than under cell-free conditions. In addition, translation proceeds faster in prokarya (15–20 amino acids s^{-1}) than in eukarya (3–4 amino acids s^{-1}), leading to an average timescale for the production of a small-/medium-size protein of ~10 s and ~65 s in prokarya and eukarya, respectively. These fairly long timescales are similar to chaperone binding/unbinding times and longer than the folding/unfolding timescales of small proteins. The above suggests that, *in vivo*, nascent chains encoding small proteins may have sufficient time to adopt preferred conformations as they are synthesized. On the other hand, very large proteins take multiple seconds to fold and may or may not attain stable conformations co-translationally. Accordingly, the absence [62] or presence [63] of *in vivo* co-translational folding in *E. coli* seems to be highly protein and codon dependent. Rare codon clusters [14], sometimes localized at interdomain junctions in large proteins, are emerging as important sites for an orchestrated pausing. This pausing is responsible for facilitating co-translational domain folding before synthesis of the following domain is initiated [92]. A proper balance between translation rates and co- and post-translational folding is important for the production of active ribosome-released multidomain proteins. For instance, mutant ribosomes displaying slower translation than wild-type *E. coli* ribosomes enhance the production of active multidomain proteins (of eukaryotic origin) in bacteria [76]. A detailed review of this topic was recently published by Zhang &

Ignatova [93].

1.5.2   Preparation and Analysis of RNCs for Model Studies on the Conformation of Nascent Proteins at Equilibrium

The highest resolution information on protein folding at the exit tunnel has so far been achieved via studies on purified arrested ribosomes bearing nascent proteins, sometimes labeled with fluorophores or NMR-active tags. These studies implicitly assume that nascent protein chains have the opportunity to conformationally equilibrate faster than the rate of translation. This assumption is likely acceptable in many cases, especially for small proteins. However, in general, caution should be exercised, and it is desirable to assess the validity of this approximation in each case.

The most common methods to prepare ribosome-bound nascent chains (RNCs) for model studies at equilibrium are outlined in Figure 1-4a. An exhaustive overview of these methodologies is beyond the scope of this review. Therefore, we simply provide general guidelines here.

RNCs can be prepared via genetic approaches exploiting addition or *in situ* production of truncated mRNAs. These methodologies have been employed *in vitro*, either in cell-free systems (e.g., from *E. coli,* wheat germ, or rabbit reticulocyte) [20, 96, 103] or via reconstituted systems containing all the necessary components for translation (e.g., PURE, protein synthesis using recombinant elements, based on prokaryotic components) [64, 75]. As an example, a procedure for the generation of RNCs in cell-free systems via coupled transcription/translation is shown in Figure 1-4b [20, 96]. Reconstituted *in vitro* expression systems have recently emerged as a convenient option because of their complete lack of nucleases, proteases, tmRNA, and other undesired

components.

Alternatively, stalled ribosomes can be generated via protein-based approaches by controlling translation arrest via special gene products, i.e., short amino acid sequences (typically ∼15–25 residues) that interact strongly with specific portions of the ribosomal tunnel and cause translation to stop (Figure 1-4a). The most popular of these approaches is based on generating very stable RNCs via the 17-residue SecM arrest sequence [26, 67, 68, 69], as shown in Figure 1-4c. The SecM approach is particularly convenient when RNCs are generated *in vivo*, given the affordability of the method and its high yields.

On the other hand, the SecM sequence is by far not the only available method to generate arrested ribosomes by protein-based approaches. Other strategies, based on both intrinsic and inducible classes of ribosome-stalling sequences (e.g., MifM, TnaC, and the Arg attenuator peptide), have been reviewed recently [41] (Figure 1-4a).
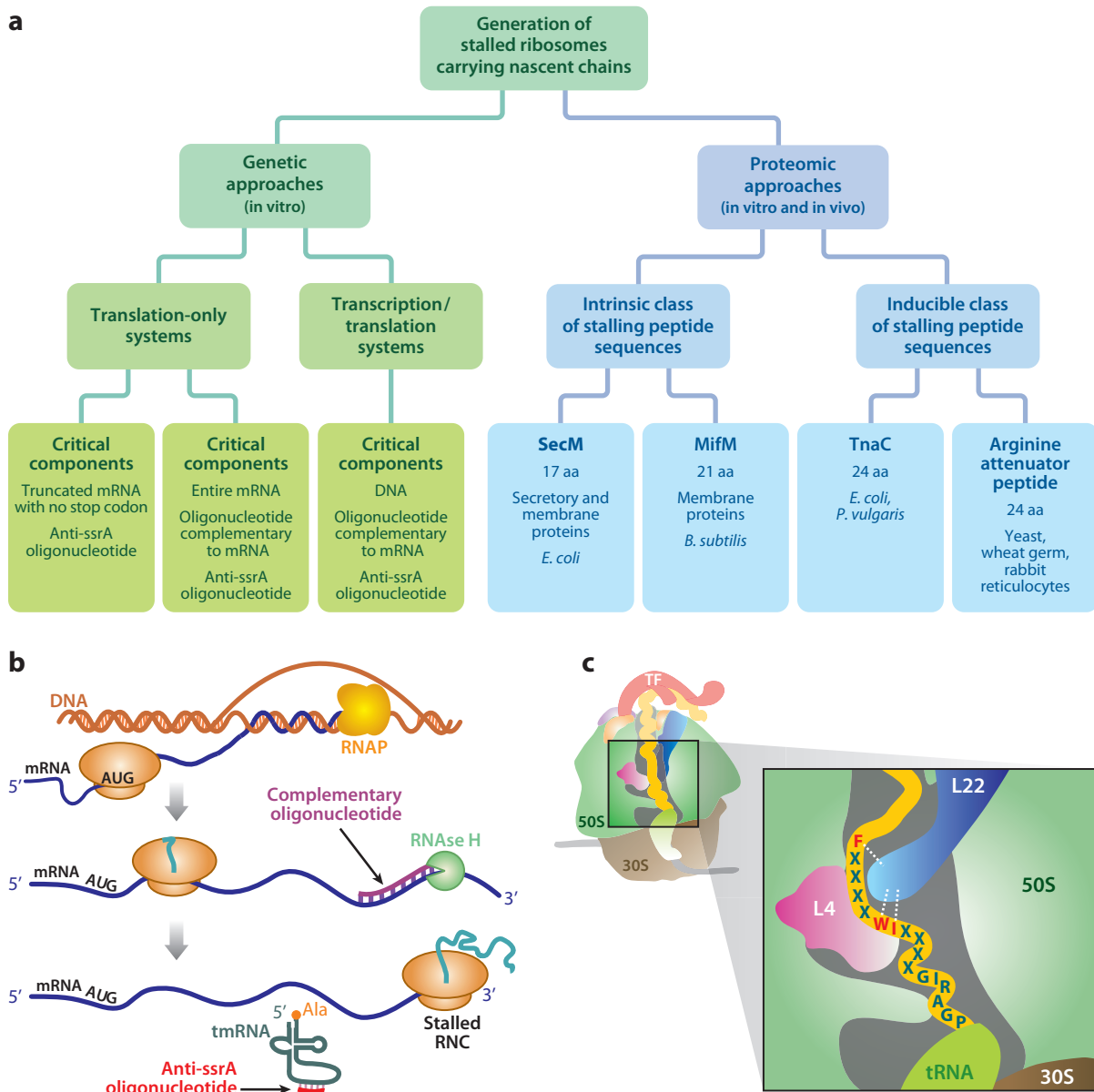
Fedyukina, D. and S. Cavagnero 2011.
Annu. Rev. Biophys. 40:337–59.

Figure 1-4. Methods for generating ribosome-bound nascent proteins. (a) Overview of currently available methods to generate RNCs of well-defined chain length. Step-by-step procedures based on (b) *in vitro* (cell-free) coupled transcription-translation [78] and (c) SecM stalling. For simplicity, cotranslationally active chaperones are omitted. The 17-residue SecM peptide-stalling sequence (FXXXXWIXXXXGIRAGP) is shown inside the

ribosomal tunnel. The underlined amino acids (in *red*) experience critical interactions with the ribosomal tunnel (white *dashed lines)* with L22. Abbreviations: RNAP, RNA polymerase; RNC, ribosome-bound nascent chain; X, any residue.

Generation of RNCs at equilibrium enabled analysis of the structure and dynamics of nascent proteins at a level of detail presently unattainable by *in vivo* studies. This analysis has allowed addressing several questions regarding the stepwise generation of 3D protein structures in nature. Figure 1-5 provides a global overview of the RNC structural and dynamic aspects that have been addressed thus far. This figure also links these specific folding-related aspects to the technique used to gain the desired information. As shown in Figure 1-5, a large variety of spectroscopy- and microscopy-based and biological techniques have been used synergistically. Perhaps even more importantly, the diagram shows that the same structural/dynamic question can often be addressed by both biological and spectroscopy/microscopy biophysical approaches. This synergism has been particularly valuable in the field of RNC folding, given the challenging features of RNC for direct biophysical analysis (e.g., large internal dynamics, conformational heterogeneity, and large size of the RNC complexes). Despite the potential of the biophysical methods to provide higher resolution insights, biological approaches [97, 98] taking advantage of properties such as protein activity and antibody response are often efficient and highly informative.
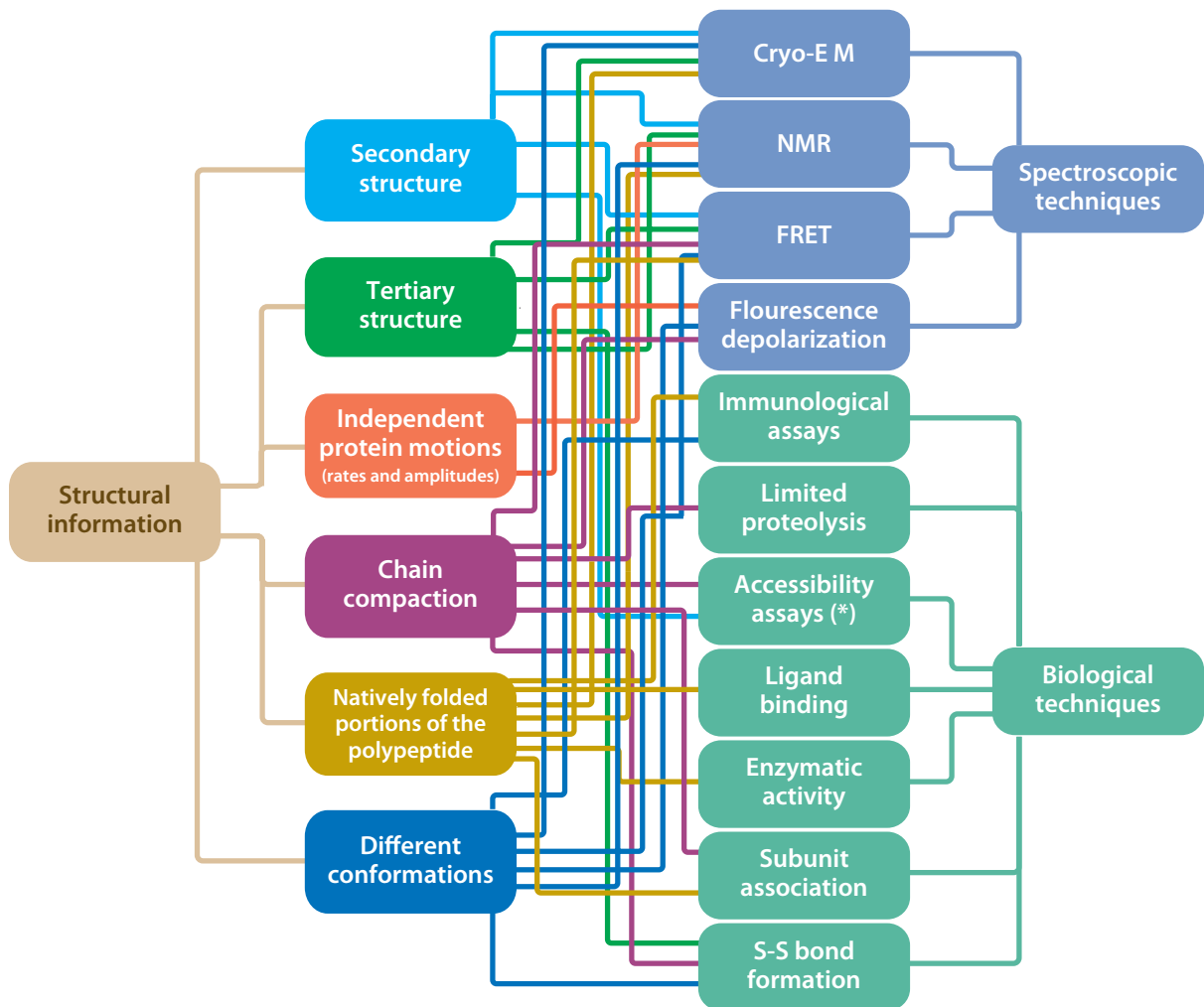
Fedyukina, D. and S. Cavagnero 2011.
Annu. Rev. Biophys. 40:337–59.

(*) e.g., pegylation, N-linked glycosylation

Figure 1-5. Relationship between specific RNC structural features and biological or spectroscopic techniques employed to elucidate them. Abbreviations: cryo-EM, cryo-electron microscopy; FRET, Förster resonance energy transfer; NMR, nuclear magnetic resonance; RNC, ribosome-bound nascent chain.

1.5.3   Investigations on Nascent Polypeptides Inside the Ribosomal Tunnel

The ribosome exit tunnel (Figure 1-3d) has a width that ranges from 10 Å (constriction site) to 20 Å (widest region). These dimensions are incompatible with tertiary structure formation within the tunnel's interior. Major tunnel dynamics, presumably accompanied by extensive ribosome rearrangements, would be required for the tunnel to host even a simple tertiary fold such as a helical hairpin. Therefore, although it is recognized that the ribosome is a dynamic entity, it is likely that no tertiary structure, and only secondary structure, can be populated in nascent chains inside the ribosomal tunnel [85]. This argument is supported by the finding of a highly spatially confined environment inside the tunnel, revealed by recent investigations showing that the N terminus of nascent polypeptides buried in the tunnel experiences narrow local motions [21].

The lower limit of the tunnel length is 80 Å, as proposed by Voss et al. [85] for the archaeal ribosome. Given this value, a fully α-helical polypeptide would bury approximately 53 residues (assuming an effective length of 1.5 Å per residue for the α-helix), and a fully extended polypeptide would bury approximately 23 residues (assuming an effective length of 3.5 Å per residue for an extended chain). These geometrical considerations prompt the question of whether any specific secondary structure is supported by the tunnel. Pioneering experiments by Malkin & Rich in 1967 [57], using *in vivo* pulse-chase techniques followed by cell lysis and proteolysis, showed that approximately 30 to 35 residues of nascent globin are protected from proteolysis in eukaryotic polysomes, implying that those residues are buried inside the ribosomal exit tunnel. These results are consistent with later investigations (reviewed in Reference 47)

showing that there are 30 to 40 protected residues in nascent proteins. The above finding supports the presence of a partially helical conformation inside the tunnel. Computational studies by Ziv et al. [95] showed that a helical conformation can be entropically favored in a cylinder that models the ribosomal tunnel's dimensions. This result suggests that even a Teflon-like non-interacting tunnel [2] may be capable of inducing helical structure, especially in the case of nascent polypeptides whose coil state is highly disordered.

Förster resonance energy transfer (FRET) investigations on peptide sequences from a soluble secretory protein and a membrane protein showed that inside the eukaryotic tunnel the former is less helical than the latter [43]. As the polypeptide chain elongates, helices can persist beyond the tunnel if they are stable in that environment. For instance, a peptide sequence from an integral membrane protein stays helical outside the tunnel as it is inserted into the membrane. However, the same sequence loses its helicity if the ribosomal surface faces bulk solution and is not bound to the membrane. Moreover, peptide sequences from soluble proteins have negligible helicity both inside and outside the tunnel [43]. In summary, the ribosomal tunnel is capable of inducing helicity in nascent polypeptides, and this phenomenon is highly sequence dependent.

Additional investigations from Deutsch and coworkers [52–54] support the above conclusion by exploiting ingenuous accessibility assays, which enabled the detection of distinct tunnel zones characterized by different (highly negative) electrostatic potential. The authors also showed that some of these tunnel regions promote polypeptide chain compaction, suggestive of helix formation [46, 83, 84].

Recent cryo-electron microscopy (cryo-EM) work by Beckman and coworkers [5,

6] provides to date the highest-resolution insights on nascent secondary structure within the exit tunnel. As shown in Figure 1-6, the authors detected helical structure for sequences with high helical propensity in distinct regions of the tunnel [5]. However, sequences with lower intrinsic helical propensity are disordered [5, 73]. This work effectively complements and supports the findings by Johnson and Deutsch [46, 52, 84, 90, 103].

Whether the secondary structure formation in distinct regions of the tunnel results from specific polypeptide-tunnel interactions or whether it is driven (or at least contributed by) by entropic effects is still unclear and in need of further investigation. In the specific case of the SecM and TnaC ribosome stalling sequences, convincing evidence for tunnel-polypeptide interactions was presented [73, 90].
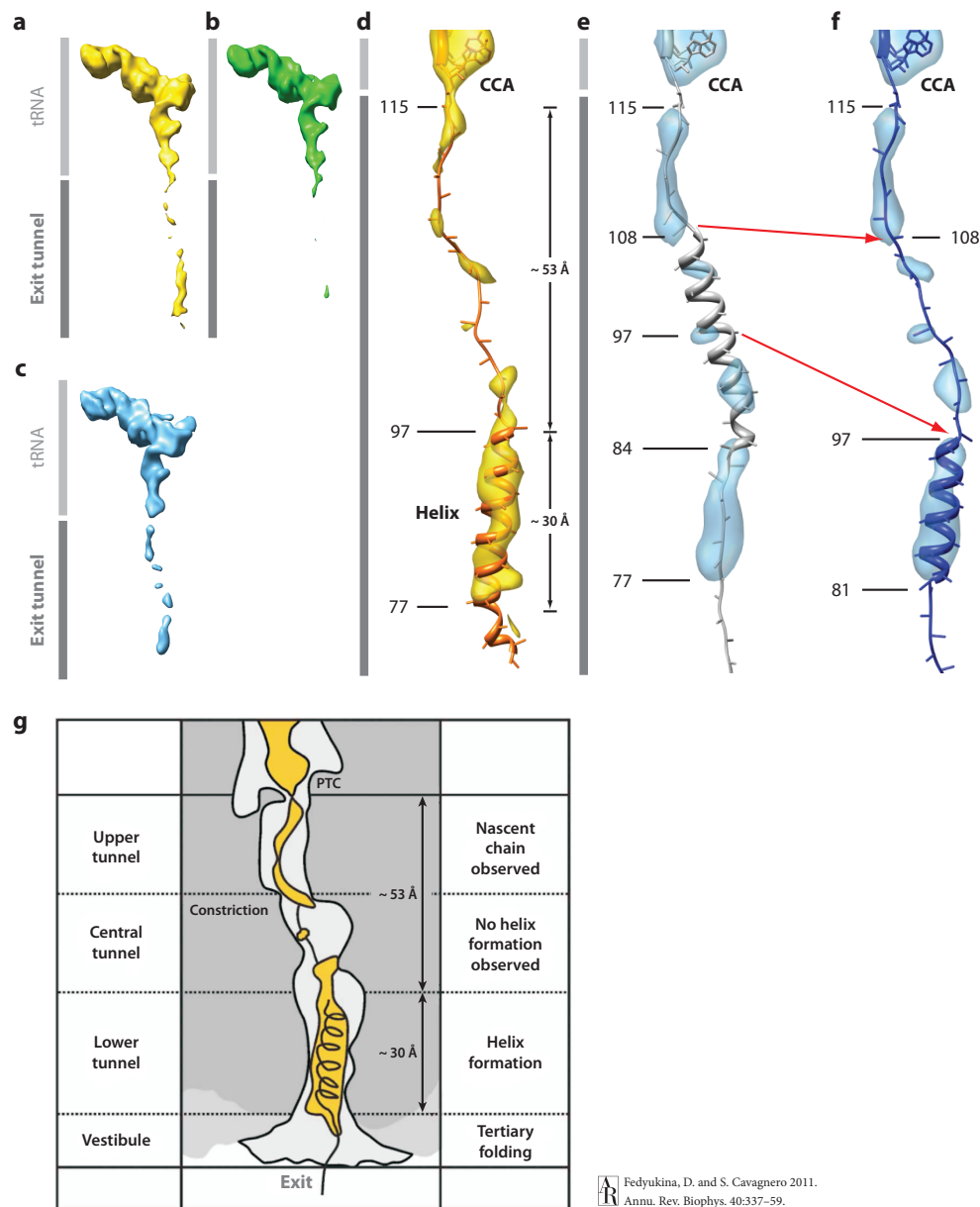
Figure 1-6. Cryo-EM maps of different peptidyl tRNAs inside the eukaryotic ribosome's P-site and exit tunnel. (a) 80S–helix 1 RNC, (b) 80S–DPAP RNC, (c) 80S–helix2 RNC, and (d) enlarged view of transparent density of panel a with fitted ukary model for tRNA and nascent chain. (e, f) Enlarged view of panel c with alternative models for helix 2 nascent chain. Red arrows indicate corresponding region (residues 97–108) modeled as helical (e) or extended (f). (g) Schematic cross-section of 80S–helix 1

RNC representing helix formation within the exit tunnel. Abbreviations: cryo-EM, cyro-electron microscopy; RNC, ribosome-bound nascent chain; DPAP, dipeptidylaminopeptidase; PTC, peptidyl transferase center. Adapted by permission from Macmillan Publishers Ltd: Nature Structural & Molecular Biology (Reference 5), copyright (2010).

1.5.4 Investigations on Nascent Proteins Emerging from the Ribosomal Tunnel

Our understanding of polypeptide conformation and dynamics as nascent proteins emerge from the ribosomal tunnel is not as advanced as our knowledge on nascent peptide structure inside the tunnel. The many experimental challenges presented by out-of-tunnel RNCs include the high conformational heterogeneity of the nascent chain and the variable effects introduced by co-translationally active chaperones. Nevertheless, considerable progress has been made, and recent technical advances hold promise for additional exciting future progress. Comprehensive reviews of earlier work are available [24, 27, 47]. Here, we focus on recent findings. In short, several examples of independent nascent structure and dynamics were discovered in RNCs emerging from the tunnel, defying the earlier proposal [81] that proteins acquire an independent conformation only after departing from the ribosome.

Cryo-EM images of polypeptides emerging from the ribosomal tunnel [34, 59] provided somewhat moderate structural detail. These studies were important to establish the possibility of tertiary structure formation outside the exit tunnel, in small single domain proteins. Additional evidence on 3D structure development comes from nascent chains from the ion channels, where tertiary structure was detected close to the ribosomal tunnel exit, via accessibility experiments based on side chain pegylation [45, 46]. These results are supported by recent computational investigations [100].

Analysis of fluorescence depolarization decays of RNCs and ribosome-released fluorophore-labeled apoMb in the frequency domain [87] enabled Ellis et al. [20] to study the dynamics of nascent apoMb's N terminus on the subnanosecond timescale and follow the formation of an independent protein domain on the nanosecond timescale, as shown

in Figure 1-7a,b. ApoMb RNCs acquire independent dynamics, indicative of compact or semi-compact species, only when a significant portion of the sequence emerges from the ribosomal tunnel. The rotational correlation time reporting on the protein's nanosecond local motions increases significantly upon nascent protein release from the ribosome, showing that the structure of the full-length RNC differs from that of the ribosome-released native apoMb. RNCs encoding the natively unfolded protein PIR (phosphorylated insulin receptor interaction region) experience no motions on the nanosecond timescale, suggesting that PIR does not fold on the ribosome. The spatial amplitude of the nascent chain local motions is very narrow inside and, surprisingly, even outside the ribosomal tunnel (Figure 1-7c,d) [21]. This is true even when RNCs are depleted of bound chaperones (TF and Hsp70). This result suggests that both the tunnel and the outer surface of the ribosome exert a severe local confinement on nascent apoMb and PIR.

The limits of NMR spectroscopy have been pushed by recent studies on RNCs at atomic resolution [10, 11, 19, 39, 40]. These investigations revealed that nascent single-domain proteins are not fully structured before they have entirely emerged from the ribosomal tunnel, consistent with the expectation that the C-terminal portion of the chain plays an important role in folding [49, 50].

Taken together, the above findings suggest that relatively small, full-length single-domain nascent proteins may adopt compact conformations outside the ribosomal tunnel. However, the nascent chains whose buried C-terminal residues are not available for folding may retain a considerable degree of disorder. Additional future studies are needed to provide more extensive evidence for these emerging trends.

a Apparent rotational correlation times for apoMb and PIR RNCs: intermediate timescale (ns) local motions

b Short RNCs (two components) → Long RNCs (three components) → Ribosome-released protein (two components)

ApoMb

BODIPY-FL

■ sub-ns
■ ns
■ μs

All RNCs (2 components)

PIR

c Cone semiangle for the local sub-ns motions of the N terminus of flourophore-labeled RNCs

d Cone semiangles of apoMb and PIR nascent chains: fast (sub-ns) local motions

Wild type { ApoMb RNCs, PIR RNCs }

Δtig TF-depleted { ApoMb RNCs (chaperone-free), PIR RNCs (chaperone-free) }

Length of ribosome-bound nascent chains (number of amino acids)

Fedyukina, D. and S. Cavagnero 2011.
Annu. Rev. Biophys. 40:337–59.

Figure 1-7. Dynamics of apoMb RNC by fluorescent depolarization. (a) Frequency domain dynamic fluorescence depolarization of ribosome-bound apoMb and PIR nascent chains generated in an *E.coli* cell-free system. Data are shown only for the nanosecond local motions that reveal the presence of a small compact or semi-compact species. (b) Scheme highlighting the motions associated with each fluorescence phase with each associated component of the motion. (c) Scheme illustrating the spatial amplitude of the

subnanosecond local motion of the N terminus of the fluorophore-labeled RNC. The symbol θ_o represents the cone semi-angle (in *red*) assessed in panel *d*. (d) Amplitude of the fast (subnanosecond) motions experienced by the N termini of nascent apoMb and natively unfolded PIR nascent polypeptides of increasing length under different conditions. Panels *a* and *b* adapted with permission from References 20 and 21, respectively. Copyright 2008 and 2009, respectively, American Chemical Society and John Wiley and Sons. Abbreviations: apoMb, apomyoglobin; PIR, phosphorylated insulin receptor interaction region; RNC, ribosome-bound nascent chain; TF, trigger factor chaperone.
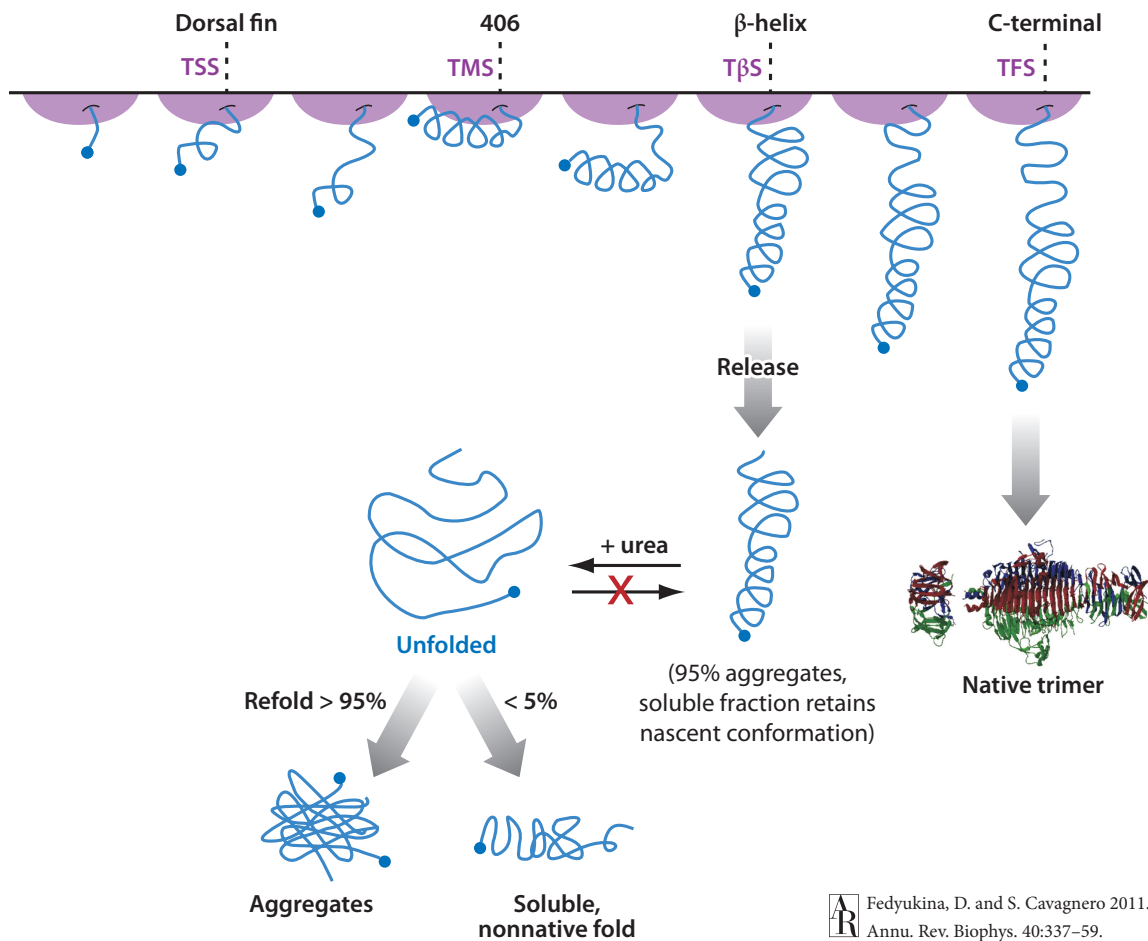
Figure 1-8. Model for the cotranslational folding of P22 tailspike nascent protein chains. Abbreviations: TSS, tailspike short stalled nascent chain; TMS, tailspike mid-length stalled nascent chain; TβS, tailspike stalled nascent chain with the entire β-helix exposed; TFS, tailspike full stalled nascent chain. Reprinted from the *Journal of Molecular Biology,* Vol. 383, Evans MS, Sander IM, Clark PL. "Cotranslational folding promotes beta-helix formation and avoids aggregation *in vivo*" pp. 683–92, Copyright (2008), with permission from Elsevier.

The influence of the ribosome on protein folding is striking particularly for very large proteins unable to fold *in vitro* in the absence or presence of molecular chaperones. For such systems (e.g., the trimeric phage P2 tailspike protein), co-translational folding is an irreplaceable requirement to attain the folded state and exploit biological activity [25]. This important concept is illustrated in Figure 1-8.

1.6 Summary points

1. Chain compaction preceding or concurrent with secondary structure formation is a dominant class of mechanisms for the *in vitro* folding of small- and medium-size proteins, starting from largely unstructured unfolded ensembles.

2. The unfolded state of full-length proteins is believed to be rather compact in aqueous solution and physiological pH. Hence, secondary structure formation from compact states may be an important motif in posttranslational protein folding in the cell. Landscapes corresponding to this process may be rather rugged.

3. Incomplete N-terminal protein fragments (lacking the C terminus) often lack much of the native structure and may aggregate in aqueous solution and physiological pH, in the absence of the ribosome and molecular chaperones.

4. What do incomplete protein chains look like before translation is complete? The answer to this question is still largely unknown but great progress has been made over the past few years. There is a lot of activity in this exciting area.

5. The ribosomal tunnel is narrow and it provides an extremely spatially constrained environment for nascent polypeptides. The tunnel is capable of inducing helical structure, even in nascent polypeptides (derived from soluble proteins) that lack independent structure in solution. However, this process is highly sequence dependent.

6. The ribosomal tunnel consists of zones that differ in chemical potential and may promote secondary structure formation to a different degree.

7. Folding-competent proteins emerging from the ribosomal exit tunnel can assume a

compact or semi-compact conformation. Small single-domain proteins experience variations in their chain dynamics (and possibly folding) as they are released from the ribosome.

8. Very large proteins such as P22 tailspike are incapable of reaching their native state unless they are allowed to fold vectorially on the ribosome.

1.7   References

1. Anfinsen CB, Redfield RR, Choate WI, Page J, Carroll WR. 1954. Studies on the gross structure, cross-linkages, and terminal sequences in ribonuclease. *J. Biol. Chem.* 207:201–10

2. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution. *Science* 289:905–20

3. Baram D, Pyetan E, Sittner A, Auerbach-Nevo T, Bashan A, Yonath A. 2005. Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. *Proc. Natl. Acad. Sci. USA* 102:12017—22

4. Batey S, Nickson AA, Clarke J. 2008. Studying the folding of multidomain proteins. *HFSP J.* 2:365–77

5. Bhushan S, Gartmann M, Halic M, Armache J-P, Jarasch A, Mielke T, Berninghausen O, Wilson DN, Beckmann R. 2010. Alpha-helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* 17:313–17

6. Bhushan S, Meyer H, Starosta AL, Becker T, Mielke T, Berninghausen O, Sattler M, Wilson DN, Beckmann R. 2010. Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol. Cell* 40:138–46

7. Brandt F, Etchells SA, Ortiz JO, Elcock AH, Hartl FU, Baumeister W. 2009. Thenative 3D organization of bacterial polysomes. *Cell* 136:261–71

8. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. 1995. Funnels, pathways, and

the energy landscape of protein folding: a synthesis. *Proteins Struct. Funct. Bioinform.* 21:167–95

9. Brockwell DJ, Smith DA, Radford SE. 2000. Protein folding mechanisms: new methods and emerging ideas. *Curr. Opin. Struct. Biol.* 10:16–25

10. Cabrita LD, Dobson CM, Christodoulou J. 2010. Early nascent chain folding events on the ribosome. *Isr. J. Chem. USA* 50:99–108

11. Cabrita LD, Hsu STD, Launay H, Dobson CM, Christodoulou J. 2009. Probing ribosome-nascent chain complexes produced *in vivo* by NMR spectroscopy. *Proc. Natl. Acad. Sci. USA* 106:22239–44

12. Cavagnero S, Dyson HJ, Wright PE. 1999. Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. *J. Mol. Biol.* 285:269–82

13. Chow CC, Chow C, Raghunathan V, Huppert TJ, Kimball EB, Cavagnero S. 2003. Chain length dependence of apomyoglobin folding: structural evolution from misfolded sheets to native helices. *Biochemistry* 42:7090–99

14. Clarke TF 4th, Clark PL. 2008. Rare codons cluster. *PLOS One* 3:e3412

15. Daggett V, Fersht AR. 2003. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* 28:18–25

16. Daggett V, Fersht A. 2003. The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell Biol.* 4:497–502

17. Deuerling E, Bukau B. 2004. Chaperone-assisted folding of newly synthesized proteins in the cytosol. *Crit. Rev. Biochem. Mol. Biol.* 39:261–77

18. Eaton WA, Munoz V, Thompson PA, Henry ER, Hofrichter J. 1998. Kinetics and dynamics of loops, alpha-helices, beta-hairpins, and fast-folding proteins. *Acc. Chem.*

Res. 31:745–53

19. Eichmann C, Preissler S, Riek R, Deuerling E. 2010. Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proc. Natl. Acad. Sci. USA* 107:9111–16

20. Ellis JP, Bakke CK, Kirchdoerfer RN, Jungbauer LM, Cavagnero S. 2008. Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem. Biol.* 3:555–66

21. Ellis JP, Culviner PH, Cavagnero S. 2009. Confined dynamics of a ribosome-bound nascent globin: Cone angle analysis of fluorescence depolarization decays in the presence of two local motions. *Protein Sci.* 18:2003–15

22. Ellison PA, Cavagnero S. 2006. Role of unfolded state heterogeneity and en-route ruggedness in protein folding kinetics. *Protein Sci.* 15:564–82

23. Epand RM, Scheraga A. 1968. Influence of long-range interactions on structure of myoglobin. *Biochemistry* 7:2864–72

24. Evans MS, Clark TF, Clark PL. 2005. Conformations of co-translational folding intermediates. *Protein Pept. Lett.* 12:189–95

25. Evans MS, Sander IM, Clark PL. 2008. Cotranslational folding promotes beta-helix formation and avoids aggregation *in vivo*. *J. Mol. Biol.* 383:683–92

26. Evans MS, Ugrinov KG, Frese M-A, Clark PL. 2005. Homogeneous stalled ribosome nascent chain complexes produced *in vivo* or *in vitro*. *Nat. Methods* 2:757–62

27. Fedorov AN, Baldwin TO. 1997. Cotranslational protein folding. *J. Biol. Chem.* 272:32715–18

28. Ferbitz L, Maier T, Patzelt H, Bukau B, Deuerling E, Ban N. 2004. Trigger factor in

complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* 431:590–96.

29. Flanagan JM, Kataoka M, Shortle D, Engelman DM. 1992. Truncated staphylococcal nuclease is compact but disordered. *Proc. Natl. Acad. Sci. USA* 89:748–52

30. Frydman J. 2001. Folding of newly translated proteins *in vivo*: the role of molecular chaperones. *Annu. Rev. Biochem.* 70:603–47

31. Fulle S, Gohlke H. 2009. Statics of the ribosomal exit tunnel: implications for cotranslational peptide folding, elongation regulation, and antibiotics binding. *J. Mol. Biol.* 387:502–17

32. Genevaux P, Keppel F, Schwager F, Langendijk-Genevaux PS, Hartl FU, Georgopoulos C. 2004. *In vivo* analysis of the overlapping functions of DnaK and trigger factor. *EMBO Rep.* 5:195–200

33. Ghosh N, Hazra K, Sarkar SN. 1996. Ribosome facilitates refolding of rhodanese and lysozyme by suppressing aggregation. *Prog. Biophys. Mol. Biol.* 65:85

34. Gilbert RJC, Fucini P, Connell S, Fuller SD, Nierhaus KH, Robinson CV, Dobson CM, Stuart DI. 2004. Three-dimensional structures of translating ribosomes by cryo-EM. *Mol. Cell* 14:57–66

35. Haber E, Anfinsen CB. 1962. Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J. Biol. Chem.* 237:1839–44

36. Harms J, Schluenzen F, Zarivach R, Bashan A, Gat S, Agmon I, Bartels H, Franceschi F, Yonath A. 2001. High-resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* 107:679–88

37. Hartl FU, Hayer-Hartl M. 2009. Converging concepts of protein folding *in vitro* and

in vivo. *Nat. Struct. Mol. Biol.* 16:574–81

38. Hoffmann A, Merz F, Rutkowska A, Zachmann-Brand B, Deuerling E, Bukau B. 2006. Trigger factor forms a protective shield for nascent polypeptides at the ribosome. *J. Biol. Chem.* 281:6539–45

39. Hsu ST, Cabrita LD, Fucini P, Christodoulou J, Dobson CM. 2009. Probing side-chain dynamics of a ribosome-bound nascent chain using methyl NMR spectroscopy. *J. Am. Chem. Soc.* 131:8366–67

40. Hsu ST, Fucini P, Cabrita LD, Launay H, Dobson CM, Christodoulou J. 2007. Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proc. Natl. Acad. Sci. USA* 104:16516–21

41. Ito K, Chiba S, Pogliano K. 2010. Divergent stalling sequences sense and control cellular physiology. *Biochem. Biophys. Res. Commun.* 393:1–5

42. Jennings PA, Wright PE. 1993. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* 262:892–96

43. Johnson AE. 2004. Functional ramifications of FRET-detected nascent chain folding far inside the membrane-bound ribosome. *Biochem. Soc. Trans.* 32:668–72

44. Katzen F, Chang G, Kudlicki W. 2005. The past, present and future of cell-free protein synthesis. *Trends Biotechnol.* 23:150–56

45. Kosolapov A, Deutsch C. 2009. Tertiary interactions within the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* 16:405–11

46. Kosolapov A, Tu L, Wang J, Deutsch C. 2004. Structure acquisition of the T1 domain of Kv1.3 during biogenesis. *Neuron* 44:295–307

47. Kramer G, Ramachandiran V, Hardesty B. 2001. Cotranslational folding— ukar

mea mecum porto? *Int. J. Biochem. Cell Biol.* 33:541–53

48. Deleted upon reproduction of the review.

49. Kurt N, Cavagnero S. 2005. The burial of solvent-accessible surface area is a predictor of polypeptide folding and misfolding as a function of chain elongation. *J. Am. Chem. Soc.* 127:15690–91

50. Kurt N, Mounce BC, Ellison PA, Cavagnero S. 2008. Residue-specific contact order and contact breadth in single-domain proteins: implications for folding as a function of chain elongation. *Biotechnol. Progr.* 24:570–75

51. Levinthal C. 1968. Are there pathways for protein folding? *J. Chim. Phys. Phys.-Chim. Biol.* 65:44–45

52. LuJ,DeutschC.2005.Foldingzonesinsidetheribosomalexittunnel.Nat.Struct.Mol.Biol.1 2:1123–29

53. Lu J, Deutsch C. 2008. Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J. Mol. Biol.* 384:73–86

54. Lu J, Kobertz WR, Deutsch C. 2007. Mapping the electrostatic potential within the ribosomal exit tunnel. *J. Mol. Biol.* 371:1378–91

55. Maier R, Eckert B, Scholz C, Lilie H, Schmid FX. 2003. Interaction of trigger factor with the ribosome. *J. Mol. Biol.* 326:585—92

56. Maier R, Scholz C, Schmid FX. 2001. Dynamic association of trigger factor with protein substrates. *J. Mol. Biol.* 314:1181–90

57. Malkin LI, Rich A. 1967. Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *J. Mol. Biol.* 26:329–46

58. Mayer MP, Bukau B. 2005. Hsp70 chaperones: cellular functions and molecular

mechanism. *Cell. Mol. Life Sci.* 62:670–84.

59. Merz F, Boehringer D, Schaffitzel C, Preissler S, Hoffmann A, Maier T, Rutkowska A, Lozza J, Ban N, Bukau B, Deuerling E. 2008. Molecular mechanism and structure of trigger factor bound to the translating ribosome. *EMBO J.* 27:1622–32

60. Millett IS, Doniach S, Plaxco KW. 2002. Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. *Adv. Protein Chem.* 62:241–62

61. Neira JL, Fersht AR. 1999. Exploring the folding funnel of a polypeptide chain by biophysical studies on protein fragments. *J. Mol. Biol.* 285:1309–33

62. Netzer WJ, Hartl FU. 1997. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* 388:343–49

63. Nicola AV, Chen W, Helenius A. 1999. Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat. Cell Biol.* 1:341–45

64. Ohashi H, Kanamori T, Shimizu Y, Ueda T. 2010. A highly controllable reconstituted cell-free system—a breakthrough in protein synthesis research. *Curr. Pharm. Biotechnol.* 11:267–71

65. Phillips DC. 1967. The hen egg-white lysozyme molecule. *Proc. Natl. Acad. Sci. USA* 57:483–95

66. Ramakrishnan V. 2008. What we have learned from ribosome structures. *Biochem. Soc. Trans.* 036:567–74

67. Rutkowska A, Beerbaum M, Rajagopalan N, Fiaux J, Schmieder P, et al. 2009. Large-scale purification of ribosome-nascent chain complexes for biochemical and structural studies. *FEBS Lett.* 583:2407–13

68. Rutkowska A, Mayer MP, Hoffmann A, Merz F, Zachmann-Brand B, et al. 2008.

Dynamics of trigger factor interaction with translating ribosomes. *J. Biol. Chem.* 283:4124–32.

69. Schaffitzel C, Ban N. 2007. Generation of ribosome nascent chain complexes for structural and functional studies. *J. Struct. Biol.* 159:302–10

70. Schmeing TM, Ramakrishnan V. 2009. What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461:1234–42

71. Schuler B, Eaton WA. 2008. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* 18:16–26

72. Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate, JHD. 2005. Structures of the bacterial ribosome at 3.5 A˚ resolution. *Science* 310:827–34

73. Seidelt B, Innis CA, Wilson DN, Gartmann M, Armache J-P, Villa E, Trabuco LG, Becker T, Mielke T, Schulten K, Steitz TA, Beckmann R. 2009. Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* 326:1412–15

74. Selmer M, Dunham CM, Murphy FV 4th, Weixlbaumer A, Petry S, Kelley AC, Weir JR, Ramakrishnan V. 2006. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313:1935–42

75. Shimizu Y, Kanamori T, Ueda T. 2005. Protein synthesis by pure translation systems. *Methods* 36:299– 304

76. Siller E, DeZwaan DC, Anderson JF, Freeman BC, Barral JM. 2010. Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *J. Mol. Biol.* 396:1310–18

77. Sinha KK, Udgaonkar JB. 2009. Early events in protein folding. *Curr. Sci.* 96:1053–

70

78. Spirin AS. 2004. High-throughput cell-free systems for synthesis of functionally active proteins. *Trends Biotechnol.* 22:538–45

79. Steitz TA. 2008. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* 9:242–53

80. Tang YF, Rigotti DJ, Fairman R, Raleigh DP. 2004. Peptide models provide evidence for significant structure in the denatured state of a rapidly folding protein: the villin headpiece subdomain. *Biochemistry* 43:3264–72

81. Taniuchi H, Anfinsen CB. 1969. An experimental approach to the study of the folding of staphylococcal nuclease. *J. Biol. Chem.* 244:3864–75

82. Tomic S, Johnson AE, Hartl FU, Etchells SA. 2006. Exploring the capacity of trigger factor to function as a shield for ribosome bound polypeptide chains. *FEBS Lett.* 580:72-76

83. Tu L, Wang J, Deutsch C. 2007. Biogenesis of the T1S1 linker of voltage-gated K^+ channels. *Biochemistry* 46:8075–84

84. Tu LW, Deutsch C. 2010. A folding zone in the ribosomal exit tunnel for Kv1.3 helix formation. *J. Mol. Biol.* 396:1346–60

85. Voss NR, Gerstein M, Steitz TA, Moore PB. 2006. The geometry of the ribosomal polypeptide exit tunnel. *J. Mol. Biol.* 360:893–906

86. Wegrzyn RD, Deuerling E. 2005. Molecular guardians for newborn proteins: ribosome-associated chaperones and their role in protein folding. *Cell. Mol. Life Sci.* 62:2727–38

87. Weinreis SA, Ellis JP, Cavagnero S. 2010. Dynamic fluorescence depolarization: a

powerful tool to explore protein folding on the ribosome. *Methods* 52:57–73

88. Williamson JR. 2008. Biophysical studies of bacterial ribosome assembly. *Curr. Opin. Struct. Biol.* 18:299– 304

89. Witt SN. 2009. Tethering creates unusual kinetics for ribosome-associated chaperones with nascent chains. *Protein Pept. Lett.* 16:631–34

90. Woolhead CA, Johnson AE, Bernstein HD. 2006. Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol. Cell* 22:587–98

91. Yusupov MM, Yusupova GZ, Albion B, Lieberman K, Earnest TN, Cate JHD, Noller HF. 2001. Crystal structure of the ribosome at 5.5 angstrom resolution. *Science* 292:883–96

92. Zhang G, Hubalewska M, Ignatova Z. 2009. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* 16:274–80

93. Zhang G, Ignatova Z. 2010. Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Curr. Opin. Struct. Biol.* 21:1–7

94. Zimmerman E, Yonath A. 2009. Biological implications of the ribosome's stunning stereochemistry. *ChemBioChem* 10:63–72

95. Ziv G, Haran G, Thirumalai D. 2005. Ribosome exit tunnel can entropically stabilize alpha-helices. Proc. *Natl. Acad. Sci. USA* 102:18956–61

96. Behrmann M, Koch HG, Hengelage T, Wieseler B, Hoffschulte HK, Muller M. 1998. Requirements for the translocation of elongation-arrested, ribosome-associated OmpA across the plasma membrane of *Escherichia coli. J. Biol. Chem.* 273:13898–

904

97. Chattopadhyay S, Das B, Dasgupta C. 1996. Reactivation of denatured proteins by 23S ribosomal RNA: Role of domain V. Proc. Natl. Acad. Sci. USA 93:8284–7

98. Das B, Chattopadhyay S, Das Gupta C. 1992. Reactivation of denatured fungal glucose 6-phosphate dehydrogenase and alkaline phosphatase with ribosome. Biochem. Biophys. Res. Commun. 183:774–80

99. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding: a perspective from simple exact models. *Protein Sci.* 4:561–602

100. O'Brien EP, Christodoulou J, Vendruscolo M, Dobson CM. 2011. New scenarios of protein folding can occur on the ribosome. *J. Am. Chem. Soc.* 133:513–26

101. Schuler B, Lipman EA, Eaton WA. 2002. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743–47

102. Uversky VN, Dunker AK. 2010. Understanding protein non-folding. *Biochim. Biophys. Acta* 1804:1231–64

103. Woolhead CA, McCormick PJ, Johnson AE. 2004. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* 116:725–36

Chapter 2 Contribution of Long-Range Interactions to the Secondary Structure of an Unfolded Globin

This chapter is derived from an article published as a communication by

Daria V. Fedyukina, Senapathy Rajagopalan, Ashok Sekhar, Eric C. Fulmer, Ye-Jin Eun

and

Silvia Cavagnero

in

Biophysical Journal **2010**, 99, L37-L39

Authors' contributions:

Daria V. Fedyukina participated in sample preparation, NMR data collection, analysis,

and interpretation, and in discussions and manuscript writing.

Senapathy Rajagopalan performed preliminary NMR data collection and analysis

Ashok Sekhar collected CD data and participated in NMR data collection.

Eric C. Fulmer participated in preliminary NMR data analysis.

Ye-Jin Eun prepared some samples.

Silvia Cavagnero participated in data analysis, discussions, and manuscript writing.

2.1 Abstract

This work explores the effect of long-range tertiary contacts on the distribution of residual secondary structure in the unfolded state of an α-helical protein. N-terminal fragments of increasing length, in conjunction with multidimensional nuclear magnetic resonance, were employed. A protein representative of the ubiquitous globin fold was chosen as the model system. We found that, while most of the detectable α-helical population in the unfolded ensemble does not depend on the presence of the C-terminal region (corresponding to the native G and H helices), specific N-to-C long-range contacts between the H and A-B-C regions enhance the helical secondary structure content of the N terminus (A-B-C regions). The simple approach introduced here, based on the evaluation of N-terminal polypeptide fragments of increasing length, is of general applicability to identify the influence of long-range interactions in unfolded proteins.

2.2   Introduction

The study of structural and dynamic characteristics of unfolded proteins has gained tremendous momentum in recent years [1,2]. The discovery of intrinsically disordered proteins in living cells has led to a deeper appreciation for the nontrivial biological role of unstructured states in substrate binding specificity, transport, regulation, and disease. Partly or fully disordered proteins under physiologically relevant conditions constitute ~30% of all eukaryotic proteins [3]. The energy-weighted ensemble character of a protein's unfolded state at equilibrium is best described by the term ''statistical coil'' [4], although ''random coil'' has historically been more commonly used. Although the most physiologically relevant unfolded state is the one populated *in vivo* at neutral pH, unfolded states generated *in vitro* under denaturing conditions are a more directly accessible alternative. Furthermore, this state is the starting condition of most *in vitro* refolding studies. Its detailed comprehension is therefore an important gateway toward a better understanding of both the unfolded ensemble and the pathways of *in vitro* protein folding.

The secondary structure distribution of unfolded proteins has been characterized, in some cases, even at atomic resolution. However, very little is known about its origin. Recent experimental studies have detected the presence of both residual secondary structure and specific native and nonnative long-range interactions in the unfolded state [5–7]. There is, however, an urgent need to clarify whether the residual secondary structure typically detected in the unfolded ensemble arises from local contacts or is a consequence of long-range interactions (i.e., contacts among residues far apart in sequence). The two aspects of this issue, which have clear implications for unfolded

chain dynamics and protein folding, are illustrated in Figure 2-1a.
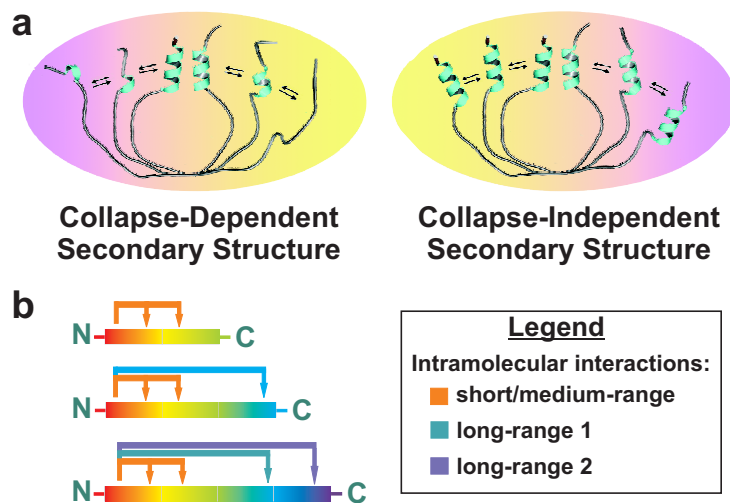
Figure 2-1. Hypotheses of secondary structure formation and chain elongation strategy. (a) General cartoon showing that chain collapse and the resulting long-range contacts may (image, left) or may not (image, right) be responsible for the residual secondary structure often observed in unfolded proteins. (Arrows) Transitions among different conformations. (b) Scheme illustrating the chain elongation strategy adopted here.

This work tackles the above question by what we consider a novel approach. In short, the evolution of backbone secondary structure of the unfolded ensemble is monitored by NMR at atomic resolution upon comparing unfolded N-terminal protein fragments of increasing length (Figure 2-1b). This procedure enables pinpointing the specific contributions of the C-terminal regions to the residual secondary structure in the N-terminal portion of the chain. Sperm whale apomyoglobin (apoMb), a very well-studied member of the ubiquitous globin fold, was selected as the target model chain. We found that, overall, most of the detectable α-helical conformation is populated irrespective of fragment length. However, specific long-range interactions involving the last (~30) C-terminal residues enhance the secondary structure content of the N-terminal region.

2.3   Results and Discussion

The acid unfolded (pH 2.3) state of full-length apoMb, here denoted as (1-153)apoMb, was characterized at atomic resolution [5] and shown to populate partial helical conformation for the amino acids corresponding to the A, D/E, and H helices. In addition, paramagnetic spin labeling led to the identification of long-range interactions involving residues corresponding to the native A and G-H helices and medium-range interactions within the A-B-C and G-H regions [6,7]. Such interactions, due to their transient nature and to extensive conformational averaging in the unfolded state, are undetectable by long-range nuclear Overhauser effects [5]. Here, we focus on unfolded (pH 2.4) (1-77), (1-119), and (1-153)apoMb.

Here, we focus on unfolded (pH 2.4) [13]C-^{15}N-labeled (1-77), (1-119), and (1-153)apoMb. The (1-77)apoMb fragment lacks long-range interactions involving the G

and H regions whereas (1-119)apoMb lacks contacts involving the H region. Previous data [8,9] show that all three unfolded apoMb chains have overall dimensions consistent with expanded polymers in a good solvent according to Flory's scaling law [10] and all three chains fit the definition of random coil [10,11]. Resonance assignments were carried out by triple-resonance methods, and chemical shift analysis was performed on the C^α, C', H^N, and N nuclei. The C^α NMR chemical shift deviations from random coil reference values, denoted as secondary chemical shifts (SCS), are employed as selective reporters of backbone secondary structure [12].
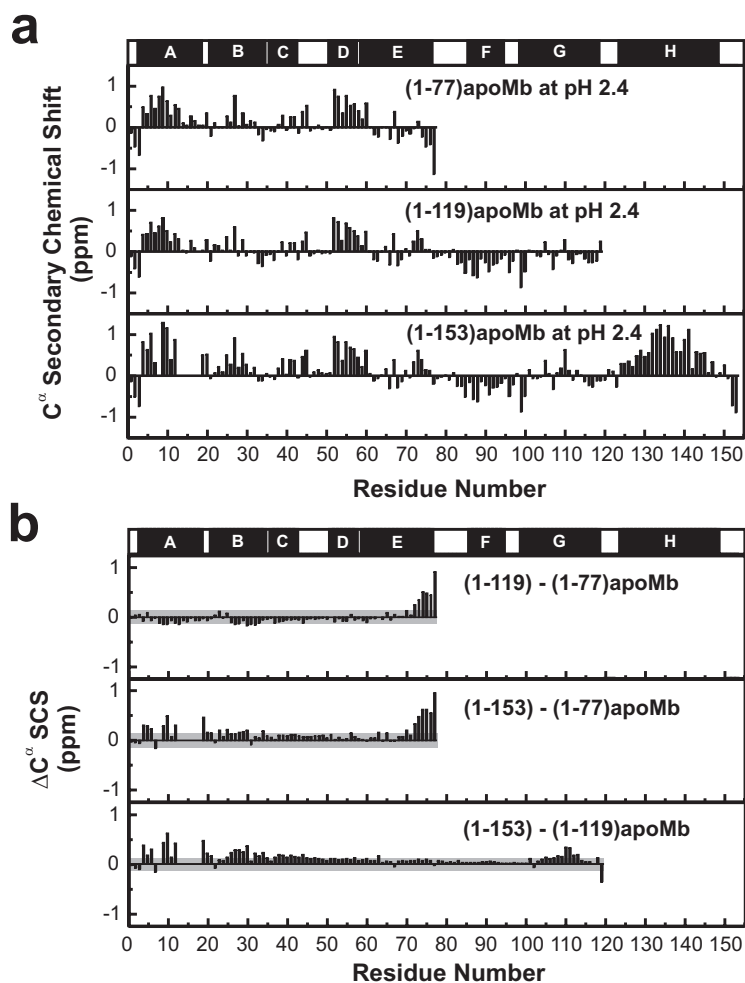
Figure 2-2. C$^{\alpha}$ secondary chemical shift analysis. (a) C$^{\alpha}$ secondary chemical shifts (SCS) and (b) differences among the C$^{\alpha}$ secondary chemical shifts (ΔC$^{\alpha}$ SCS) of (1-77), (1-119), and (1-153)apoMb at pH 2.4 and 25°C in 5 mM CD$_3$COOH, 5% D$_2$O. Native apoMb helices (solid bars) are mapped above the graph. (Gray-shaded regions) ΔC$^{\alpha}$ SCSs falling within experimental error, assessed from backbone assignments on two independent samples (see the Supporting Material).

Figure 2-2a shows that the C^α SCS values for the same residues in each fragment are mostly independent of chain length. Therefore, overall, the distribution of residual secondary structure in the unfolded ensemble does not vary significantly upon progressive extension of the polypeptide chain from 77 amino acids to full-length protein. However, a more detailed comparison provided by the difference between C_α secondary chemical shifts (ΔC^α SCS) among individual chain lengths (Figure 2-2b) indicates that specific long-range interactions modify the secondary structure of the N-terminal portion of the chain.

The C^α SCSs for the residues corresponding to the native A, B, and C helices (i.e., amino acids 3–42) are nearly identical in (1-77) and (1-119)apoMb (Figure 2-2b). Thus, the helicity of this portion of the polypeptide does not vary in the presence and absence of long-range interactions involving the G region. Such contacts are known to exist in full-length unfolded apoMb [6,7]. However, they cannot be present in (1-77)apoMb. In essence, either long-range contacts involving the G residues do not affect secondary structure (in the A-B-C region), or the population experiencing long-range contacts with G is small and undetectable. The former scenario is more likely, given that 1), the overall populations interacting with either G or H regions are similar (3.7 and 3.6%, respectively) [7]; and 2), the effect of interactions with H region is explicitly detectable (see below). However, caution should be exercised, considering that the estimated populations [7] do not explicitly take into account the potential distance-dependence of the long-range contacts.

The slightly negative ΔC^α SCS in the A-B-C region may reflect some small (i.e., within experimental error) second order effects (see the Supporting Material). The

carbonyl carbon SCSs (C' SCS) are also effective reporters of secondary structure, particularly in the unfolded state [5]. The trends followed by C' SCS and ΔC' SCS (Figure 2-3) parallel those of C$^\alpha$ SCS (Figure 2-2). H^N and N SCSs, known to be less diagnostic of secondary structure, are reported in the Supporting Material.

The ΔC$^\alpha$ SCS values of residues 74–77 fall far outside the estimated propagated error (~50.1 ppm) and are consistently positive (Figure 2-2b). We attribute them to end-effects due to the proximity of (1-77)apoMb's residues 74–77 to the C-terminus.
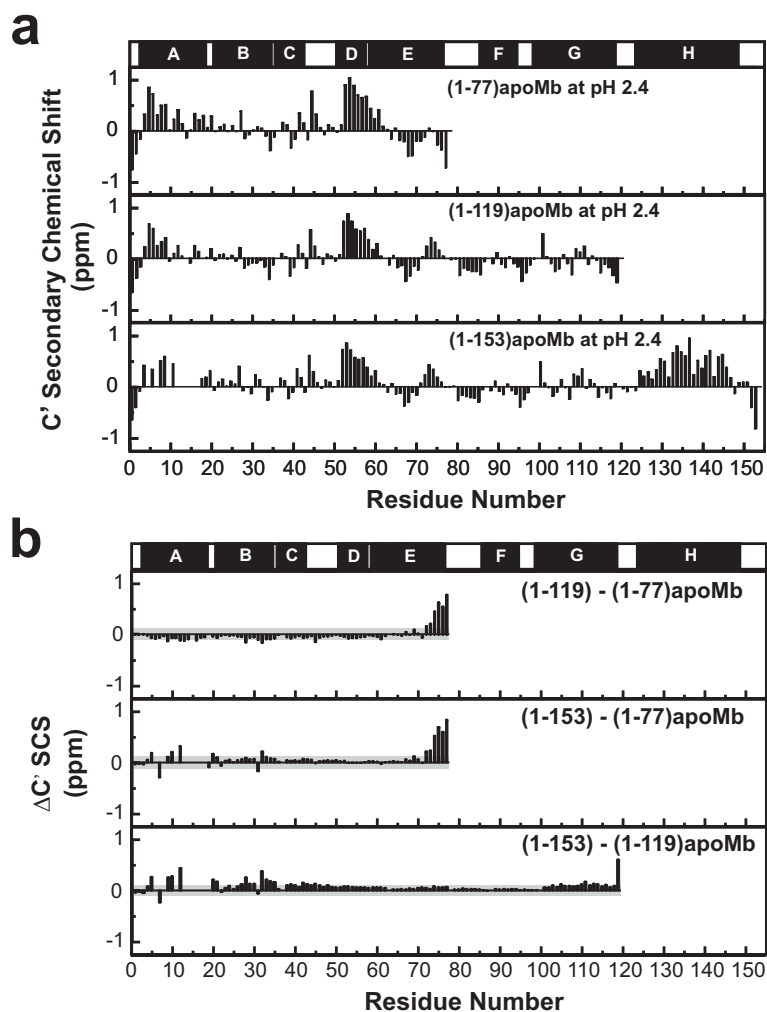
Figure 2-3. C' secondary chemical shift analysis. (a) C' SCS and (b) ΔC' SCS of (1-77), (1-119), and (1-153)apoMb at pH 2.4 and 25°C in 5 mM CD$_3$COOH, 5% D$_2$O. The native apoMb helices are mapped above the graph. (Gray-shaded regions) ΔC' SCS values' experimental errors, determined as described in the legend for Figure 2-2.

The two lower panels of Figure 2-2b highlight the effect of the H region, which is the major finding of this work. Clearly, the C-terminal residues corresponding to the native H helix enhance the residual helicity of the amino acids in the A-B-C cluster, from 5 to 12% (see Figure 2-10 in the Supporting Material). This result demonstrates that specific long-range contacts are capable of increasing local chain helicity in the unfolded state.

The H-region-dependent enhancement in secondary structure in the A-B-C region goes hand-in-hand with selective line-broadening beyond detection in the A region (Figure 2-11). Hence, the observed tertiary contact-driven increase of secondary structure in the A region appears linked to the establishment of slow dynamic processes (on the micro-second-to-millisecond timescale; see also Yao et al. [5]).

Residues in the D, E, and F regions are known not to experience medium- and long-range interactions with other portions of the chain [6,7]. Despite the absence of such contacts, C^{α} SCS reveal the presence of local clusters spanning 6–12 residues. Therefore, Flory's isolated pair hypothesis [10], stating that backbone dihedral angles are independent on those of the neighboring residues, does not apply here. Our results agree with the computational prediction of more-extended local backbone biases [13,14] in the unfolded ensemble, and with the fact that apoMb C^{α} SCSs depend on the position of each residue type (see Figure 2-12).

The long-range ΔC^{α} SCS effects observed here may either be due to small populations of compact species experiencing a large increase in helicity (due to A-B-C to H contacts) or to a small increase in helical content by relatively larger populations of semi-compact species. A combination of the above effects is also possible.

In summary, this work shows that only specific long-range interactions between the N-terminal A-B-C cluster and the C-terminal H region lead to detectable increased helicity in the N-terminal region. Thus, some long-range interactions lead to enhanced secondary structure. The resulting collapsed species are likely important, as they may provide preferential kinetic escape routes to the native state, upon switching to folding conditions.

2.4   References

1. Dyson, H. J., and P. E. Wright. 2004. Unfolded proteins and protein folding studied by NMR. Chem. Rev. 104:3607–3622.

2. McCarney, E. R., J. E. Kohn, and K. W. Plaxco. 2005. Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. Crit. Rev. Biochem. Mol. Biol. 40:181–189.

3. Fink, A. L. 2005. Natively unfolded proteins. Curr. Opin. Struct. Biol. 15:35–41.

4. Vila, J. A., D. R. Ripoll, H. A. Baldoni, H. A. Scheraga. 2002. Unblocked statistical coil tetrapeptides and pentapeptides in aqueous solution: a theoretical study. J. Biomol. NMR. 24:245–262.

5. Yao, J., J. Chung, D. Eliezer, P. E. Wright, H. J. Dyson. 2001. NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. Biochemistry. 40:3561–3571.

6. Lietzow, M. A., M. Jamin, H. J. Dyson, P. E. Wright. 2002. Mapping long-range contacts in a highly unfolded protein. J. Mol. Biol. 322:655–662.

7. Felitsky, D. J., M. A. Lietzow, H. J. Dyson, P. E. Wright. 2008. Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. Proc. Natl. Acad. Sci. USA. 105:6278–6283.

8. Kurt, N., S. Rajagopalan, and S. Cavagnero. 2006. Effect of hsp70 chaperone on the folding and misfolding of polypeptides modeling an elongating protein chain. J. Mol. Biol. 355:809–820.

9. Gast, K., H. Damaschun, R. Misselwitz, M. Mueller-Frohne, D. Zirwer, G. Damaschun. 1994. Compactness of protein molten globules: temperature-induced

structural changes of the apomyoglobin folding intermediate. Eur. Biophys. J. 23:297–305.

10. Flory, P.J. 1969. Statistical Mechanics of Chain Molecules. Wiley, New York.

11. Kohn, J. E., I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach, K. W. Plaxco. 2004. Random-coil behavior and the dimensions of chemically unfolded proteins. Proc. Natl. Acad. Sci. USA. 101:12491–12496.

12. Wishart, D. S., B. D. Sykes, and F. M. Richards. 1991. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J. Mol. Biol. 222:311–333.

13. Fitzkee, N. C., and G. D. Rose. 2004. Reassessing random-coil statistics in unfolded proteins. Proc. Natl. Acad. Sci. USA. 101:12497–12502.

14. Tran, H. T., X. Wang, and R. V. Pappu. 2005. Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins. Biochemistry. 44:11369– 11380.

2.5 Appendix: Supporting Information for Chapter 2

Contribution of Long-Range Interactions to the Secondary Structure of an Unfolded Globin

2.5.1   Supplementary Discussion

2.5.1.1   Far-UV circular dichroism

The far-UV circular dichroism (CD) of apoMb, fragments and circular permutants [1, 2] has been studied before at pH ca. 6.0. On the other hand, no prior studies on acid-unfolded apoMb N-terminal fragments were performed before.

Our far-UV CD data for pH unfolded (1-77), (1-119) and (1-153)apoMb are consistent with the NMR results. The CD spectra (Figure 2-4) show that the helical content of (1-77)apoMb is nearly identical to that of (1-119)apoMb, on a per-residue basis. The random coil content, however, is larger for (1-119)apoMb due to the introduction of the highly unstructured FG region. Overall helicity increases and random coil content decreases, as chain length progresses from 1-119 to 1-153, upon addition of the residues encompassing the native H helix. Based on the NMR results, the increase in helicity upon chain elongation from 1-119 to 1-153 is likely attributable to the additional induced helicity in other regions of the protein (A, B, and C regions) due to long-range interactions.

2.5.1.2   NMR data analysis

Figures 2-5, 2-6, and 2-7 show the ^{1}H-^{15}N-HSQC spectra of (1-77), (1-119), and (1-153)apoMb with annotated N and H^N resonance assignments. Chemical shift values for the C$^\alpha$, H^N, N, C' and selected ΔC$^\beta$ assigned resonances were deposited in BioMagResBank (see Supplementary Methods).

Secondary chemical shifts for the H^N and N backbone resonances of (1-77), (1-119) and (1-153)apoMb are provided in Figure 2-8. While C$^\alpha$ and C' secondary chemical shifts are significantly diagnostic for the presence of secondary structure [1-4], H^N and N

shifts are known to be influenced by other structural and environmental factors [3, 4] and are therefore not good reporters for the presence of residual secondary structure in (1-77), (1-119), and (1-153)apoMb. However, H^N and N secondary chemical shifts can still be regarded as qualitative reporters of variations in the surrounding electronic environment.

The secondary chemical shifts for the H^N and N backbone resonances (Figure 2-8) for (1-77) and (1-119)apoMb are remarkably similar. While this result does not shed specific light on the chain length dependence of secondary structure, it supports the fact that the ensemble is populated with similar three dimensional conformations in both (1-77) and (1-119)apoMb, given that the electronic environment is similar. This suggests that the previously reported tertiary contacts involving the A and the G helix [5, 6] may not be as extensive.

The equilibrium unfolded ensemble ΔH^N SCS and ΔN SCS values (Figure 2-8 and Figure 2-9) confirm this result. Comparisons between each of the N-terminal fragments and the full-length protein show that introduction of the amino acids corresponding to the H helix into the polypeptide chain causes small perturbation in the chemical shift of the second portion of the B region. While it is hard to interpret the nature of the structural variations resulting from this effect, due to the heterogeneity of the unfolded ensemble, it is tempting to speculate that there may be significant B-H contacts underlying the increased helicity found in the A-B-C region as a result of introducing the H helix into the polypeptide chain.

2.5.1.3   Determination of percent helicity enhancements in the N-terminal region

The estimated helical populations for the A, B, and C regions of the 1-119 fragment of apoMb and the full-length protein are shown in Figure 2-2c. Percent

helicities were evaluated from the assigned C^α and C' secondary chemical shifts and the expected shifts for a 100% helix, as described in the figure legend. The data of Figure 2-2c explicitly illustrate the fact that the presence of the C-terminal residues corresponding to the native H helix enhance the apparent helicity of the ABC cluster by ca. 7%, i.e., from 5 to 12% in the absence and presence of the H region, respectively.

2.5.1.4 Additional comments on ΔC^α SCS values

It is interesting to notice that the ΔC^α and ΔC' SCS values smaller than the experimental error (see Figure 2-2b) do not randomly fluctuate around zero but lay mostly on one side. We want to emphasize that this trend is clearly hard to interpret on solid ground, given that these ΔC^α and ΔC' SCS values are smaller than the reported uncertainties. On the other hand, we speculate that such monotonic trends may still be of some experimental significance, due to their nonrandom nature, and may possibly be due to extremely small second order effects, leading to tiny variations in secondary structure induced by long-range interactions. For instance, based on the data in Figure 2-2b, addition of the residues belonging to the G region may lead to extremely small decreases in helicity throughout the rest of the sequence (residues 1 to 77). Conversely, addition of the residues belonging to the native H helix may lead to extremely small increases in helicity throughout the rest of the sequence (residues 1 to 119).

2.5.1.5 Percent helicity enhancements in the N-terminal region

Estimated helical populations for the A, B, and C regions of the 1-119 fragment of apoMb and the full-length protein are shown in Figure 2-10. Percent helicities were evaluated from the assigned C^α and C' secondary chemical shifts and the expected shifts for a 100% helix, as described in the figure legend. The data of Figure 2-10 explicitly

illustrate the fact that the presence of the C-terminal residues corresponding to the native H helix enhance the apparent helicity of the ABC cluster by ca. 7%, i.e., from 5 to 12% in the absence and presence of the H region, respectively.

2.5.1.6   Linewidth analysis

We performed full width at half height (FWHH) analysis of the backbone ^{15}N nuclei of all three unfolded apoMb constructs analyzed in this work. The results are shown in Figure 2-11.

The sequence-specific FWHH trends observed for the full-length protein, (1-153)apoMb, are qualitatively consistent with the R2 profile reported by Yao *et al.* in reference 6 of the main text.

Interestingly, Figure 2-11 shows that the H-region-dependent enhancement in the secondary structure of the A-B-C region correlates with a selective line-broadening beyond detection in the A portion of the sequence. This broadening is primarily observed for the full-length (1-153)apoMb, bearing the residues corresponding to the native H helix and, most importantly, it is absent in (1-77) and (1-119)apoMb. Hence, the above tertiary structure-driven increase in secondary structure appears to be linked with the establishment of specific slow dynamic processes on the µs-ms timescale.

No other major line-broadening effects are observed in other regions, as chain elongates. As shown in Figure 2-11, the full length construct does not display any line-broadening beyond detection in the H region, despite the presence of the A-to-H contacts in some of the unfolded population and despite the observed extensive line-broadening in the A region.

The H region resonances 143 and 144, on the other hand, are fairly broad. The above observations suggest that line-broadening effects resulting from long-range contacts may, in general, not necessarily be symmetrical, i.e., they may not need to be present to the same degree for all the involved interaction counterparts.

While a more complete relaxation analysis is desirable, to more specifically identify slow-timescale R_{ex} contributions reporting on slow time-scale motions (μs-ms) and reduced spectral densities, the FWHH data presented here are sufficient to highlight a correlation between slow motions and long-range-contact-driven secondary structure formation in the unfolded state.

2.5.1.7   Secondary structure analysis of acid-unfolded apoMb by amino acid type

The C^{α} secondary chemical shifts of pH-unfolded full-length apoMb were collected by amino acid type, as shown in Figure 2-12. The plot shows that, for each occurrences of the same amino acid in the sequence, the secondary chemical shift, and thus also the secondary structure, assume very different values. This result clearly indicates that the backbone secondary structure of each unfolded residue is not determined by amino acid type, and it is influenced by the presence of other residues in the sequence.

2.5.2   Supplementary Methods

2.5.2.1   Expression and purification of N-terminal apoMb fragments

Uniformly $^{13}C,^{15}N$-labeled N-terminal apoMb fragments were overexpressed in *E.coli* using Tuner DE3 pLacI (in the case of (1-77)apoMb) and BL21 DE3 cells (in the case of (1-119)apoMb and (1-153)apoMb) (Novagen, San Diego, CA) in M9 minimal medium containing ^{15}N-NH$_4$Cl (1.5g/L) and ^{13}C-glucose (2g/L). Cell growth and purification were carried out as described [7].

2.5.2.2   Circular dichroism

Far-UV CD spectra were recorded at room temperature on an MOS-450 spectropolarimeter (Bio-Logic Science Instruments, Claix, France) using a 1 mm path length cuvette. All samples were in water adjusted to pH 2.4 with HCl. Sample concentrations ranged from 6 to 18 µM. The spectra displayed in Figure 2-4 are the average of two independent experiments. Two scans were averaged for each experiment. Each scan was performed with a 2 nm band-width, 1 nm step-size and 10 s averaging time per point.

2.5.2.3   NMR sample preparation

$^{13}C,^{15}N$-labeled lyophilized polypeptides (i.e., (1-77), (1-119), or (1-153)apoMb) were dissolved in a solution containing 5 mM CD$_3$COOH at pH 2.40. The pH was adjusted to 2.40 with HCl. The samples were then eluted through a 3-mL Sephadex G-25 (fine grade, Amersham Biosciences) spin column pre-equilibrated with 5 mM CD$_3$COOH at pH 2.40. After the spin column treatment, D$_2$O was added to each sample (final concentration: 5% v/v). The apparent pH was then readjusted to 2.40 with HCl. The final

concentrations of (1-77), (1-119), and (1-153)apoMb were 313, 280, and 304 μM, respectively. Solutions containing 2,2-dimethyl-2-silapentane-5-sufonic acid (DSS) for external chemical shift referencing were prepared under identical solution conditions.

2.5.2.4 NMR data collection and processing

General considerations and NMR data collection. All the NMR experiments were performed on a Varian INOVA 600 MHz NMR spectrometer equipped with a Varian ^{1}H{^{13}C, ^{15}N} triple resonance probe with triple axis gradients. Acquisition parameters for the 2D high resolution ^{1}H-^{15}N- HSQC [8] experiments and the triple-resonance experiments used for resonance assignments (HNCA [9-12], HN(CO)CA [13], HNCACB [11, 12, 14], and HNCO [9-12]) are provided in Table 1. Relaxation delays were set to 1 s for all experiments. Chemical shift referencing in the ^{1}H dimension was done with by a 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) external standard. The ^{15}N and ^{13}C dimensions were referenced indirectly as described [15] and gyromagnetic ratios of ^{13}C/^{1}H = 0.251449530; ^{15}N/^{1}H = 0.101329118 were used [16]. In all experiments, the ^{1}H carrier frequency was placed in correspondence of the HDO frequency. The spectrometer temperature was calibrated with neat methanol [17].

Resonance assignments were carried out in close succession within 1.5 months for all three species in a single NMR session, at 25.0°C.

NMR data processing. The NMRPipe [18] and NMRDraw [18] software packages were used for NMR data processing. Linear prediction was performed once (36 points) on the ^{15}N dimension, in all the triple-resonance experiments. Time domain data were apodized with a 90°-shifted sine-bell squared window function in both the ^{13}C and ^{15}N dimensions, in all 3D experiments. Free induction decays were apodized with an

unshifted Gaussian window function in the ^{1}H dimension of all 3D experiments and in both dimensions of the 2D experiments. A 90°-shifted sine-bell square window function was applied to ^{13}C and ^{15}N dimensions in all 3D experiments. Triple- and double-resonance time-domain data were zero-filled once and twice in all dimensions, respectively.

Backbone assignments were carried out by the Sparky-3 software package [19].

The H^N, N, C′ and C$^\alpha$ chemical shift assignments for (1-77), (1-119), and (1-153)apoMb, and assignments of selected C$^\beta$ resonances, were deposited in the BioMagResBank (http://www.bmrb.wisc.edu). The BMRB accession numbers are 16499, 16500, and 16501 for (1-77), (1-119), and (1-153)apoMb, respectively. Previous assignments for (1-77) and (1-119)apoMb, used to calculate standard deviation for chemical shifts values (Figures 2-4 and 2-9), have accession numbers 7076 and 7077, respectively.

2.5.2.5 NMR secondary chemical shift analysis of acid-unfolded apoMb fragments

Reference random coil chemical shift (RCCS) values, including corrections for neighboring residues, were calculated according to Schwarzinger *et al.* [20, 21], following the relations

$$\delta_{\text{random coil-corrected}} = \delta_{\text{random coil}} + A + B + C + D \; ;$$

where A, B, C and D are corrections for amino acids *i-2, i-1, i+1* and *i+2*, respectively.

Secondary Chemical Shifts (SCS) were calculated from experimental chemical shifts (CS) according to the simple relation *SCS* (ppm) = *experimental CS* (ppm) − *sequence-corrected RCCS* (ppm).

2.5.2.6   Error analysis of NMR chemical shifts

For (1-77)apoMb and (1-119)apoMb, the propagated uncertainties for NMR chemical shift differences (see horizontal gray bars in Figure 2-4 and Supplementary Figure 2-9) were calculated from two independent resonance assignments, carried out on two separate samples of each species. In each case, NMR data were collected with the same acquisition parameters at the same temperature, at comparable pH and protein concentrations. For each assigned chemical shift of (1-77) and (1-119)apoMb, the standard deviation of the mean was calculated from the above repeats. The standard deviation for (1-153)apoMb was estimated as the average of the (1-77) and (1-119)apoMb standard deviations. The uncertainty in NMR chemical shift differences was then determined according to standard error propagation procedures [22] as ± the square root of the sum of the squares of the standard deviation for each relevant species. No covariance was applied, as judged to be irrelevant in this case.

Table 2-1. Experimental acquisition parameters for the 2D HSQC and triple resonance experiments used for the resonance assignments of pH-unfolded (1-77) and (1-119) N-terminal apoMb fragments and pH-unfolded full-length apoMb (i.e., (1-53)apoMb). Data were collected on a 600 MHz NMR spectrometer as described in the Supplementary Materials and Methods.

Experiment	Number of complex points (t1 x t2 x t3), t1 : ^{13}C ; t2 : ^{15}N ; t3: ^{1}H.	Sweep Widths (Hz)		
		^{1}H	^{13}C	^{15}N
(1-77)apoMb				
2D HSQC	512 x 2048	6600	-	1500
3D HNCA	80 x 36 x 2048	6600	3620	1500
3D HN(CO)CA	64 x 36 x 2048	6600	3620	1500
3D HNCACB	64 x 36 x 2048	6600	8450	1500
3D HNCO	80 x 36 x 2048	6600	1000	1500
(1-119)apoMb				
2D HSQC	512 x 2048	6600	-	1500
3D HNCA	80 x 37 x 2048	6600	3620	1500
3D HN(CO)CA	64 x 36 x 2048	6600	3620	1500
3D HNCACB	68 x 36 x 2048	6600	8450	1500
3D HNCO	64 x 36 x 2048	6600	1000	1500
(1-153)apoMb				
2D HSQC	512 x 2048	6600	-	1500
3D HNCA	80 x 38 x 2048	6600	3620	1500
3D HN(CO)CA	64 x 36 x 2048	6600	3620	1500
3D HNCACB	64 x 36 x 2048	6600	8450	1500
3D HNCO	64 x 36 x 2048	6600	1000	1500

Figure 2-4. Far-UV CD spectra of (1-77), (1-119), and (1-153)apoMb at room temperature and pH 2.5. Data are shown as (a) ellipticity per unit concentration ($[\theta]$/Conc.) and (b) mean residue ellipticity ($[\theta]_{MRE}$).
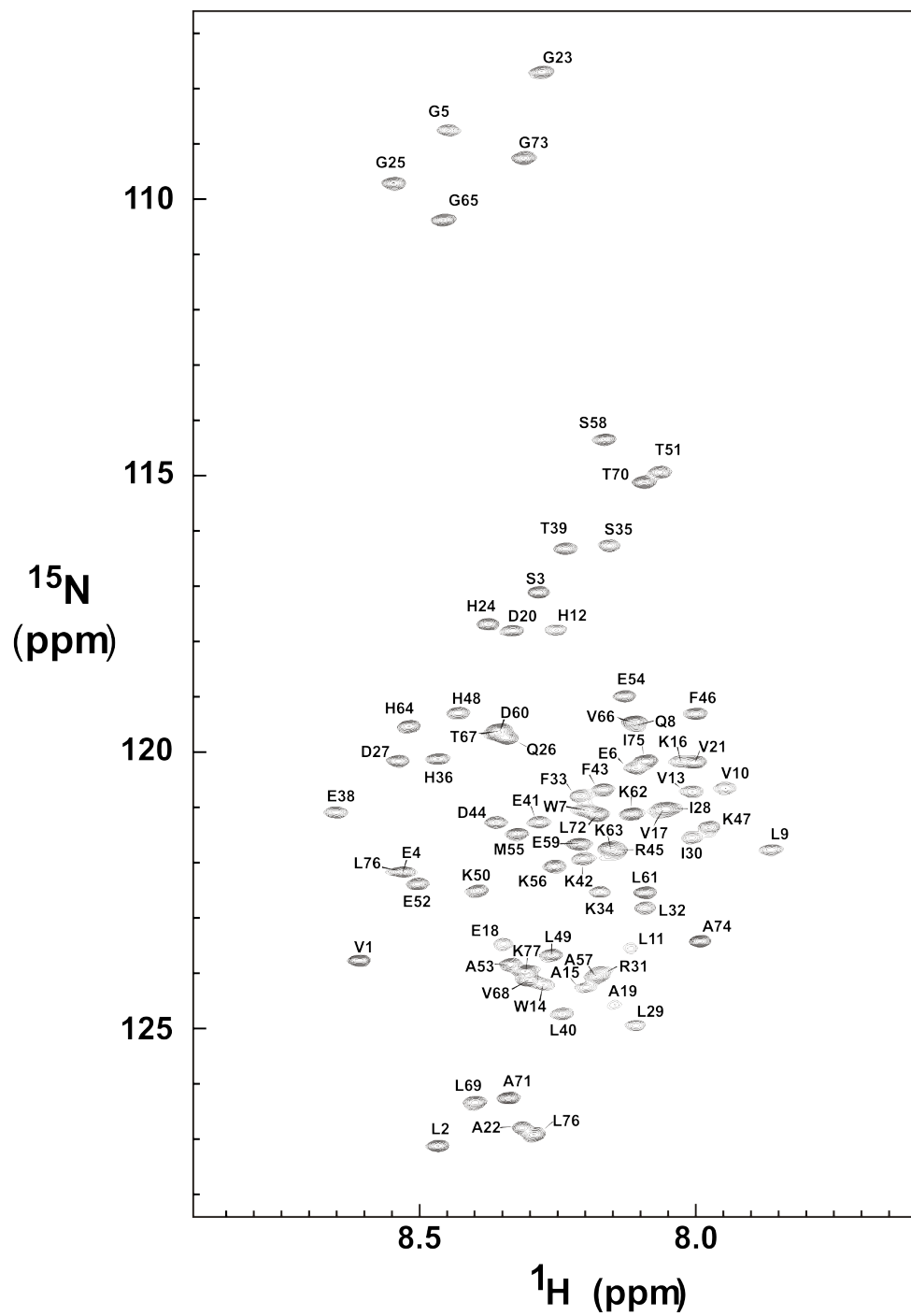
Figure 2-5. ^{1}H,^{15}N-HSQC spectrum of (1-77)apoMb at pH 2.4 and 25°C with annotated resonance assignments.
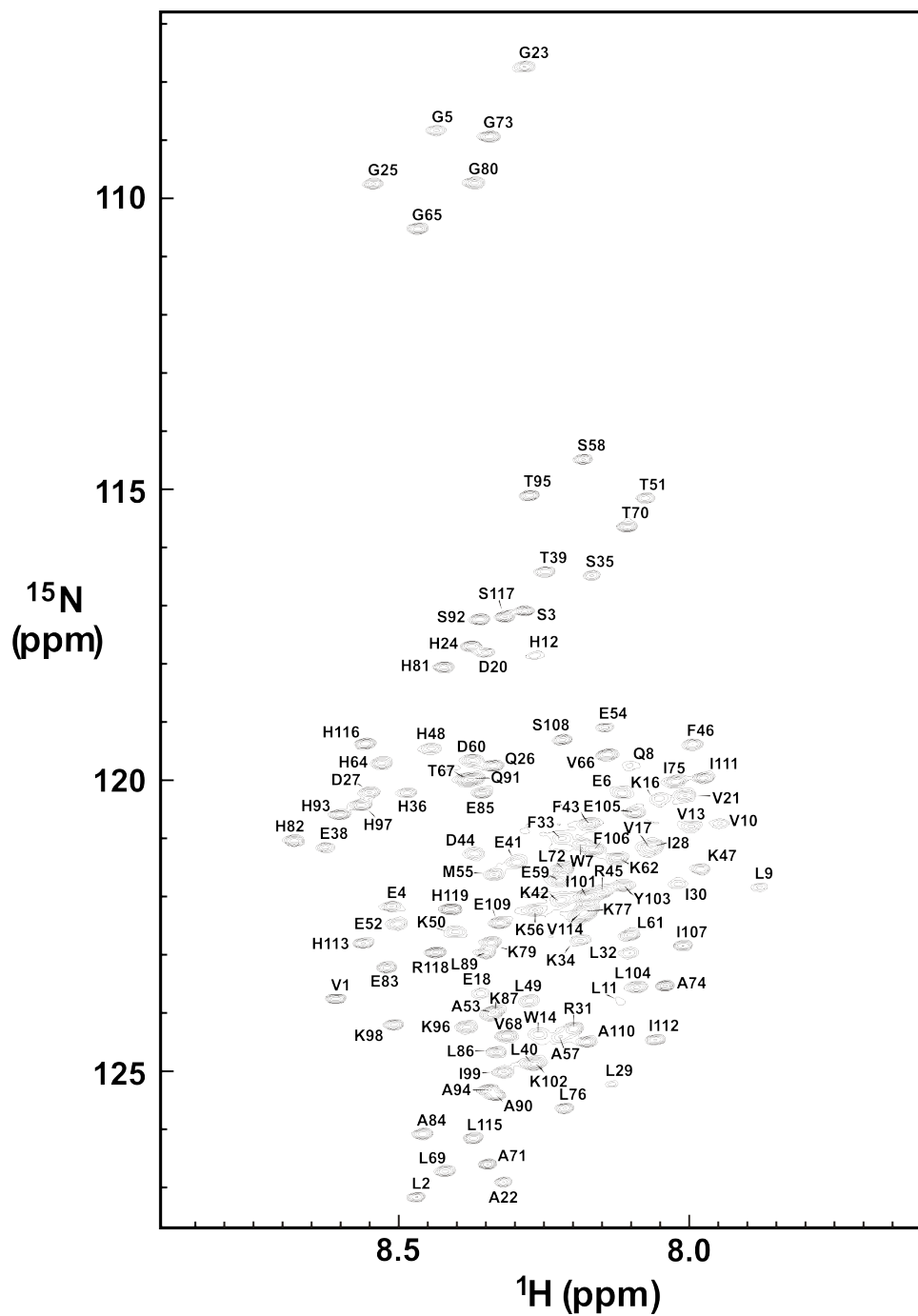
Figure 2-6. ^{1}H,^{15}N-HSQC spectrum of (1-119)apoMb at pH 2.4 and 25°C with annotated resonance assignments.
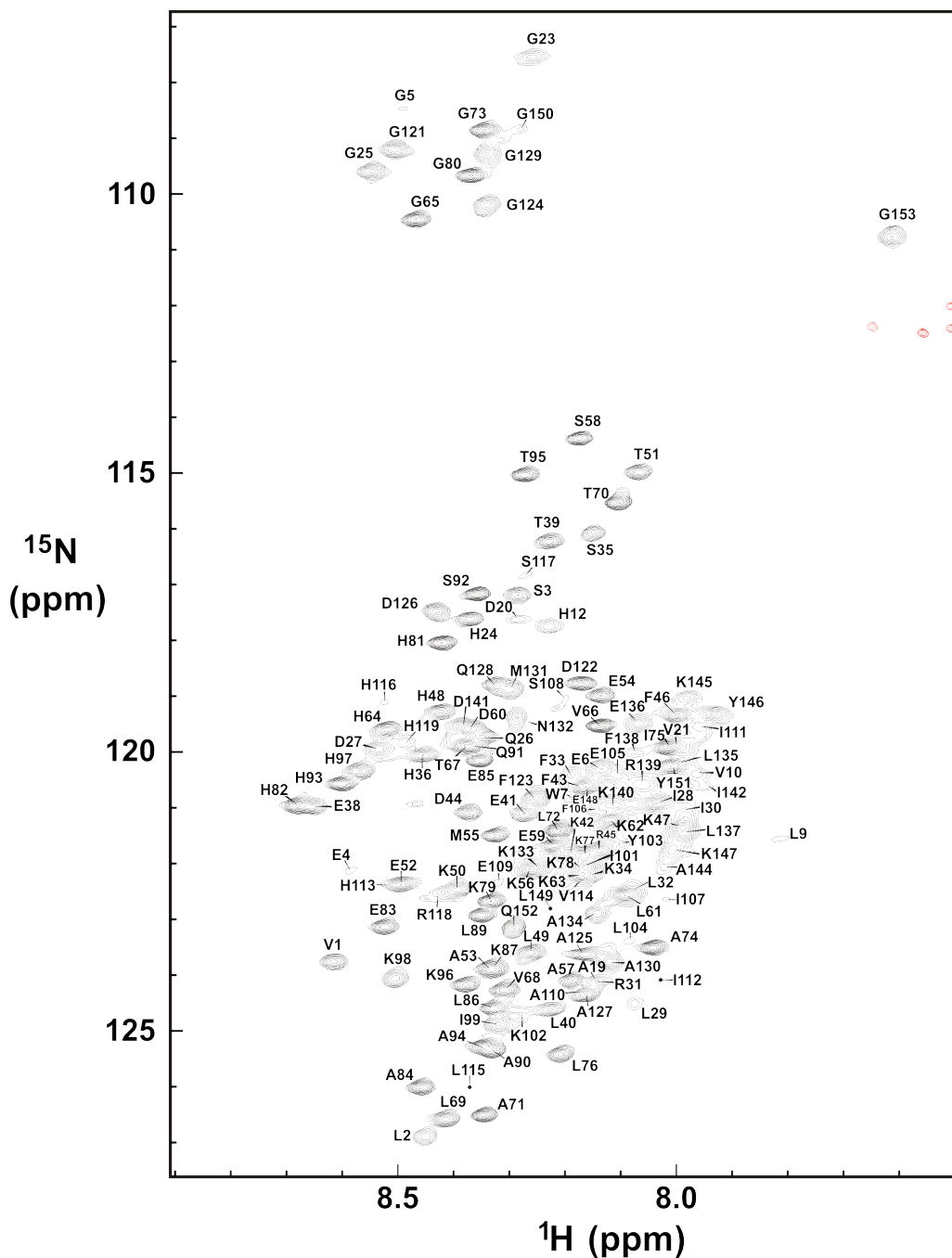
Figure 2-7. ^{1}H,^{15}N-HSQC **spectrum of (1-153)apoMb at pH 2.4 and 25°C with annotated resonance assignments.** Red peaks indicate folded peaks originating from the side chain amides.
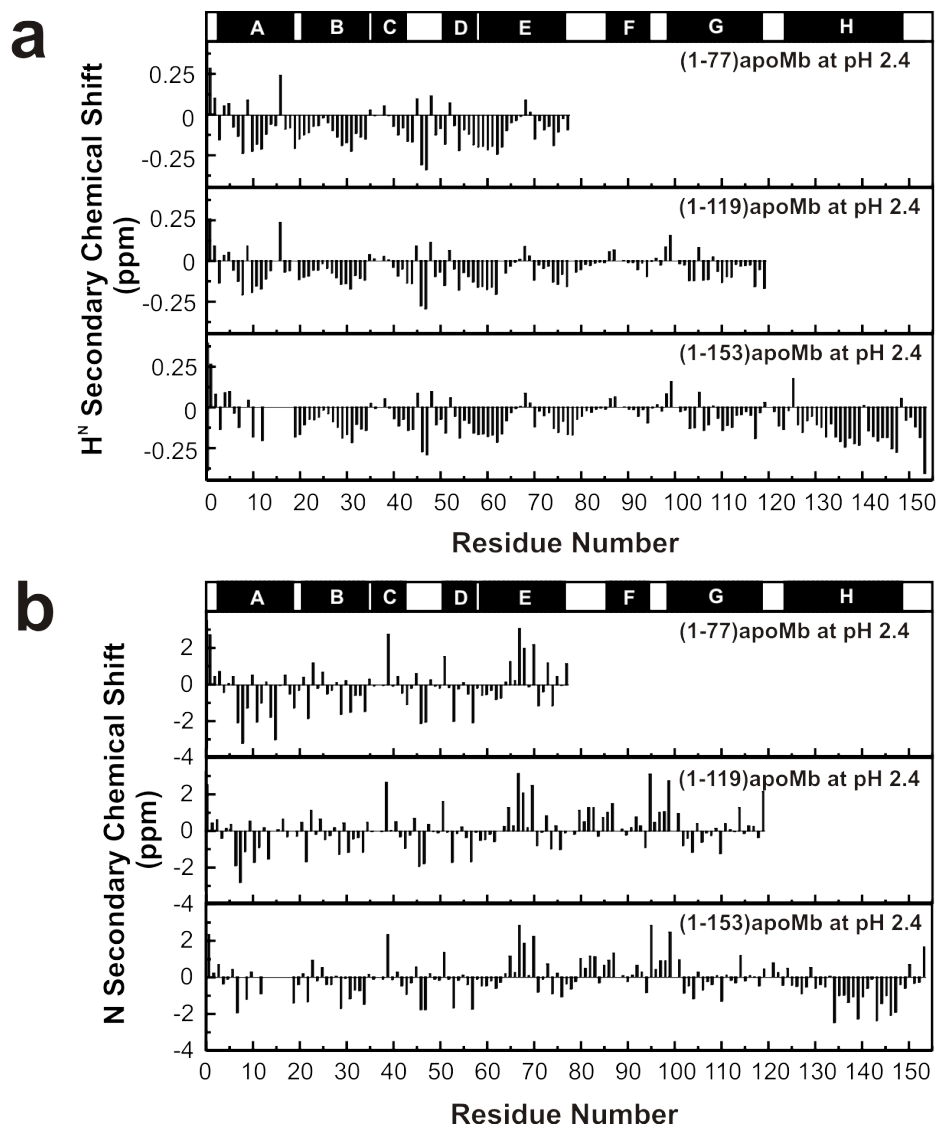
Figure 2-8. (a) H^{N} and (b) N secondary chemical shifts of (1-77), (1-119) and (1-153)apoMb at pH 2.4 and 25°C. The helices of native full-length apoMb are mapped above the graph as black bars.
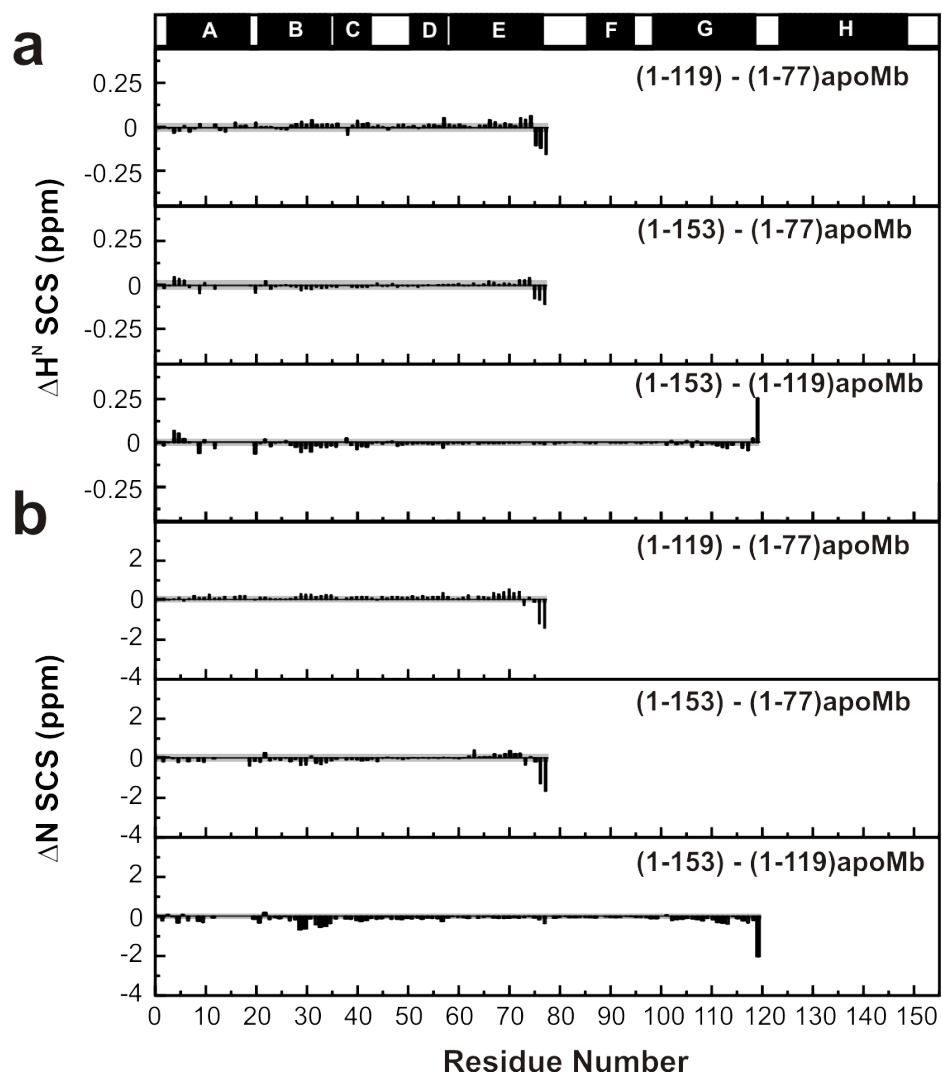
Figure 2-9. Differences between the (a) H^N and (b) N secondary chemical shifts of the species analyzed in this work, i.e., (1-119)apoMb – (1-77)apoMb, (1-153)apoMb – (1-77)apoMb, and (1-153)apoMb – (1-119)apoMb at pH 2.4 and 25°C. The native apoMb helices are mapped above the graph as black bars. The horizontal gray bars denote experimental uncertainties (see Supplemental Materials and Methods).

Figure 2-10. Estimated helical populations in the N-terminal region of acid-unfolded (a) (1-77)apoMb, (b) (1-119)apoMb and (c) (1-153)apoMb expressed as percent helicities for the A, B and C regions of the sequence. Percent helicities were estimated from the assigned secondary chemical shifts, averaged over the pertinent portion of the sequence, divided by the experimental chemical shift expected for a 100% helix. The secondary chemical shifts corresponding to a 100% helix were taken to be 2.8 and 2.1 ppm for C^{α} and C', respectively [23]. Estimated percent helicities are shown for data derived from C^{α} (light blue) and C' (blue) nuclei. Averages over the values derived from C^{α} and C' secondary chemical shifts are shown in black. The percent helicity of the ABC cluster was calculated by averaging over the secondary chemical shifts of all the cluster's residues, defined as amino acids 4-12 (A region), 20-35 (B region) and 46-52 (C region).
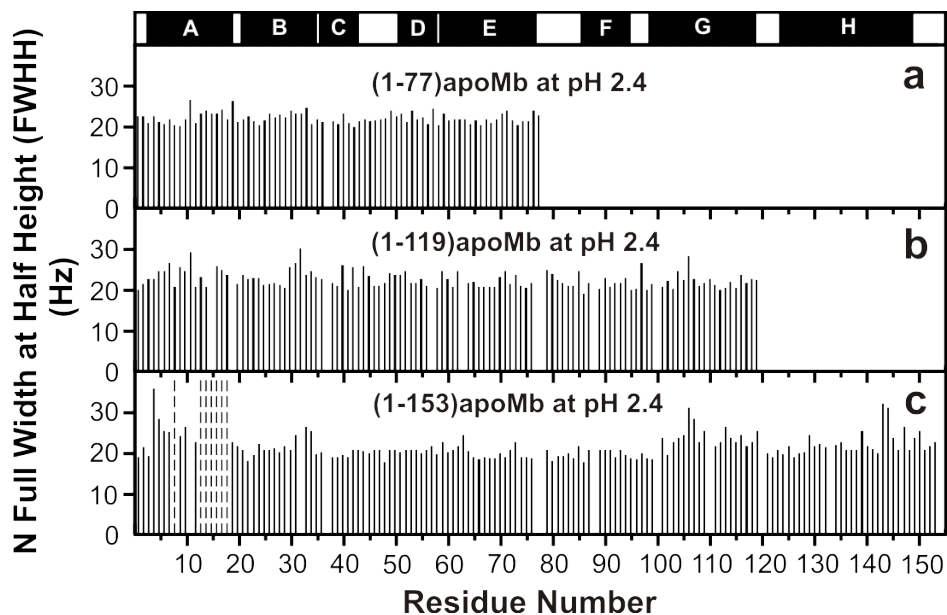
Figure 2-11. Linewidth analysis, reported as full width at half height (FWHH), of the ^{15}N resonances of (a) (1-77)apoMb, (b) (1-119)apoMb and (c) (1-153)apoMb at pH 2.4 and 25°C. Dashed lines correspond to resonances broadened beyond detection. The missing resonances correspond to either prolines or peaks whose FWHH could not be reliably assessed due to spectral overlaps.
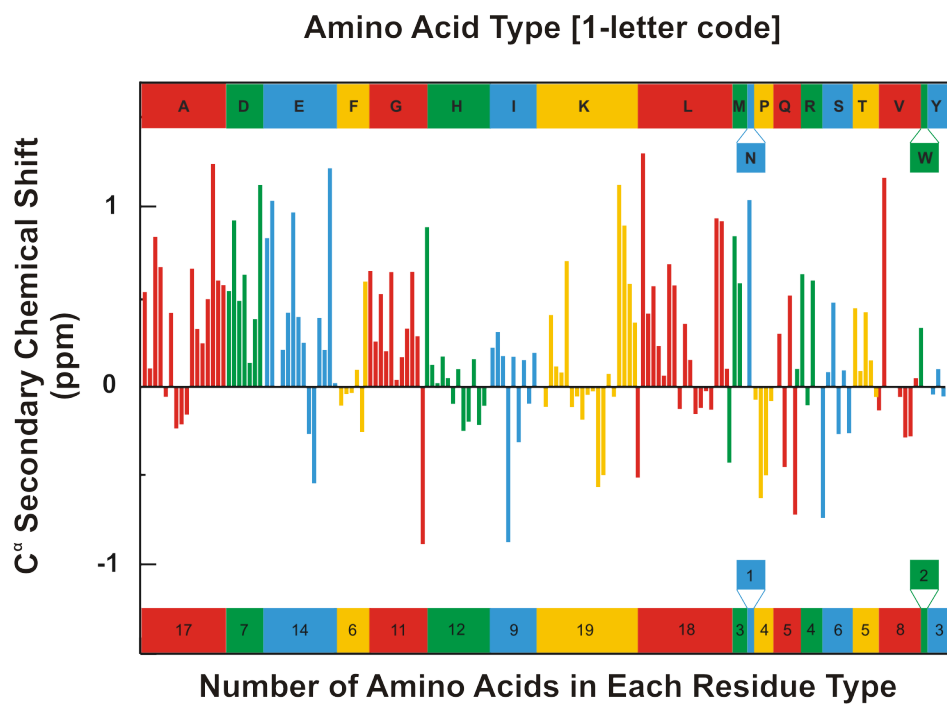
Figure 2-12. Cᵅ secondary chemical shifts for full length apoMb at pH 2.4 and 25°C grouped according to amino acid type.

2.5.3 References

1. Ribeiro, E. A., Jr.; Ramos, C. H. I. Circular Permutation and Deletion Studies of Myoglobin Indicate that the Correct Position of Its N-Terminus Is Required for Native Stability and Solubility but Not for Native-like Heme Binding and Folding. *Biochemistry* **2005**, 44, (12), 4699-4709

2. Chow, C. C.; Chow, C.; Raghunathan, V.; Huppert, T. J.; Kimball, E. B.; Cavagnero, S. Chain Length Dependence of Apomyoglobin Folding: Structural Evolution from Misfolded Sheets to Native Helices. Biochemistry, **2003**, 42, (23), 7090-7099

3. Wishart, D. S.; Sykes, B. D.; Richards, F. M., Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **1991,** 222, (2), 311-33.

4. Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D., 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **1995,** 6, (2), 135-40.

5. Lietzow, M. A.; Jamin, M.; Jane Dyson, H.; Wright, P. E., Mapping long-range contacts in a highly unfolded protein. *J. Mol. Biol.* **2002,** 322, (4), 655-662.

6. Felitsky, D. J.; Lietzow, M. A.; Dyson, H. J.; Wright, P. E., Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** 105, (17), 6278-6283.

7. Kurt, N.; Rajagopalan, S.; Cavagnero, S., Effect of Hsp70 Chaperone on the Folding and Misfolding of Polypeptides Modeling an Elongating Protein Chain. *J. Mol. Biol.* **2006,** 355, (4), 809-820.

8. Kay, L. E.; Keifer, P.; Saarinen, T., Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *J. Am. Chem. Soc.* **1992,** 114, (26), 10663-10665.

9. Ikura, M.; Kay, L. E.; Bax, A., A Novel-Approach for Sequential Assignment of H-1, C-13, and N-15 Spectra of Larger Proteins – Heteronuclear Triple-Resonance 3-Dimensional Nmr-Spectroscopy – Application to Calmodulin. *Biochemistry* **1990,** 29, (19), 4659-4667.

10. Grzesiek, S.; Bax, A., Improved 3D Triple-Resonance NMR Techniques Applied to a 31-Kda Protein. *J. Magn. Reson.* **1992,** 96, (2), 432-440.

11. Muhandiram, D. R.; Kay, L. E., Gradient-Enhanced Triple-Resonance 3-Dimensional Nmr Experiments with Improved Sensitivity. *J. Magn. Reson. Ser. B* **1994,** 103, (3), 203-216.

12. Kay, L. E.; Xu, G. Y.; Yamazaki, T., Enhanced-Sensitivity Triple-Resonance Spectroscopy with Minimal H2O Saturation. *J. Magn. Reson. Ser. A* **1994,** 109, (1), 129-133.

13. Yamazaki, T.; Lee, W.; Arrowsmith, C. H.; Muhandiram, D. R.; Kay, L. E., A Suite of Triple-Resonance Nmr Experiments for the Backbone Assignment of N-15, C-13, H-2 Labeled Proteins with High-Sensitivity. *J. Am. Chem. Soc.* **1994,** 116, (26), 11655-11666.

14. Wittekind, M.; Mueller, L., HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins. *J. Magn. Reson. Ser. B* **1993,** 101, (2), 201-205.

15. Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wuthrich, K., Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J. Mol. Biol.* **1998,** 280, (5), 933-52.

16. Wishart, D. S.; Nip, A. M., Protein chemical shift analysis: a practical guide. *Biochem. Cell Biol.* **1998,** 76, (2/3), 153-163.

17. Van Geet, A. L., Calibration of methanol nuclear magnetic resonance thermometer at low temperature. *Anal. Chem.* **1970,** 42, (6), 679-80.

18. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* **1995,** 6, (3), 277-93.

19. Goddard, T.D. and Kneller, D.G., SPARKY 3, University of California, San Francisco

20. Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Wright, P. E.; Dyson, H. J., Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J. Biomol. NMR* **2000,** 18, (1), 43-48.

21. Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J., Sequence-Dependent Correction of Random Coil NMR Chemical Shifts. *J. Am. Chem. Soc.* **2001,** 123, (13), 2970-2978.

22. Bevington, P. R.; Robinson, D. K., *Data Reduction and Error Analysis for the Physical Sciences*. Second ed.; McGraw-Hill, Inc.: New York, **1992**.

23. Wishart, D.S. and Sykes, B.D., Chemical shifts as a tool for structure determination. *Methods Enzymol.*, **1994**, 239, 363-392.

Chapter 3   Development of Hydrogen-Deuterium Exchange Methodology for Folding Studies of Ribosome-Bound Nascent Proteins

Daria V. Fedyukina and Silvia Cavagnero

Unpublished results

2012

3.1   Abstract

A fundamental understanding of protein folding *in vivo* is a gateway to protein design and disease prevention. Several decades of research led to a fairly clear picture of how purified proteins attain their functional structure when refolded in a test tube. However, little is known about protein folding and unfolding in living cells. One of the factors influencing folding in the cell is the ribosome, i.e., the protein-synthesizing machinery. This chapter develops methodologies to study ribosome-bound protein conformation at atomic resolution. Our model system includes a full length SecM-stalled ribosome-bound apomyoglobin and its N-terminal incomplete chains. Ribosome-bound nascent proteins (RNCs) are highly charged large complexes (~2.5 Mda) comprising both proteins and RNAs. Nuclear magnetic resonance (NMR), a very attractive and powerful solution-based atomic resolution technique, is desirable but not ideal for the direct analysis of intact RNCs. Direct NMR detection on the ribosome is currently not possible because the molecular weight of the ribosome exceeds the molecular weight limit of NMR spectroscopy. Here, we develop an NMR-detectable H/D-exchange (HDX) methodology for RNC protein folding studies that circumvents obstacles posed by the direct NMR detection of RNCs. This memory-based approach provides amino acid-specific information about RNCs by performing the NMR analysis after RNCs are released from the ribosome. The two main goals of this chapter are to develop HDX to render it suitable to study RNCs and to increase the RNC concentration to enable NMR detection. This research sets the foundation for utilizing NMR spectroscopy to answer questions about 1) the structure of RNC, 2) the interaction of the nascent chain with the

ribosomal surface, and 3) the key differences between protein folding from denatured states and protein folding on the ribosome.

3.2 Hydrogen/deuterium exchange is a gateway to NMR studies of ribosome-bound nascent proteins

The goal of this project is to develop a technique that will enable protein folding studies of a ribosome-bound nascent protein chains (RNCs) at amino acid resolution. RNCs are large macromolecular complexes (~2.5 Mda) comprising both proteins and RNAs. Solution-based atomic resolution techniques such as NMR are not suitable for the direct analysis of intact RNCs because the size of RNC (and the ribosome) exceeds NMR molecular weight limit. We proposed using NMR-detectable hydrogen/deuterium exchange (HDX) as a powerful tool to obtain structural information about each individual amino acid in the protein. We hypothesized that hydrogen/deuterium exchange monitored by nuclear magnetic resonance (NMR) spectroscopy allows us to obtain structural information about each amino acid in an RNC and, thus, to investigate the folding of a nascent protein at atomic resolution.

In this chapter, we explored three main areas. First, we rationalized how the developed HDX procedure would allow one to identify the degree of structure of RNCs derived from sperm whale apomyoglobin. Second, we determine optimal experimental conditions for the quenching of HDX and the behavior of RNCs under these conditions. We aimed at generating soluble ribosome-released nascent chains (NC) under quenching conditions. Third, we explored ways to increase the RNC concentration to enable the NMR detection. Utilization of Sec-M stalling approach promises high enough concentrations of RNC to employ NMR detection [1].

Over the last half-century scientists have been striving to establish the basic rules of protein folding [2-4]. These rules are required to (i) provide mechanistic understanding

of protein structure formation in a given environment; (ii) determine protein function; (iii) explain the defects in protein folding such as misfolding, aggregation, and amyloid formation; and, finally, (iv) create a foundation for devising strategies to combat diseases related to protein misfolding and aggregation such as Parkinson's, Alzheimer's, and amyloidosis. Many experiments showed how proteins fold and refold *in vitro*, i.e. in the test tube, from denaturant or low pH. However, little is known about how proteins fold in the living cell [5]. The molecular chaperones, the ribosome, and crowded cellular environment are known to modulate kinetics and folding pathways in the cell [2, 6-8]. In this Chapter, we are taking a step towards a more physiologically relevant system for studying protein folding, specifically an RNC. Investigations of the ribosomal tunnel and surface have already shown that the ribosome is indeed an interaction-rich machinery influencing folding of a synthesized polypeptide [6, 9-12]. Chapter 1 of this thesis highlights the scarce findings in this area [2]. The details of protein folding of RNCs can only be revealed by high resolution tools, yet to be developed and optimized. Atomic information provided by NMR is critical for the elucidation of the secondary and tertiary structure of NC, their compaction, and interaction with other species.

3.3   Optimization of HDX to study RNCs in solution

3.3.1   HDX methodology

We have chosen HDX as a primary technique for probing the conformation of RNCs. NMR-detected HDX enables the identification of buried and solvent-exposed regions of the protein at atomic level [13]. Procedures for performing HDX experiments on proteins are widely available in the literature and have been successfully used to study protein folding *in vitro* [14]. The general scheme for HDX experiments is shown in

Figure 3-1A. A native protein is exposed to D_2O for a variable period of time, under the conditions where folding/unfolding is faster than HDX. Solvent-exposed backbone amide protons are exchanged to the deuterium. More detailed information about HDX mechanisms can be found in a comprehensive review by Englander, S.W. *et al.* [15]. The quenching of HDX is achieved by lowering pH to 2-3 because at this pH the rates of exchange are minimal, i.e. a half-life time is more than 1 hour at 0 ^{o}C [15]. Such low rates allow one to analyze the sample of the H/D-exchanged protein, for instance, to detect NMR spectra (Figure 3-1B). From the spectroscopy data, one can obtain information about the regions or residues that are protected or buried in the initial state at pH 7. If a residue is exposed to the solvent, it is expected to exchange with D_2O resulting in the deuterated backbone nitrogen atoms and the absence of NMR signal. If a residue is not solvent exposed, for example, if it is buried in a hydrophobic core or forms salt bridges with other residues, then the backbone nitrogen atoms are expected to remain protonated. As a result, they are detectable by NMR spectroscopy, for example, by the ^{1}H,^{15}N-HSQC experiment. The detection of the H/D-exchanged protein can be also done by mass spectrometry as shown in Figure 3-1. However, the interpretation of the mass spectrometry analysis data for RNCs is undoubtedly very complex because, in addition to the nascent chain, RNCs contain at least 55 ribosomal proteins [16] and possibly co-translational chaperones [6, 7, 17]. In this Chapter, we developed methodology for potential NMR detection of the HDXed nascent chain [16].
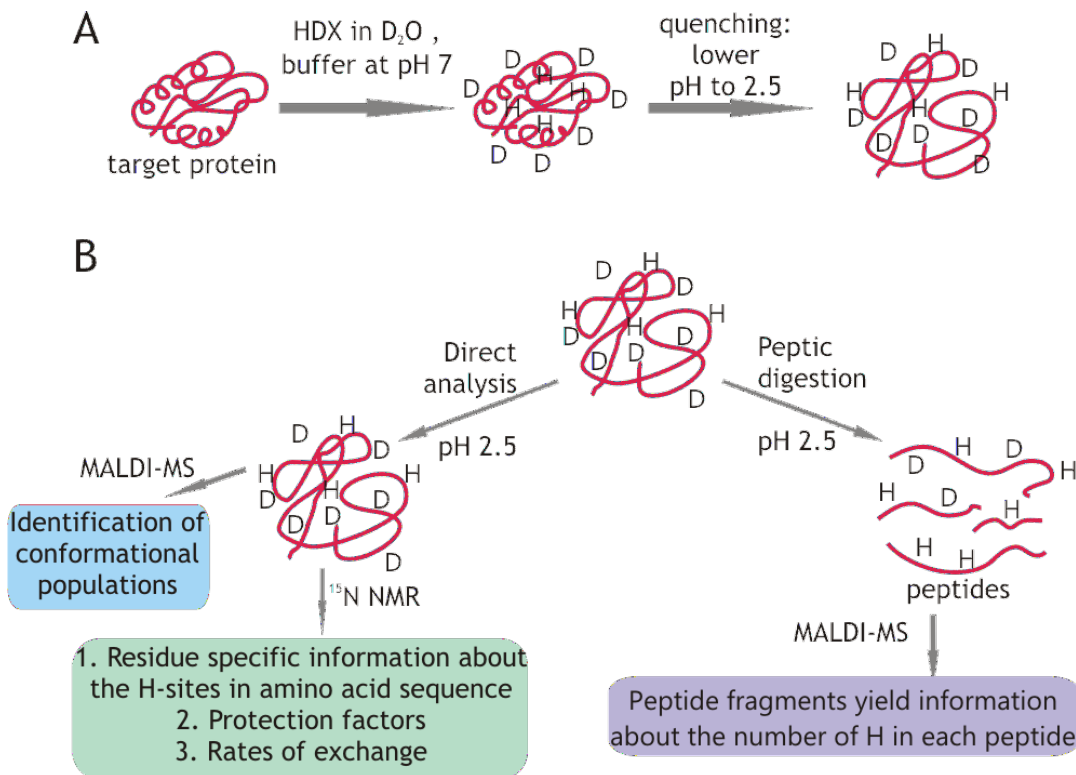
Figure 3-1. Conventional Hydrogen/Deuterium Exchange (HDX). (A) The exchange occurs in the protein solution at neutral pH and is followed by rapid quenching via lowering the temperature and pH. (B) The exchange is detected by either direct or indirect methods using atomic resolution techniques such as NMR or mass spectrometry.

3.3.2 Salt concentration dependence of HDX

HDX is dependent on pH of solution [18], electrostatic properties of the medium, and the nature of the target protein [19-21]. These dependences may have crucial effects on HDX of the large, non-covalent, and highly charged RNC complexes. The most important factor is the highly charged ribosomal surface. Pioneering studies on the influence of charge on HDX showed that the change in salt concentration affects the rates of backbone N-H HDX of charged amino acids and does not affect the rates of neutral amino acids [20]. Kim et al. showed that amide proton exchange of poly(DL-alanine) is independent on NaCl concentration in the range $0 - 2$ mM. On the other hand, rates of both acid- and base-catalyzed exchange (k_{ex}) in poly(DL-lysine) are sensitive to salt concentrations because the increase in salt concentration increases counterion competition (less H^+ or OH^- surface concentration, local pH $\neq$ bulk pH) and enhances electrostatic screening (from reactive species like H^+ and OH^-) at the surface of a macromolecule [20]. Therefore, the neutral residues are good probes of the nascent chain structure. We hypothesized that these residues have k_{ex} that do not depend on the salt concentration in the solution.

Figure 3-2 illustrates the kinetic behavior of the backbone NH groups under HDX conditions. Based on these trends we concluded that HDX on RNC experiments should be conducted in the following manner. First, the HDX experiments should be carried out at a constant salt concentration and the protection factors (i.e., the ratio of $k_{ex,random coil}/k_{ex, nascent chain}$) of neutral residues should be examined. Next, the salt concentration of the medium should be increased (from 0 to 2 M NaCl) to detect the changes in the protection factors. The changes should be detected on residues with neutral side chains because they

are not affected by the increase of salt concentration according to Kim *et al* [20]. If changes in protection factors are observed, they are most likely due to elimination of ribosome-peptide interactions, but not due to the intra-peptide conformational changes. These interactions, if significant, are expected to influence the structure and compaction of the RNC that can be captured by the proposed HDX experiments.
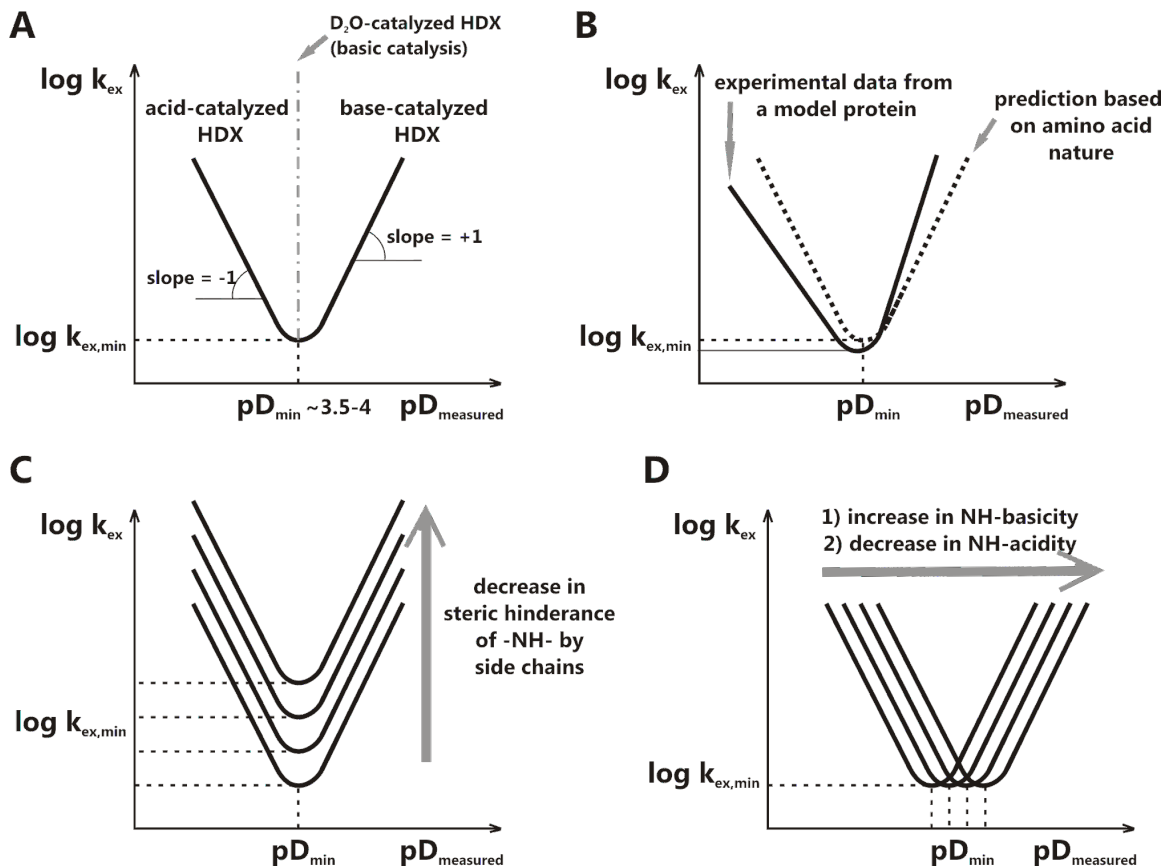
Figure 3-2. Kinetic behavior of peptide –NH- groups under HDX conditions. (A) Dependence of the rate constant of exchange on the measured pD (i.e., $-\log[D^+]$). (B) Comparison of predicted and experimental kinetic behavior of a model protein under HDX conditions. (C and D) The change of the kinetic behavior based on the nature of amino acid and its close neighbors.

Back-exchange under quenching conditions is an inevitable complication associated with the HDX method. Even at quenching conditions the rate of exchange is catalyzed by water molecules ($k_{w,min} \sim 10^{-2} - 10^{-3} min^{-1}$) [18]. The rate of back exchange is residue specific and can be taken into account [22, 23]. The initial degree of protection for most residues may be found from a series of $^{1}H,^{15}N$-SOFAST-HMQC spectra. In order to decrease the rate of back-exchange, the detection may be carried out in an organic solvent, such as DMSO [22] or CH_3OH [23] to effectively lower the catalysts' concentrations (i.e. [H^+] and [OH^-]).

3.3.3   Model system

We chose to study RNCs formed from full-length sperm whale apomyoglobin (apoMb), a 153-amino-acid-long protein (Figure 3-3). ApoMb was characterized at atomic resolution at pH 2.3 [24], pH 4 [25-27], and pH 6 [28]. At low pH, apoMb populates partial helical conformation for the amino acids corresponding to the A, D/E, and H helices. Also, long-range interactions involving residues corresponding to the native A and G-H helices and medium-range interactions within the A-B-C and G-H regions were identified in paramagnetic spin-labeling experiments [29, 30]. ApoMb's shorter N-terminal fragments have also been characterized at atomic resolution at pH 2.4 [31]. In Chapter 2, we showed that the long-range interactions of N-terminus (A-region) and C-terminus (H-region) induce additional secondary structure in the N-terminus. Therefore, apomyoglobin is a promising nascent chain because its N-terminus does not have a chance to interact with its C-terminus until after the chain is completely synthesized and released from the ribosome.

Recently, the dynamics of the ribosome-bound nascent chains of apomyoglobin and its shorter N-terminal fragments were studied by dynamic fluorescence depolarization [32]. This protein has an intermediate motion component while on the ribosome [32], is soluble at quenching conditions of HDX (pH 2.5, 4$^\text{o}$C) [31], and is small enough (~17 kDa) to be analyzed by NMR spectroscopy. The Cavagnero group expertise on the folding behavior of apomyoglobin [31] in solution sets a foundation for discovering the role of the ribosome in the folding of emerging apoMb NC.
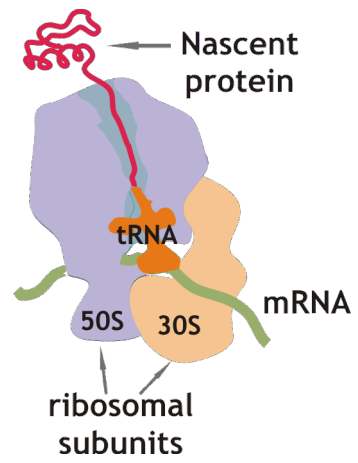
Figure 3-3. Ribosome-bound Nascent protein Chain (RNC).
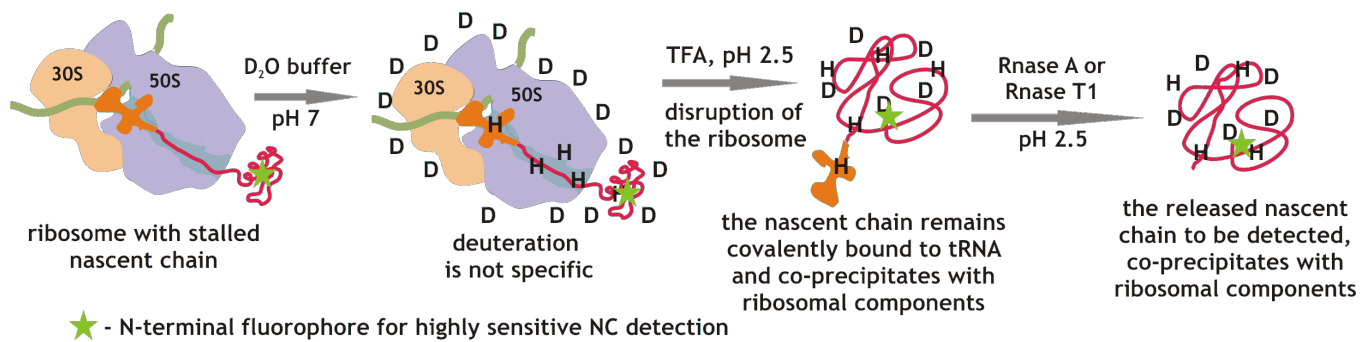
Figure 3-4. HDX on RNCs. The RNCs are generated in the cell free system with the N-terminal BODIPY-fl fluorophore (4,4-difluoro-4-bora-3a,4a-diaza-s-indacene) for sensitive fluorescence detection.

3.3.4   Optimization of HDX conditions

3.3.4.1   Large size of RNC.

RNC is very large and complex. It contains NC, mRNA, tRNA, and two ribosomal subunits 50S and 30S comprising 55 ribosomal proteins and 3 ribosomal RNAs (Figure 3-3). This complex has a molecular mass of more than 2.5 Mda and is intractable to study at atomic resolution in intact form by solution spectroscopic techniques. Even the NC-tRNA covalent complex is too large (~55kDa) to be successfully detected by NMR. We circumvented this problem using a two-step approach. First, we selectively labeled the NC with a fluorescent dye and with ^{15}N-labeled amino acids to allow for the fluorescence and NMR detection, respectively. Under these labeling conditions, 55 ribosomal proteins remain silent during spectroscopic measurements. Second, we perform HDX followed by the acidic quenching as discussed earlier in Figure 3-1. Acidic quenching does not break covalent bond between NC and tRNA. Therefore, we cleaved this covalent bond by treating the quenched RNC solution with Rnase A or Rnase T1 (Figure 3-4) [33, 34]. We showed that these highly active enzymes Rnases are capable of digesting RNA in less than a minute at pH 2.5 (Figure 3-5).

3.3.4.2   Highly charged ribosomal surface.

Upon HDX quenching conditions, a precipitate formed that contained NC. Even after the Rnase A (or T1) cleavage step, NCs remained in the precipitate (Figure 3-5). Based on highly electrostatic nature of the ribosome [35], we believe that this precipitate is formed due to strong electrostatic interactions between positively charged proteins and negatively charged RNAs. The majority of ribosomal proteins have high isoelectric

points (pI ~ 9-11) [17] and apoMb's NC has pI 8. Therefore, at pH 2.5 both ribosomal and nascent proteins are highly positively charged, whereas ribosomal RNAs and tRNA are highly negatively charged. We calculated that, at pH 2.5, there is still one negative charge per four phosphodiester bonds in rRNA.

The resulting precipitation is undesirable for two reasons. First, the precipitate may lead to isotope scrambling and local changes in pH. Local changes in pH may increase the rates of HDX leading to the premature back-exchange and incorrect NMR data interpretation. Second, the precipitate drastically reduces the NC concentration available for the NMR detection. To circumvent the problems that may result from RNCs' precipitation, we developed HDX quenching conditions for RNC.

We found the conditions under which the ribosome-released NCs are soluble under quenching conditions at pH 2.5. Our preliminary experiments showed that organic solvents (such as methanol, acetone, acetonitrile, etc.), an increased salt concentration of the solution, the competition effect by protamine sulfate [36], and the ionic denaturant 6M guanidinium chloride were not effective in improving the solubility of the released NC. Our experiments showed that 8M urea is a reproducibly good candidate for solubilization (Figure 3-6). However, the removal of 8M urea from the released NC solution turned out to be very time consuming and impractical for HDX experiments.
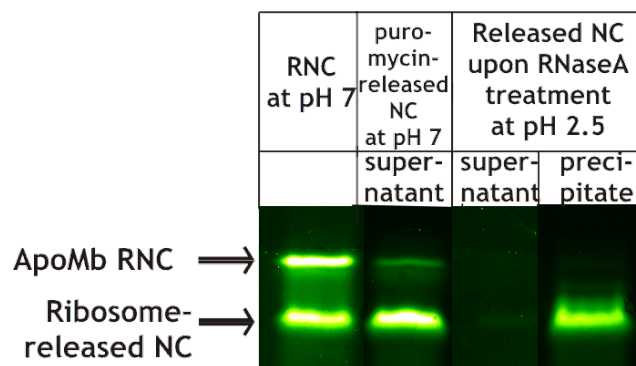
Figure 3-5. Behavior of ApoMb RNCs under HDX quenching conditions. The fluorescent gel detects only RNCs with the N-terminal BODIPY-fl fluorophore. Positions of released and bound NCs are indicated by arrows.

	no denaturant		8M Urea		6M GdnCl	
	precip	super	precip	elute 80%B	precip	elute 80%B

Ribosome-released NC →
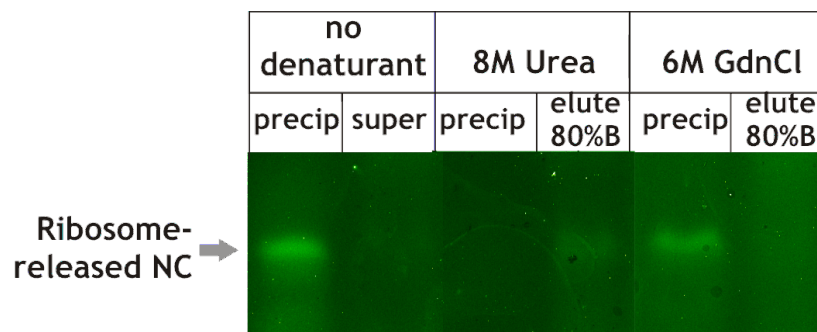
Figure 3-6. Denaturant effect on solubilization of RNCs at pH 2.5. The lanes titled "elute 80%" correspond to the fractions of released RNCs that were purified from small molecules and RNA fragments on the C18 spin column and eluted with aqueous solution of 80% CH_3CN and 0.1% TFA.
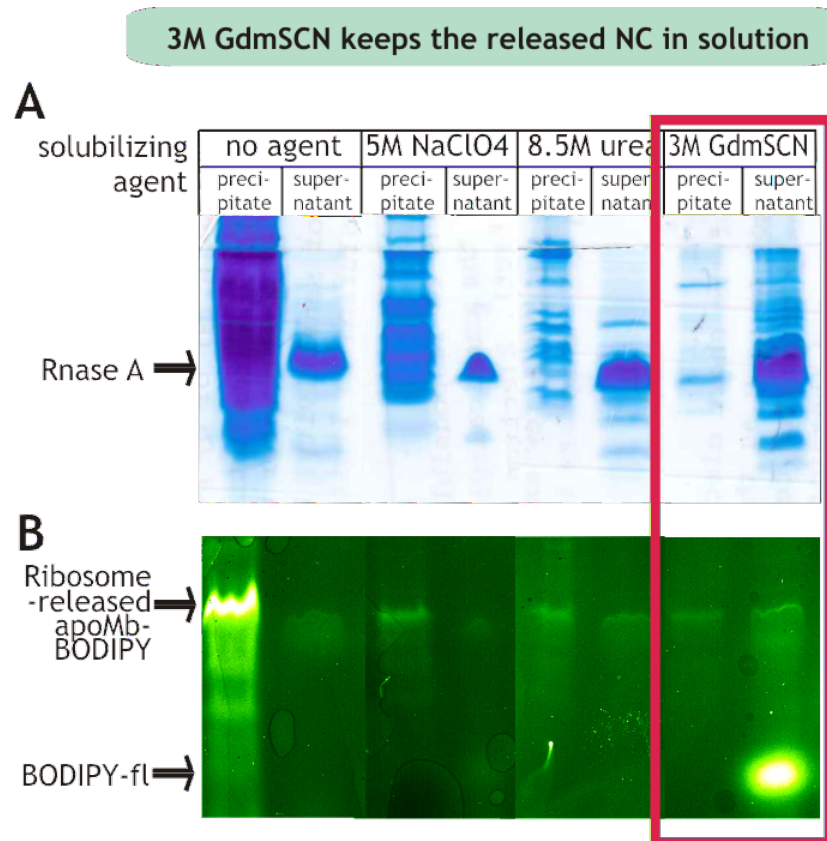
Figure 3-7. PAGE detection (pH 5) of RnaseA-digested RNCs of apoMb at pH 2.5.
Several solubilizing agents were used to bring nascent apoMb back into solution. (A) Gel stained with Coomassie blue. (B) Same gel analyzed by fluorescence with 512BP filter.

We explored Hofmeister series (or lyotropic series) to find the best salting in agent [37-39]. These series classify ions based on their ability to salt out or salt in proteins. Anions appear to have a larger effect than cations, and are usually ordered as follows:

$$F^- \sim SO_4^{2-} > HPO_4^{2-} > CH_3COO^- > Cl^- > NO_3^- > Br^- > ClO_3^- > I^- > ClO_4^- > SCN^-,$$
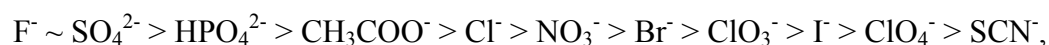
where F^- and SO_4^{2-} are the strongest kosmotropes (they salt out, stabilize proteins) and SCN^- is the strongest chaotrope (it salts in, destabilizes proteins).

The order of cations is the following:

$$NH_4^+ > K^+ > Na^+ > Li^+ > Mg^{2+} > Ca^{2+} > Guanidinium\ (Gdm^-)$$
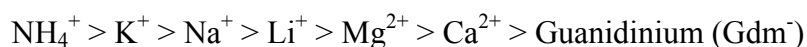
Based on these series, we chose guanidinium thiocyanate (GdmSCN) as a solubilizing agent because it consists of ions with the highest denaturing ability. Figure 3-7 shows that even at 3M concentration it is effective at solubilizing almost all ribosomal proteins and the NCs in contrast to 8.5M urea solution or ineffective GdmCl (data not shown). Even though GdmSCN solubilizes all proteins, the fluorescent band from the supernatant (Figure 3-7, highlighted in a red box) is faint. We ascribed it to the loss of the BODIPY-fl fluorophore upon GdmSCN treatment manifesting in a bright spot on the gel. It is unlikely that this fluorophore hydrolysis affects the released NC because NC's band from the supernatant fraction corresponds to the expected molecular weight of the released NC.

We concluded that the treatment of the released NC with GdmSCN at the HDX quenching conditions is the solution for precipitation problem. The next issue we tackled was the low concentration of RNC generated in the cell-free system. Based on our experiments, this concentration is 300 times less than that required for the NMR

experiments. Therefore, we chose the SecM-stalling approach to address this issue [40, 41].

3.3.5   SecM-stalled RNCs for achieving NMR-enabling concentrations.

NMR experiments for HDX detection have two requirements. First, the yield of apoMb RNC must be at least 300 μM in 10 μL. Second, the RNC should be stable and uniformly ^{15}N-labeled. A SecM stalling approach for the large scale *in vivo* RNC production and labeling is a very promising procedure both for generating NMR-required concentrations of RNC and for their uniform labeling [1, 40-42]. This approach is based on the ability of a specific C-terminal 17-amino-acid sequence to stall the translating ribosomes (see Figure 3-8A). The stalling occurs when three conserved amino acids (Phe, Trp, and Ile) in the 17-residue SecM stalling region interact tightly with two ribosomal proteins, L22 and L4. These proteins protrude inside the ribosomal tunnel and, upon interaction with the SecM region, change the conformation narrowing the tunnel [2, 42].

In order to produce the SecM-stalled RNC, we sub-cloned the apomyoglobin gene into the pBat vector containing the SecM gene to create a construct shown on Figure 3-8B. The pBat vector was kindly provided by Prof. Bukau [1, 43]. This construct has two additional components necessary for RNC purification: a TEV cleavage site and a triple Strep-tag region. The latter is critical for fast and easy purification of RNC via affinity chromatography. This allows us to separate RNC from the empty ribosomes efficiently. The former is needed to cleave purification tags from purified RNC. TEV protease removes almost all residues attached to the apoMb RNC N-terminus, retaining only Ser and Gly.

We expressed and purified the full-length apoMb RNC. This procedure enabled us to achieve 10 μM concentration of unlabeled and 1.2 μM of ^{15}N-labeled RNC in 250 μL solution. The experimental data are presented in Figure 3-9. The concentration of labeled RNC is not sufficient to prepare 300 μl of 10 μM NMR sample. In addition, the production of labeled RNC was not reproducible. We decided to halt this project.
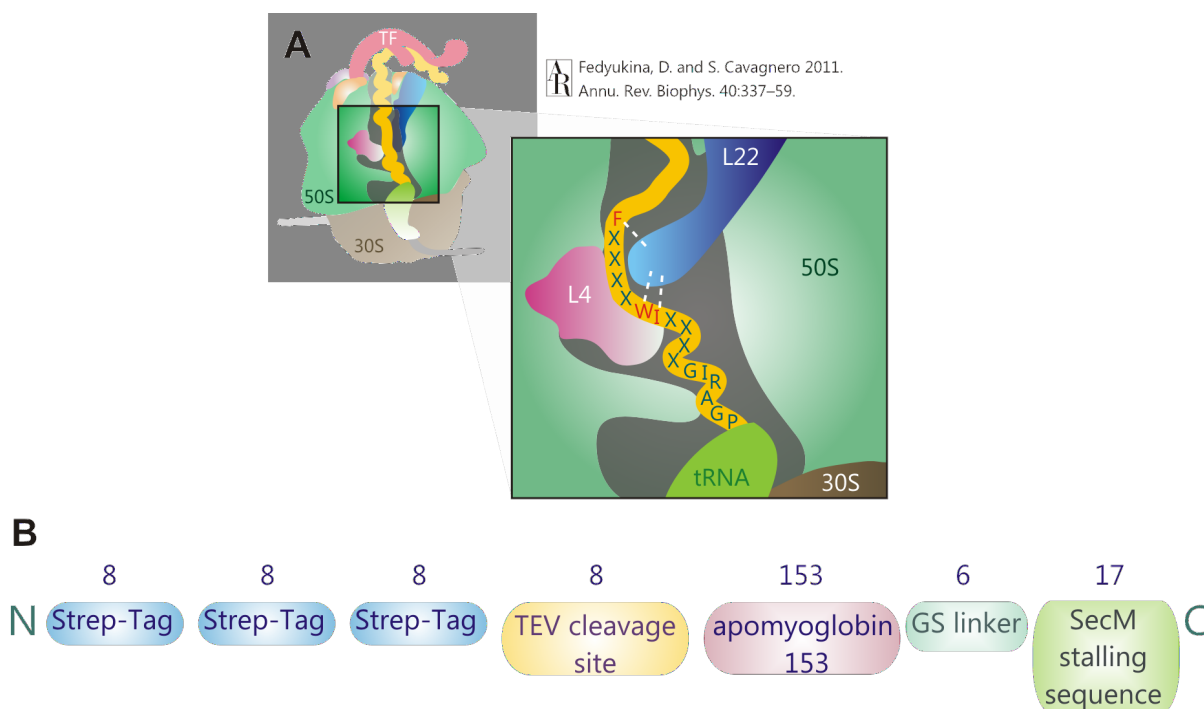
Figure 3-8. SecM stalling approach. (A) The SecM stalling mechanism is based on the tight binding of the three conserved residues in 17-residue SecM sequence to the L22 and L4 proteins protruding into the ribosomal tunnel. Critical interactions are shown in the enlarged area. The figure is reproduced with permission from the publication by Daria V. Fedyukina and Silvia Cavagnero, Annual Reviews Biophysics, 2011, 40, 337-359 [2]. (B) The construct for the large scale production of ApoMb RNCs in vivo. Numbers above the sequence correspond to the amount of amino acids in each fragment.
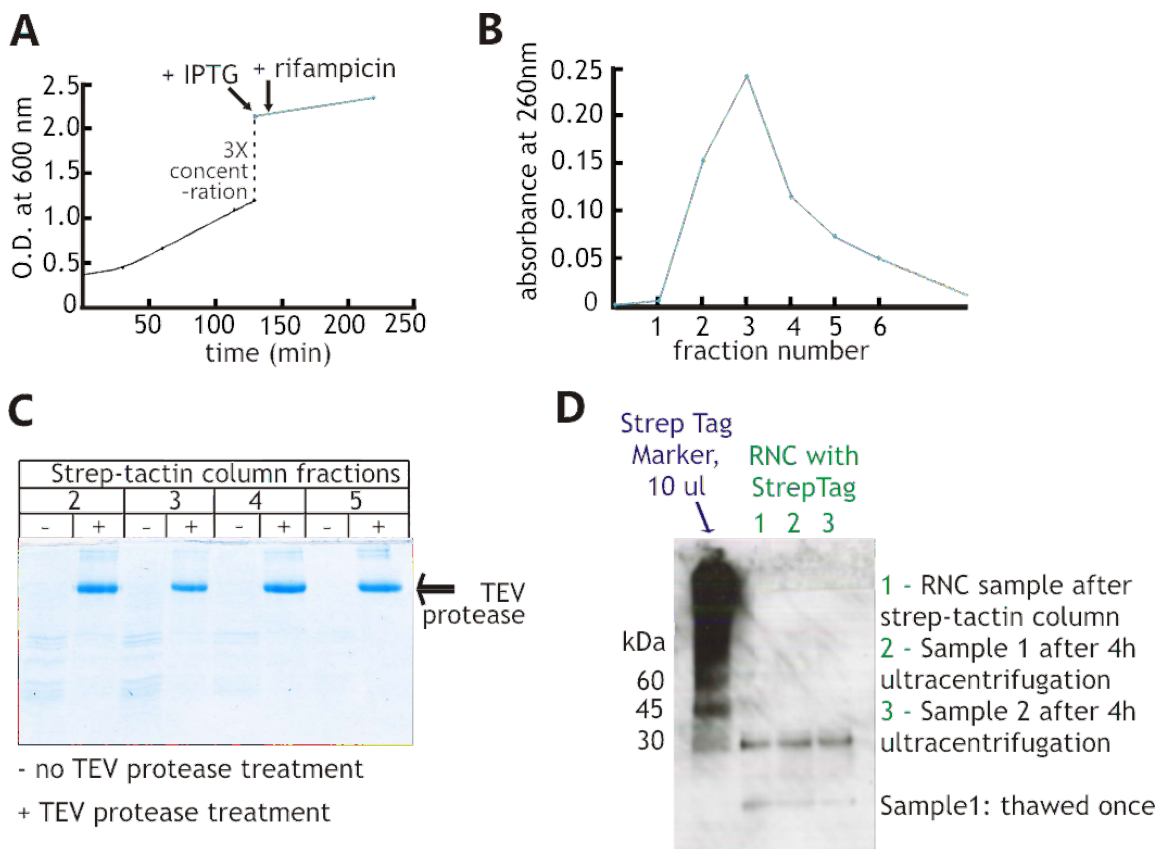
Figure 3-9. SecM-stalled RNC production. (A) Expression of ^{15}N-labeled RNC. Cells were grown in LB media to optical density ~1.3. The cell culture (3 L) was centrifuged and the cells were re-suspended in 1 L of M9 media with ^{15}NH$_4$Cl. IPTG was added immediately, rifampicin was added after 10 min. (B) Purification of ^{15}N-labeled RNC: Strep-tactin column elution profile. Absorbance at 260 nm in each fraction. (C) Purification of ^{15}N-labeled RNC: TEV cleavage of the N-terminal Strep3 tag. SDS-PAGE of the indicated fractions after Rnase A treatment [1]. (D) Western Blotting of RNC sample using Strep Tag monoclonal antibody as a primary antibody. Samples 2 and 3 are tests for the stability of RNC over the period of more than 10 hours.

3.4 Future Directions

We have not succeeded in achieving the reproducibility of the ^{15}N-labeled RNC production. Given that the HDX study requires a large number of highly reproducible ^{15}N-labled RNC samples, we reconsidered our approach to study protein folding on the ribosome at the atomic resolution. One of the future directions to be pursued in Cavagnero research group is the use of Trp-, Tyr-, or Phe-auxotrophic cell strains to selectively label RNC with ^{13}C and/or ^{15}N at Trp, Tyr, or Phe positions. We have received the corresponding cell strains from the Prof. Inoyue [44] that are suitable for the *in vivo* RNC production. The detailed discussion of the proposed direction of the RNC folding studies can be found in Chapter 5.

3.5 References

1.   Rutkowska, A., et al., *Large-scale purification of ribosome-nascent chain complexes for biochemical and structural studies.* FEBS Letters, 2009. **583**: p. 2407-2413.

2.   Fedyukina, D.V. and S. Cavagnero, *Protein Folding at the Exit Tunnel.* Annual Review of Biophysics, 2011. **40**(1): p. 337-359.

3.   Hartl, F.U. and M. Hayer-Hartl, *Converging concepts of protein folding in vitro and in vivo.* Nat Struct Mol Biol, 2009. **16**(6): p. 574-581.

4.   Dyson, H.J. and P.E. Wright, *Unfolded proteins and protein folding studied by NMR.* Chem. Rev. (Washington, DC, U. S.), 2004. **104**(8): p. 3607-3622.

5.   Gershenson, A. and L.M. Gierasch, *Protein folding in the cell: challenges and progress.* Current Opinion in Structural Biology, 2010. In Press, Corrected Proof.

6.   Hoffmann, A., et al., *Trigger Factor Forms a Protective Shield for Nascent Polypeptides at the Ribosome.* Journal of Biological Chemistry, 2006. **281**(10): p. 6539-6545.

7.   Svetlov, M.S., et al., *Effective cotranslational folding of firefly luciferase without chaperones of the Hsp70 family.* Protein Science, 2006. **15**(2): p. 242-247.

8.   Wyttenbach, A., V. O'Connor, and R.J. Ellis, *Protein Aggregation: Opposing Effects of Chaperones and Crowding*, in *Folding for the Synapse*2011, Springer US. P. 9-34.

9.   Bhushan, S., et al., *[alpha]-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel.* Nat Struct Mol Biol, 2010. **17**(3): p. 313-317.

10.   Kosolapov, A. and C. Deutsch, *Tertiary interactions within the ribosomal exit tunnel.* Nat Struct Mol Biol, 2009. **16**(4): p. 405-411.

11.   Lu, J. and C. Deutsch, *Folding zones inside the ribosomal exit tunnel.* Nat. Struct. Mol. Biol., 2005. **12**: p. 1123-1129.

12.   Tu, L.W. and C. Deutsch, *A Folding Zone in the Ribosomal Exit Tunnel for Kv1.3 Helix Formation.* Journal of Molecular Biology, 2010. **396**(5): p. 1346-1360.

13.   Englander, S.W. and N.R. Kallenbach, *Hydrogen exchange and structural dynamics of proteins and nucleic acids.* Quarterly reviews of biophysics, 1983. **16**(04): p. 521-655.

14.   Krishna, M.M.G., Hoang, L., Lin, Y., Englander, S.W., *Hydrogen exchange methods to study protein folding.* Methods, 2004. **34**: p. 51-64.

15.   Englander, S.W., et al., *Mechanisms and uses of hydrogen exchange.* Current Opinion in Structural Biology, 1996. **6**(1): p. 18-23.

16.   Korobeinikova, A.V., M.B. Garber, and G.M. Gongadze, *Ribosomal proteins: Structure, function, and evolution.* Biochemistry (Moscow), 2012. **77**(6): p. 562-574.

17.   Wilson, D.N., K.H. Nierhaus, and M.M. Cox, *Ribosomal Proteins in the Spotlight.* Critical Reviews in Biochemistry & Molecular Biology, 2005. **40**(5): p. 243-267.

18.   Yawen, B., et al., *Primary structure effects on peptide group hydrogen exchange.* Proteins: Structure, Function, and Genetics, 1993. **17**(1): p. 75-86.

19. Fogolari, F., et al., *pKa Shift Effects on Backbone Amide Base-Catalyzed Hydrogen Exchange Rates in Peptides.* Journal of the American Chemical Society, 1998. **120**(15): p. 3735-3738.

20. Kim, P.S. and R.L. Baldwin, *Influence of charge on the rate of amide proton exchange.* Biochemistry, 1982. **21**(1): p. 1-5.

21. Perrin, C.L., J.-H. Chen, and B.K. Ohta, *Amide Proton Exchange in Micelles.* Journal of the American Chemical Society, 1999. **121**(11): p. 2448-2455.

22. Nishimura, C., H.J. Dyson, and P.E. Wright, *Enhanced picture of protein-folding intermediates using organic solvents in H/D exchange and quench-flow experiments.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(13): p. 4765-4770.

23. Rand, K.D., et al., *Investigation of amide hydrogen back-exchange in Asp and His repeats measured by hydrogen (1H/2H) exchange mass spectrometry.* International Journal of Mass Spectrometry, 2010. In Press, Corrected Proof.

24. Yao, J., et al., *NMR Structural and Dynamic Characterization of the Acid-Unfolded State of Apomyoglobin Provides Insights into the Early Events in Protein Folding.* Biochemistry and Cell Biology, 2001. **40**(12): p. 3561-3571.

25. Eliezer, D., et al., *Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin.* Biochemistry, 2000. **39**(11): p. 2894-2901.

26. Eliezer, D., et al., *Populating the equilibrium molten globule state of apomyoglobin under conditions suitable for structural characterization by NMR.* FEBS Letters, 1997. **417**(1): p. 92-96.

27. Jamin, M. and R.L. Baldwin, *Two forms of the pH 4 folding intermediate of apomyoglobin.* Journal of Molecular Biology, 1998. **276**(2): p. 491-504.

28. Kitahara, R., et al., *High pressure NMR reveals that apomyoglobin is an equilibrium mixture from the native to the unfolded.* Journal of Molecular Biology, 2002. **320**(2): p. 311-319.

29. Lietzow, M.A., et al., *Mapping long-range contacts in a highly unfolded protein.* J. Mol. Biol., 2002. **322**(4): p. 655-662.

30. Felitsky, D.J., et al., *Modeling transient collapsed states of an unfolded protein to provide insights into early folding events.* Proc. Natl. Acad. Sci. U. S. A., 2008. **105**(17): p. 6278-6283.

31. Fedyukina, D.V., et al., *Contribution of Long-Range Interactions to the Secondary Structure of an Unfolded Globin.* Biophysical Journal, 2010. **99**(5): p. L37-L39.

32. Ellis, J.P., et al., *Chain Dynamics of Nascent Polypeptides Emerging from the Ribosome.* ACS Chem. Biol., 2008. **3**(9): p. 555-566.

33. Osterman, H.L. and F.G. Walz, *Subsites and catalytic mechanism of ribonuclease T1: kinetic studies using GpA, GpC, GpG, and GpU as substrates.* Biochemistry, 2002. **17**(20): p. 4124-4130.

34. Pace, C.N., et al., *Ribonuclease T1: Structure, Function, and Stability.* Angewandte Chemie International Edition in English, 1991. **30**(4): p. 343-360.

35. Baker, N.A., et al., *Electrostatics of nanosystems: Application to microtubules and the ribosome.* Proceedings of the National Academy of Sciences, 2001. **98**(18): p. 10037-10041.

36. Anton, P., J. Paul, and E.P. Peter, *Hydrogen/deuterium exchange on protein solutions containing nucleic acids: utility of protamine sulfate.* Rapid Communications in Mass Spectrometry, 2008. **22**(16): p. 2423-2428.

37. Zhang, Y. and P.S. Cremer, *Interactions between macromolecules and ions: the Hofmeister series.* Current Opinion in Chemical Biology, 2006. **10**(6): p. 658-663.

38. Baldwin, R.L., *How Hofmeister ion interactions affect protein stability.* Biophysical Journal, 1996. **71**(4): p. 2056-2063.

39. Goto, Y., N. Takahashi, and A.L. Fink, *Mechanism of acid-induced folding of proteins.* Biochemistry, 1990. **29**(14): p. 3480-3488.

40. Nakatogawa, H. and K. Ito, *Intraribosomal Regulation of Expression and Fate of Proteins.* ChemBioChem, 2004. **5**(1): p. 48-51.

41. Nakatogawa, H. and K. Ito, *Secretion monitor, SecM, undergoes self-translation arrest in the cytosol.* Molecular Cell, 2001. **7**(1): p. 185-192.

42. Nakatogawa, H. and K. Ito, *The ribosomal exit tunnel functions as a discriminating gate.* Cell, 2002. **108**(5): p. 629-636.

43. Rutkowska, A., et al., *Dynamics of Trigger Factor Interaction with Translating Ribosomes.* Journal of Biological Chemistry, 2008. **283**(7): p. 4124-4132.

44. Vaiphei, S.T., et al., *Use of Amino Acids as Inducers for High-Level Protein Expression in the Single-Protein Production System.* Applied and Environmental Microbiology, 2010. **76**(18): p. 6063-6068.

Chapter 4   Electrostatic and Nonpolar Properties of

Ribosomal Proteins

The content of this chapter is being developed as a manuscript in preparation:

Daria V. Fedyukina and Silvia Cavagnero 'Electrostatic properties of ribosomal proteins',

in preparation (2013)

4.1   Abstract

Ribosomes are large macromolecular complexes consisting of RNA and proteins. This composition renders the ribosome a highly charged species. The aim of this study is to investigate the electrostatic and nonpolar properties of ribosomal proteins. These properties are important because they influence interactions with other proteins and various co-solvents and ions in solutions, eventually contributing to ribosomal assembly and protein folding on the ribosome. We examined 50S ribosomal subunits from ten species, based upon structural and amino acid sequence data from the Protein Data Bank and the UniProt Knowledgebase, respectively. We found a clear distinction between the net charge of ribosomal proteins from halophilic and non-halophilic organisms. Specifically, ca. 67% of ribosomal proteins from halophiles are negatively charged while only up to ca. 15% of ribosomal proteins from non-halophiles share this property. Conversely, hydrophobicity tends to be lower for ribosomal proteins from halophiles than for the corresponding proteins from non-halophiles. The surface electrostatic potential of ribosomal proteins has distinct positive and negative regions across all the examined species. Hence, the majority of ribosomal proteins experiences a significant degree of charge segregation under both low- and high-salt conditions. This important property is a key feature that enables the ribosome to accommodate proteins regardless of their overall net charge.

4.2 Introduction

Several crystal structures of large and small ribosomal subunits became available over the past decade [1-4]. This advancement has provided the opportunity to closely examine the peculiar conformational features of ribosomal RNA and proteins. Reviews on this topic are available, including the analysis of ribosomal components [5, 6], the structural and functional [7, 8] features of ribosomal proteins and the mechanisms of ribosomal assembly [9, 10].

The electrostatic properties of ribosomal proteins, however, have not been explored in detail so far, despite their fundamental importance. These properties affect how the ribosome macromolecular complex is assembled and stabilized, and how it interacts with several biological cofactors (e.g., molecular chaperones), the newly synthesized nascent chain during translation, and various ions [11-13].

This chapter focuses on the electrostatic properties of ribosomal proteins from ten organisms: three from bacteria (*E. coli*, *T. ukaryote es*, *D. radiodurans*), two from ukaryote (*S. cerevisiae*, *T. ukaryote e*), two from non-halobacterial archaea (*M. thermautotrophicus*, *S. solfataricus*), and three from the halobacteria class of the archaea domain (*H. marismortui*, *H. jeotgali*, *H. archaeon*). The physiological environment of ribosomes from the former three groups include moderate salt content (i.e. ca. 150 mM) while the environment of archaeal ribosomes from halophilic organisms is characterized by high salt concentrations, up to 2-5 M [14].

Due to the high-salt physiological environment, the isoelectric point (pI) of proteins from halophiles is known to be generally lower [15, 16] than the pI of proteins from non-halophilic organisms. This is because low pI proteins are rich in aspartate and

glutamate residues that are abundant in proteins from halophiles [14]. Surprisingly, it is not known whether this trend is also followed by the corresponding ribosomal proteins, which serve the special task of interacting with the highly negatively charged rRNA. In essence, it is not clear how a stable ribosomal assembly could result from the interaction between highly negatively charged (i.e., ribosomal proteins and rRNA) particles. In addition, it is generally not known what the electrostatic features of the ribosomal proteins from all organisms are, and whether they share any common trends.

The motivation to pursue this study in the context of my thesis is threefold. First, the investigation of protein folding on the ribosome (e.g., see Chapter 3) strongly depends on understanding the ribosomal electrostatic properties. For instance, under hydrogen-deuterium exchange (HDX) quenching conditions (i.e., pH 2.5), the ribosome can no longer be treated as an inert macromolecular complex. The electrostatic properties of the ribosome, its surface, and its proteins affect how the ribosome interacts with other proteins and various co-solvents and ions in solutions [11, 12]. Second, electrostatic properties of the ribosomal surface as a whole and of surfaces of ribosomal proteins may help select appropriate nascent chains for co-translational protein folding studies. Third, knowledge of the physical features of ribosomal proteins contributes to our understanding of RNA-protein interactions.

We found that ribosomal proteins from halophilic bacteria follow the overall trend of the parent organism and have a higher percent of low-pI, acidic proteins than ribosomal proteins from non-halophilic organisms. We show that the majority of ribosomal proteins from all studied organisms exhibits a large degree of intramolecular charge segregation. This property supports tight binding to ribosomal RNA and ensures

better hydration of the solvent-exposed sides of ribosomal proteins. Better hydration results from the fact that condensed solvent-exposed negative charges are capable of competing for water with RNA phosphate groups on the ribosomal surface [14, 17]. We hypothesize that, overall, the ribosomal surface is characterized by a similar microenvironment across all organisms.

4.3   Results and Discussion

4.3.1   Role of net charge

We analyzed all available proteins from the large ribosomal subunit of the ribosomes from *E. coli*, *T.   ukaryote es*, *D. radiodurans*, *S. cerevisiae*, *T.   ukaryote e*, *M. thermautotrophicus*, *S. solfataricus*, *H. marismortui*, *H. jeotgali*, *H. archaeon* in terms of their pI and average net charge per residue (MNC). Using UniProt Knowledgebase [18] and online ProtParam software [19], we calculated pI values for all studied 50S ribosomal proteins from ten different organisms listed above. MNC values were calculated as described in Materials and Methods. These results are presented in Table 4-1. Two clear trends emerged differentiating ribosomal proteins from non-halophilic and halophilic organisms. First, ribosomal proteins from non-halophiles have a large fraction of high-pI proteins (i.e., positively charged proteins at the physiological pH 7.4), ranging from 86 to 100% (see Table 4-1). This result is consistent with the expectation that ribosomal proteins experience strong electrostatic interactions with the highly negatively charged ribosomal RNA [6]. In contrast, the overall percentage of high-pI proteins in the corresponding proteomes fluctuates between 27% and 60%, leaning towards a higher abundance of acidic proteins. Second, ribosomal proteins from

halophilic organisms have on average a much lower pI, i.e., they are much more acidic, than ribosomal proteins from non-halophilic organisms. The ratio of low-pI to high-pI ribosomal proteins is ~ 2:1. Average MNC values (i.e., MNC averaged over all studied 50S ribosomal proteins from each species) of ribosomal proteins in Table 4-1 are consistent with the trends in pI described above across all studied species, i.e., these values are positive for ribosomal proteins from non-halophiles and negative for those from halophiles.

The low average pI of ribosomal proteins from halophiles is consistent with the known general properties of proteins from halophilic organisms, given that even non-ribosomal proteins from halophiles have, on average, low pI [15] (see Table 4-1). For this class of proteins, it is known that the high degree of protein acidity arises as a natural response to the high salt concentration of the physiological medium. Acidic proteins have a larger fraction of negatively charged amino acids at the physiological pH of 7.4, i.e. Asp (D) and Glu (E). Proteins with a high D and E content are more highly hydrated than proteins from non-halophilic organisms, which have a lower D and E content [14]. Proteins from halophiles have to compete with the high salt content of the medium to maintain an effective hydration shell. Proteins from other organisms that live in normal salt conditions (i.e., 150 mM salt concentration) do not need to compete for water [14, 20-22].

The mere existence of some positively charged ribosomal proteins, ca. ~ 33%, in high-salt environments is intriguing. This percent is considerably larger than the percent of negatively charged proteins from the corresponding whole organism (close to 0 %) [15]. To date, it is not yet understood how the high-pI proteins from halophiles could

even be biosynthesized and survive in a high-salt medium, before being incorporated into the ribosome. More research needs to be done to shed light on this topic. We suggest that the 2:1 ratio of positively to negatively charged ribosomal proteins results from a balance between the need to prevent protein destabilization at high salt concentration [23] and the requirement to tightly bind to rRNA [6].

Table 4-1. Mean net charge per residue (MNC) and pI of proteins from the large ribosomal (Rb) subunits from ten organisms. MNC and pI are unitless parameters. The average MNC is defined as the MNC averaged over all 50S ribosomal proteins of each species studied in this work.

Species	Fraction of low pI proteins in large Rb subunits	Average MNC of proteins in large Rb subunits	Fraction of low-pI proteins in proteome[15]
E. coli K12	7%	0.08	64%
T. thermophilus	6%	0.12	~ 58%
D. radiodurans	9%	0.10	~ 65%
S. cerevisiae	11%	0.13	~ 50%
T. thermophila	0%	0.15	~ 40%
S. solfataricus	3%	0.12	~ 43%
M. thermautotrophicus	14%	0.09	~ 73%
H. marismortui	66%	-0.04	~ 93%
H. jeotgali	67%	-0.03	N/A
H. archaeon DL31	70%	-0.05	N/A

4.3.2   Role of hydrophobicity

We plotted the mean hydrophobicity per residue (MH) as a function of pI for ribosomal proteins from different types of organisms, as shown in Figure 4-1. Figure 4-1 illustrates the dramatic shift of the ribosomal proteins in halophiles towards low pI. The figure also shows that, within each type of organism, the hydrophobicity of the positively charged proteins is generally lower than that of negatively charged proteins. However, ribosomal proteins from halophilic archaea are overall less hydrophobic than ribosomal proteins from non-halophilic organisms.

We further explored the role of hydrophobicity by using MNC and MH data to generate Uversky-type plots for ribosomal proteins from ten different species (Figure 4-2) [24]. According to Uversky *et al.* [24], MH and MNC are excellent predictors of protein folding status (ordered, partially ordered, or fully disordered). For instance, disordered proteins tend to have high net charge, which leads to strong charge-charge repulsion, and low hydrophobicity. This implies an insufficient driving force for hydrophobic collapse, which is a requirement for autonomous folding. Chen *et al.* suggested that numerous families of ribosomal proteins contain conserved regions of predicted disorder [25]. In Figure 4-2, panels A-C show Uversky-type plots for ribosomal proteins from seven non-halophilic organisms and panel D shows the corresponding plots for proteins from halophilic species. The solid line represents the putative discriminating edge between intrinsically disordered (IDPs, to the left) and independently folded (to the right) proteins, based on prior work from Uversky *et al* [24]. Regions enclosed within the dashed lines correspond to sections of the graph indiscriminately hosting proteins from both structural classes [26]. After careful examination of the trends of Figure 4-2, we

calculated four hydrophobicity-related parameters to better characterize ribosomal proteins from ten organisms. These data are summarized in Table 4-2.

First, according to the definition of intrinsically unfolded proteins [24], all proteins with MH less than 0.37 are incapable of forming ordered structures independently no matter how high their MNC. This is due to the fact that their hydrophobicity is not sufficient for hydrophobic collapse. The fraction of proteins with MH < 0.37 in Table 4-2 shows that ribosomes from halophiles *H. marismortui* and *H. jeotgali* have the highest fraction of this type of proteins. We speculate that the very low hydrophobicity of these proteins evolved in response to the need to penetrate and induce order within the rRNA structure upon inter-molecular interactions during ribosomal assembly [9, 17, 27, 28].

Second, the fraction of fully disordered proteins in Table 4-2 shows that all ribosomes contain many (from 16.7% to 39.5%) IDPs, confirming the importance of disorder in ribosomal assembly across all organisms [25]. Additionally, only IDPs from halophiles are both comparably distributed between negatively and positively charged, while ribosomal IDPs from non-halophiles are largely positively charged. This observation supports the idea of a larger degree of order in negatively charged ribosomal proteins and parallels the finding of positively charged disordered extensions and negatively charged ordered parts of the *H. marismortui* ribosomal proteins [6, 17, 25].
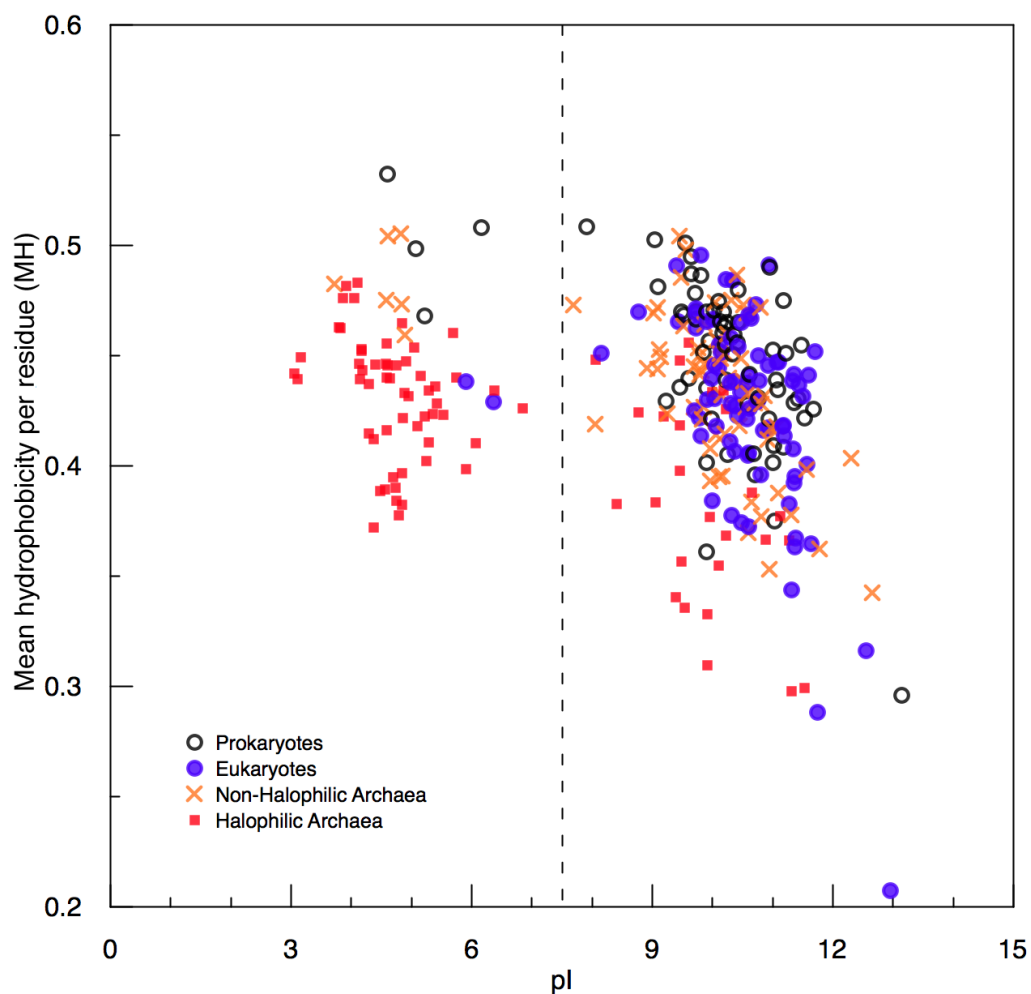
Figure 4-1. Hydrophobicity and pI of ribosomal proteins in the large subunit of the ribosome from nine organisms. Red solid squares: 50S ribosomal proteins from halophilic archaea *H. marismortui*, *H. jeotgali*, and *H. archaeon*. Blue solid circles: eukaryotic 60S ribosomal proteins from *S. cerevisiae* and *T.   ukaryote e*. Black open circles: bacterial 50S ribosomal proteins from *E. coli*, *T.   ukaryote es*, and *D. radiodurans*. Orange crosses: 50S ribosomal proteins from non-halophilic archaea *S. solfataricus* and *M. thermautotrophicus*. Hydrophobicity and pI are unitless parameters. The vertical dotted line denotes the physiological pH of 7.4.

Third, no ribosomal proteins from halophilic organisms are fully ordered. They are either disordered or partially ordered. This trend is illustrated by the fraction of fully ordered proteins. While for all non-halophilic species the fraction of fully folded ribosomal proteins fluctuates between 4.3% (*S. cerevisiae*) and 20% (*E. coli*), there are no proteins from halophiles with fully foldable sequences.

Fourth, we calculated the average MH of each of the studied species and showed that it is indeed lower in halophilic species than in non-halophilic organisms (Table 4-2, column five). This trend parallels the fact that the MH of the proteome of halophilic organisms is lower than that of non-halophilic species [29].
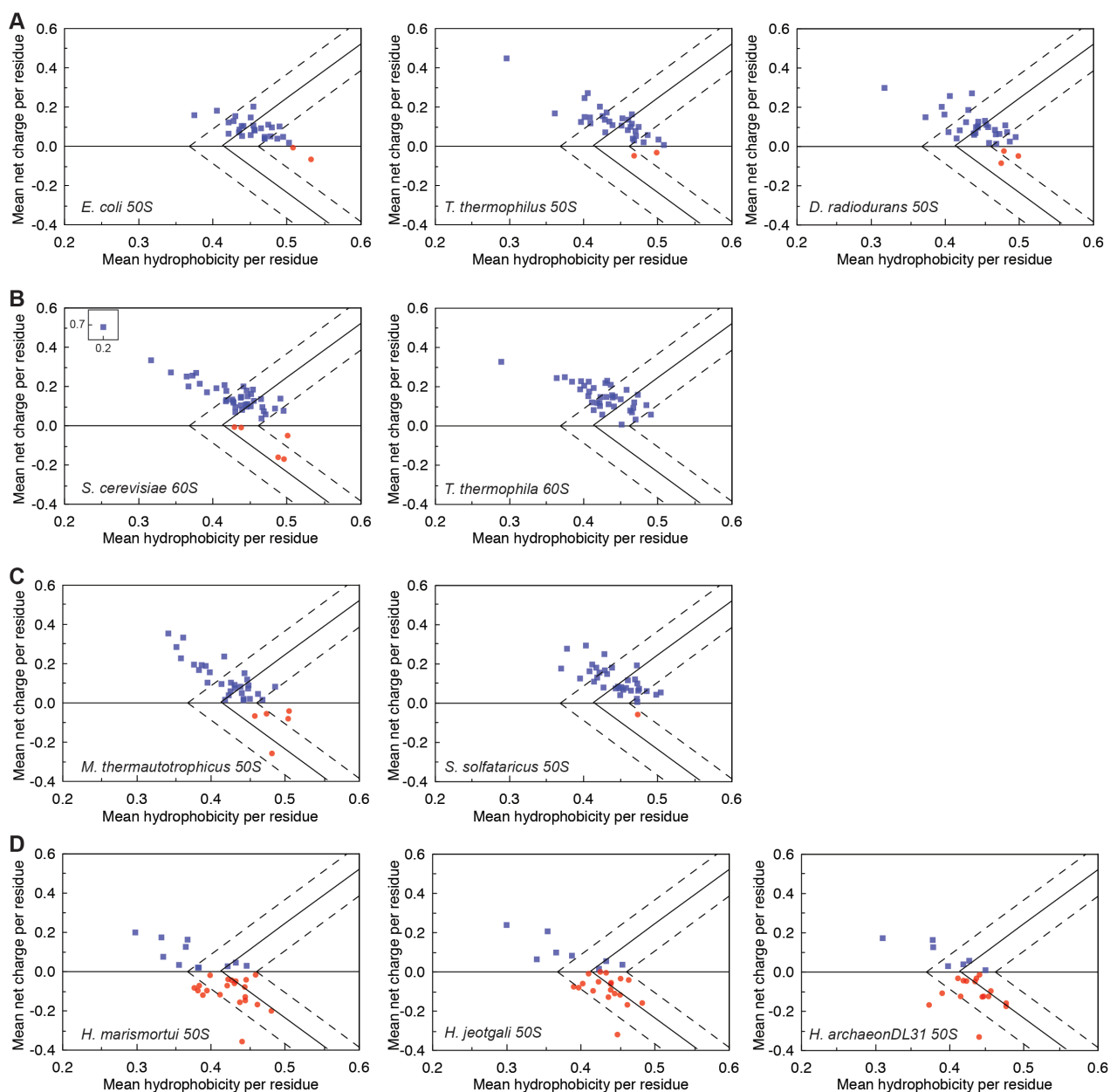
Figure 4-2. Uversky-style (i.e., MNC *vs* MH) plots [24] of proteins from the large ribosomal subunits of individual bacteria (A), ukaryote (B), non-halophilic archaea (C), and halophilic archaea (D). The solid line separates intrinsically disordered proteins (IDPs, on the left) from independently folded proteins (on the right). The regions on the left, right, and in-between the dashed lines host IDPs, folded, and partially ordered proteins, respectively. Blue squares and red circles denote proteins with

positive and negative mean net charges per residue, respectively. Calculations were performed based on the UniProtKB sequence information (see Materials and Methods).

Table 4-2. Hydrophobicity of proteins from the large ribosomal (Rb) subunits of ten organisms.

Species	Fraction of proteins in large Rb subunits with MH < 0.37 (MNC – any)	Fraction of fully disordered proteins in large Rb subunits	Fraction of fully ordered proteins in large Rb subunits	Average MH of proteins in large Rb subunits
Non-halophiles				
Prokaryotes				
E. coli K12	0.0%	17%	20%	0.46
T. thermophilus	6.1%	27%	15%	0.44
D. radiodurans	3.0%	21%	15%	0.44
Eukaryotes				
S. cerevisiae	11%	26%	4.3%	0.43
T. thermophila	5.3%	40%	5.3%	0.43
Non-halophilic Archaea				
S. solfataricus	0.0%	26%	14%	0.44
M. thermautotrophicus	11%	31%	8.6%	0.43
Halophiles				
H. marismortui	16%	28%	0.0%	0.41
H. jeotgali	15%	30%	0.0%	0.42
H. archaeon DL31	4.3%	26%	0.0%	0.42

The above observations support our hypothesis that there is some charge segregation within the majority of ribosomal proteins. To test this hypothesis, we calculated and visualized the electrostatic surface potential of 193 50S ribosomal proteins and six 50S ribosomal subunits using the Adaptive Poisson-Boltzmann Solver (APBS) software package (see Materials and Methods). Figure 4-3 compares the ribosomal electrostatic surfaces of *E. coli* (on the left) and *H. marismortui* (on the right) 50S ribosomal subunits. Ribosomal proteins make up approximately 40% of the ribosomal molecular weight [5]. In non-halophilic organisms, these proteins have pI ranging from 9 to 12. Thus, the surface of *E. coli* ribosomes is expected to be much more positively charged than that of *H. marismortui* ribosomes. However, we did not observe striking differences in surface charge (Figure 4-3). Therefore, we hypothesize that charge is ubiquitously segregated in ribosomal proteins so that the bulk of the positive charge is buried inside the ribosome upon assembly to optimize electrostatic interactions with the rRNA phosphate groups (Figure 4-4).

Figure 4-3. Exit tunnel view of the electrostatic potential mapped on the 50S molecular surface of *E. coli* (left) and *H. marismortui* (right). Electrostatic potentials are obtained upon solution of the Linearized Poisson-Boltzmann Equation (LPBE) at 150 mM KCl with a solute dielectric of 2.0 and solvent dielectric of 78.0 by using 3D structures from 2AW4 [30] and 2QA4 [31] Protein Data Bank (PDB) entries for *E. coli* and *H. marismortui*, respectively. A blue color indicates regions with positive potential (> +3 kT/e) values and a red color indicates regions with negative potential (< -3 kT/e) values, where kT – a unit of energy equal to 4.11×10^{-21} Joules at room temperature (k – Boltzmann constant, T – temperature in Kelvin) and e – electric charge in Coulombs.

Figure 4-4. Schematic representation of the charge segregation concept. The negative

charges enclosed in circles denote charges due to rRNA phosphate groups.

4.3.3 Amino acid composition of ribosomal proteins

We studied the specific amino acid composition of ribosomal proteins from halophilic and non-halophilic organisms to better understand the origin of the shifted MNC and MH. Figure 4-5 shows that the increased acidity of the ribosomal proteins from halophiles is due an increase in aspartate (D) and glutamate (E) residues, as well as the decrease in lysine (K) residues. Interestingly, the fraction of arginine I residues remains approximately constant (or even slightly increases). Fukuchi *et al.* showed that, in the proteome of halophiles, negatively charged residues dominate on the protein surface (1.7 times more frequent) in contrast to proteomes from non-halophiles, where the composition of the proteins' exterior and interior is very similar (1.1 times more frequent) [29]. Leucine (L), isoleucine (I), and valine (V) contribute the most to the overall hydrophobicity of a protein. Their fraction in ribosomal proteins is lower than in non-ribosomal ones. The preferential decrease in K content (in contrast to R content) is dictated by the higher hydrophobicity of the K side chain than that of R [32]. Taking into account the process of ribosomal assembly [17, 33], we propose that the slight increase in the R content of ribosomal proteins across halophilic and non-halophilic species supports chain flexibility, tight binding to rRNA, and helps to prevent aggregation.
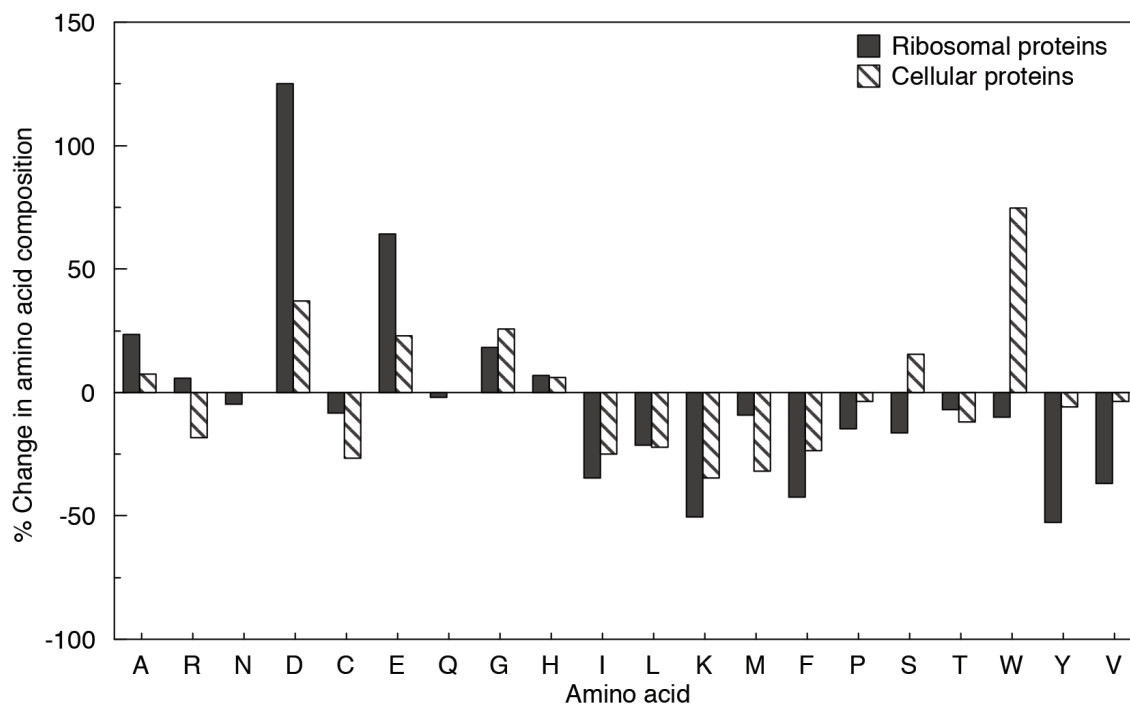
Figure 4-5. Change in amino acid composition upon transition from non-halophilic to halophilic species. Black bars: ribosomal proteins. Pattern bars: to some cellular proteins [22]. No data are available for asparagine (N) and glutamine (Q) in cellular proteins. The data were calculated as described in Materials and Methods using amino acid sequences of ribosomal proteins from seven non-halophilic and three halophilic organisms.

4.3.4 **Additional insights into the electrostatic surface potential and charge segregation of ribosomal proteins**

In general, it is not clear how ribosomal proteins interact with the highly negatively charged rRNA. Proteomes from halophilic species are highly acidic, with low pI proteins estimated to comprise more than dominating majority of the total protein content (Table 4-1) [15]. Hence, this question is particularly compelling in the case of proteins from halophiles. In halophiles, the 2:1 ratio of positively to negatively charged ribosomal proteins appears to be optimal to prevent salting-out effects [23] and preserve binding to rRNA [9]. The above question is rendered more intriguing by the fact that the rRNA's charge density is nearly identical across halophilic and non-halophilic organisms, yet rRNAs from both types of organisms are comparably effective at interacting with ribosomal proteins of widely different electrostatic properties.

A possible answer to the above dilemma is that many, if not all, ribosomal proteins may exhibit charge segregation to preserve effective binding to rRNA despite their difference in overall net charge. For instance, it is known that some ribosomal proteins of *H. marismortui* have long positively charged extensions penetrating deep into the RNA during ribosomal assembly [17, 28]. We hypothesize that charge segregation is a common feature of ribosomal proteins from all organisms.

First, to estimate the extent of charge segregation in large ribosomal subunits, we visualized six available Protein Data Bank (PDB) structures looking for distinct patches of positive and negative electrostatic potentials. The data are summarized in Table 4-3. We noticed visible electrostatic surface potential segregation in ribosomal proteins from both halophiles and non-halophiles. Figure 4-6 provides four examples of low pI and high

pI proteins in halophilic (*H. marismortui*) and non-halophilic (*E. coli*) organisms. We found that the majority (62 – 100%) of the proteins from large ribosomal subunits exhibit charge segregation. This segregation is not limited to the nonglobular extensions, as had been proposed by Klein *et al.* [17], but is also manifested in extensive negatively or positively charged patches on the surface of globular ribosomal proteins. Even though non-halophilic species do not usually have a statistically significant number of negatively charged proteins, based on the data in Table 4-3, we conclude that negatively charged proteins are more likely to exhibit charge segregation than positively charged ones.

In order to more quantitatively characterize charge segregation, we propose to evaluate electrostatic potentials (Eps) from APBS calculations and corresponding accessible surface areas (ASA) for each atom (see Materials and Methods). We selected the L3 protein from the 50S ribosomal subunit to test the concept. In *E. coli*, this protein corresponds to L3 [30], in *H. marismortui* – to L3p [31]. The choice of the protein is dictated by a couple of its properties. First, it exhibits obvious charge segregation in both species, as visualized by PyMOL (see Figure 4-6) [12, 34]. Second, it is has a negative MNC in *H. marismortui* and a positive MNC in *E. coli*. The difference in MNCs is critical to demonstrate the ubiquitous charge segregation of ribosomal proteins regardless of their pI. As part of future research, we will expand the investigation to all ribosomal proteins studied in this chapter.

We explored two characteristics of L3p and L3. First, for each of the two proteins, we calculated the fraction of positively and negatively charged surfaces that are solvent-exposed on the ribosome. Second, we visualized the charged groups of Lys, Arg, Asp,

and Glu on 3D plots and computed the length of positive and negative regions of these proteins.

4.3.4.1 Positively and negatively charged regions of L3

We identified negative and positive surface electrostatic potential (EP) values for the L3 and L3p atoms by multiplying the fraction of accessible surface area (ASA) of each atom by the corresponding EP (see Material and Methods). In order to calculate the fraction of positive and negative EP exposed to the ribosomal surface, we calculated the total surface EP of the entire L3 and L3p proteins out of the context of the ribosome (the resulting values were denoted as "isolated") and the surface EP of these proteins exposed to the ribosomal surface (called "on Rb"). We then divided the former by the latter to obtain the fraction of surface-exposed charge of each protein. All numerical values illustrating the above calculations are summarized in Table 4-4. As expected, L3 (a positively charge protein) has a smaller fraction (56%) of negative ASA, which is solvent-exposed within the ribosome, than L3p (a negatively charged protein, 61%). However, both of these proteins follow the predicted trend – i.e., they expose a larger fraction of negative EP to the surface than the positive EP.

In Table 4-4, the EP data are obtained by solving Linearized Poisson-Boltzmann Equation (LPBE). This equation is not valid for very highly charged systems like RNA and ribosomes. It is possible to solve Nonlinear Poisson-Boltzmann Equation (NLPBE), but the solution would be still incorrect because divalent cations like Mg^{2+} and putrescine^{2+} are present in high concentrations on the ribosomal surface and must be taken into account. Therefore, we decided to focus on the ASA values of the charged groups ($-COO^-$ in Asp and Glu, $-NH^+_3$ in Lys, and $-NH-C^+(NH_2)_2$ in Arg) to assess the

degree of charge segregation. Table 4-5 shows the calculated fractions of positively and negatively charged ASA solvent-exposed on the ribosomal surface for proteins from 50S subunits of *H. marismortui* and *D. radiodurans*. It is clear that the fractions of negatively charged ASA are much higher for both organisms (average values: 86% for *H. marismortui* and 84% for *D. radiodurans*) than the fraction of positively charged ASA (average values: 49% for *H. marismortui* and 57% for *D. radiodurans*). This difference is especially pronounced in proteins from halophilic *H. marismortui*. This result confirms our early hypothesis that the ribosomal proteins from halophiles exhibit a stronger degree of charge segregation than ribosomal proteins from non-halophiles.
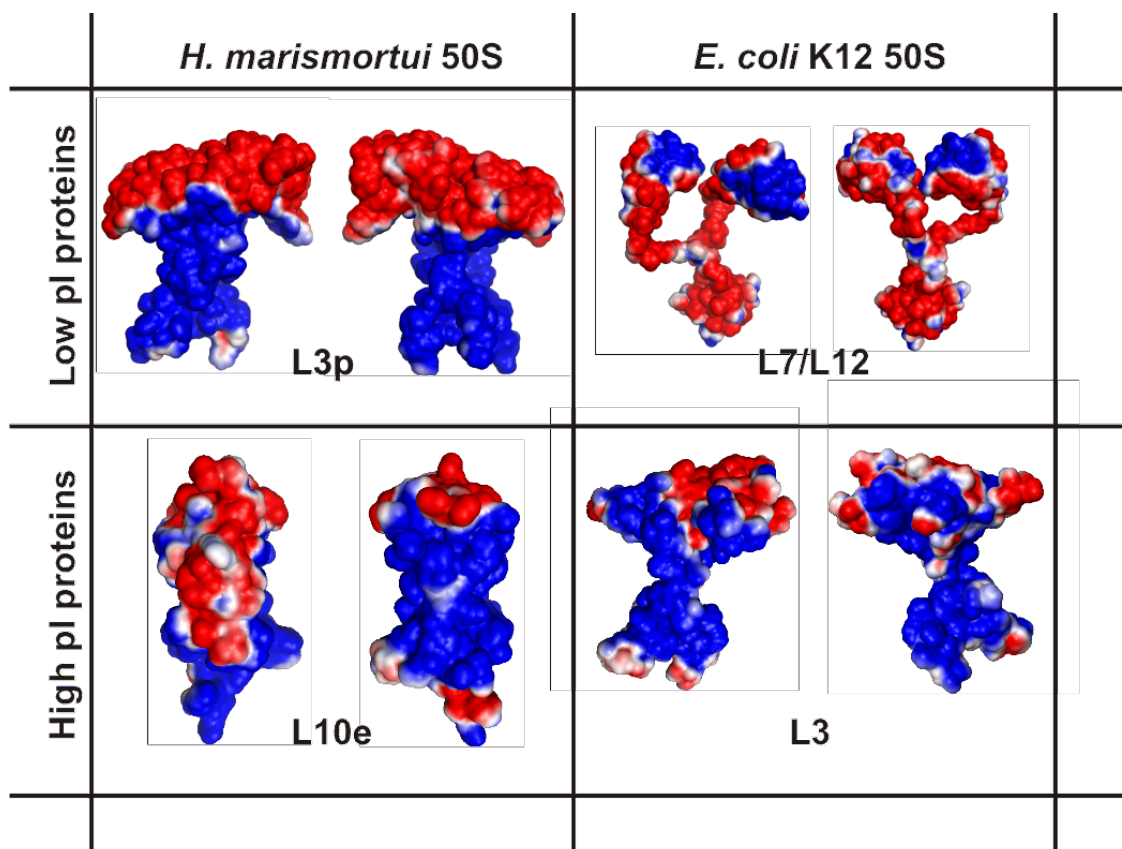
Figure 4-6. Selected images of electrostatic potential mapped on molecular surfaces of four 50S ribosomal proteins (L3p and L10e from *H. marismortui*, L7/L12 and L3 from *E. coli*). Shown are both sides of each protein. PDB structural data: L3p – chain B in 2QA4 [31], L7/L12 – 1RQU [35], L10e – chain H in 2QA4, L3 – chain D in 2AW4 [30].

Table 4-3. Fraction of proteins exhibiting charge segregation. This analysis is performed upon inspection of the electrostatic surface of 193 ribosomal proteins. Note that not all existing ribosomal proteins are present in the available crystal structures and the corresponding PDB files [36]. (*) Note the small sample size. (**) PDB files for P0, P1 alpha, and P2 beta are chains q, r, and s of 4B6A [37], respectively. Rb – ribosome.

Species	PDB file	Proteins in large Rb subunits			
		Low pI		High pI	
		Total number of proteins	Fraction of proteins exhibiting charge separation	Total number of proteins	Fraction of proteins exhibiting charge separation
Non-halophiles					
Bacteria					
E. coli	2AW4	2	100%*	27	85%
T. thermophilus	3I8I	2	100%*	29	90%
D. radiodurans	2ZJR	1	100%*	24	83%
Eukaryota					
S. cerevisiae	3U5E	5**	80%	37	70%
T. thermophila	4A1D 4A1E	0	N/A	39	62%
Halophile, Archaeal					
H. marismortui	2QA4	18	100%	9	100%

Table 4-4. Computed electrostatic parameters for the L3 (*E. coli*) and L3p (*H. marismortui*) ribosomal proteins. Symbols are defined as follows: Rb – ribosome, pI – isoelectric point, MNC – mean net charge, ASA – accessible surface area, EP – electrostatic potential.

Calculated parameters	L3		L3p	
	on Rb	isolated	on Rb	Isolated
pI	9.9		5.9	
MNC	+0.05		-0.02	
ASA of atoms with positive EP, Å^2	0.77×10^3	1.9×10^3	0.65×10^3	1.6×10^3
Total positive surface EP, kT/e	+753	+1,712	+911	+2,272
Positive EP per ASA, kT/e/Å^2	+0.98	+1.1	+1.4	+1.5
ASA of atoms with negative EP, Å^2	1.1×10^3	1.6×10^3	1.6×10^3	2.7×10^3
Total negative surface EP, kT/e	-6,836	-12,101	-12,859	-21,087
Negative EP per ASA, kT/e/Å^2	-6.3	-6.5	-8.2	-7.7
Fraction of ASA with positive EP, exposed on the ribosomal surface	44%		40%	
Fraction of ASA with negative EP, exposed on the ribosomal surface	56%		61%	

4.3.4.2   Size of the negatively charged solvent-exposed regions of ribosomal proteins

In order to provide an additional characterization of the degree of charge separation, we carried out the following preliminary work. We created 3D plots mapping the charge distribution of the L3 and L3p proteins within the ribosome and in the isolated forms (Figure 4-7). These plots focus on the surface-accessible atoms based on their coordinates. The number of the spheres per unit volume is higher in panels C and D than in panels A and B, respectively. Thus, this figure confirms that a high fraction of negatively charged Glu and Asp is exposed to the ribosomal surface. It also allows us to estimate the legths of the charged regions. We approximate that the stretch of negatively charged surface is 30 Å for L3p (*H. marismortui*) and 25 Å for L3 (*E. coli*). These are very significant values when compared with the diameter of the atoms in the charged groups, ranging from 2.8 to 4 Å. In future work we plan to more accurately assess the degree of charge segregation so that the ribosomal proteins can be classified into strongly-, poorly-, and non-charge-segregated.

Table 4-5. Fractions of negatively and positively charged accessible surface area (ASA) solvent-exposed on the ribosomal surface. Rb – ribosome.

Ribosomal proteins from 50S *D. radiodurans*	Fraction of positively charged ASA solvent-exposed on the ribosome	Fraction of negatively charged ASA solvent-exposed on the ribosome	Ribosomal proteins from 50S *H. marismortui*	Fraction of positively charged ASA solvent-exposed on the ribosome	Fraction of negatively charged ASA solvent-exposed on the ribosome
L2	58%	67%	L2p	44%	76%
L3	45%	86%	L3p	36%	86%
L4	53%	93%	L4p	34%	87%
L5	75%	86%	L5p	55%	79%
L6	75%	100%	L6p	63%	94%
L11	84%	99%	L7Ae	74%	90%
L13	52%	82%	L10e	54%	90%
L14	64%	70%	L11p	92%	96%
L15	49%	83%	L13p	49%	89%
L16	55%	60%	L14p	56%	85%
L17	43%	81%	L15e	30%	73%
L18	64%	93%	L16e	34%	86%
L19	57%	73%	L18p	39%	93%
L20	40%	60%	L19e	39%	89%
L21	37%	91%	L21e	43%	87%
L22	50%	79%	L22p	54%	92%
L23	58%	90%	L23p	46%	87%
L24	69%	100%	L24e	66%	75%
L25	74%	96%	L24p	37%	84%
L27	56%	58%	L29p	51%	89%
L28	55%	100%	L30p	45%	88%
L29	74%	85%	L31e	57%	91%
L30	64%	98%	L32e	36%	91%
L32	32%	77%	L37Ae	68%	96%
L36	41%	90%	L37e	29%	74%
Average	57%	84%	L39e	39%	52%
			L44e	60%	94%
			Average	49%	86%

Figure 4-7. 3D plots mapping the spatial distribution of the charged groups of amino acids Lys, Arg, Asp, and Glu. Red spheres: Glu and Asp, blue spheres: Lys and Arg. (A) Solvent-exposed charged groups in the isolated L3p protein (*H. marismortui*). (B) Solvent-exposed charged groups in the isolated L3 protein (*E. coli*). (C) Solvent-exposed charged groups in the L3p protein within the ribosome. (D) Solvent-exposed charged groups in the L3 protein within the ribosome. The area of the blue and red spheres is proportional to the average accessible surface area (ASA) of charged groups of given amino acids. The ratio of blue and red areas is defined as follows: 1.2:1 (A), 1.5:1 (B), 0.89:1 (C), and 1.4:1 (D).

4.3.5   How does the ribosome cope with the highly negative charge density of its surface (contributed by both rRNA and ribosomal proteins)?

The literature shows that high salt environments lead to an increase in the pKa of aspartate and glutamate from ~3 and ~4 at 5 mM NaCl to 4.9 and 5.3 at 5 M NaCl, respectively [38, 39]. Two scenarios are possible: the ribosome has a counterion layer whose features either (1) depend or (2) do not depend on the bulk salt concentration. According to scenario one, the negative charge on the ribosomal surface should be more shielded in the high-salt environment of halophiles than in the low-salt environment of non-halophiles. According to scenario two, the extent of negative charge shielding should be similar for both halophiles and non-halophiles.

The pKa is usually affected by the proximity of an acid to a charged surface. Assuming that the ribosomal surface is fully negatively charged due to the dominating presence of RNA phosphates, one can reason that the pKa of acidic and basic residues may increase due to RNA-protein proximity [40]. This prediction leads to an expected decrease of the negative charge of aspartates and glutamates and increasing positive charge of lysines and arginines.

Lastly, it was shown that the high local concentration of counterions near the polyion surface of DNA is relatively insensitive to changes in salt concentration in solution [41]. Therefore, it is possible that ribosomes from non-halophilic organisms have layers of counterions of similar composition as ribosomes form halophiles. Thus, we propose that, regardless of the bulk salt content (within values typical for halophiles and non-halophiles), the ribosome maintains its own microenvironment (i.e., a similar degree

of charge shielding) to sustain protein-RNA interactions. This view is illustrated in Figure 4-8.
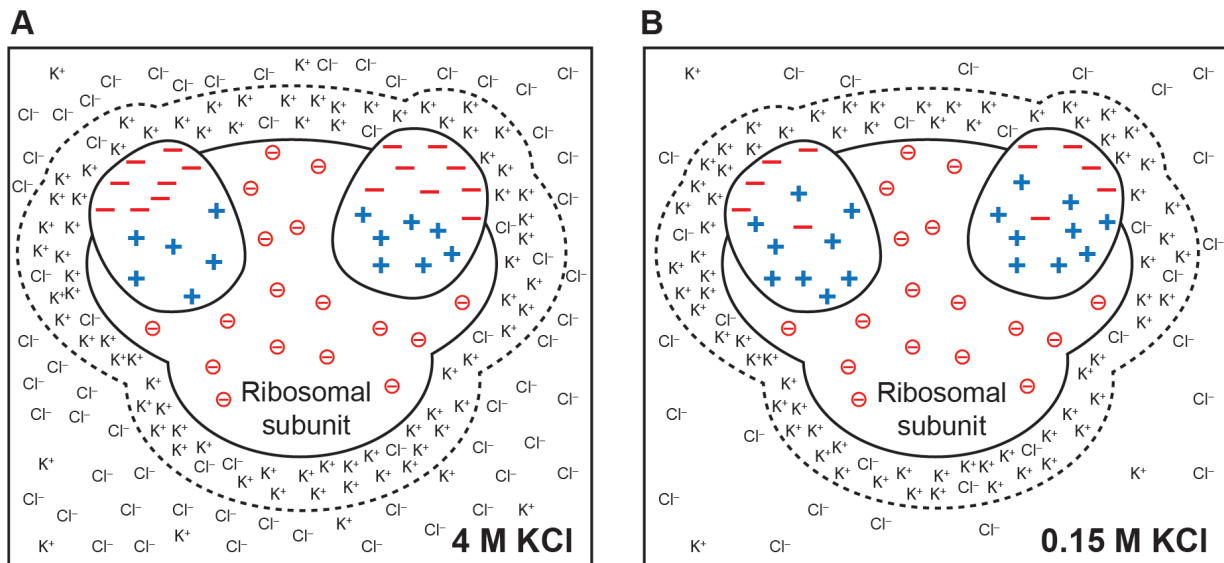
Figure 4-8. Schematic representations of ribosomal models in two different environments. (A) Ribosome from a halophile in 4 M KCl. (B) Ribosome from a non-halophile in 150 mM KCl. The dashed lines around the ribosomal surfaces enclose counterion layers and emphasize postulated similarities between these layers in halophilic and non-halophilic organisms.

In Chapter 3 of this dissertation, we discovered that 3 M guanidinium thiocyanate (GdmSCN) provides the optimal solubilizing condition for ribosomal proteins and the nascent chain under HDX quenching. No other highly charged co-solvents were successful in solubilizing the nascent chain. For example, protamine sulfate, a polycation capable of disrupting protein-DNA interactions at low pH [42], has no solubilizing effect on aggregated ribosomes at pH 2.5. Our findings in this Chapter 4 may explain these experimental observations. First, it is well known that the sulfate ion is a strong salting out agent [43], thus, protamine sulfate should contribute to precipitation and structure stabilization of ribosomal proteins and the nascent chain [44]. Second, even though protein-RNA interactions in ribosomes are primarily electrostatic, high salt concentrations are not always able to disrupt and outcompete these interactions. Proteins from halophilic organisms are able to maintain active conformations in high salt environments as long as these salts are not chaotropes [45]. Destabilization of protein conformations is achieved when chaotropes are present in solutions, with GdmSCN being the strongest one [43, 46, 47]. Third, to prevent aggregation, ribosomal proteins must be well hydrated. According to halophilic adaptation, negatively charged groups are solvent exposed to compete for water with other macromolecules and, particularly, with salt ions. Under HDX quenching conditions at pH 2.5, hydration is not efficient for the following reasons. Side chains of Glu and Asp are protonated and lose negatively charged $-COO^-$ groups. Additionally, Glu and Asp are more efficient at hydration than Lys and Arg because they attract 8 molecules of water per residue, while Lys and Arg attract only 5 and 3, respectively [20]. Therefore, at HDX quenching conditions, ribosomal proteins together with the nascent chain experience massive dehydration, and, consequently,

aggregation. GdmSCN, a strong chaotrope, is capable of salting in most of these proteins [48].

4.4 Conclusions

The findings of this chapter can be summarized as follows.

(1) Among the ribosomal proteins studied here, those from halophilic organisms are overall more negatively charged and less hydrophobic than those from non-halophiles. In other words, overall, MNC and MH values are lower for proteins from halophiles than from non-halophiles. These trends parallel the reported pI and hydrophobicity trends of the corresponding proteomes [15, 29]. Thus, we have established that ribosomal proteins follow trends similar to those of their parent organisms.

(2) Ribosomes from the halophiles studied here have no fully ordered proteins, while ribosomes from non-halophiles have 4 - 20% fully ordered proteins.

(3) Within any single organism, high pI proteins tend to be more disordered than low pI proteins. However, across organisms, 50S ribosomal proteins from halophiles are more disordered than those from non-halophiles.

(4) Regardless of the dominant fraction of high-pI proteins in the ribosomes from non-halophilic organisms studied here, the solvent-exposed surfaces of the ribosomal proteins from halophiles and non-halophiles have highly negative electrostatic surface potential.

(5) The ribosome copes with the above trends via charge segregation. The ribosomal proteins studied here have positively charged regions that interact with rRNA and

negatively charged regions that are exposed to the solvent. By electrostatic surface calculations followed by the inspection of all studied ribosomal proteins, we found that halophiles have a slightly higher fraction of charge-segregated proteins than non-halophiles. Interestingly, all studied organisms have more than half of their ribosomal proteins charge-segregated.

(6) Based on preliminary calculations of solvent-exposed ASA of proteins from *H. marismortui* and *D. radiodurans*, charge segregation in the ribosomal proteins is more pronounced in halophiles.

(7) Based on calculations from Anderson and Record [41] on DNA, it is possible that the counterion layer density on the ribosomal surface is indeed very similar across environments with very different bulk salt concentrations, i.e., across halophilic and non-halophilic organisms. This concept is consistent with recent experimental findings on nascent-chain properties from the Cavagnero group [13].

4.5 Materials and Methods

4.5.1 Organisms studied in this work

Several organisms were chosen for the analysis of ribosomal proteins. Table 4-4 below summarizes database resources used in this work. Protein sequences were obtained from the UniProt Knowledgebase (protein database) [18]. Protein structures and coordinate files were obtained from the Protein Data Bank (PDB) database [36]. We chose indicated ribosomal PDB files because they had the largest set of proteins with more than 90% atoms to allow for further electrostatic potential calculations (see below). Strains in UniProtKB were matched to the PDB files available.

4.5.2   Charge and hydrophobicity calculations

(i) Charge. We obtained pI values of each ribosomal protein using the ProtParam tool [19]. Percent of negatively charged (i.e., low pI) proteins in various proteomes were estimated based on the plots of bimodal distribution of pI [15] for all species except *E. coli*. Fraction of negatively charged proteins in *E. coli* proteome was calculated from the "pI bias" value equal to -27% [15]. The relationship between pI bias and percent of negatively charged proteins is as follows:

$$\% \ negatively \ charged \ proteins = \frac{100\% - pI \ bias}{2} \ (1)$$

A mean net charge per residue (MNC) is calculated as an average of all charges in a protein sequence. Charges of Arg, Lys, Asp, and Glu are assigned to +1, +1, -1, and -1, respectively. Charges of all other amino acid residues are considered negligible at physiological pH of 7.4.

Table 4-6. List of organisms and corresponding ribosomal data files studied in this work.

Domain		Name	Strain	UniProt	PDB files [Reference]
Bacteria		*Escherichia coli*	K12	ECOLI	50S – 2AW4 [30]; 30S – 2AVY [30]; L7/L12 – 1RQU [35]; L9 – 1DIV [49]
		Thermus thermophilus	HB8/ATCC27634/DSM579	THET8	50S – 3I8I [50] 30S – 3I8H [50]
		Deinococcus radiodurans	ATCC13939/DSM20539/JCM16871/LMG4051/NBRC15346/NCIMB9279/R1/VKMB-1422	DEIRA	50S – 2ZJR and 2ZJQ (with L7/L12) [51]
Eukaryota		*Saccharomyces cerevisiae*	ATCC204508/S288c	YEAST	60S – 3U5E (proteins) and 3U5D (RNA) [52]; 40S – 3U5C (proteins) and 3U5B (RNA) [52]; P0 – chain q in 4B6A; P1alpha – chain r in 4B6A; P2beta – chain s in 4B6A [37]
		Tetrahymena thermophila	SB210	TETTS	60S – 4A1D and 4A1E [53]; 40S – 2XZM [54]
Archaea	Other classes	*Methanothermo-bacteria thermauto-trophicus*	ATCC29096/DSM1053/JCM10044/NBRC100330/DeltaH	METTH	N/A
		Sulfolobus solfataricus	ATCC35092/DSM1617/JCM11322/P2	SULSO	N/A
	Class Halobacteria	*Haloarcula marismortui*	ATCC43049/DSM3752/JCM8966/VKMB-1809	HALMA	50S – 2QA4 [31]; 30S – N/A
		Halalkalicoccus jeotgali	DSM18796/CECT7217/JCM14584/KCTC4019/B3	HALJB	N/A
		Halophilic archaeon	DL31	9ARCH	N/A

(ii) MNC and MH Plots. According to V. Uversky *at el.* [24], proteins can be grouped into "natively folded" (NF) and "natively unfolded" (NU) by plotting the dependence of MNC of a protein on its mean hydrophobicity (MH). The line that separates NFs from NUs has the following equation:

$$|MNC| = 2.785 \times MH - 1.151 \qquad (2)$$

To allow for existence partially ordered proteins, the authors suggest adding an MH cushion of 0.045 that results in two additional equations:

$$|MNC| = 2.785 \times MH - 1.026 \qquad (3)$$

$$|MNC| = 2.785 \times MH - 1.276 \qquad (4)$$

Equation 3 separates NU proteins from partially ordered proteins and Equation 4 separates partially ordered proteins from NF proteins. Equations 3 and 4 correspond to dashed lines on all Uversky-style plots.

In order to calculate MH and MNC, I used a Python script, called ProtAnalyst, written by Rudy Clausen (Cavagnero lab, unpublished). The hydrophobicities of individual residues were taken from Kyte & Doolittle [55] (determined using water-vapor transfer free energies and the interior-exterior distribution of amino acid side chains) and normalized to a scale of 0 to 1 in our calculations. MH is defined as the sum of the normalized hydrophobicities of all residues divided by the number of residues in the polypeptide. MNC is calculated as explained above. Values for both MNC and MH are rounded to two decimal places.

4.5.3   Amino acid composition

Figure 4-3 shows change in amino acid composition. This change is calculated as follows:

$$change\ in\ AA = \frac{\%\ AA\ in\ halophilic\ species - \%\ AA\ in\ nonhalophilic\ species}{\%\ AA\ in\ nonhalophilic\ species} \times 100\% \qquad (5)$$

where *AA* is an amino acid. The calculations were performed separately for ribosomal proteins and for cellular proteins. The data for cellular proteins were taken from the work of Rao *et al.* [22]

4.5.4 Electrostatic potential calculations

PDB files were converted to PQR files using PDB2PQR software [56, 57]. PQR files have a necessary format for electrostatic potential calculations using Adaptive Poisson-Boltzmann Solver (APBS) software package developed by Baker *et al.* [12]. The calculations of ribosomal subunits and their proteins were performed by solving the Linearized Poisson-Boltzmann Equation (LPBE) [40]. The ribosome is treated as a dielectric continuum with dielectric constant 2.0 embedded in the solvent of dielectric constant 78.0. All calculations use 150 mM salt concentration of the solvent. Surfaces of all proteins and ribosomes are colored. A blue color indicates regions with positive potential values (> +1 kT/e, unless otherwise specified) and a red color indicates regions with negative potential values (< -1 kT/e, unless otherwise specified). kT is a unit of energy equal to 4.11×10^{-21} Joules at room temperature (k – Boltzmann constant in Joule/Kelvin, T – temperature in Kelvin) and e is an electric charge in Coulombs. Table 4-5 lists parameters of electrostatic calculations.

To view electrostatic surface potential, we used PyMOL visualization software [34] equipped with APBS plug-in.

Table 4-7. Parameters used in the electrostatic calculations employing the Adaptive

Poisson-Boltzmann Solver (APBS) software package.

Parameter	Value
Resolution	~ 0.5 Å
Dielectric constant of species	2.0
Dielectric constant of solvent	78.0
Ion concentration in solvent	150 mM
Number of grid points	PDB file-specific
Solvent radius	1.4 Å
Temperature	310.0 K
Vacuum sphere density (grid points per $Å^2$)	10.0

4.5.5 Charge segregation calculations

To gather numerical values for the surface electrostatic potential of each protein, we identified surface atoms in the APBS output files containing atomic Cartesian coordinates and values of electrostatic potential. Accessible surface area (ASA) of each atom in a protein was calculated using NACCESS 2.1.1 software package [58]. Simple Python codes were written to eliminate incompatibility of APBS and NACCESS output files. NACCESS's algorithm uses Lee's & Richards' "rolling a ball" approach to calculate ASA [59]. In order to calculate surface electrostatic potential of an entire molecule (a ribosomal protein or an entire 50S ribosomal subunit), we determined percent of ASA (see Equation 7) for each atom in a PDB file and multiplied it by a corresponding electrostatic potential value. Thus, buried atoms did not contribute to the value of surface electrostatic potential because their ASA is zero.

$$Total\ SA = 4 * \pi * (r_{atom} + r_{probe})^2 \ (\mathbf{6})$$

$$\%ASA = \frac{ASA(output)}{Total\ SA} * 100\% \ (\mathbf{7})$$

where radius of the "ball" is r_{probe} = 1.4 Å.

4.6 References

1. Harms, J., et al., *High Resolution Structure of the Large Ribosomal Subunit from a Mesophilic Eubacterium.* Cell, 2001. **107**(5): p. 679-688.

2. Selmer, M., et al., *Structure of the 70S Ribosome Complexed with mRNA and tRNA.* Science, 2006. **313**(5795): p. 1935-1942.

3. Yusupov, M.M., et al., *Crystal Structure of the Ribosome at 5.5A Resolution.* Science, 2001. **292**(5518): p. 883-896.

4. Ramakrishnan, V., *The Eukaryotic Ribosome.* Science, 2011. **331**(6018): p. 681-682.

5. Melnikov, S., et al., *One core, two shells: bacterial and eukaryotic ribosomes.* Nat Struct Mol Biol, 2012. **19**(6): p. 560-567.

6. Wilson, D.N., K.H. Nierhaus, and M.M. Cox, *Ribosomal Proteins in the Spotlight.* Critical Reviews in Biochemistry & Molecular Biology, 2005. **40**(5): p. 243-267.

7. Korobeinikova, A.V., M.B. Garber, and G.M. Gongadze, *Ribosomal proteins: Structure, function, and evolution.* Biochemistry (Moscow), 2012. **77**(6): p. 562-574.

8. Brodersen, D.E. and P. Nissen, *The social life of ribosomal proteins.* FEBS Journal, 2005. **272**(9): p. 2098-2108.

9. Woodson, S.A., *RNA Folding Pathways and the Self-Assembly of Ribosomes.* Accounts of Chemical Research, 2011. **44**(12): p. 1312-1319.

10. Shajani, Z., M.T. Sykes, and J.R. Williamson, *Assembly of Bacterial Ribosomes.* Annual Review of Biochemistry, 2011. **80**(1): p. 501-526.

11. Guttman, H.J., et al., *K+-Ribosome interactions determine the large enhancements of 39K NMR transverse relaxation rates in the cytoplasm of Escherichia coli K-12.* Biochemistry, 1995. **34**(4): p. 1393-1404.

12. Baker, N.A., et al., *Electrostatics of nanosystems: Application to microtubules and the ribosome.* Proceedings of the National Academy of Sciences, 2001. **98**(18): p. 10037-10041.

13. Culviner, P.H., et al., *Electrostatic Effect of the Ribosomal Surface on Nascent Polypeptide Dynamics*, in *ACS Chemical Biology*2013.

14. Lanyi, J., *Salt-dependent properties of proteins from extremely halophilic bacteria.* Bacteriological Reviews, 1974. **38**: p. 272 - 290.

15. Kiraga, J., et al., *The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms.* BMC Genomics, 2007. **8**(1): p. 163.

16. Elevi Bardavid, R. and A. Oren, *Acid-shifted isoelectric point profiles of the proteins in a hypersaline microbial mat: an adaptation to life at high salt concentrations?* Extremophiles, 2012. **16**(5): p. 787-792.

17. Klein, D.J., P.B. Moore, and T.A. Steitz, *The Roles of Ribosomal Proteins in the Structure Assembly, and Evolution of the Large Ribosomal Subunit.* Journal of Molecular Biology, 2004. **340**(1): p. 141-177.

18. The UniProt Consortium, *The Universal Protein Resource (UniProt).* Nucleic Acids Res, 2007. **35**: p. D193 - D197.

19.	Gasteiger, E., et al., *Protein Identification and Analysis Tools on the ExPASy Server*, in *The Proteomics Protocols Handbook*, J. Walker, Editor 2005, Humana Press. p. 571-607.

20.	Kuntz, I.D., *Hydration of macromolecules. III. Hydration of polypeptides.* Journal of the American Chemical Society, 1971. **93**(2): p. 514-516.

21.	Kuntz, I.D., Jr., et al., *Hydration of Macromolecules.* Science, 1969. **163**(3873): p. 1329-1331.

22.	Rao, J.K.M. and P. Argos, *Structural stability of halophilic proteins.* Biochemistry, 1981. **20**(23): p. 6536-6543.

23.	Elcock, A.H. and J.A. McCammon, *Electrostatic contributions to the stability of halophilic proteins.* Journal of Molecular Biology, 1998. **280**(4): p. 731-748.

24.	Uversky, V.N., J.R. Gillespie, and A.L. Fink, *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins: Structure, Function, and Bioinformatics, 2000. **41**(3): p. 415-427.

25.	Chen, J.W., et al., *Conservation of Intrinsic Disorder in Protein Domains and Families:‚Äâ II. Functions of Conserved Disorder.* Journal of Proteome Research, 2006. **5**(4): p. 888-898.

26.	Oldfield, C.J., et al., *Comparing and Combining Predictors of Mostly Disordered Proteins‚Ä†.* Biochemistry, 2005. **44**(6): p. 1989-2000.

27.	Sayers, E.W., et al., *Structural Preordering in the N-Terminal Region of Ribosomal Protein S4 Revealed by Heteronuclear NMR Spectroscopy‚Ä†.* Biochemistry, 2000. **39**(44): p. 13602-13613.

28. Timsit, Y., et al., *The Role of Disordered Ribosomal Protein Extensions in the Early Steps of Eubacterial 50 S Ribosomal Subunit Assembly.* International Journal of Molecular Sciences, 2009. **10**(3): p. 817-834.

29. Fukuchi, S., et al., *Unique Amino Acid Composition of Proteins in Halophilic Bacteria.* Journal of Molecular Biology, 2003. **327**(2): p. 347-357.

30. Schuwirth, B.S.F., et al., *Structures of the bacterial ribosome at 3.5 A resolution.* Science, 2005. **310**(5749): p. 827-834.

31. Kavran, J.M. and T.A. Steitz, *Structure of the Base of the L7/L12 Stalk of the Haloarcula marismortui Large Ribosomal Subunit: Analysis of L11 Movements.* Journal of Molecular Biology, 2007. **371**(4): p. 1047-1059.

32. Tadeo, X., et al., *Structural Basis for the Aminoacid Composition of Proteins from Halophilic Archea.* PLoS Biol, 2009. **7**(12): p. e1000257.

33. Burton, B., et al., *A Computational Investigation on the Connection between Dynamics Properties of Ribosomal Proteins and Ribosome Assembly.* PLoS Comput Biol, 2012. **8**(5): p. e1002530.

34. Schrodinger, LLC, *The PyMOL Molecular Graphics System, Version 1.3r1*, 2010.

35. Bocharov, E.V., et al., *From Structure and Dynamics of Protein L7/L12 to Molecular Switching in Ribosome.* Journal of Biological Chemistry, 2004. **279**(17): p. 17697-17706.

36. *Protein Data Bank*: http://www.pdb.org.

37. Greber, B.J., et al., *Cryo-EM structures of Arx1 and maturation factors Rei1 and Jjj1 bound to the 60S ribosomal subunit.* Nat Struct Mol Biol, 2012. **19**(12): p. 1228-1233.

38. Abe, Y., et al., *Effect of Salt Concentration on the pKa of Acidic Residues in Lysozyme.* Journal of Biochemistry, 1995. **118**(5): p. 946-952.

39. Nozaki, Y. and C. Tanford, *[84] Examination of titration behavior*, in *Methods in Enzymology*, C.H.W. Hirs, Editor 1967, Academic Press. p. 715-734.

40. Dill, K.A., *Molecular Driving Forces: Statistical Thermodynamics in Chemistry & Biology*2002: Routledge.

41. Anderson, C.F. and M.T. Record, *Salt-Nucleic Acid Interactions.* Annual Review of Physical Chemistry, 1995. **46**(1): p. 657-700.

42. Anton, P., J. Paul, and E.P. Peter, *Hydrogen/deuterium exchange on protein solutions containing nucleic acids: utility of protamine sulfate.* Rapid Communications in Mass Spectrometry, 2008. **22**(16): p. 2423-2428.

43. Zhang, Y. and P.S. Cremer, *Interactions between macromolecules and ions: the Hofmeister series.* Current Opinion in Chemical Biology, 2006. **10**(6): p. 658-663.

44. Timasheff, S.N., *The Control of Protein Stability and Association by Weak Interactions with Water: How Do Solvents Affect These Processes?* Annual Review of Biophysics and Biomolecular Structure, 1993. **22**(1): p. 67-97.

45. Bonnete, F., D. Madern, and G. Zaccai, *Stability against Denaturation Mechanisms in Halophilic Malate Dehydrogenase "Adapt" to Solvent Conditions.* Journal of Molecular Biology, 1994. **244**(4): p. 436-447.

46. Cacace, M.G., E.M. Landau, and J.J. Ramsden, *The Hofmeister series: salt and solvent effects on interfacial phenomena.* Quarterly Reviews of Biophysics, 1997. **30**(03): p. 241-277.

47.   Baldwin, R.L., *How Hofmeister ion interactions affect protein stability.* Biophysical Journal, 1996. **71**(4): p. 2056-2063.

48.   Pegram, L.M., et al., *Why Hofmeister effects of many salts favor protein folding but not DNA helix formation.* Proceedings of the National Academy of Sciences, 2010. **107**(17): p. 7716-7721.

49.   Hoffman, D., et al., *Crystal structure of prokaryotic ribosomal protein L9: a bilobed RNA-binding protein.* The EMBO journal, 1994. **13**(1): p. 205.

50.   Jenner, L.B., et al., *Structural aspects of messenger RNA reading frame maintenance by the ribosome.* Nat Struct Mol Biol, 2010. **17**(5): p. 555-560.

51.   Harms, J.M., et al., *Translational Regulation via L11: Molecular Switches on the Ribosome Turned On and Off by Thiostrepton and Micrococcin.* Molecular cell, 2008. **30**(1): p. 26-38.

52.   Ben-Shem, A., et al., *The structure of the eukaryotic ribosome at 3.0 A resolution.* Science, 2011. **334**(6062): p. 1524-1529.

53.   Klinge, S., et al., *Crystal Structure of the Eukaryotic 60S Ribosomal Subunit in Complex with Initiation Factor 6.* Science, 2011. **334**(6058): p. 941-948.

54.   Rabl, J., et al., *Crystal Structure of the Eukaryotic 40S Ribosomal Subunit in Complex with Initiation Factor 1.* Science, 2011. **331**(6018): p. 730-736.

55.   Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* Journal of Molecular Biology, 1982. **157**(1): p. 105-132.

56.   Dolinsky, T.J., et al., *PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations.* Nucleic Acids Research, 2007. **35**(suppl 2): p. W522-W525.

57. Dolinsky, T.J., et al., *PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations.* Nucleic Acids Research, 2004. **32**(suppl 2): p. W665-W667.

58. Hubbard, S.J. and J.M. Thornton, *NACCESS.* Computer Program, Department of Biochemistry and Molecular Biology, University College London., 1993.

59. Lee, B. and F.M. Richards, *The interpretation of protein structures: Estimation of static accessibility.* Journal of Molecular Biology, 1971. **55**(3): p. 379-IN4.

Chapter 5   Future Directions

5.1 Protein folding on the ribosome

To take the proposed in this dissertation research to the next level, several future efforts pertinent to the protein folding on the ribosome research (Chapter 3) may be focused on secondary structure analysis of the N-terminal region of the apomyoglobin ribosome-bound nascent chain (RNC). This may enable a solid understanding of how the ribosome modulates NC's secondary structure formation and may inform mechanisms of protein folding *in vivo*.

5.1.1 Secondary structure

To identify the degree of secondary structure in RNC, one may collect secondary chemical shift (SCS) data for $^{13}C^{\alpha}$ and $^{13}C(O)$ of tryptophan (Trp) residues in apomyoglobin RNC as a function of chain elongation. There are three reasons for this. First, SCS of $^{13}C^{\alpha}$ and $^{13}C(O)$ are known to be reliable reporters of secondary structure in proteins (see Chapter 2). Second, both Trp residues (Trp7 and Trp14) are present in the N-terminus of apomyoglobin (A-helical region in the natively folded protein, Figure 5-1). Therefore, one will not need to assign the ^{13}C resonances to be able to collect SCS and identify the degree of secondary structure in the A region. Third, the chain elongation approach is useful for determining when long-range interactions are contributing to the formation of secondary structure in the A region. As we have demonstrated in Chapter 2, the only interactions that induce a significant amount of secondary structure in the A region are the long-range interactions between the A and H regions (N-terminal and C-terminal regions, respectively). 1D ^{13}C NMR experiments on RNC may be carried out, using ^{13}C-Trp-labeled RNC.

5.1.1.1   Sample preparation

According to Chapter 3, the most efficient way to obtain stable and labeled RNC is to use SecM stalling and overexpression of RNC *in vivo* [1]. For the selective Trp labeling, one may use Trp-auxotrophic cell strains. These BL21(DE3)Δtrp::kan cells are widely used in the literature [2] and available in our laboratory. The RNC expression in the auxotrophic cells is carried out in minimal M9 medium supplemented with unlabeled Trp before IPTG induction. ^{13}C-labeled Trp is added concomitantly with IPTG.
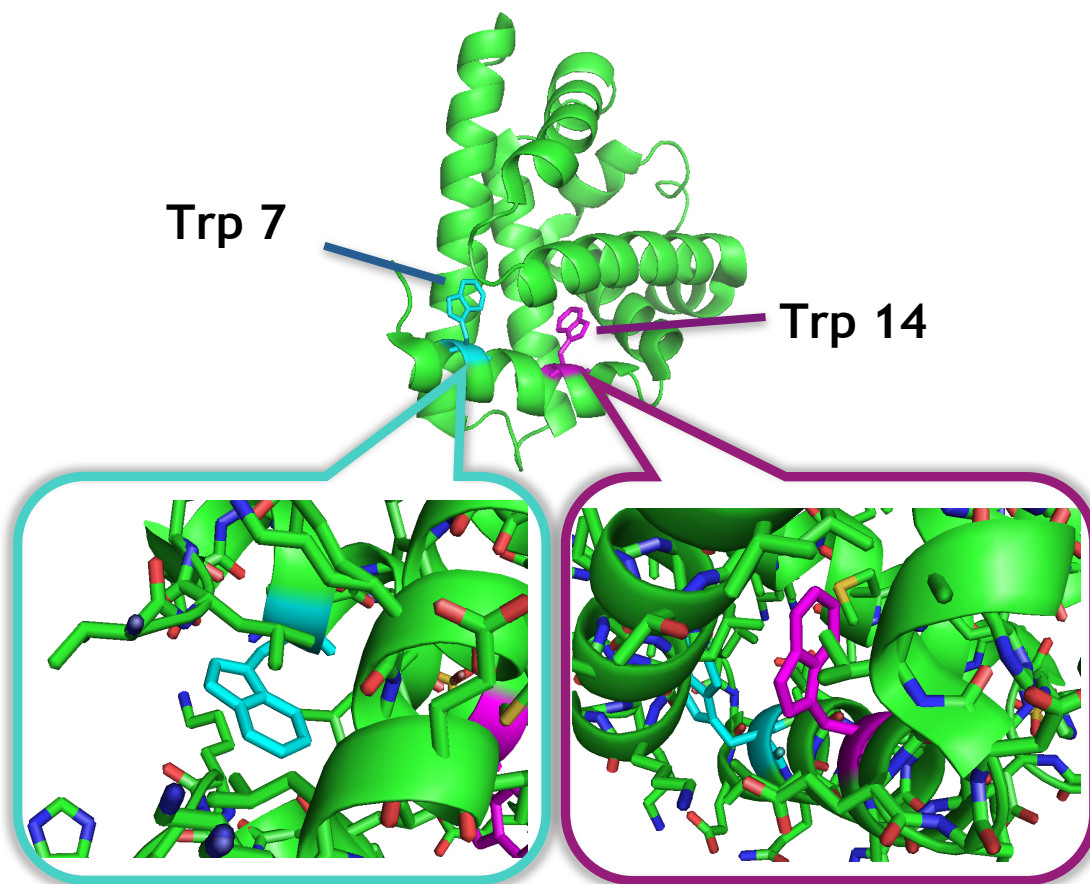
Figure 5-1. Structure of sperm whale apomyoglobin highlighting tryptophan residues. The environment of each tryptophan is shown in the zoomed areas. The Protein Data Bank (PDB) file: 2W6W [3], UniProt Knowledgebase (KB) file: P02185.

5.1.1.2 NMR Data Collection and interpretation

The data collection is designed to accommodate the following RNC features:

1. Large size of the *E. coli* ribosome, i.e. ~ 2.4 MDa;

2. Presence of 55 ribosomal proteins in the solution in addition to the nascent chain (NC) of interest;

3. Burial of ~35-40 C-terminal amino acids of NC at each chain length.

Large size of the ribosome

^{13}C NMR experiments may be set up to detect produced ^{13}C-labeled RNC (Figure 5-2). Given the low concentration of RNC (~ 10 µM), 1D spectra promise better sensitivity than 2D and 3D. ^{13}C NMR experiments allows one to collect C^{α} chemical shifts that are very selective reporters of secondary structure. Alternatively, carbonyl carbons could be labeled and detected by NMR.

Presence of 55 ribosomal proteins

Exclusive Trp labeling may help eliminate spectral interference of 55 ribosomal proteins. Trp residues are not present in the most flexible *E. coli* ribosomal protein, L7/12, known to give resonances in HSQC spectra of the ribosomes and RNCs [4]. These resonances overlap with the NC resonances and impede assignments. It is very likely that the only resonances that appear in the ^{13}C spectrum will correspond to Trp residues of apomyoglobin NC. To test this hypothesis, one needs to prepare and perform NMR experiments on the ribosomes grown under the identical conditions as ^{13}C-labeled RNC with the exception of IPTG induction.

Burial of C-terminal amino acids and chain elongation on the ribosome.

To answer the questions 1) *at what length of NC does apomyoglobin start to form the secondary structure in the A region on the ribosome?* and 2) *to what extent is this structure formed when the chain is fully synthesized?*, one may monitor the change of these Trp chemical shifts upon chain elongation. The chain elongation approach is demonstrated in Chapter 2. To generate NCs of different lengths corresponding to WT, Mu5, and Mu3 in Chapter 2, one should insert 33-amino-acid SecM fragment right after the C terminus of WT (existing plasmid, see Chapter 3), Mu5, and Mu3.
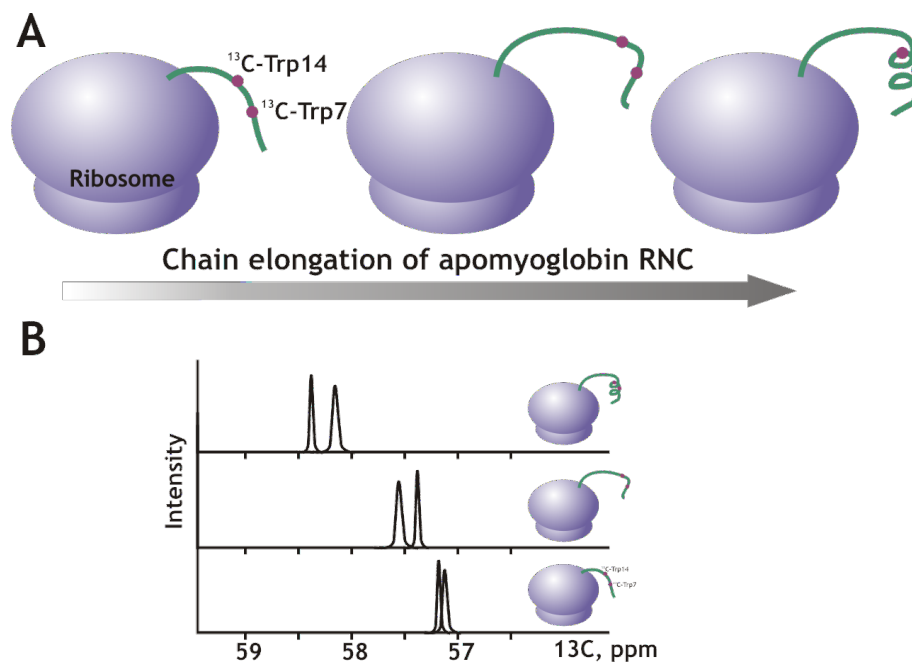
Figure 5-2. Elongation strategy coupled with ^{13}C selective labeling and ^{13}C NMR detection. (A) Chain elongation of RNC ^{13}C-labeled at Trp7 and Trp14. (B) General trend expected on 1D ^{13}C NMR spectra as RNC elongates. The ^{13}C$^{\alpha}$ chemical shifts move from 57.3 ppm (random coil) to higher ppm (within ~ 3 ppm range) as Trp7 and Trp14 attain ϕ, ψ angles corresponding to the helical structure of the A region.

5.1.1.3 Potential pitfalls and proposed solutions

The ability to detect NMR signals from RNC depends on many factors including the interactions of the NC with

1) the ribosomal surface and

2) co-translational chaperones.

First, based on the insights from Chapter 4, the surface of the ribosome is highly negatively charged. It is wise to use a nascent chain that has negative mean net charge per residue (MNC) and will repulse from the ribosomal surface. This repulsion may promise sharp resonances in NMR spectra due to faster tumbling in solution as was demonstrated in case of low pI ribosomal protein L7/12 [4]. Apomyoglobin RNCs have MNC ≤ 0 at lengths of 138 and shorter (see Figure 5-3A). The studied in Chapter 2 Mu3 (apoMb1-77) has MNC = -0.013, Mu5 (apoMb1-119) has MNC = 0.017, and WT (the full-length apoMb1-153) has MNC = 0.013. Both Mu5 and the full-length apoMb may not be amenable to NMR spectroscopy due to the potential electrostatic attraction to the ribosomal surface. Mu5 may be modified to have 126 residues to reach MNC = 0 while still lacking region H, as desired (see Chapter 2). If Mu5 and WT, nevertheless, remain unsuitable for NMR investigations, one may carry out the RNC folding study on folding-competent low-pI NC. An example of a protein-folding-relevant low pI protein is human alpha synuclein (UniProtKB 37840) with pI 4.67 (closest to 4.60 of *E. coli* ribosomal protein L7/12) and molecular weight of 14.5 kDa (similar to 17kDa for apomyoglobin, UniProtKB P02185) [5]. As shown in Figure 5-3B, synuclein has MNC ≤ 0 at the length of 115 and higher, which allows one to design several neutral or negatively charged mutants for the chain elongation approach. This protein has no Trp; therefore, it will

require Phe-auxotrophic (Phe positions are 3 and 93) and/or Tyr-auxotrophic (Tyr positions are 38, 124, 132, 135) cell strains for the NMR studies. Additionally, the biological role of synuclein in Parkinson's disease renders it a very attractive model to gain insights into protein folding/misfolding on the ribosome.

Second, the removal of the co-translational chaperones can be achieved in two ways [6]. DnaK can be removed by adding the GrpE chaperone and ATP to the RNC solution followed by dialysis. Trigger Factor can be eliminated either by using cell strains depleted from TF gene *tig* (Δ*tig* strain) or by extensive salt washes and dialysis of the RNC solution [6].
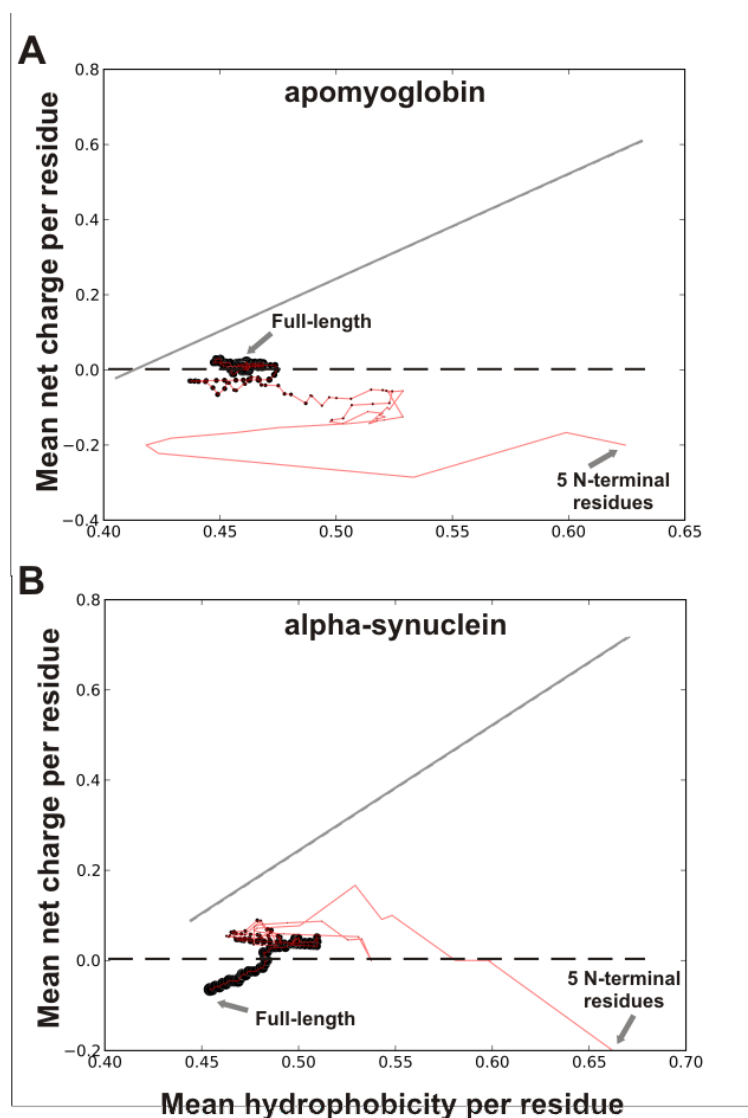
Figure 5-3. Elongation plots for sperm whale apomyoglobin (A) and human alpha-synuclein (B). The plots show how mean net charge per residue (MNC) and mean hydrophobicity per residue (MH) changes as protein chains elongate [7]. Each black dot corresponds to a new combination of MNC and MH values upon addition of one more residue to a growing chain. The solid grey line separates intrinsically disordered proteins (IDP, above the line) from ordered ones (below the line). The dashed black line separates negatively charge chains (below) from positively charged chains (above).

5.2 Electrostatics of ribosomal proteins

In an effort to prepare the findings in Chapter 4 for publishing in a peer reviewed journal, I propose to enhance the research in the following ways:

(1) Apply the calculations of fraction of Lys, Arg, Glu, and Asp that are solvent-exposed in the isolated forms of the ribosomal proteins and within the ribosome to the entire set of 193 ribosomal proteins structures studied in Chapter 4.

(2) Define the degree of charge segregation and classify 193 studied proteins according to this degree.

(3) Explore the hydrophobicity of positively charged and negatively charged regions.

5.2.1 Fraction of ASA of Lys, Arg, Glu, and Asp

I will to expand the calculations in Table 4-5 to all 193 proteins studied in Chapter 4. This will allow me to find out how many ribosomal proteins expose higher percentage of charged Lys, Arg, Asp, Glu side chains within the ribosome than in the isolated forms. Based on the findings in Chapter 4, my hypothesis is that all the ribosomal proteins that exhibit charge segregation expose higher fraction of charged groups within the ribosome than in their isolated forms.

5.2.2 Definition of charge segregation

I propose to define the degree of charge segregation (DCS) and use this parameter to classify the ribosomal proteins as strongly-, poorly-, and non-charge-segregated. First, I suggest to evaluate DCS using values of ASA for each atom in charge groups of Lys, Arg, Glu, and Asp (see Chapter 4's Materials and Methods). I propose to define DCS as follows:

$$DCS = \frac{\frac{\sum_p negative \ or \ positive \ patch \ area \ in \ Å^2}{\sum p}}{Surface \ area \ of \ the \ protein \ in \ square \ Å^2} \qquad \textbf{(1)}$$

where p – the number of patches/regions with negative or positive charges, $Å$ – Angstrom. This may allow me to identify an average surface area of each patch, to normalize it to the total surface area of each protein, and to make sure that the size of the protein is taken into account. An alternative algorithm could use a radius of gyration as a measurement of DCS.

5.2.3 Hydrophobicity of positively and negatively charged regions in ribosomal proteins

In Chapter 4, we found that, within a single organism, negatively charged proteins have higher hydrophobicity than positively charged proteins. I suggest to evaluate hydrophobicity in positively and negatively charged regions of proteins exhibiting charge segregation and classified as strongly segregated. The hypothesis being tested is that the regions corresponding to the negatively charged surface area have higher hydrophobicity than the regions corresponding to the positively charged surface area. This hypothesis is in line with the finding of Klein *et. al.* [8] that hydrophobicity is lower in the positively charged extensions and higher in the negatively charged globular regions of *H. marismortui* 50S ribosomal proteins.

5.3   References

1.   Rutkowska, A., et al., *Large-scale purification of ribosome-nascent chain complexes for biochemical and structural studies.* FEBS Letters, 2009. **583**: p. 2407-2413.

2.   Vaiphei, S.T., et al., *Use of Amino Acids as Inducers for High-Level Protein Expression in the Single-Protein Production System.* Applied and Environmental Microbiology, 2010. **76**(18): p. 6063-6068.

3.   Savino, C., et al., *Pattern of cavities in globins: The case of human hemoglobin.* Biopolymers, 2009. **91**(12): p. 1097-1107.

4.   Christodoulou, J., et al., *Heteronuclear NMR investigations of dynamic regions of intact Escherichia coli ribosomes.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(30): p. 10949-10954.

5.   The UniProt Consortium, *The Universal Protein Resource (UniProt).* Nucleic Acids Res, 2007. **35**: p. D193 - D197.

6.   Ellis, J.P., et al., *Chain Dynamics of Nascent Polypeptides Emerging from the Ribosome.* ACS Chem. Biol., 2008. **3**(9): p. 555-566.

7.   Uversky, V.N., J.R. Gillespie, and A.L. Fink, *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins: Structure, Function, and Bioinformatics, 2000. **41**(3): p. 415-427.

8.   Klein, D.J., P.B. Moore, and T.A. Steitz, *The Roles of Ribosomal Proteins in the Structure Assembly, and Evolution of the Large Ribosomal Subunit.* Journal of Molecular Biology, 2004. **340**(1): p. 141-177.