



universität
wien

DISSERTATION

Structural and Biophysical studies
of RNA-Chaperone Hfq from *E. coli*

Verfasser

Mads Beich-Frandsen

angestrebter akademischer Grad

Doktor der Naturwissenschaften (Dr. rer.nat.)

Wien, 2011

Studienkennzahl

A 091 449

Dissertationsgebiet

Pharmazie

Betreuerin / Betreuer:

Univ.-Prof. Dipl.-Ing. Dr. Kristina Djinovic-Carugo

Acknowledgements

"Wisdom begins in wonder"
Socrates

I dedicate this work to my granddad.

I need to express my gratitude to all the people who participated in the process, who supported me along the way, or were merely present.

To my mother, my father, and the rest of the family – thank you for your love, patience and support. To my dear friends Quanah & Karin, Tobias & Maria, Jan & Laura, Nelson & Tracey, and Bjørnen Schulz, thank you for your kindness and friendship.

I want to thank professors **Georg Kontaxis** and **Robert Konrat** for their invaluable enthusiasm and persisting genius, and for insisting to believe in me. The rest of the NMR-group at MFPL I thank for good collegial support – above all especially **Karin Ledolter**.

I would like to thank the head of my PhD-committee Prof. **Michael Freissmuth** for getting involved beyond the scope of his appointment, for listening, and for his wise counsel. Also, I owe thanks to the RNA-community at Max F. Perutz Laboratories, mentioning Jenny, Bob, Martina, Kasia and especially Sabine Stamfl. I would also like to express my appreciation to Prof. **Oliviero Carugo** on the notion, “..know physical chemistry – and you know everything!”.

Dr. **Quanah Hudson** I thank for numerous read-overs, and Dr. Renate Auer for inspiration. I give thanks for numerous good colleagues and friends – Ulrich, Georg, Kira, Connie, Stephanie, Florian, Martha, Dr. Jakob, Beastian, Oliver, Elisabeth, Greg and the CIBIV-gang. And Dr. **Barbara Hamilton** – we should all have one!

And finally I must thank my supervisors Prof. **Udo Bläsi** and Prof. **Kristina Djinovic-Carugo** and their respective research-groups, without whom the scope of the study would have never been conceivable.

Wien, Sept. 2011

Mads Beich-Frandsen



Curriculum Vitae

MADS BEICH-FRANDSEN

PERSONAL DATA

- Born on 17. April 1977, in Hillerød, Denmark
- Nationality: Danish

EDUCATION

- | | |
|-----------|--|
| 2005-2011 | Vienna University, Max F. Perutz Laboratories,
Dept. for Structural and Computational Biology
PhD, Biomolecular structure-/-function studies
Dissertation: "Structural and biophysical
studies of RNA-chaperone Hfq" |
| 2004-2005 | Royal Danish School of Pharmaceutical Sciences,
Institute of Medicinal Chemistry
MSc, Pharmaceutical Sciences
Thesis: "X-ray crystallographic studies of iono-
tropic glutamate receptor type 2" |
| 2003-2004 | Columbia University, New York, USA
Dept. of Biochemistry and Molecular Biophysics
Research stay in the laboratory of Prof. Eric Gouaux |
| 2002-2003 | Royal Danish School of Pharmaceutical Sciences,
Institute of Medicinal Chemistry |
| 1998-2001 | Royal Danish School of Pharmaceutical Sciences
BSc, Pharmaceutical Sciences |
| 1997 | 2. Engineer Battalion, Royal Danish Engineer Regiment
Military service |
| 1993-1996 | Skt. Knuds Gymnasium in Odense, Denmark
High-school, mathematic line w. focus on natural sciences |
| 1983-1993 | Elementary school, Giersings Realskole, Odense, Denmark |

CONFERENCE CONTRIBUTIONS

- 2009 – *Neutrons in Biology satellite meeting of the EBSA Biophysics Congress, Lund University, Sweden*. Poster: **“Structural studies in bacterial post-transcriptional regulation: RNA-chaperone Hfq”** (M. Beich-Frandsen, B. Večerek, G. Kontaxis, D. Svergun, U. Bläsi, K. Djinovic-Carugo).
- 2009 – *XIV International Conference on Small-Angle Scattering Oxford, UK*. Poster: **“Structural studies on the RNA-chaperone Hfq - a bacterial post-transcriptional regulator”** (M. Beich-Frandsen, B. Večerek, G. Kontaxis, P. Callow, D. Svergun, U. Bläsi, K. Djinovic-Carugo).
- 2011 – *1. Niels Bohr International Academy meeting on ESS science, Copenhagen, Denmark*. Poster: **“Structural Studies on the RNA-chaperone Hfq”** (M. Beich-Frandsen, B. Večerek, P. V. Konarev, L. Rajkowitsch, P. Callow, G. Kontaxis, B. A. Wallace, D. I. Svergun, R. Konrat, U. Bläsi, K. Djinovic-Carugo).

PUBLICATIONS

1. Beich-Frandsen M, Večerek B, Sjöblom B, Bläsi U, Djinović-Carugo K. **Structural analysis of full-length Hfq from *Escherichia coli***. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2011 May 1;67(Pt 5):536-40.
2. Beich-Frandsen M, Večerek B, Konarev PV, Sjöblom B, Kloiber K, Hämmerle H, Rajkowitsch L, Miles AJ, Kontaxis G, Wallace BA, Svergun DI, Konrat R, Bläsi U, Djinovic-Carugo K. **Structural insights into the dynamics and function of the C-terminus of the *E. coli* RNA chaperone Hfq**. *Nucleic Acids Res*. 2011 Jun;39(11):4900-15.
3. Večerek B, Beich-Frandsen M, Resch A, Bläsi U. **Translational activation of *rpoS* mRNA by the non-coding RNA DsrA and Hfq does not require ribosome binding**. *Nucleic Acids Res*. 2010 Mar;38(4):1284-93.
4. Beich-Frandsen M, Pickering DS, Mirza O, Johansen TN, Greenwood J, Vestergaard B, Schousboe A, Gajhede M, Liljefors T, Kastrup JS. **Structures of the ligand-binding core of iGluR2 in complex with the agonists (R)- and (S)-2-amino-3-(4-hydroxy-1,2,5-thiadiazol-3-yl)propionic acid explain their unusual equipotency**. *J Med Chem*. 2008 Mar 13;51(5):1459-63.
5. Armstrong N, Jasti J, Beich-Frandsen M, Gouaux E. **Measurement of conformational changes accompanying desensitization in an ionotropic glutamate receptor**. *Cell*. 2006 Oct 6;127(1):85-97.

Deutsche Zusammenfassung

Diese Dissertation wurde an der Universität Wien in den Max F. Perutz Laboratories von Mag. Mads Beich-Frandsen von 2005 bis 2011 durchgeführt und betreut von Prof. Kristina Djinović-Carugo vom Department für Strukturbiologie und Computational Biology in Kollaboration mit Prof. Udo Bläsi vom Department für Mikrobiologie, Immunbiologie und Genetik.

Der Titel verweist auf die strukturellen Untersuchungen am RNA-bindenden Protein Hfq aus *Escherichia coli*. Durch die Erkenntnis, dass nur ein Bruchteil des gesamten Genoms für Protein kodiert, verlagerte sich der Forschungsschwerpunkt in der Biologie hin zu RNA-basierter Regulation. Mit der wachsenden Anzahl vollständiger Transkriptomprofile kristallisiert sich das Sm-like Protein Hfq als zentrale Schaltstelle zur Genregulation durch small regulatory RNAs (sRNAs) in Bakterien heraus.

In *E. coli* und anderen gram-negativen Pathogenen ist der konservierte Sm-like Kern von Hfq um eine carboxyterminale Domäne erweitert, deren Länge 30% der Sequenz des gesamten Proteins beträgt. Seine kurze aminoternale Region ist hingegen in höherem Maße konserviert. Sowohl N- als auch C-Terminus von Hfq, beide gekennzeichnet durch intrinsisches Fehlen einer geordneten Struktur, tragen nachweislich zur Funktionalität des Proteins bei. Der Schlüssel-mechanismus der Hfq-vermittelten Regulation besteht darin, trans-kodierte sRNAs mit ihrer Ziel-mRNA zu hybridisieren. Hfq agiert somit als RNA-Chaperon, welches die Sekundärstruktur von RNA modifiziert.

Die Aufgabenstellung des Projekts bestand darin, die Funktion von Hfq aus einer strukturellen Perspektive zu beleuchten. Im Rahmen dieser Forschungsarbeit wurden Röntgenkristallographie, Kleinwinkelstreuung, NMR Spektroskopie, Zirkulardichroismus mit Synchrotronstrahlung, kombiniert mit und integriert in bioinformatische und funktionelle Studien, angewendet.

Aus dieser Arbeit gingen zwei Publikationen hervor, die strukturelle Aspekte von Hfq in *E. coli* beschreiben. Die Analyse jener Ergebnisse geschah im Kontext biophysikalischer und funktioneller Resultate, welche den intrinsisch unstrukturierten Termini von *E. coli* Hfq Funktionalität zuweisen. Es konnte festgestellt werden, dass die Termini, ausgelöst durch die Interaktion mit RNA, Strukturen ausbilden. Die Interpretation dieser Resultate folgt dem „Entropietransfermodell“, welches vorschlägt, dass intrinsisch unstrukturierte Sequenzen durch isothermische Enthalpie/Entropie-Kompensation die Entfaltung von Zielstrukturen begünstigen können.

Das Zusammenspiel von strukturierter und ungeordneter Sequenz in *E. coli* Hfq ermöglicht es diesem Protein, mit einer Vielzahl an RNAs zu interagieren und diese zu regulieren. Die zentrale Funktion von Hfq ist hierbei, die RNA in einem ungefalteten Zustand zu erhalten. Sie kann in folgendem Dogma zusammengefasst werden: *Bindung fördert Entfaltung – Entfaltung fördert Hybridisierung – Hybridisierung fördert Loslösung von Hfq!*

English Abstract

This dissertation was conducted at the University of Vienna, Max F. Perutz Laboratories in the years 2005-2011. Here is reported on research performed by Mag. Mads Beich-Frandsen, supervised by Prof. Kristina Djinovic-Carugo at the Department of Structural and Computational Biology, in collaboration with Prof. Udo Bläsi at the Department of Microbiology, Immunobiology and Genetics.

The title refers to the biostructural investigations conducted for the RNA-binding protein Hfq from *Escherichia coli*. Upon the understanding that only a fraction of a genome encodes protein, focus has been shifted to RNA-based regulation in biology. With the increasing number of transcriptome profiles being completed, the Sm-like protein Hfq emerges as the central switchboard of gene regulation, as mediated by small regulatory RNAs (sRNAs) in Bacteria.

In *E. coli* and other gram-negative pathogens, the conserved Sm-like core of Hfq is extended 30% in sequence length, by a C-terminal domain. The short N-terminal region of Hfq is conserved to higher degree. Both the N- and C-terminus of Hfq have been demonstrated of functional importance for the protein, and are characterized as intrinsically disordered. The key mechanism of Hfq-mediated regulation is by annealing trans-encoded sRNAs to target mRNA. Here Hfq acts as an RNA-chaperone, with ability to alter the secondary structure of RNA.

The scope of the project was to elucidate the function of *E. coli* Hfq from the perspective of structural biology. The research presented here employs X-ray crystallography, Small Angle Scattering, Nuclear Magnetic Resonance, Synchrotron-Radiation Circular-Dichroism, in an integrated approach with bioinformatics and functional studies.

The work resulted in two publications, reporting on structural aspects of *E. coli* Hfq. These results were analyzed in context of acquired biophysical and functional results, which annotates function to the intrinsically disordered N- and C-terminus of *E. coli* Hfq. Interaction with RNA was found to induce structure upon the termini of Hfq. This was interpreted in line of the 'entropy-transfer' model, which proposes intrinsically disordered sequence to have a function in unfolding targets by isothermal entropy/enthalpy compensation.

The interplay between the structured and disordered sequence in *E. coli* Hfq provides the protein with the ability to interact with and exert regulation on a wide variety of RNAs. Hfq functions to keep the RNA unfolded, following the dogma: *Binding promotes unfolding – unfolding promotes annealing – annealing promotes release of Hfq!*

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	I
CURRICULUM VITAE	II
GERMAN ABSTRACT	IV
ENGLISH ABSTRACT	V
1 PREFACE	1
1.1 SMALL REGULATORY RNA IN BACTERIA	1
1.2 BASE PAIRING sRNAs IN BACTERIA	3
1.2.1 CIS-ACTING sRNAs	3
1.2.2 TRANS-ACTING sRNAs	4
1.3 HFQ AS PLEIOTROPIC REGULATOR	5
1.4 HFQ AS RNA-CHAPERONE	7
1.5 RNA-BINDING TO HFQ: THE FUNCTIONAL PERSPECTIVE	8
1.6 RNA-BINDING TO HFQ: THE STRUCTURAL PERSPECTIVE	11
1.6.1 AN SM-LIKE FOLD	12
1.6.2 HFQ FORMS A COMPACT RNA-BINDING HEXAMER	13
1.6.3 RNA BINDING SITES IN <i>E. COLI</i> HFQ	14
1.6.4 FLEXIBLE TERMINI	17
1.7 AIM OF THE STUDY	18
2 STRUCTURAL AND BIOPHYSICAL STUDY OF <i>E. COLI</i> HFQ	19
SYNOPSIS	19
2.1 SUMMARY PAPER I: SOLUTION STRUCTURE OF <i>E. COLI</i> HFQ	19
2.2 SUMMARY PAPER II: CRYSTAL STRUCTURE OF <i>E. COLI</i> HFQ	20
3 SOLUTION STRUCTURE OF HFQ:RNA-COMPLEXES	21
SYNOPSIS	21
3.1 PREAMBLE	21
3.2 METHODS & MATERIALS	24
3.2.1 EXPERIMENTAL APPROACH	24
3.2.1.1 INTRODUCTION TO APPLIED SCATTERING THEORY	24
3.2.1.2 CONTRAST VARIATION EXPERIMENTS	26
3.2.2 PREPARATION OF HFQ:RNA-COMPLEX	28
3.2.3 SMALL ANGLE X-RAY SCATTERING	28
3.2.4 NMR-EXPERIMENTS	30
3.2.5 SMALL ANGLE NEUTRON SCATTERING	30
3.3 RESULTS & DISCUSSION	32
3.3.1 PREPARATION OF HFQ:RNA-COMPLEX	32
3.3.2 SHAPE-RECONSTRUCTION BY SAXS	35
3.3.3 ANALYSIS OF NMR-EXPERIMENTS SUPPORT SAXS-RESULTS	37
3.3.4 SPECULATIONS ON SMALL ANGLE NEUTRON SCATTERING	41
3.3.4.1 MOLECULAR MASS ESTIMATE	43

3.3.4.2	INTERPRETING SCATTERING CURVES USING PERCEPTUAL CRITERIA	44
3.3.4.3	SHAPE-RECONSTRUCTION ASSUMING MONODISPERSITY	48
3.3.4.4	VOLUME-RESTRAINED SHAPE-RECONSTRUCTION	49
3.3.5	SUMMARY & CONCLUDING REMARKS	51
4	STRUCTURE OF HfQ_{Ec65} COMPLEX WITH ATP	54
	SYNOPSIS	54
4.1	PREAMBLE	54
4.2	RESULTS & DISCUSSION	55
4.2.1	R-SITE COORDINATION OF ATP	55
4.2.2	COMPARISON TO COORDINATION OF RNA	57
4.3	METHODS & MATERIALS	59
4.2.3	PROTEIN EXPRESSION & PURIFICATION	59
4.2.4	CRYSTALLIZATION & DATA COLLECTION	60
4.2.5	PHASING, MODEL BUILDING & REFINEMENT	60
4.2.6	STRUCTURAL ANALYSIS AND SUPERPOSITIONS	61
5	DISCUSSION & PERSPECTIVES	62
6	BIBLIOGRAPHY	67
7	ABBREVIATIONS	79
8	PAPER I	80
8.1	SUPPLEMENTARY MATERIAL	97
8.1.1	PURIFICATION OF HfQ PROTEINS	97
8.1.2	SUPPLEMENTARY FIGURE S1	99
8.1.3	SUPPLEMENTARY FIGURE S2	100
9	PAPER II	102

1 Preface

In the central dogma of molecular biology first proposed by Crick in 1958¹, RNA is placed snug in between DNA and protein, providing this biopolymer a unique strategic position in cellular regulatory events. The regulatory systems on the level of RNA in higher eukaryotic organisms have been demonstrated to be incredibly diverse and highly developed, with multiple roles served by un-translated non-protein-coding RNAs (ncRNAs). This is contrasted by the fact that only ~1.5% of the human genome encodes protein², the remainder being a huge potential source of RNA genes. Though of a much simpler nature, bacteria have retained regulation on the level of RNA. At present stage, in *E. coli* around 80 small ncRNA have been verified as functional RNA-genes with regulatory properties, comprising so far about ~2% of all genes in this organism³.

1.1 Small regulatory RNA in Bacteria

Small regulatory RNAs (sRNAs) in Bacteria have been classified to three main groups³. RNA-molecules that regulate by interacting with specific proteins make up a diverse group of conserved sRNAs, and the recently discovered DNA-binding 'clustered regulatory interspaced short palindromic repeats' (CRISPRs) sRNAs comprise the newest group. The third group classifies sRNAs acts through a process known as riboregulation, exerting regulation through base-pairing with other RNAs. In addition to providing an energy-efficient regulation, this also achieves a high degree of specificity due to the requirement of target sequence complementarity. Fig.1.1 gives a graphic representation of the three groups and their respective functions.

Protein binding sRNAs can act by two mechanisms, either by structural mimicry of nucleic acid targets by binding to regulatory proteins and sequestering them from their function, or as an essential component for catalytic activity of an RNA-protein complex (Fig.1.1). The 6S-RNA will bind and sequester the σ^{70} -containing RNA polymerase by mimicking the open complex of promoter DNA⁴, and the regulation of *E. coli* carbon storage can be de-repressed by sequestering of the master repressor protein CsrA by the CsrB/C sRNAs⁵. The 4.5S sRNA of RNase P is the catalytic active component of this RNA-protein complex, in which a fold that is conserved through all kingdoms of life coordinates the 2'-OH scissile ribonucleotide⁶. However, protein subunits are also essential parts of RNase P, which provides possible insights to the evolution from the RNA-world⁷.

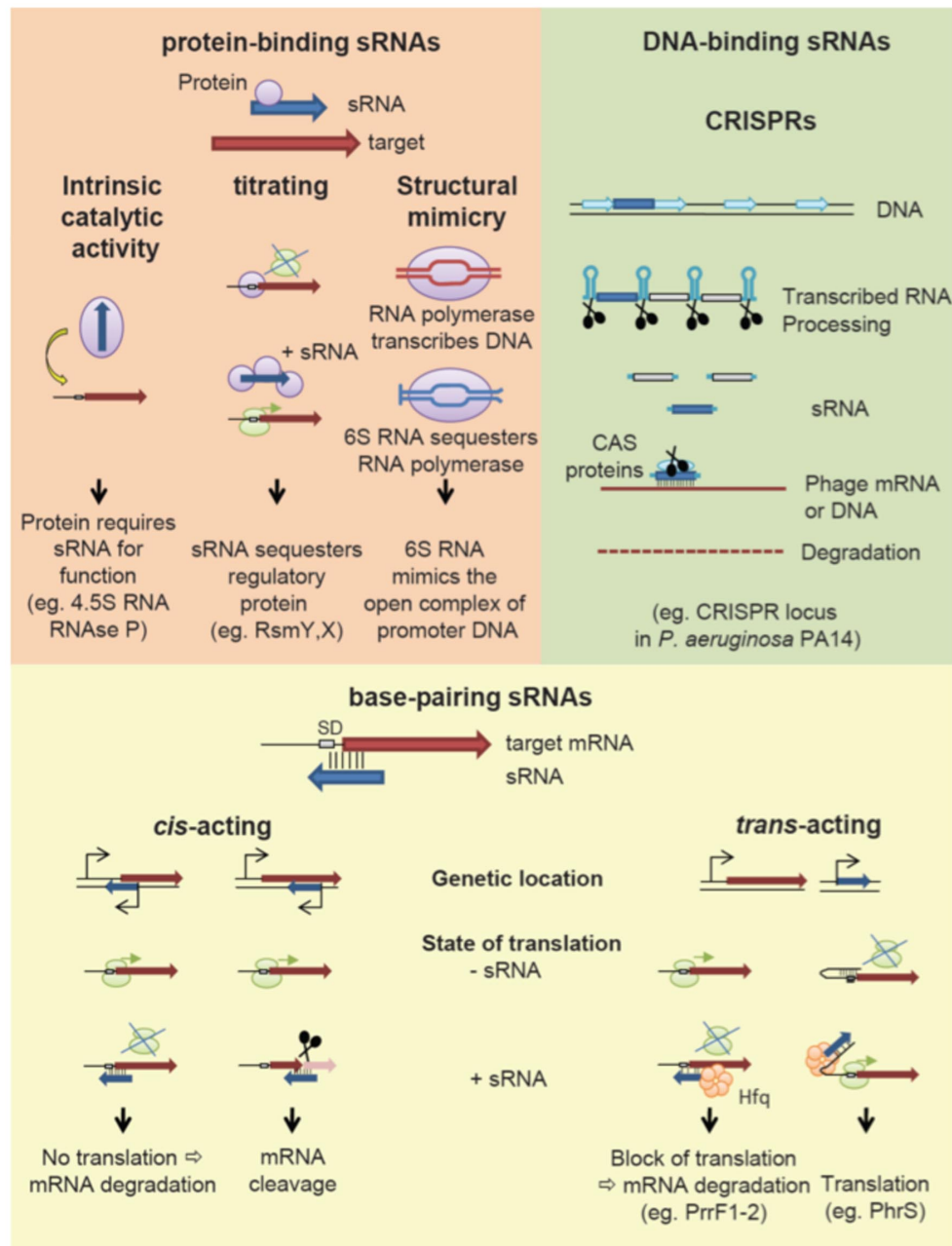


Fig.1.1: Bacterial sRNAs are classified in three groups with distinct mechanisms of function (figure adapted by courtesy of Elisabeth Sonnleitner/Udo Bläsi).

Recently it has been established that Bacteria contain an RNA-processing system that targets DNA that is mechanistically similar to the eukaryotic RNAi-pathway⁸. ‘Clustered regulatory interspaced short palindromic repeats’ (CRISPRs) sRNAs are found in many bacterial strains and most of Archea. Comprising arrays of palindromic repetitive DNA-elements interspaced by sequence homologous to DNA from phages and plasmids, the whole array is transcribed as a single long

RNA (crRNA) and processed by specific CRISPR-associated (CAS-) gene functions to produce multiple short sRNAs (Fig.1.1). These sRNAs then recruit the CAS-system to the foreign DNA by base-pairing, leading to DNA-degradation⁹. The DNA-homologous spacer sequences are heritable, and new spacers are added to the 5'-region of the CRISPs after each new infection¹⁰.

The regulatory strategies in the examples mentioned above have found to be conserved in multiple organisms (in the case of RNase P across phyla), and these mechanisms are therefore considered to be of ancient origin. The group of sRNAs acting through base-pairing is generally less conserved. However, different bacterial strains have developed similar mechanisms of riboregulation of specific events¹¹.

1.2 Base Pairing sRNAs in Bacteria

Genetic location with respect their target gene, divides bacterial riboregulatory sRNAs in two groups (Fig.1.1). Cis-acting sRNAs are encoded anti-sense to their target gene, are therefore fully complementary and their function is hence highly specific. Trans-acting sRNAs are encoded distantly from their target genes, which provide less sequence complementarity, but allow a single trans-acting sRNA to act on multiple specific targets.

1.2.1 Cis-acting sRNAs

The majority of the known cis-acting sRNAs are encoded in accessory DNA-elements like plasmids, transposons and phages. Though the known examples regulate important processes such as replication, conjugation, toxin-antitoxin modules, transposition, mRNA degradation and translation initiation, few have yet been found to be chromosomally encoded¹². This number though is expected to grow, as suggested by genomic SELEX-studies^{13,14}.

Cis-acting sRNAs are transcribed from convergent promoters, and due to their small size (50-300 nt) and high degree of homology, no complementary factors are needed to facilitate RNA-RNA interaction. The specific mechanisms by which cis-acting sRNAs function is as diverse as the regulatory events they target, and examples include transcription attenuation, translation inhibition, inhibition of primer maturation, promotion or inhibition of mRNA degradation and inhibition of RNA pseudoknot formation¹². Recent advances have brought examples of cis-acting sRNAs that overlap with the 5'- or 3'-untranslated region (utr) of target mRNAs and long (>1000 nt) sRNAs that span multiple genes¹⁵, in addition to sRNAs that facilitate premature termination of transcription of the cis-encoded mRNA target¹⁶.

1.2.2 Trans-acting sRNAs

Trans-acting sRNAs on the other hand are encoded distantly from their target gene, and therefore conserve less sequence homology. Whereas negative regulation dominates the known trans-acting sRNAs, the relatively few examples of positive regulation act mainly by a mechanism of “anti-antisense”-regulation, alleviating inhibitory structures in the target mRNA. Of particular interest are the kind of sRNAs that are able to act on several different mRNA targets, especially those capable of both positive and negative regulation, turning these into regulatory junctions of the post-transcriptional regulatory network.

Trans-acting sRNAs generally are involved in cellular stress-response¹⁷, and are particularly transcribed under conditions of nutritional stress¹⁸. *E. coli* sRNAs have been shown to be specifically up-regulated during changes in glucose concentration, elevated glucose-phosphate levels, elevated glycine levels, iron depletion, but also in oxidative stress and outer membrane stress¹⁹⁻²³.

In negative regulation, trans-acting sRNAs act by a mechanism of obstructing ribosome binding to the mRNA, as reviewed by Kaberdin and Bläsi 2006²⁴. The sRNAs recognize complementary sequence in vicinity of the ribosomal binding site (rbs), and it has been shown how very few basepairs can be required for this interaction²⁵. The sRNA prevents ribosome-loading onto the mRNA, leading to rapid degradation of the mRNA. Conversely, in positive regulation sRNAs act by making the rbs available for ribosome-loading. Mechanistically the sRNA will interact with the mRNA in an ‘anti-antisense’ manner, by basepairing with selfcomplementary regions of the mRNA 5'-utr. Hereby, the sRNA melt or compete with formation of inhibitory intramolecular secondary structures covering the rbs²⁶⁻²⁸.

Several trans-acting sRNAs have been reported to have a promiscuous behavior in their ability to interact with multiple mRNA targets, as reviewed by Papenfort and Vogel 2009²⁹. The best-studied example of this type of sRNA is DsrA from *E. coli*, a particular interesting specimen in its ability to exert translational regulation on transcriptional systems. DsrA is an 87-nt global regulator with several specific riboregulatory properties due to discreet domains of its secondary structure, which comprises three stem-loop structures termed domain 1-3 (Fig.1.2). Through ‘anti-antisense’ regulation by domain 1, DsrA up-regulates translation of the mRNA *rpoS* that encodes the stress-response RNA-polymerase sigma-factor, and genes upstream hereof. Domain 2 of DsrA binds and sequesters the *hns* mRNA through base-pairing with the 5'- and 3'-regions of this mRNA, hereby down-regulating expression of the major nucleoid structuring protein H-NS. Domain 3 comprises the transcriptional terminator of DsrA, which is not directly involved in regulation. It has been shown that domain 3 can be substituted by a

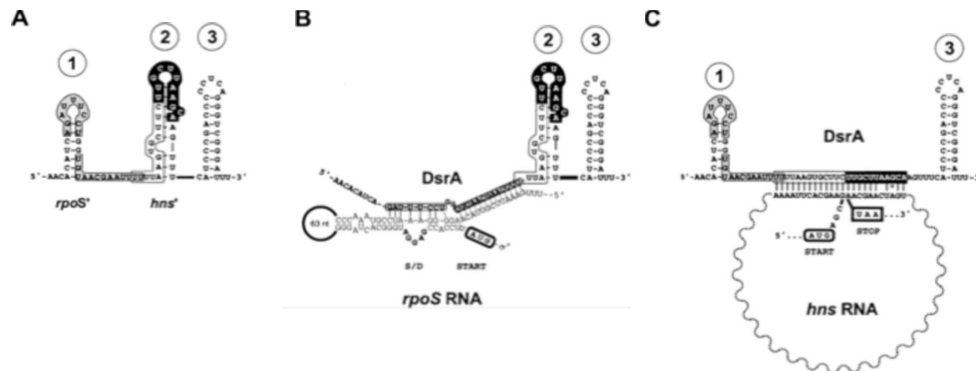


Fig.1.2: (A) Secondary structure of the trans-acting sRNA DsrA from *E. coli* comprising domain 1-3. (B) Domain 1 exerts 'anti-antisense' regulation on the *rpoS* mRNA. (C) Domain 2 regulates by sequestering the *hns* mRNA (figure adapted from Lease and Belfort 2000^{30,31}).

heterologous terminator sequence, and it is thought to be essential for proper fold of the DsrA sRNA^{30,31}.

Trans-acting sRNAs are generally characterized by non-continuous and low target sequence complementarity, a property that enables the above-mentioned promiscuity among targets. The relatively low sequence complementarity provides the binding event with an enthalpic penalty, which can be alleviated by complementary binding factors, a role that in Bacteria for the majority of the trans-acting sRNAs is carried out by the RNA-chaperone protein Hfq.

1.3 Hfq as pleiotropic regulator

When Hfq was first identified in *E. coli* by Franze de Fernandez *et al.* in 1968³², as being pre-requisite for the replication of the bacterial RNA-phage Q β (hence the name "host factor I-Q" – Hfq), the protein would co-purify with RNA in a 6:1 molar-ratio. Soon hereafter zone-sedimentation experiments by ultracentrifugation, established the protein to adopt the quaternary structure of a hexamer³³, which is the functional form of Hfq.

The protein has been established as one of the most abundant in Bacteria with some 30 - 60000 copies (5 - 10000 functional hexamers) per cell, therefore Hfq can be considered to be constitutively present in all regulatory events required³⁴. The level of Hfq in the cell under different growth conditions has been disputed, and the *hfq* gene, which is located in a tightly regulated super-operon, has been shown to be subject to both transcriptional and translational regulation³⁵⁻³⁹.

Knock-out mutants of Hfq display diverse phenotypes, placing the protein at center in bacterial homeostasis⁴⁰. Among the different observations made is decreased growth rates and yields, decreased negative supercoiling of plasmids in stationary phase, increased cell

size, osmosensitivity, increased oxidation of carbon sources and increased sensitivity to ultraviolet light. Subsequent studies have established that knock-out of the *hfq* gene effectively attenuates the virulence of several gram-negative pathogens, making the protein a potential target for antibiotics⁴¹⁻⁴³.

In seminal studies, Hfq was implicated in positive regulation of the *rpoS* gene that encodes the RNA-polymerase σ -factor under stress (σ^S), a central regulator of stationary phase and general stress response^{44,45}. Experiments suggested that Hfq most likely would alleviate an inhibitory secondary structure covering the rbs in the translation initiation region of the *rpoS* mRNA^{44,46,47}. The discovery that Hfq binds the sRNA OxyS⁴⁸, and that Hfq competes for binding to the rbs of *ompA* mRNA³⁹, suggested that Hfq is involved primarily with post-transcriptional regulation. Since then it has been established that Hfq is predominantly involved in sRNA-mediated regulation, and a recent study employing deep-sequencing analysis show Hfq to be involved with regulation of as many as 20% of the genes in *Salmonella thypimurium*⁴⁹.

The best studied Hfq-dependent sRNA-mediated regulation by far, is the riboregulation of *rpoS* by DsrA⁵⁰⁻⁵³. Many other examples have also been well characterized, and the involvement of Hfq has been characterized in a variety of systems. For example the spot42 sRNA mediated translational regulation of *galK* mRNA translation encoding galactokinase⁵⁴, or the OxyS sRNA regulation of *fhlA* mRNA the gene product of which affects more than 40 genes in response to oxidative stress⁵⁵. In response to phosphorsugar stress, silencing is enforced on *ptsG* mRNA, encoding the membrane component of the glucose-specific phosphoenolpyruvate phosphotransferase system, by the SgrS sRNA together with RNaseIII and Hfq²⁵. Similarly, the RybB sRNA promotes degradation of mRNAs encoding outer membrane porins (OMP)^{21,56}. Hfq has been shown to be involved with RyhB regulation of bacterial iron homeostasis, through silencing of the *sodB* mRNA⁵⁷, and also in regulation of ABC-transport systems, through the sRNA GcvB acting on the mRNAs *oppA* and *dppA*, which encode periplasmic components of major peptide transport systems⁵⁸. Furthermore, Hfq has been demonstrated to autoregulate repression of its own translation by sequestration of its own mRNA³⁸.

Recently it was demonstrated how Hfq can be depleted in the cell, by sequestering of the protein by induced over-expression of sRNAs⁵⁹. The finding that despite Hfq's abundance it can be in short supply for required regulatory events, leads to questions regarding the competition between RNA binding partners as a means of regulation. In addition, auto-assembly of self-associating sRNAs have also been considered as a possible method of regulation^{60,61}.

Hfq has also been shown to be binding DNA-fragments of primarily membrane proteins⁶². However, further studies needs to be undertaken to establish a specific function for Hfq in this area.

1.4 Hfq as RNA-chaperone

As coined by D. Herschlag in 1995⁶³, an RNA-chaperone is a protein that aids in RNA-folding. The original concept of a RNA-chaperone arose from studies of the HIV nucleocapsid protein and hnRNP A1 protein, both of which Herschlag *et al.* in 1994 characterized to enhance the hammerhead ribozyme reaction⁶⁴.

Similarly, Moll and co-workers used stimulation of group I intron splicing in vivo⁶⁵ and a modified in vitro toeprinting assay, to assess the RNA-chaperone activity of Hfq⁶⁶. This demonstrated that Hfq facilitates refolding of a misfolded splicing defective T4 phage *td* gene in vivo, and that Hfq-induced structural changes in *ompA* mRNA are stable after proteolytic digestion of the protein. In the original definition of the term, this classifies Hfq as a bona-fide RNA-chaperone.

Rajkowitsch and Schroeder further investigated the concept of RNA-chaperone activity in 2007, defining it as a functional property of “disrupting RNA-RNA interactions and ‘loosening’ RNA-structure”, hereby giving emphasis to the assays employed for characterization⁶⁷. In their definition, an RNA-chaperone should be able to displace double-stranded RNA, as well as promote the annealing of complementary sequence, and to do so on a wide variety of targets. In an effort to diversify the rather large group of RNA-chaperones, they introduce the concept of an ‘RNA-annealer’, to describe a protein which does not show general strand-displacement activity. In the definition of Rajkowitsch and Schroeder Hfq does not display the required strand-displacement to warrant the classification as RNA-chaperone.

However, the known biological function of Hfq is not to displace two annealed RNA-strands. On the contrary, it is to facilitate annealing of complementary sequence, which in respective RNA-molecules can be buried inside the intramolecular secondary structure. For annealing to occur, the complementary sequence of the binding site in both RNA-molecules must be exposed. As the binding site is buried in a strand folding up on itself, the ‘un-folding’ or local strand-displacement facilitated by the RNA-chaperone activity of Hfq could therefore for example be done by pulling the two separate ends of the folded RNA. This event might actively involve the corresponding regulatory or target RNA-molecule, which would then serve the role of an “RNA-co-chaperone”. In analogy, consider the definition of protein co-chaperones as exemplified by DnaJ, which apart from having its own chaperone activity is involved in recruiting the DnaK-chaperone to specific misfolded protein substrates⁶⁸. Similarly, trans-encoded sRNAs

can also anneal to their target mRNA and function independently, some do not even need Hfq to do so⁶⁹⁻⁷¹; but most - if not all, will accelerate their annealing when Hfq is recruited to the process⁷²⁻⁷⁴.

Speculation has been raised about the ATP-binding properties of Hfq, and the purpose of a suggested ATPase activity. The ATPase activity was first proposed by Sukhodolets and Garges in 2003⁷⁵ and followed up by Arluison *et al.* in 2007⁷⁶, in a study proposing ATP-binding and hydrolysis to be involved in the chaperone activity of Hfq. In the study by Arluison *et al.*, data obtained from electron microscopy (EM) experiments indicated that ATP-binding would cause a conformational change in a complex between Hfq and a heptameric A₇-RNA oligonucleotide, disrupting the hexameric symmetry of the complex. These findings could not be reproduced with non-hydrolysable ATP-analogs, linking the ATP-hydrolysis to the disruption of symmetry suggested by the EM-results. However, Hämmerle and Bläsi have studied the ATPase-activity of Hfq extensively and found the hydrolysis to be caused by impurities in the preparation of Hfq (unpublished results). In addition, Hfq was shown to function in conjunction with the ATP-dependent RNA-helicase CsdA⁷⁷. In the study by Arluison *et al.*⁷⁶ excess of ATP was also found to promote dissociation of Hfq and a longer oligo nucleotide, which incorporated a A₁₈-tail in its 3'-end that "...mimics the basic structures found in the 3'-termini of mRNAs". Competition of ATP with binding of poly-A RNA on Hfq is quite plausible given the chemical similarity of these ligands.

1.5 RNA-binding to Hfq: the functional perspective

As mentioned above, considering of the abundance of Hfq in the cell, it is reasonable to imagine Hfq to be present for all the events where it is required. Yet however, it is possible to deplete Hfq in un-physiological extreme cases. This dualism provides an insight into a careful fine-tuned network of trans-encoded post-transcriptional regulation, a network with Hfq at its center. Furthermore, the preference for binding of specific sRNAs has recently been implicated as a possible hierarchical approach to riboregulation⁷⁸. One could think of Hfq as the major transcript structuring protein, being bound on either mRNAs or sRNAs and acting as a "molecular force-field", sampling for a suitable binding partner to induce a regulatory event. On occurrence of a suitable match, Hfq will dissociate from the RNA-duplex to be recycled in a new regulatory event⁷⁹.

To add to the complexity, Hfq has also been implicated in both protection against and stimulation of RNA degradation. Hfq was demonstrated to turn poly(A)-polymerase I (PAP-I) into a processive

enzyme, hereby enhancing poly-adenylation and increasing the length of poly(A) tails. The length of poly(A) tails of bacterial mRNAs is thought to result from competition between PAP-I and exoribonucleases like polynucleotide phosphorylase (PNPase) and RNase II, which target the 3' ends of RNAs⁸⁰⁻⁸². Both Hfq and PNPase were shown to co-purify with His-tagged PAP-I under native conditions, indicating a potential complex among these proteins that was confirmed by immunoprecipitation experiments⁸³.

In *E. coli*, Hfq has been found in complex with a wide variety of RNA-binding proteins, implying a cooperative function for Hfq in these systems. Among the protein binding partners are RNA polymerase, ribosomal protein S1⁷⁵, RNase E⁸⁴ and RNA-helicase CsdA⁸⁵. The C-terminus of *E. coli* Hfq is highly enriched in serine residues, a feature also found in intrinsically disordered hub-proteins that are central to eukaryotic interactomes⁸⁶, and it has therefore been speculated that Hfq also serves to assemble and organize a variety of multi-component RNA-targeting protein complexes. However, focused studies could not find direct protein-protein interactions between Hfq and RNase E⁸⁷, ribosomal protein S1⁸⁸ or CsdA⁷⁷, leading to the likely conclusion that the observed complexes are RNA-mediated. Hfq has also been shown to protect against endonucleolytic RNase E mediated RNA degradation, in that the Hfq binding site coincides with the AU-rich recognition site of RNase E. This was shown specifically for the two sRNAs RyhB and DsrA, and is likely valid for all Hfq-dependent uridine-rich regulatory RNAs^{89,90}.

In a study by Møller *et al.* from 2002⁹¹, the binding site for Hfq on the spot42 sRNA was probed by hydroxyl radical foot-printing, which confirmed the previous findings that Hfq, similar to the Sm-proteins of higher organisms, shows preference for binding to uridine-rich sequence stretches. Brescia and co-workers studied the interaction between Hfq and DsrA in detail in 2003⁹², and showed that Hfq recognizes U-stretches adjacent to stem-loop structures. The major-groove of RNA stem-loops is known to coordinate α -helices⁹³. In 2004 Geissmann and Touati found interaction with Hfq to specifically require a stem-loop structure succeeding an A/U-rich sequence, and to subsequently partly open the structure to allow for binding of the sRNA RyhB, which provided a novel functional mechanism for Hfq⁹⁴.

A second binding site, specific for poly(A)-sequence, was characterized by Mickulecky *et al.* in 2004⁹⁵. Mutations located at distance to the uridine-specific binding site reduced the affinity for and translation of *rpoS* mRNA, and established a model whereby two binding sites and dual binding of sRNA and target mRNA is required for the function of Hfq.

In a study by Sonnleitner and co-workers in 2004, site-directed mutagenesis showed how a single mutation of Lys-3 in the N-terminus of *E. coli* Hfq is detrimental for the function of the protein, and also that the C-terminus stabilizes DsrA sRNA when bound to Hfq⁹⁶. In *E. coli* Hfq both the N- and C-terminal regions are intrinsically disordered, which is known to be an inherent property of RNA-chaperone proteins, and probably related to refolding initiation by entropy-transfer^{97,98}. In 2008, Vecerek *et al.* published studies, indicating that the C-terminal domain of Hfq contains a previously unrecognized binding site with specificity for mRNA⁹⁹.

The mechanism of Hfq's role in positive riboregulation has been the subject of extensive studies by Soper and Woodson, investigating Hfq's involvement in the interaction between sRNA DsrA and mRNA target *rpoS*. In a 2008 study⁵³ Hfq was demonstrated to bind both RNAs simultaneously and hereby increase their local concentration. Hfq interacts with and alleviates an inhibitory structure in the *rpoS* mRNA 5'-leader, and subsequently promotes annealing of DsrA. The Hfq binding site in DsrA coincides with the *rpoS* complementary sequence and contains a uridine-rich stretch, providing DsrA to interact with Hfq according to this specificity. The Hfq binding site in *rpoS* was mapped by mutational analysis to an adenine-rich stretch upstream of the DsrA complementary region, which is then expected to interact with the poly(A)-specific binding site on Hfq, characterized by Mickulecky *et al.*⁹⁵. Upon annealing the RNA-duplex dissociates from Hfq, although Hfq apparently remains bound to the mRNA. In a follow-up study from 2010¹⁰⁰ these observations were found to apply to three separate instances of positive sRNA mediated regulation of *rpoS*. These findings gives perspective to recent results, which suggest how competition among different sRNAs for binding on Hfq could be involved in the riboregulatory network⁷⁸.

In 2011, two studies both applying fluorescence based methods have been published on the mechanistic aspects of Hfq's interaction in DsrA:*rpoS*-duplex formation. Arluison and co-workers⁵² for the first time applied single-molecule fluorescence energy transfer (smFRET) to study the function of Hfq in real-time, whereas Woodson and co-workers⁵¹ applied stopped-flow fluorescence assays to measure kinetics of the individual steps in Hfq-mediated duplex formation between DsrA and *rpoS*.

The RNA-substrates used in both of these studies are very short (16 and 37 nucleotides), and none of the *rpoS*-fragments include the upstream A-rich region mentioned above. Interestingly, Arluison and co-workers found that the DsrA and *rpoS* RNA-fragments, which both include a U-rich region natively, compete for the U-specific bindingsite on Hfq. In

the study by Woodson and co-workers a RNA molecular-beacon emulates the regulatory strand of *rpoS*, and is therefore complementary to the utilized DsrA-fragment, and hence carrying an A-rich region. Here no competition was found, and taken together with results from mutation studies, the authors concluded that the two RNA-fragments simultaneously and independently interact with the two established, and respectively A- and U-specific, binding sites.

The dualisms in these findings signify the recurring paradox in the pleiotropic function of Hfq. In the recent fluorescence study by Woodson *et al.* where it was reported how the Y25D mutation in the poly-A bindingsite, not alone effect the binding of RNA, but also the annealing properties of *E. coli* Hfq. The oligomeric nature of Hfq means a single point-mutation changes six aminoacids in the functional hexamer, effectively changing Hfq's overall properties and possibly the interaction with RNA. This makes it virtually impossible to achieve results on single subunit resolution level from such studies. However, these fluorescence studies (and also other mutation studies) provide readily analyzable data. Despite the at times contradictory conclusions, these results can serve to provide circumstantial evidence for the molecular function of *E. coli* Hfq. In this regard, it could be interesting to see the result of these two fluorescence studies combined, in a study employing smFRET with *E. coli* Hfq labeled with specific fluorophores in one, maybe two monomers. From such a study one might be able to deduce how the two RNAs interact with *E. coli* Hfq, in what order and binding site, and if there is any cooperativity between the A- and U-specific binding sites. A number of crystal structures of Hfq have been solved, some in complex with RNA, and it would be very informative to see the crystal structure of a ternary complex, with both the A- and U-specific binding sites in *E. coli* Hfq occupied simultaneously.

1.6 RNA-binding to Hfq: the structural perspective

At time of writing, 15 crystal structures of Hfq were available from the Protein Data Bank (www.rcsb.org), with 10 more waiting to be released. The available structures represent Hfq homologs from eight different species, with four comprising complexes between Hfq and RNA oligo nucleotides.

A general dogma of Hfq functionality was initially based on cryo EM-studies¹⁰¹ and subsequent crystal structure of the human spliceosomal U1-snRNP¹⁰², demonstrating U1 RNA to protrude from the central pore of a human Sm-protein heptamer. Evidence for a similar function in Hfq has yet to be established.

The structures of Hfq all reveal the protein to oligomerize as a hexamer, with an overall shape of a doughnut. The Hfq subunits adopt a fold in

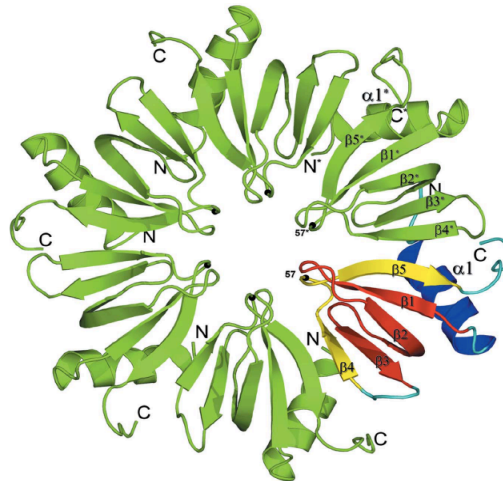


Fig.1.3: The hexameric oligomerization in Hfq forms a circular and continuous 30-stranded β -sheet, which in part explains the high thermostability of this protein. Here the Hfq hexamer from *Pseudomonas Aeruginosa* is shown from the distal face, in center the location of His-57 is noted by black dots. The secondary structure motifs of the Sm-like fold are noted, blue: N-terminal helix, red: Sm1-motif, yellow: Sm2-motif (figure adapted from¹⁰³).

close resemblance to the classical (oligonucleotide/oligosaccharide binding) OB-fold; a five-stranded anti-parallel bent β -sheet forming an SH3-like β -barrel, capped by an N-terminal α -helix. By conserved backbone interactions between adjacent SH3-like β -barrel domains, a circular 30-stranded β -sheet is formed, which in part can explain the high thermostability of this protein (Fig.1.3).

1.6.1 An Sm-like fold

In a difference to eukaryotic and archaeal Sm/Lsm-proteins that function as hetero heptamers¹⁰⁴, Hfq adopts the quaternary structure of a homo hexamer^{55,91}. Lacking some distinct Sm-protein features Hfq is classified as an Sm-like protein, belonging to the Sm-protein superfamily. Sm-proteins conserve the Sm1- and Sm2-motifs comprising hydrophobic residues that maintain the core of the Sm-fold and highly conserved residues that bind RNA. Clear homology can be found for the Sm1-motif in Hfq, whereas identification of the Sm2-motif is less obvious, and is the reason why Hfq initially evaded bioinformatic classification (see Fig.1.3). A hallmark difference of the Sm2-motif in Hfq, is the YKHA-segment that make up loop-5 between the β -strands 4 and 5¹⁰⁵. Despite sequence variability (-RGXX- in the eukaryotic Sm-proteins) this loop maintains its topology, likely due to its conserved involvement in RNA-binding¹⁰⁶. In addition, the Hfq-family is characterized by a slightly longer N-terminal α -helix and an almost complete absence of the 'variable region' that links the Sm1- and Sm2-motifs. Two residues extend the N-terminal α -helix, a feature that is only found in one of eight human Sm-proteins, the human Sm-D2. The

eight variants of the human Sm-protein, which dynamically form two compositions of hetero-oligomers, are characterized by their uniqueness in the 'variable region'¹⁰⁷.

1.6.2 Hfq forms a compact RNA-binding hexamer

The hexameric oligomerization of Hfq has been suggested to be possible due to the absence of the 'variable region'¹⁰⁸. This, in combination with the YKHA-segment of the Sm2-motif that conserves an inter-subunit hydrogen bond between the imidazole-ring of His-57 and carbonyl-oxygen of Ile-59' of the adjacent subunit¹⁰⁵, allows the subunits to pack closer compared to Sm-proteins, giving preference to a hexameric organization for Hfq¹⁰⁹. The quaternary structure is stabilized by a conserved hydrogen bonding pattern between β -strand 4 and β -strand 5' of the adjacent monomer, further supported by a hydrophobic interface between the N-terminal α -helix and loop 3' and between β -strand 2' and 3' of the adjacent monomer.

In *Staphylococcus aureus* Hfq (Hfq_{Sau}) an additional H-bond between Tyr-56 and Tyr-63 is found to stabilize the oligomer. In other common bacteria like *E. coli*, Tyr-56 (Tyr-55 in *E. coli*) is unpaired and thought to increase the RNA-binding properties of Hfq¹⁰⁵.

The importance of His-57 for stabilization of the oligomer was investigated by mutational studies, which indicated that the H-bond donor properties of His-57 could be mediated by water. In an H57T-mutant of *Pseudomonas aeruginosa* Hfq (Hfq_{pae}), lower stability of the oligomer was measured by thermal-melting, and the oligomer was further destabilized in a H57A-mutant, where no water-mediated H-bond was possible¹⁰³.

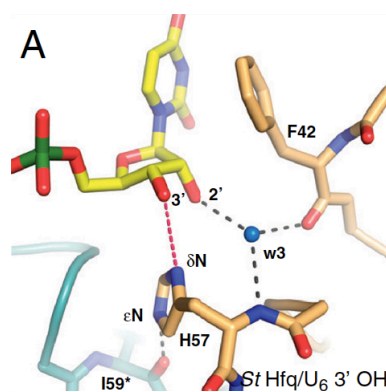


Fig.1.4: The residue His-57 is completely conserved in the Hfq-family, and has been implicated in RNA-binding. The N_ε of His-57 H-bonds to Ile-59', making the N_δ of His-57 more prone to act as a H-bond acceptor, here shown to specifically coordinate the 3'-OH of a terminal uridine (figure adapted from ¹¹⁰).

In the most recent structure of *Salmonella typhimurium* Hfq (Hfq_{St}) in complex with RNA, the highly conserved His-57 of the YKHA-segment was identified as implicated in recognition of the terminal uridine 3'-OH of Rho-independent transcripts, as an effect of its importance for oligomerization¹¹⁰. As can be seen in Fig.1.4, the interaction between His-57 and Ile-59' stabilizes the imidazole ring of His-57 in a orientation suitable for H-bonding to the 3'-OH. The N_ε of the imidazole ring H-bonds to the Ile-59' carbonyl oxygen of the adjacent Hfq monomer, making the N_δ of His-57 more prone to act as a H-bond acceptor, as a consequence of the dielectric properties. . This feature is completely conserved throughout the Hfq-family, and gives emphasis to the importance of ring-formation for the function of Hfq, other LSm-proteins and possibly the dynamics of Sm-proteins.

1.6.3 RNA binding sites in *E. coli* Hfq

In Hfq two specific RNA-binding sites have been characterized structurally, localizing respectively to each side of the doughnut hexamer⁷³ (Fig.1.6A), and both formed by residues from adjacent subunits.

In accordance with the recognition that Hfq shows preference for binding of U-rich RNA-sequence, Hfq_{Sau} was first crystallized in complex with a AUUUUUG-oligo nucleotide, which was found to bind in a conserved A/U-specific binding site on what came to be known as the proximal face of the doughnut¹⁰⁸ (Fig.1.5A+C-D). Located on the rim of the central pore, the binding site is comprised primarily of residues from the loops between β -strand 2 and 3, and β -strand 4 and 5. Discrimination towards A/U-nucleotides is governed by the completely conserved residue Gln-8 in the α -helix (Fig.1.5B), which form H-bonds with the uracil bases. A hallmark feature of the proximal binding site is the π -stacking interaction, which coordinates the uracil bases in a π -sandwich between Phe-42 (as numbered in Hfq_{St}) from adjacent Hfq-subunits (Fig.1.5B). The phosphate backbone of the RNA rests on a circular positively charged patch formed by His-58 (as numbered in Hfq_{Sau}, His-57 in Hfq_{St}) of the YKHA-segment (Fig.1.5E+F).

The proximal binding site has recently been suggested to also specifically recognize 3'-uridines present in the end of Rho-independent transcripts, as described above. The high-resolution crystal structures of Hfq_{St} and its complex with an U6-oligo nucleotide, was correlated with isothermal calorimetric (ITC) measurements of the proteins affinity for a variety of RNA-substrates, and compared with the original Hfq_{Sau}:RNA complex structure.

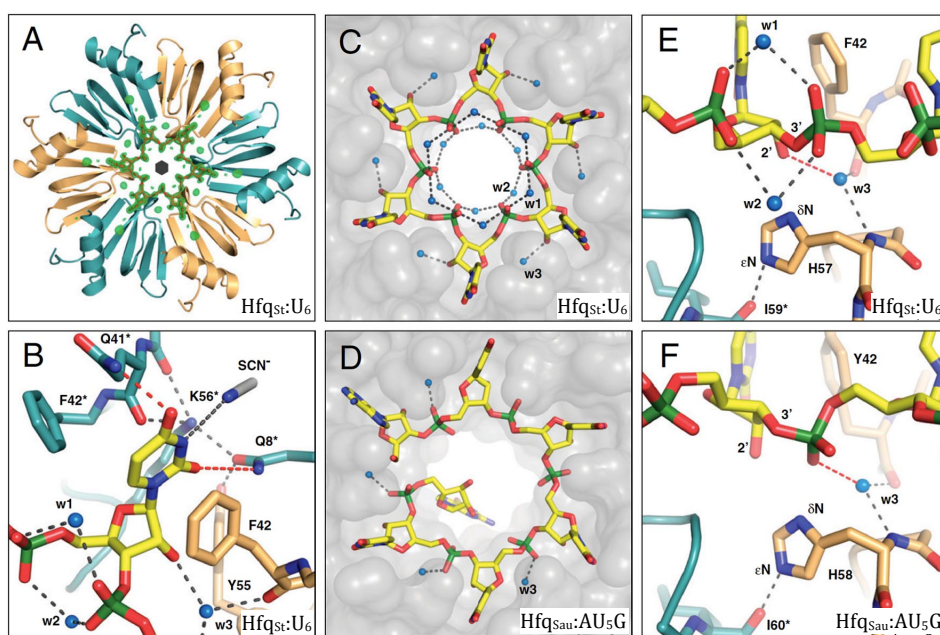


Fig.1.5: The proximal A/U-specific binding site of Hfq_{St} (A, B, C, E) is compared to that of Hfq_{Sau} (D+F). **(A)** Clear density for an U6-oligo located in the pore on proximal face of Hfq_{St}. **(B)** Close-up of a single uridine coordination in complex with Hfq_{St}, which was crystallized in presence of thiocyanate (SCN⁻). A similar coordination is found in Hfq_{Sau}. **(C+D)** The difference between the 'constricted' and 'dilated' modes of binding, as depicted by the structures of Hfq_{St}:U₆ and Hfq_{Sau}:AU₅G. **(E+F)** The difference in phosphate-backbone coordination and consequently conformation leads to either the 'constricted' or 'dilated' mode of binding (figure adapted from ¹¹⁰).

This led to the conclusion that there exist at least two modes of interaction for the proximal U-specific binding site. The 'constricted mode' (Fig.1.5C) that binds tight and possibly coordinates the 3'-end of the RNA, and the 'dilated mode' (Fig.1.5D) that bind the RNA less specifically, but possibly allows for interaction with the internal sequence of RNA-molecules¹¹⁰.

In addition Hfq's involvement in 3'-adenylation of mRNA, and specific binding of poly(A)-RNA has been the subject of structural studies. The crystal structure of Hfq from *Escherichia coli* Hfq_{Ec} in complex with A₁₅-oligo nucleotide RNA revealed a hitherto unknown 'tri-partite' mode of RNA-protein interaction. Located on the opposite, distal face of Hfq, the RNA is recognized in A-R-N-triplets (A=adenine, R=purine, N=any nucleotide), in a purine-specific crevice (the R-site) formed between adjacent monomers (Fig.1.6A+B). The specific recognition also incorporates an adenine recognition-site, which is comprised by residues from a single monomer alone (the A-site)¹¹¹. In the complex between *E. coli* Hfq and poly(A)₁₅-oligo nucleotide, the RNA weaves back-and-forth as it traverses the distal face of the hexamer (Fig.1.6C). The same distal binding site has since been demonstrated to bind an

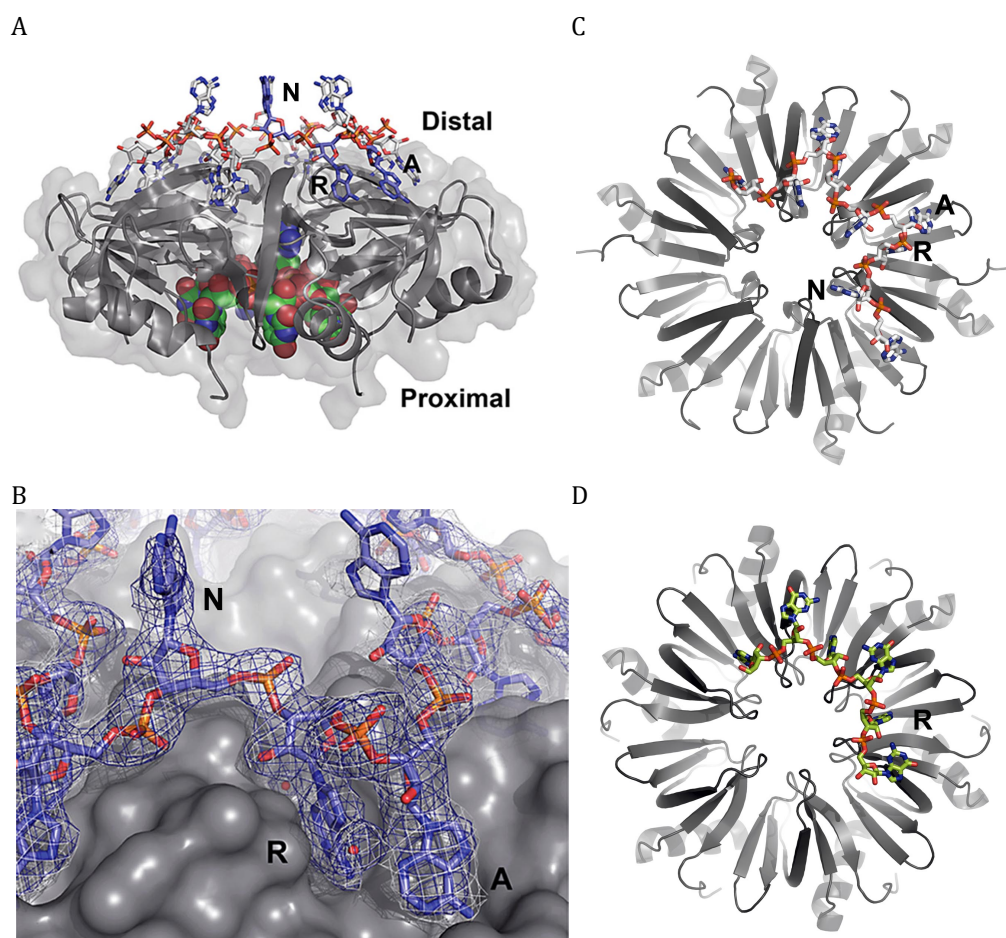


Fig.1.6: **(A)** The distal tri-partite A-R-N binding site on Hfq_{Ec} (in grey cartoon) with the poly(A)₁₅-RNA (upper, in CPK) located opposite the proximal Sm-site (lower, RNA modeled in CPK). Location of the three parts are noted by their specificity, A=adenine, R=purine, N=Any nucleotide. **(B)** Close-up of the A-R-N sites with composite omit-maps $F_o - F_c$ density contoured at 0.8σ (blue), and $2F_o - F_c$ at 1.0σ (white). Protein shown as grey surface. **(C+D)** Comparison between the modes of RNA-interaction at the distal binding site in (C) Hfq_{Ec}:poly(A)₁₅ (PDBid:3GIB) and (D) Hfq_{Bac}:AGAGAGA (PDBid:3AHU). Figures was adopted from¹¹¹, or made with MacPymol¹¹³.

AGAGAGA-oligo nucleotide. In the crystal structure of Hfq from *Bacillus subtilis* (Hfq_{Bac}) in complex with the oligo nucleotide 5'-AGAGAGA-3'¹¹², the RNA traverses the distal face of Hfq in more direct line (Fig.1.6D), with every second nucleotide inserted in the purine-specific crevice. The structure provides new insights to Hfq's ability to interact with RNA, and demonstrates how the purine-specific crevice shows preference for binding adenine above guanine¹¹². Common to both modes of binding is the 2'-OH coordination with the carboxyl-group of the completely conserved Gly-29 that is thought to select RNA above DNA, in addition to the π -stacking interaction with Tyr-25 upon purine-insertion into the crevice.

1.6.4 Flexible termini

The C-terminal domain of *E. coli* Hfq has been proposed as an additional RNA binding site, with specificity for mRNA⁹⁹. Question remains however, to what extent the distinction between mRNA and sRNA molecules is conceivable on the level of chemical biology of interaction between protein and RNA. In addition, the C-terminal domain of *E. coli* Hfq has recently also been implicated in the binding of DNA⁶².

Only the β - and γ -proteobacteria carry Hfq proteins with an extended C-terminal domain¹⁰⁶, which Jousselin *et al.* suggest involved in bacterial riboregulation¹¹⁴. The functional importance of the termini of Hfq was first demonstrated by mutation and truncation studies, showing that a single lysine mutation (Lys-3) in the N-terminus of *E. coli* Hfq impairs the protein in complementing replication of the RNA-phage Q/ β ⁹⁶. Both the N- and the C-termini of Hfq have been characterized as intrinsically disordered⁹⁹, with the N-terminus of *E. coli* Hfq conserved to a higher degree than the C-terminus.

None of the published crystal structures of Hfq, account for the full sequence of the Hfq protein. Only an EM-study by Arluison *et al.* provides information on the location of the C-termini, correlating the crystal structure of C-terminally truncated *E. coli* Hfq with EM-micrographs¹¹⁵. This indicated that the C-termini reside on the proximal face of the doughnut, on the interface between monomers, inferring a structural rearrangement on the Sm2-motif of Hfq. The study also showed the C-termini to increase the stability of the hexamer, and employed Fourier-transformed infrared spectroscopy (FT-IR) to show that the C-termini increase the β -strand content of the protein.

Terminal extensions are common in the Hfq-family, and Hfq from *Methanococcus jannaschii* carry an extended and disordered N-terminus preceding the α -helix, as compared to *E. coli* Hfq. A PSI-PRED analysis¹¹⁶ predicts the N-terminus of this protein to form an additional short helix interrupted by an extended region, similar to human Sm-protein D2 (hSm-D2) in the crystal structure of the human U1 snRNP-complex¹⁰². This structure comprises a full Sm-protein heptamer, and displays RNA-induced folding in the intrinsically disordered N-terminus of hSm-D2. The N-terminus forms a kinked α -helix that is interrupted by a short loop, upon coordinating to the minor groove of the core helix H of the four-helix-junction of U1 snRNA. This could suggest a degree of similarity of Hfq proteins and hSm-D2. Furthermore, as the only human Sm-protein, hSm-D2 and Hfq share the similarity of a two-residue extension of the N-terminal α -helix as compared to other human Sm-proteins¹⁰⁵.

As mentioned above, the C-terminal domain of *E. coli* Hfq shares features with eukaryotic hub-proteins. In its intrinsically disordered character it could further be implicated in protein-protein interaction. Intrinsically disordered protein sequence has been implicated in protein-protein interaction¹¹⁷. Regulatory proteins, and especially RNA-chaperones, tend to retain a higher amount of disordered sequence, which has a functional implication in an entropy-transfer model⁹⁸. The model describes how an intrinsically disordered chaperone binds and unfolds a (mis-)folded target, utilizing intrinsic high entropy to gain thermodynamic control over a kinetically trapped target.

1.7 Aim of the study

Studies have been undertaken to elucidate structural and biophysical aspects of the involvement of *E. coli* Hfq in riboregulation.

2 Structural and biophysical study of *E. coli* Hfq

Synopsis

A variety of structural and biophysical studies have been undertaken, to elucidate the functional behavior and structure of full-length Hfq from *E.coli*. Combining bioinformatics and functional studies (biochemical, FRET) with results from biophysical experiments (SR-CD, NMR), function has been annotated to the disordered domains of *E. coli* Hfq, visualized by the solution structure established by SAXS-experiments. Subsequently, the crystal structure of full-length *E. coli* Hfq provided further insights to the function and behavior of the C-terminal domain. This work resulted in two publications, which in common provide insights to the versatile nature of the C-terminal domain of this partly disordered RNA-chaperone. The papers are included in section 8 and 9.

2.1 Summary paper I: Solution structure of *E. coli* Hfq

The solution structure of full-length *E.coli* Hfq has been reconstructed with P6-symmetry, based on SAXS-experiments. The model shows the C-termini to be extending laterally away from the structured hexameric (L)Sm-core of Hfq, with structural variability predominantly in the latter half of the C-terminal domain. Functional studies indicate the latter half to be of functional importance, which bioinformatics predict to be intrinsically disordered. In accordance, the solution structure was complemented with studies by SR-CD spectroscopy, recorded from constructs carrying C-terminal truncations of decreasing lengths. The resulting spectra display an overall blue-shift with increasing length of the C-terminus, suggesting this to be completely disordered. Previous results from FT-IR experiments indicated the C-terminal domain to increase the β -strand content of the protein, which would be in agreement with the deconvolution of the SR-CD spectra. However, the deconvolution of SR-CD has presently been interpreted as the result of an experimental artifact, due to the stronger signal of disordered protein compared to β -strand. Furthermore, NMR-experiments provide insights to the interaction between the C-terminal domain and the (L)Sm-core, and suggest a dynamic transition between an ordered and disordered state for the C-terminal domain. Finally, SR-CD spectra recorded in the presence of RNA-fragments of increasing length support the hypothesis that the C-terminal domain becomes structured upon interaction with RNA-molecules above a certain length.

Beich-Frandsen M et al. Structural insights into the dynamics and function of the C-terminus of the E. coli RNA chaperone Hfq. Nucleic Acids Res. 2011 Jun;39(11):4900-15.

2.2 Summary paper II: Crystal structure of *E. coli* Hfq

Following the study of solution behavior of *E. coli* Hfq, a crystal structure was obtained that indicates an additional mode of interaction for the C-terminal domain. Despite problems with degradation of the full-length protein, an optimized purification procedure enabled the crystallization of the non-degraded full-length protein. Little to no electron density of the C-termini is visible from the resulting data, which was collected from a highly anisotropic P1-crystal, thought to be due to a lattice translocation defect. However, the result appears in agreement with previous EM-studies that suggested the C-termini to reside 'on top' of the proximal face of Hfq. Interestingly, compared to previous crystal structures of *E.coli* Hfq, the presence of the C-terminal domain in the crystal appears to coordinate between adjacent hexamers, which could suggest residual structure to reside in this intrinsically disordered domain.

Beich-Frandsen M, Večerek B, Sjöblom B, Bläsi U, Djinović-Carugo K. Structural analysis of full-length Hfq from Escherichia coli. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 2011 May 1;67(Pt 5):536-40.

3 Solution structure of Hfq:RNA-complexes

Synopsis

To investigate the interaction of Hfq with longer RNA, the solution structure of C-terminally truncated *E.coli* Hfq in complex with a longer RNA-oligonucleotide has been reconstructed. The initial study was based on SAXS-data, which raised questions about the stoichiometry in the complex. NMR relaxation-experiments could confirm the intended 1:1 stoichiometry of the complex, and comparative NMR-analysis provided a chemical shift pattern, for which no disagreement with the proposed SAXS solution structure could be found. In addition, NMR indicated structuring of the N-terminus of *E.coli* Hfq in complex with RNA. SANS-experiments were performed with initial intent of showing a shape-change in full-length Hfq, under the hypothesis that the disordered sequence become structured in the presence of RNA. The sample preparation procedure was compromised by addition of D₂O for contrast variation, and data could only be acquired from the RNA-complex with the C-terminally truncated protein. Analysis of the SANS-data was done while taking the results from SAXS and NMR into account, and provided some proof of concept, in addition to elucidating aspects about the possible organization of a complex between Hfq and RNA.

3.1 Preamble

In continuation of small angle X-ray scattering measurements and shape-reconstructions of the *E. coli* Hfq-protein in its apo form as described in M. Beich-Frandsen *et al.* 2010¹¹⁸, we desired to study complexes of Hfq with specific RNAs. The overall aim of these studies was to identify a specific mode of interaction, which could sustain an understanding of the Hfq-protein's ability to process the wide variety of RNA-molecules, with the high degree of specificity it is known for²⁹.

Crystal structures have elucidated interaction between RNA and Hfq from three different organisms. The structures of *Escherichia coli* Hfq (aa 2-69) in complex with poly-A₁₅ RNA¹¹¹, *Bacillus subtilis* Hfq (aa 1-73) in complex with RNA-heptamer (5'-AGAGAGA-3')¹¹², and *Staphylococcus aureus* Hfq (aa 1-77) in complex with poly-U RNA (5'-AUUUUUG-3')¹⁰⁸ all elucidate unique binding sites and modes, and explain important aspects about Hfq:RNA-interaction, all commonly providing insight into the function of Hfq. However, none are conclusive with regards to the binding mode required for longer RNA-molecules.

E. coli Hfq is thought to show preference for binding uridine-rich stretches of RNA adjacent to a hairpin-loop⁹². Studies have attempted to

elucidate the binding of longer RNAs on Hfq by a mutational analysis⁹⁵, and taking the crystal structures in view in light of these functional results provides a more nuanced picture⁷³. Interpretation of such results, as well as those of a standard cross-linking / mass-spectrometric (MS) approach¹¹⁹, are compromised by the 6-fold symmetry of the hexameric Hfq protein. A single point-mutation in Hfq effectively changes the chemical environment at six positions in the functional oligomeric molecule. Such results should be regarded keeping in mind the possibility that these mutations may alter the way the RNA molecule binds to Hfq. Additionally, when analyzing results of cross-linking experiments by means of MS, the peptides resulting from protein digestion can essentially stem from six different positions in the assembled hexamer, complicating the analysis of such results.

The solution-structure of a longer RNA-molecule in complex with the native protein, even with a fragment of a native RNA-molecule, is regarded as a closer view on functional Hfq¹²⁰. By shape reconstruction of the solution-structure of RNA complexes with Hfq, we aimed to elucidate how RNA-molecules bind on the oblate shape of the hexameric Hfq from *E.coli*.

Hfq binding sites in RNAs are known from analysis of protection studies, combining chemical foot-/toe-printing⁷⁹ and enzymatic degradation⁹⁰. These studies have revealed single RNA-molecules to have several binding sites for Hfq. Based on this knowledge we designed fragments of specific mRNA and sRNA comprising single binding-sites for Hfq (Branislav Vecerek; personal communication, and^{121,38,122}), to achieve complexes of 1:1 molar ratio between RNA and hexameric Hfq-protein.

Small angle X-ray scattering (SAXS) was performed for *E. coli* Hfq in complex with a range of RNA-molecules. Complexes were formed and data acquired with two different construct lengths of the protein: with and without the C-terminal domain, as described in Beich-Frandsen M. *et al.* 2010¹¹⁸. Full-length *E. coli* Hfq (Hfq_{Ec}) was studied in complex with fragments of mRNAs hfq, rpoS and sodB, domain 2 of sRNA DsrA and the full-length sRNAs DsrA and ryhB. The C-terminally truncated *E. coli* Hfq (aa 1-65) (Hfq_{Ec65}) was studied in complex with domain 2 of sRNA DsrA and the full-length of the sRNAs DsrA and ryhB.

Preliminary data-analysis revealed the majority of the recorded datasets to be compromised by aggregation and polydispersity. These datasets could be analyzed only with respect to the oligomeric composition of the solution, but the overall goal of shape-reconstruction could not be achieved, and analysis of these datasets was therefore not pursued.

The SAXS-data of the complex between Hfq_{Ec65} and DsrA domain 2 (DsrA₃₄) was of a quality allowing for shape-reconstruction. The

resulting model of the complex (Hfq_{DsrA}) is supported by data obtained by nuclear magnetic resonance (NMR), and the combined X-ray and NMR data opened to the possibility of acquiring small angle neutron scattering (SANS) data. Through the methodological approach of contrast variation experiments, utilizing the innate differences in chemical composition of the two bio-polymers¹²³, SANS-experiments have proven to be a powerful method for studying low-resolution shapes of and interactions in dynamic protein:nucleic-acid complexes^{124,125}.

Here are reported procedures and outcome model of a combined approach of SAXS, NMR and SANS, in studies of a complex between the C-terminally truncated Hfq_{Ec65} protein and a fragment of the trans-encoded sRNA DsrA from *E. coli*.

3.2 Methods & Materials

3.2.1 Experimental approach

3.2.1.1 Introduction to applied scattering theory

In a scattering event the atoms interacting with the incident radiation, be it X-ray or neutrons, become sources of spherical waves, scattered in all directions. In crystal diffraction these atoms are perfectly ordered in the crystal lattice, and when coherent radiation is scattered from a crystal, the ordering gives rise to constructive interference between waves scattered in perfect phase from 'scattering planes' inside the crystal (Fig.3.1). By virtue of the constructive interference the amplitude of the scattered waves add, which result in discrete observations - diffractions.

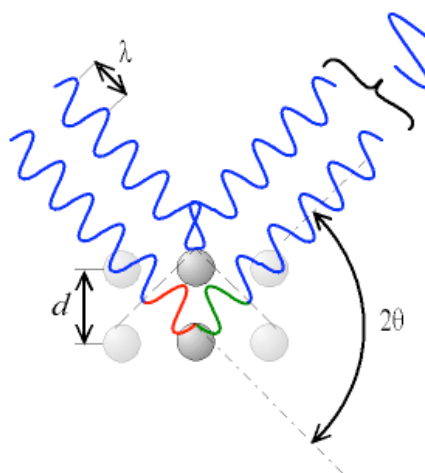


Fig.3.1: Scattering in crystal diffraction. The small X-ray wavelength (λ) allows atoms to become the origin of constructive interference of scattered X-rays. Bragg's law ($n\lambda = 2d \sin \theta$) denotes that the inter-atomic distance (d) should be exactly an integer (n) wavelength (λ) for this to occur at a given scattering angle (2θ) (adapted from www.wikipedia.com).

In solution scattering the molecules are oriented completely at random, and the scattering of the coherent radiation occurs from all possible orientations of the molecules, leading to no constructive interference. The observed result is a radially symmetric scattering around the center of the incident beam (Fig.3.2), leading to summation of the intensities of the scattered waves rather than their amplitudes¹²⁶. The radial average of the scattering intensity is recorded as a function of the scattering angle with respect to the incident beam, expressed in terms of the momentum transfer (s) as described by $s = (4\pi \sin \theta) / \lambda$; where λ is

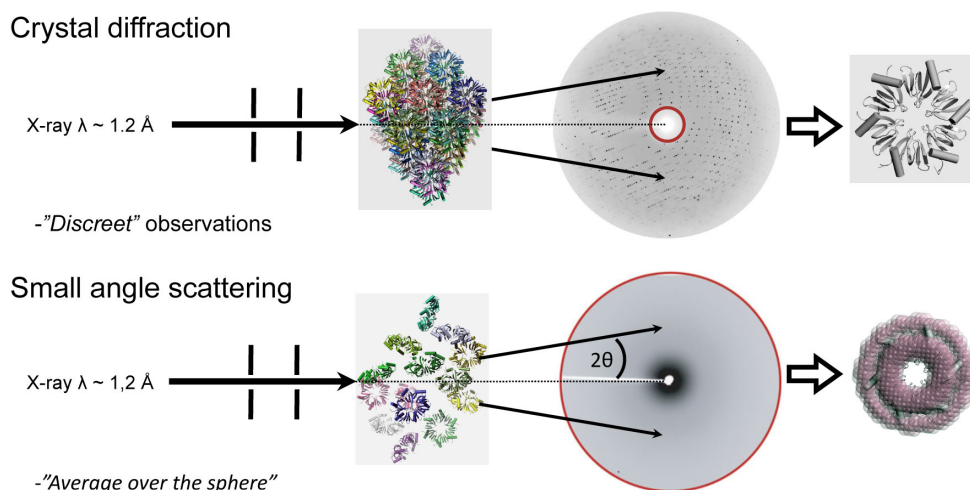


Fig.3.2: Schematic diagram depicting the differences between diffraction and solution scattering experiments. The red circle denotes the same scattering angle with respect to the incident beam, signifying the relative resolution achievable from either technique (adapted from ¹²⁷).

the wavelength and (2θ) is the scattering angle. The experimental scattering curve will then be a semi-logarithmic plot of the scattering intensity as a function of the momentum transfer.

Small angle scattering is inherently a contrast-based technique. Experiments are recorded for the solution containing the molecule in question, and subtracted by the scattering curve of the solvent alone. Hereby a scattering curve representing the excess scattering density of the molecule alone is obtained.

The excess scattering density of a molecule is governed by its chemical and isotopic composition. Neutron and X-ray radiation interact differently with matter, resulting in different potential of scattering for each chemical element, for each type of radiation. X-ray scattering is dependent on the electron shells and orbital configuration in an atom, whereas neutron scattering is dependent on the isotopic composition in the nucleus¹²⁶. The scattering length measures the scattering potential of the atoms and has been tabulated for common elements, the amino acids and the building blocks of other biopolymers, for both neutron and X-ray radiation¹²³. This allows for summing the theoretical excess scattering density of a specific protein or nucleic acid, with regards to the incident radiation. Table I show examples of the scattering lengths of some common chemical elements.

As can be seen from Table I, hydrogen scatters X-rays about an order of magnitude weaker than the other common biological elements. Most notable is the sign difference in scattering length with respect to neutrons, between hydrogen (^1H) and its heavier isotope deuterium ($^2\text{H} \sim \text{D}$). Due to the high abundance of hydrogen in biological material,

Table I. X-ray and neutron scattering length of common elements.

Atom	H	D	C	N	O	P	S	Au
Atomic mass	1	2	12	14	16	30	32	197
N electrons	1	1	6	7	8	15	16	79
$b_{\text{X-ray}}, 10^{-12} \text{ cm}$	0.282	0.282	1.69	1.97	2.16	3.23	4.51	22.3
$b_{\text{Neutron}}, 10^{-12} \text{ cm}$	-0.374	0.667	0.665	0.940	0.580	0.510	0.280	0.760

Reprinted from Svergun/Koch 2003 ¹²⁶

neutron scattering from dilute solutions of macromolecules is governed hydrogen scattering, a phenomenon that is exploited in neutron scattering contrast variation.

3.2.1.2 Contrast variation experiments

Fig.3.3 shows a plot of the neutron average excess scattering density ($\bar{b}$) for some common biopolymers, as a function of the composition of the solvent, in this case a mixture of H₂O and D₂O. Note how for pure H₂O the average scattering density is negative, reflecting the negative neutron scattering length of hydrogen (Table I).

In following of the principle, of subtracting the ‘solvent scattering’ to obtain the ‘solute sample scattering’ (which is the difference between the average scattering density of the solvent and the cumulated scattering density of the solute), the difference in average scattering density is termed the ‘contrast’ (denoted $\bar{\rho}$), quantifying the solute scattering length compared to the solution. In this regard, the solution

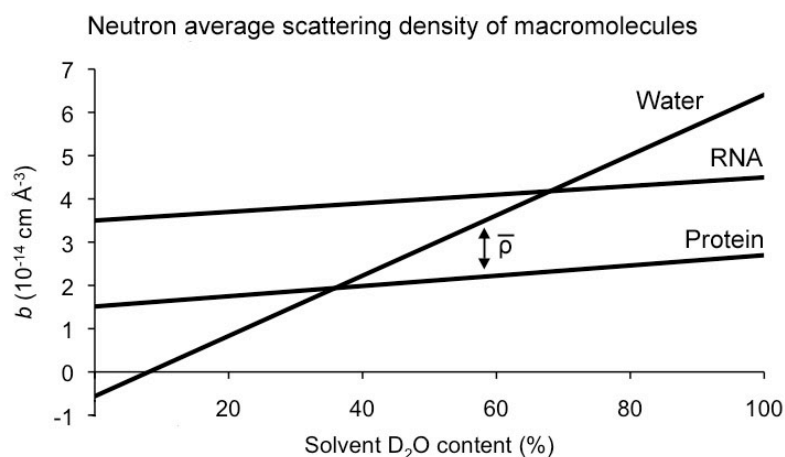


Fig.3.3: Neutron average scattering density of protein and RNA compared to that of water, indicating the ‘match-points’ of ~40% and ~70% respectively. The contrast between protein and H₂O is denoted by $\bar{\rho}$.

composition can be titrated to establish zero contrast between solute and solution, making the solvent scattering match the solute scattering, which by subtraction then cancels. Essentially this allows masking the scattering contribution by a specific component in solution. By experimentally establishing the sample “match-point” of solution composition, it is then possible to retrieve scattering contribution from a specific solute component.

Consider a given complex between protein and nucleic acid, where recording small angle neutron scattering at ~40% D₂O will mask the protein scattering contribution. This allows recording in-situ scattering from the nucleic acid alone (still in complex with protein). Similarly, recording scattering from the same sample at ~70% D₂O will mask the nucleic acid scattering contribution, and retrieve scattering from the protein alone. With the same approach, and employing uniform deuteration of specific binding partners at specific contrast levels (D₂O-content), it is readily conceivable to acquire scattering curves from a complex between multiple proteins, or a protein in complex with one or several nucleic acid binding partners^{128,129}. Recording scattering curves for the protein alone and in complex with different binding partners, at the D₂O-content (contrast-level) matching the binding partner(s), will allow for evaluation of the effect the binding partner has on protein shape. Differences in the scattering curve between the apo-state and the bound form will reflect a shape-change, which can be interpreted as a change in conformation when bound in complex.

Acquiring a similar H₂O/D₂O-experiment with X-rays will return a scattering curve representing both binding partners simultaneously, since there is no significant difference in the scattering density of the binding partners with respect to X-rays – i.e. no contrast between the two components of the complex. Complexes between protein and nucleic acid can be studied by X-ray scattering contrast variation. In place of titrating the D₂O-content of the solvent, a contrast-generating component can be added to the solution, typically a stabilizing agent like glycerol or sucrose is employed. However, it is important to take into account that this compound will be present in increasing amounts through the contrast series, which potentially could have an affect on the sample under study.

In present study SANS contrast variation experiments are employed to support the model of a complex between protein and RNA, which was obtained based on SAXS and NMR-experiments.

3.2.2 Preparation of Hfq:RNA-complex

Full-length *E. coli* Hfq (Hfq_{Ec} ; aa1-102) and C-terminally truncated construct comprising residues 1-65 (Hfq_{Ec65}) were expressed and purified as described previously¹¹⁸.

Chemically synthesized RNA was generously provided by Ronald Micura's group at Innsbruck University, and in larger quantities acquired commercially from Dharmacon. A minimal fragment required for binding of the sRNA DsrA to Hfq had previously been identified by Brescia *et al.*⁹², and further truncation led to a 34-nucleotide fragment of DsrA domain 2, 5'-G₂₄A-A-U-U-U-U-U-A-A-G-U-G-C-U-U-C-U-U-G-C-U-U-A-A-G-C-A-A-G-U-U-U₅₉-3', which we termed DsrA₃₄.

The protein concentration was assessed prior to RNA-addition to allow for a 1:1 molar-ratio of Hfq_{Ec65} (hexamer) to RNA. The protein and RNA were mixed at elevated temperature (~60°C), and incubated 5 hrs in a 65°C incubator, and allowed to slowly cool to room temperature. Subsequently the complexes were centrifuged for 1 hour at 10000 x G, at room temperature. Prior studies by dynamic light scattering (DLS) performed on a Wyatt Dynapro DLS, had indicated this procedure to result in a highly monodisperse sample. Samples were temperature labile, and prone to aggregation on sudden temperature or pressure change.

3.2.3 Small Angle X-ray Scattering

Small angle X-ray scattering experiments were acquired at beamline X33^{130,131} at the DORIS III synchrotron storage-ring (DESY, Hamburg, Germany), using a MAR345 image-plate detector and X-ray wavelength of $\lambda = 1.5 \text{ \AA}$, with a sample to detector distance of 2.7 m. This setup covers a range of momentum transfer $0.012 < s < 0.045 \text{ \AA}^{-1}$ ($s = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle). Data were acquired at 37°C from four concentrations of a C-terminally truncated Hfq core-construct Hfq_{Ec65} (19.0, 9.5, 4.8, 2.4 mg/ml), and five samples in complex with DsrA₃₄ (3.9, 7.7, 15.5, 23.2, 30.9 mg/ml). Sample concentration was measured at 260 nm and 280 nm on a Nanodrop 1000 UV/VIS spectrophotometer, immediately prior to data-acquisition. The data were averaged and normalized to the intensity of the incident beam prior to subtraction of buffer scattering. The data were processed in the ATSAS program-package¹³². The difference data curves were scaled, merged and initial analysis performed using PRIMUS¹³³. Invariants were derived by standard approaches for comparison¹²⁷. The intrinsic scattering (I_0) and radius of gyration (R_g) were derived by Guinier analysis for the data range $s < 1.3 / R_g$, approximating $I(s) = I_0 \exp(-(s R_g)^2 / 3)$. The reverse transform program GNOM¹³⁴ was employed for calculation of the distance-distribution function, under the assumption of correct estimation of the particle maximum dimension (D_{MAX}). The program also outputs a value for I_0 and R_g . The particle excluded

volume was calculated from the Porod equation¹³⁵, where $I_{\text{exp}}(s)$ is the experimental data.

$$V = 2\pi^2 I_0 / \int_0^\infty s^2 I_{\text{exp}}(s) ds ; \text{ (equation 1)}$$

In addition, the program Dammif¹³⁶ run without symmetry restraints, outputs the particle volume. The solute molecular mass was further evaluated by the standard approach of comparison of the intrinsic scattering (I_0) with that recorded from a reference solution of known molecular mass and concentration. In the present study a single sample of 3.68 mg/ml bovine serum albumin ($M_r = 66.4$ kDa) was used.

Low resolution shape-models of the SAXS data were generated using the program BUNCH¹³⁷. Employing a combined rigid body and ab-initio modeling approach, the program searches for a spatial arrangement of domains with known high-resolution structure (in present study the Hfq protein), and probable conformations of the RNA best fitting the scattering curve are represented as dummy-residues. Similar to the program Dammif¹³⁶, the program represents the protein as a collection of densely packed dummy-residues inside a sphere with the diameter D_{MAX} . Each dummy-residue belongs either to the particle or the solvent, and the shape is described by a binary string of length M . Starting from a random string, simulated annealing is employed to search for a compact model that fits the shape scattering curve $I(s)$ to minimize discrepancy χ^2 :

$$\chi^2 = \frac{1}{N-1} \sum_j \left[\frac{I(s_j) - cI_{\text{calc}}(s_j)}{\sigma(s_j)} \right]^2 ; \text{ (equation 2)}$$

where N is the number of experimental points, c is a scaling factor, and $I_{\text{calc}}(s_j)$ and $\sigma(s_j)$ are the calculated intensity and the experimental error at the momentum transfer s_j respectively.

The N-termini of Hfq_{Ec65} was first reconstructed, utilizing the known structure of the Hfq_{Ec65} monomer subunit (aa 6-65) (PDBid: 1HK9). The Bunch program¹³⁷ was run with an overall P6 symmetry, retaining the hexameric organization of the Hfq core. Theoretical scattering from Hfq core encompassing residues 6-65 was calculated using the program CRY SOL¹³⁸. Subsequently, the output model was used in shape reconstruction of complexes with RNA. The 34-nucleotide sRNA fragment DsrA₃₄ was represented as a chain of 102 dummy residues. Due to increased electron density in phosphate and hence the higher scattering length (see Table I), each nucleotide was represented by three dummy-residues, which is acceptable for representation of an average nucleotide form-factor (P. Konarev - personal communication,

and ¹³⁹). The program employs a simulated annealing procedure to manipulate the local arrangement of the chain representing the RNA, to minimize the χ^2 -discrepancy as described in equation 2.

3.2.4 NMR-experiments

All NMR experiments were performed at 310 K on a Varian Inova 600 MHz spectrometer equipped with 5 mm triple resonance cryo-probe and pulsed field gradients, and a Varian Inova 800 MHz spectrometer equipped with 5 mm conventional triple resonance probe equipped with pulsed field gradients. NMR spectra were processed with NMRPipe¹⁴⁰ and analyzed with the Sparky software. The protein samples for NMR-experiments were all prepared by size-exclusion chromatography in 50 mM Na-PO₄ pH~7.2, 200 mM NaCl, and concentrated to ~1 mM Hfq_{Ec65} (with respect to the monomer). Samples were supplemented ~10% (v/v) D₂O to provide the deuterium signal for the field-frequency lock, as well as 0.1–0.2% (w/v) NaN₃ to inhibit bacterial growth. Titration experiments of Hfq by DrsA₃₄ monitored by NMR ¹⁵N-HSQC were conducted to map residues in Hfq_{Ec65} affected in binding of DsrA₃₄. By comparison with assignments for Hfq_{Ec65}¹¹⁸ the interacting residues could be identified, and 3-dimensional HNCO¹⁴¹ and HNCA¹⁴² experiments were additionally employed to resolve ambiguity. NMR ¹⁵N-transverse relaxation experiments were conducted to validate the stoichiometry of the complex between Hfq_{Ec65} and DsrA₃₄. The measured relaxation rate ($R_2 = 1/T_2$) of a globular protein is approximately proportional to the rotational correlation time (τ_c), or inversely proportional to the rotational diffusion constant (D_r). The rotational diffusion constant and the hydrodynamic radius to the third power (i.e. the effective hydrodynamic volume, reflecting the molecular mass) are related to each other by the Einstein/Smoluchowski-relation.

$$D_r = \frac{k_B T}{(8\pi\eta)R_h^3} ; \text{ (equation 3)}$$

3.2.5 Small Angle Neutron Scattering

Neutron scattering experiments were measured at beamline D22, at the high-flux reactor, Institut Laue-Langevin, Grenoble, France; in collaboration with Dr. Phillip Callow. Data was acquired from two instrument configurations (2 m and 5.6 m detector distance) with incident wavelength $\lambda = 6 \text{ \AA}$, allowing for a range of momentum-transfer of $0.008 < s < 0.35 \text{ \AA}^{-1}$ ($s = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle). Data was collected at 37°C, with a sample acquisition-time of 900 and 1800 sec in the short and long distance respectively.

Images were fitted with beamstop-mask, radially averaged and integrated with the Grasp program (www.ill.fr/lss/grasp). Data was

normalized to the respective neutron transmission, merged and buffer scattering subtracted using same program.

Protein and RNA were purified in water, and the buffer changed into the designated D₂O content. The protein was buffer-changed by two consecutive runs over PD-10 desalting columns (GE Healthcare) prior equilibrated with designated D₂O containing buffer. The RNA was lyophilized and re-dissolved similarly. 100% D₂O-buffer was made in H₂O and pH titrated to 6.8, subsequently lyophilized to dryness and the residual re-dissolved in D₂O. The kinetic isotope effect of substituting hydrogen for deuterium causes the pD to increase 0.4 units above the H₂O nominal value¹⁴³. The 100% H₂O-buffer was treated similar at pH 7.2. For performing contrast variation experiments, Protein:RNA complexes were then formed by same approach as for SAXS and NMR, using components of specific D₂O-content.

Samples of Hfq_{DsrA} were measured at an intended concentration of 7 mg/ml. Concentration was assessed spectrophotometrically by the direct method after the scattering experiment. Data was acquired at three levels of contrast, 0%, 75% and 100% D₂O. For the purpose of experimental matchpoint determination, a concentration of 2 mg/ml of the DsrA₃₄ RNA-fragment was measured at 0%, 40% and 100% D₂O.

The data curves were scaled and initial analysis performed using PRIMUS¹³³. A constant was subtracted from all data curves, to ensure homogenous scattering at high angle. The intrinsic scattering (I_0) and radius of gyration (R_g) was derived by Guinier analysis for the data range $s < 1.3 / R_g$, approximating $I(s) = I_0 \exp(-(s R_g)^2 / 3)$. The particle excluded volume was calculated from the Porod equation¹³⁵, and the solute molecular mass was evaluated by the standard approach of normalization of the intrinsic forward scattering (I_0) with that recorded from water¹⁴⁴.

The reverse transform program GNOM¹³⁴ was employed for calculation of the distance-distribution function, under the assumption of correct estimation of the particle maximum dimension (D_{MAX}), this program also outputs a value for I_0 and R_g .

The program MONSA^{145,146} was used to generate low-resolution shapes *ab-initio*, based on the SANS-data. MONSA is an extension of the original DAMMIN-program allowing for simultaneously fitting multiple phases, represented in multiple scattering curves with specific contrast levels¹⁴⁷. Here the particle is represented as a collection of densely packed dummy-residues, each belonging to the solvent or to a designated particle phase. Imposing an overall symmetry inside the global parameter of D_{MAX} , the program minimizes the discrepancy between the experimental scattering curve and calculated model scattering curve by means of simulated annealing (equation 2).

3.3 Results & Discussion

3.3.1 Preparation of Hfq:RNA-complex

A problem in forming the complex between Hfq_{Ec65} and DsrA₃₄ is the inevitable occurrence of aggregation at room temperature when mixing protein and RNA. The components in the complex were therefore mixed from solutions of equal molar concentration in order to minimize the effect of the mixing.

The desired 1:1 complex between Hfq and RNA dominates at elevated temperature. However, aggregation is an inherent problem that leads to loss of material by centrifugation. Furthermore, in the samples used for SANS-experiments a 'kinetic isotope effect' of the D₂O in the samples for contrast variation measurements, caused the precipitation to increase as a function of the D₂O-content.

In small angle scattering measurements it is important to know the concentration of the sample with high accuracy, in order to assess the molecular mass. Due to the unique sample composition of the complex, estimating the concentration by conventional methods is complicated. In spectrophotometric determination of the concentration, the extinction coefficients of components of a complex are cumulative, and under assumption of a correct stoichiometric ratio the complex extinction coefficient can be summed from those of the components. However, since nucleic acid absorbs much stronger than protein it is problematic to assume a stoichiometric ratio based on the absorption alone. Given the relatively very small absorption of the protein, it is impossible to assess whether a complex is formed in a 1:1 or a 2:1 ratio between RNA and protein. This will lead to dramatic changes in the extinction coefficient and orders of magnitude errors in the concentration of the complex – and subsequent errors in estimation of the molecular mass from scattering experiments. Conversely, experimental error in estimating the molecular weight of the complex from analysis of small-angle scattering, which could be more than 10%, can also lead to faulty assumptions about the stoichiometry.

Based on prior studies, the complex between Hfq_{Ec65} and DsrA₃₄ (Hfq_{DsrA}) was formed in a 1:1 molar ratio. Fig.3.4A display a gel result of an electrophoretic mobility shift assay (EMSA), in comparison with analytical gel-filtration results (B-F) of similar molar ratios between full-length Hfq_{Ec} and DsrA₃₄. As can be seen, complexes form in different ratios as a function of the protein concentration. The reported molecular mass of the peaks are under assumption of a globular shape, and the increased size compared to the expected (Hfq:RNA 1:1 ~80 kDa, 1:2 ~90 kDa) was interpreted as anisometric shape of the complex.

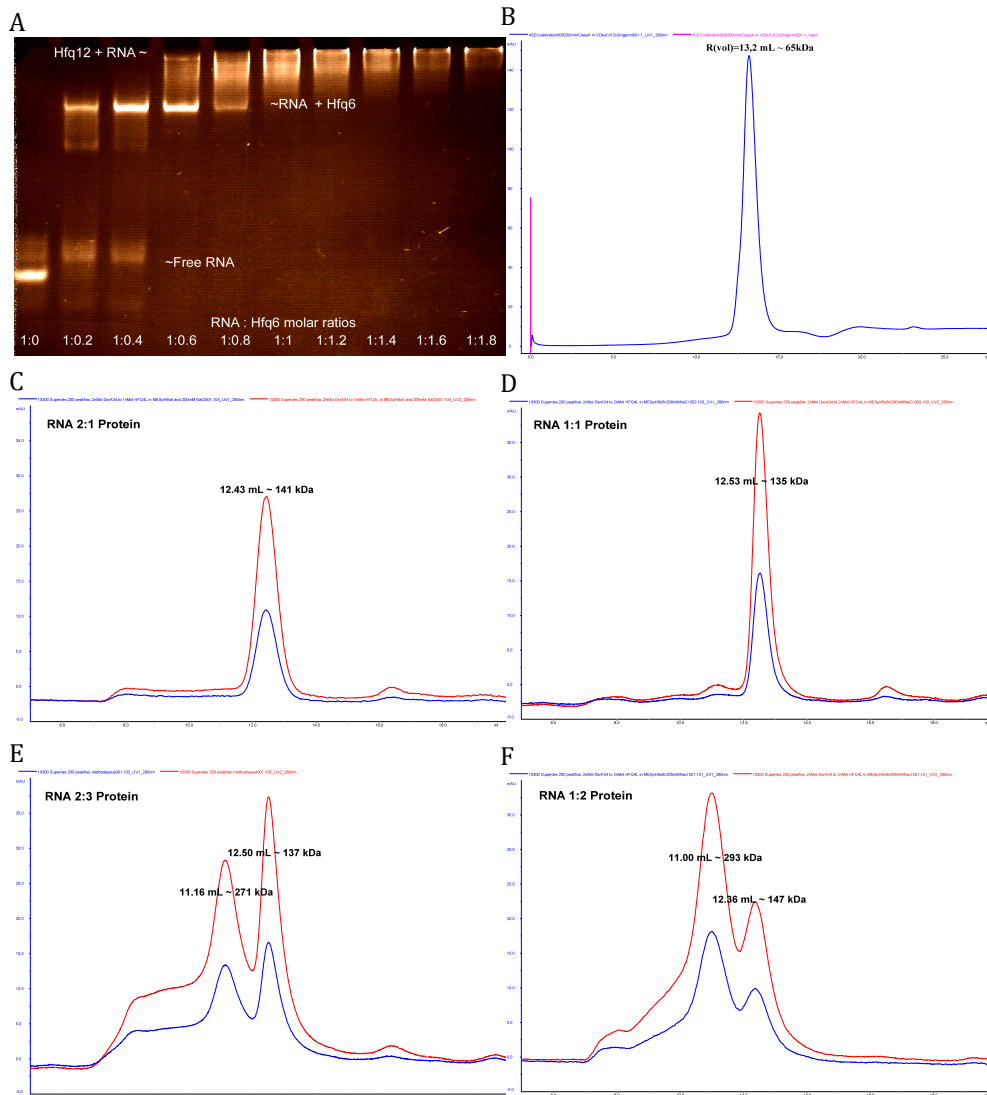


Fig.3.4: **(A)** 8% PAGE-gel of complexes of different molar ratios of Hfq_{Ec} to fixed amount of DsrA₃₄ (EMSA). **(B-F)** Chromatograms of similar ratios analyzed by gel-filtration on a GE Healthcare Tricorn 10/300 Superdex 200 column, by reference to the GE Healthcare molecular mass standardization kit of globular protein standards. Absorption was detected at 260 nm (red curve) and 280 nm (blue curve), and plotted as a function of retention volume. The apparent increased size of the complex is the effect of an anisometric shape, of a complex of smaller molecular mass.

Fig.3.5A shows EMSA-results comparing the binding of DsrA₃₄ to full-length Hfq_{Ec} and Hfq_{Ec65}, displaying a change in affinity on truncating the C-terminal of the protein. The full-length Hfq_{Ec} appears to be binding DsrA₃₄ with increased affinity when compared to C-terminally

truncated Hfq_{Ec65}, however the complex between Hfq_{Ec65} and DsrA₃₄ was much less prone to aggregation. Dynamic light scattering (DLS) analysis performed on a complex between Hfq_{Ec65} and DsrA₃₄ (Fig.3.5B) show a hydrodynamic behavior indicating a highly monodisperse sample, formed in batch – without purification. Similar samples of Hfq_{Ec65} in 1:1 complex with DsrA₃₄ were used in all small angle scattering experiments.

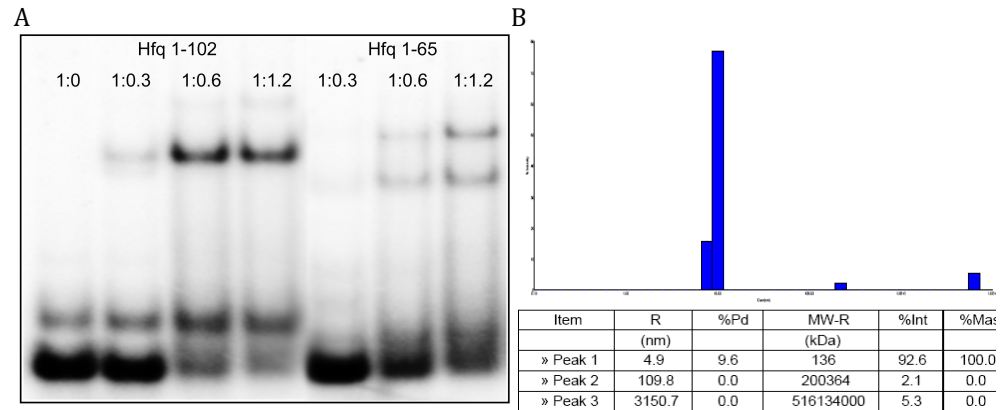


Fig.3.5: **(A)** 8%-page gel analysis of EMSA-studies depicting weaker interaction between the C-terminally truncated Hfq_{Ec65} compared to full-length Hfq_{Ec}. **(B)** DLS analysis of complex between Hfq_{Ec65} and DsrA₃₄ displays high monodispersity at elevated temperature (37°C).

3.3.2 Shape-reconstruction by SAXS

Data was acquired for a complex between Hfq_{Ec65} and DsrA₃₄, and models were established in line with procedures employed in Beich-Frandsen *et al.* 2010¹¹⁸. The missing N-terminal residues of Hfq_{Ec65} were built onto the known structure of *E.coli* Hfq (PDBid: 1HK9) and the resulting model was used in rigid-body modeling of the RNA-complex. Fig.3.6A shows the fit between experimental scattering data and the calculated scattering curve¹³⁸, confirming the close agreement between crystal and solution structure of the highly conserved and stable homo-hexameric core of Hfq. An improvement of the fit ($\chi^2 \sim 1.55 \rightarrow 1.40$) was observed upon modeling the missing N-termini (Fig.3.6B), indicating a modest contribution to the overall scattering of these $6 \times 5 = 30$ unstructured residues.

Guinier analysis of the experimental curve¹³³ established for Hfq_{Ec65} the R_g of $\sim(24.2 \pm 0.5) \text{ \AA}$, which is in agreement with that calculated for the crystal structure, $R_{g(\text{Calc})} \sim 23.5 \text{ \AA}$.

In Fig.3.7A the experimental SAXS-data of Hfq_{DsrA34} is shown. Guinier-analysis estimated $R_g \sim 31.1 \text{ \AA}$, indicating a rather dramatic change in the particle in presence of the RNA. By calculation of the pair-distribution functions using the program GNOM¹³⁴, the estimated particle largest dimension were for Hfq_{Ec65} $D_{\text{MAX}} \sim 67 \text{ \AA}$ with $R_g \text{ } 23.1 \text{ \AA}$, and for Hfq_{DsrA} $D_{\text{MAX}} \sim 102 \text{ \AA}$ with $R_g \text{ } 32.3 \text{ \AA}$ (Fig.3.7A insert).

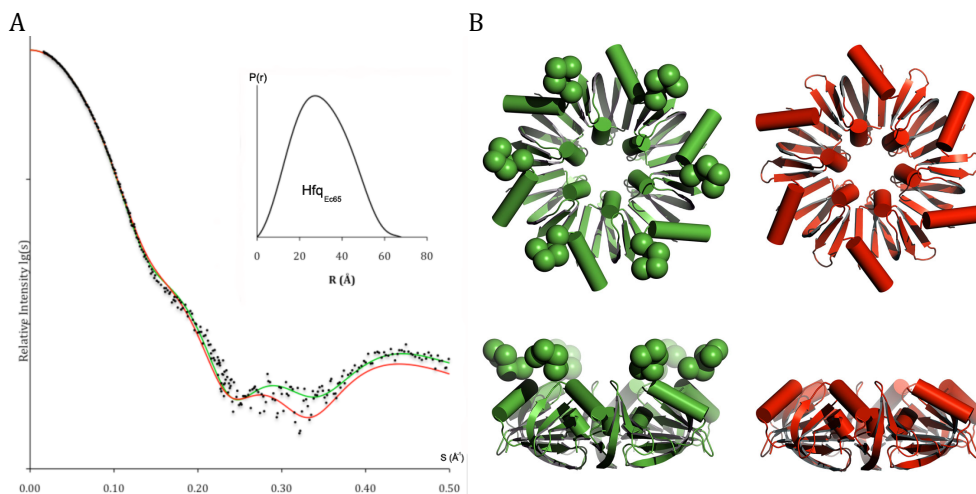


Fig.3.6: **(A)** Calculated SAXS scattering-curves for Hfq_{Ec65} (aa 1-65; green) and N-terminally truncated Hfq similar to the crystal structure (aa 6-65; red) display good overall fit to experimental SAXS data recorded from Hfq_{Ec65} (black). Inserted is the pair-distribution function representing the experimental data of Hfq_{Ec65}. **(B)** Models of Hfq_{Ec65} (aa 1-65; green) and N-terminally truncated (aa 6-65; red).

The molecular mass was estimated to 44 ± 5 kDa for Hfq_{Ec65} and 54 ± 5 kDa for Hfq_{DsrA}, by normalization of the intrinsic scattering (I_0) to a single standard sample of bovine serum albumin (BSA)¹⁴⁸. Based on primary sequence we can calculate the molecular mass for hexameric Hfq_{Ec65} ~ 43.2 kDa, and for the complex Hfq_{DsrA} ~ 54.0 kDa. In corroboration with estimates of the solute excluded volume (Porod's volume) for Hfq_{Ec65} $\sim (70 \pm 5) \cdot 10^3 \text{ \AA}^3$ and Hfq_{DsrA} $\sim (85 \pm 5) \cdot 10^3 \text{ \AA}^3$, this indicated the complex to be formed in the expected 1:1 molar-ratio, between hexameric Hfq_{Ec65} and DsrA₃₄. In addition the program Dammif¹³⁶, run without symmetry restraints, output the particle excluded volume, and for Hfq_{Ec65} $(73 \pm 5) \cdot 10^3 \text{ \AA}^3$ and Hfq_{DsrA} $(82 \pm 5) \cdot 10^3 \text{ \AA}^3$ were retrieved, respectively.

Under the assumption that the samples comprise a monodispersed 1:1 complex, the solution structure of Hfq_{DsrA} was then reconstructed by the program Bunch¹³⁷.

Utilizing the previously reconstructed solution-structure of Hfq_{Ec65} and the knowledge of uridine binding in the conserved core bindingsite (PDBid: 1KQ2¹⁰⁸), 18 dummy-residues were manually placed in the input model, representing the 6 uridine nucleotides expected to be

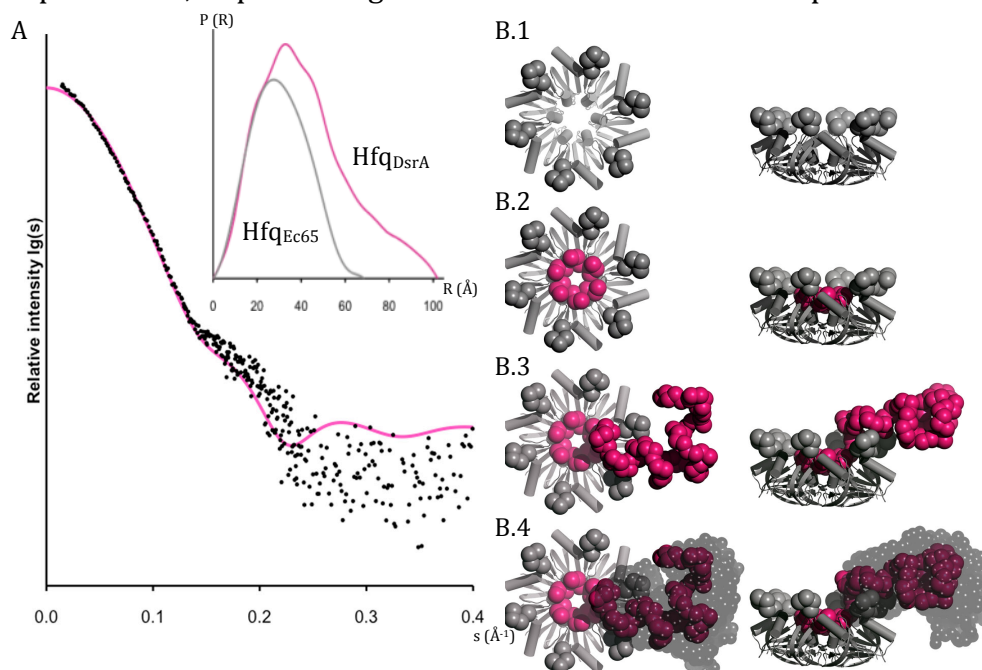


Fig.3.7: X-ray small angle scattering and rigid-body model of Hfq_{Ec65} in complex with DsrA₃₄. **(A)** The model fits the data to $s=0.4$ with χ^2 of ~ 1.7 . Notice the deviation at low s . Inserted is the pair-distribution function of Hfq_{Ec65} and Hfq_{DsrA}, displaying a large increase in the particle dimension in the presence of the RNA. **(B.1-4)** The DsrA₃₄ RNA is represented as (3×34) 102 dummy residues. *B.1*: The solution-structure of Hfq_{Ec65} with the N-terminal reconstructed by SAXS. *B.2*: 18 dummy-residues were placed in the core uridine-binding site. *B.3*: Reconstruction of DsrA₃₄ by the program Bunch. Top-fitting model ($\chi^2 \sim 1.7$) shows the RNA extending away from Hfq. *B.4*: Iterative modeling suggests an extended shape of the RNA.

bound in the core binding site of Hfq. The program Bunch would then model the remaining 28 nucleotides by representing each nucleotide by three dummy-residues (a total of 84), to satisfy the increased scattering-length of the phosphate backbone (in accordance with other models of nucleic acid¹³⁹; P. Konarev, personal communication).

Fig.3.7A displays the fit to the experimental data, inserted is the pair-distribution function for Hfq_{Ec65} and Hfq_{DsrA}, indicating the dramatic increase in the particle dimensions in the presence of the RNA. The stages of model building are illustrated in Fig.3.7B.

On average this modeling approach would fit the data with a best X^2 of 1.7, to a momentum transfer of $s = 0.4 \text{ \AA}^{-1}$. Iterative modeling revealed no distinct shape of the RNA (Fig.3.7B.4), which could lead to the assumption that in complex with Hfq_{Ec65} the DsrA₃₄ RNA is extended and lacks defined secondary structure. Overall, the fit shows the tendency to follow the behavior of the experimental data, displaying reasonable agreement between the shape of the model and that of the complex in solution. However, in this model each nucleotide is represented by three dummy-residues, which establishes a maximum volume of the RNA. Careful inspection of the fit to experimental data reveals a lack of fit in the region of low momentum transfer, indicating that the model does not satisfy the intrinsic scattering of the sample. This would suggest the sample particle to contain more material than the reconstructed model. However, another interpretation could be that a small amount of aggregated material is present in the sample solution.

3.3.3 Analysis of NMR-experiments support SAXS-results

NMR titration experiments are a powerful tool for investigating interactions in a protein. The change in resonance in a ¹⁵N-HSQC spectra in presence of increasing amounts of binding partner is significant for residues involved in the interaction. For a protein with a symmetric quaternary structure like Hfq this is particularly powerful. In its apo form, the residues in the six symmetrical subunits are equivalent, giving rise to a resonance pattern similar of a single subunit in the ¹⁵N-HSQC. In the presence of a binding partner the symmetry is broken and the monomers are no longer equivalent, which thus results in potentially multiple (theoretically up to six) different resonances for each residue.

We recorded ¹⁵N-HSQC-spectra of Hfq_{Ec65} in the apo form and in the presence of increasing amounts of DsrA₃₄, Fig.3.8 shows the spectra of apo Hfq_{Ec65} in red, overlaid onto spectra in the presence of one molar equivalent of DsrA₃₄ in green. For the apo spectra, sequential backbone resonance assignment was possible for residues 6-65. The N-terminal residues 1-5 were not observed due to exchange broadening, which can be reasoned by conformational flexibility in this region.

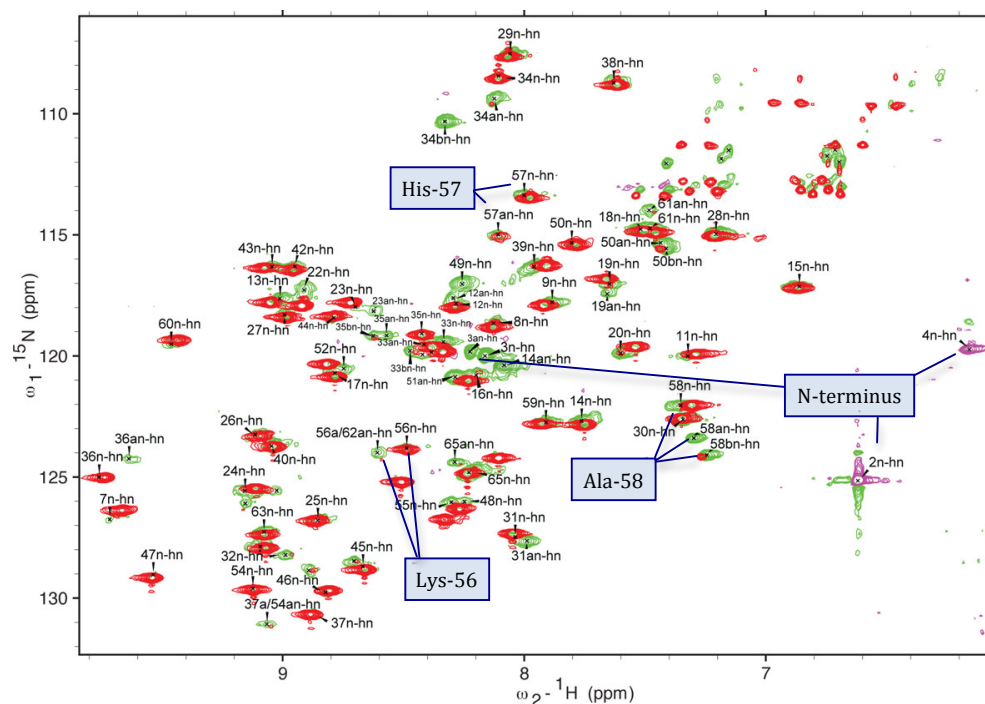


Fig.3.8: Comparative NMR-analysis of Hfq_{Ec65} in the absence and presence of one molar equivalent of the DsrA₃₄ RNA-fragment. ¹⁵N-HSQC-spectra of Hfq_{Ec65} (red) overlaid that of Hfq_{DsrA} (green). In the presence of RNA additional and shifted peaks can be observed. Peaks corresponding to the core uridine binding site and the N-terminus are indicated.

Upon addition of RNA, additional shifted signals can be observed next to the original set of resonances (Fig.3.8), indicating a specific interaction between the RNA and the hexameric Hfq_{Ec65} protein occurring with high affinity (slow exchange on the NMR timescale).

We observed significant shifts in the region corresponding to the conserved uridine bindingsite around His-57¹⁰⁸, indicating the RNA indeed to be bound on the rim of the central pore as expected. Fig.3.9 displays a plot of the ¹⁵N-¹HN chemical shift differences in the presence of RNA. We observed multiple peak resonances for several residues in the region around Gly-34 and around Val-50 illustrating the before mentioned reduction in symmetry. However, no particular property for RNA interaction is known for glycines or valines¹⁴⁹.

Gly-34 is one of the completely conserved residues in the core Sm-fold. Owing to its property of unrestricted dihedral angle, this residue is involved in forming the characteristic 'bent' β -stand 2 of the Sm1-motif. Val-50 is located in the so-called 'variable region', which in Hfq comprise only a short solvent exposed loop on the outside of the torus. Among the eukaryotic Sm-proteins, insertions of up to 25 amino acids can be found here. Furthermore, an unusually strong shift is observed for Ala-14, which is placed in the middle of the N-terminal α -helix. In Fig.3.10 the chemical shift differences between apo and RNA-bound

protein has been mapped onto the proximal surface of the Hfq_{Ec65} (PDBid: 1HK9; aa 6-65).

Most interestingly, some of the previously unobservable N-terminal residues of Hfq_{Ec65} become observable in the presence of the RNA. Ala-2 and Gly-4 are observed with unusual ¹⁵N-chemical shift values, probably due to aromatic stacking with the RNA-bases. For Lys-3, two non-equivalent signals are observed, again in accordance with the broken symmetry. In the crystal structure of Hfq_{Ec72} (aa 1-72; PDBid: 1HK9) the N-terminal of Hfq_{Ec} is not observed, and this domain is characterized as intrinsically disordered^{118,105,150,99}. This unexpected observation in the presence of RNA indicates an interaction between the RNA and the N-terminal residues of Hfq, which results in a stabilization of their structure. As mentioned, a single mutation in the highly conserved N-terminal of *E.coli* Hfq (Lys-3 to Ala) has proven detrimental to the protein's function⁹⁶.

The spectra were closely examined for signals corresponding to nucleic acid imino-hydrogens, the presence of which would indicate base-pairing in the RNA. No signals could be found in any of the spectra recorded, which supported our hypothesis from SAXS that the RNA is without defined secondary structure. However, another possible explanation is that a dynamic behavior of the RNA broadens the imino-signals. Recurrent formation and melting of secondary structure in the RNA would lead to line broadening of the imino-hydrogen signals, causing signal-reduction beyond detection. Part of the RNA-chaperone function of Hfq is exactly to melt such secondary structure.

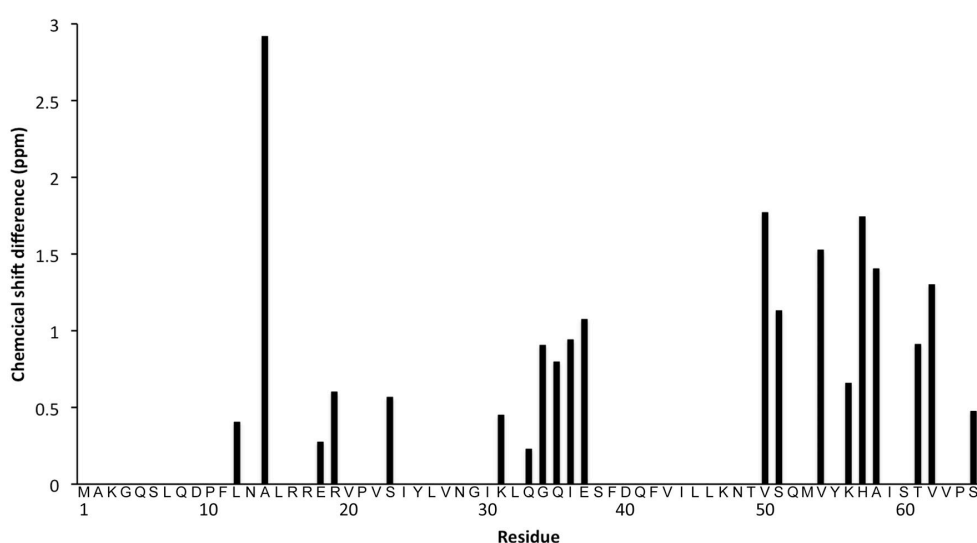


Fig.3.9: Comparative NMR-analysis of Hfq_{Ec65} in the absence and presence of DsrA₃₄ RNA-fragment. ¹H/¹⁵N- chemical shift differences for assigned peaks

$(\Delta\omega_{1H15N}=25*(\Delta\omega_{1H}^2+\Delta\omega_{15N}^2)^{1/2})$ is plotted for residue position.

Furthermore, ^{15}N transverse relaxation experiments suggest the stoichiometry in the complex to be in agreement with a 1:1 complex as indicated by the SAXS-experiments. For Hfq_{Ec65} we measure a relaxation time T_2 of 39.9 ± 2.1 ms ($R_2 = 25 \pm 0.5$ s⁻¹), and for Hfq_{DsrA} 30.5 ± 6.5 ms ($R_2 = 33 \pm 0.2$ s⁻¹). The relaxation time increased by approximately ~25% in the presence of the RNA, reflecting a corresponding increase in particle size and thus molecular mass. Under assumption of a globular shape, the difference in molecular mass of 25% would indicate complex between Hfq_{Ec65} and DsrA₃₄ to be formed in their intended 1:1 molar ratio.

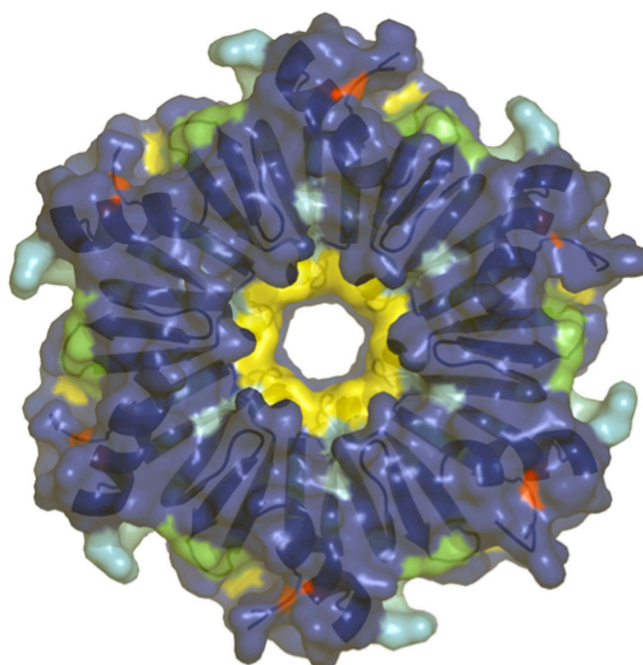


Fig.3.10: Interaction with DsrA₃₄ RNA mapped on the proximal surface of Hfq_{Ec65}. Highest chemical shift change (ppm) is shown in red, going towards blue. Primary interaction site map clearly to the conserved uridine binding site in the center of the hexamer core.

3.3.4 Speculations on Small Angle Neutron Scattering

Small angle neutron scattering data were acquired for Hfq_{Ec65} and Hfq_{DsrA} in 0%, 75% and 100% D₂O. Fig.3.11 show the six data-curves, shifted for clarity along the intensity scale in accordance with expected relative intensity.

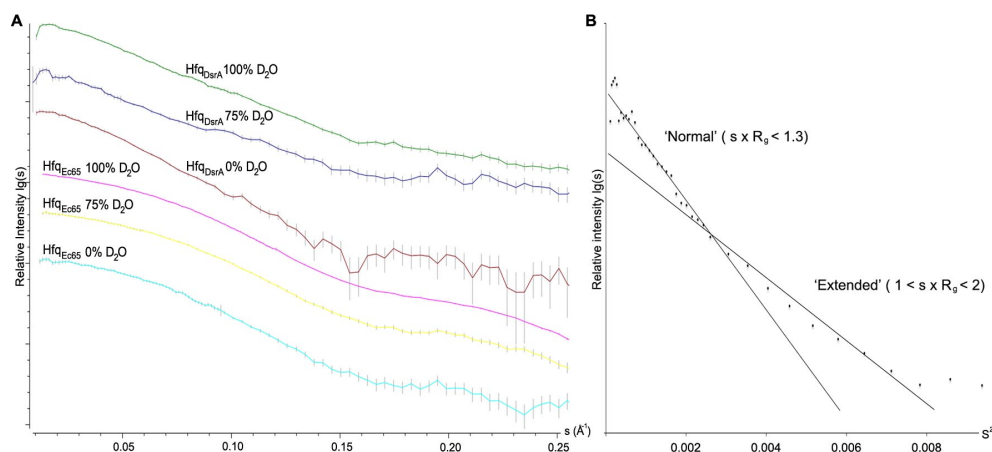


Fig.3.11: **(A)** SANS-data recorded at three contrast levels of 0%, 75% and 100% D₂O for Hfq_{Ec65} and Hfq_{DsrA}. The curves have been shifted along the intensity scale. Notice the lower slope of the curve for Hfq_{DsrA} at 75% D₂O as compared to the same sample at 0% and 100% D₂O. **(B)** Guinier-plot of Hfq_{DsrA} at 75% D₂O showing the difference of deriving parameters from the 'Normal' and 'Extended' range of $s \times R_g$.

Guinier analysis of the data using two different ranges of momentum transfer derived two unique sets of parameters, as summarized in Table II. In Fig.3.11.B is shown the Guinier-plot of Hfq_{DsrA} at 75% D₂O using the two Guinier-ranges. From the conventional approach of assuming a Guinier-range ($s \times R_g$) < 1.3 a set of values for R_g for the Hfq_{DsrA} complex was observed, which indicate the RNA to not be matched at 75% D₂O ($R_g \sim 31.0 \pm 1.3$ $\text{\AA}$). At 75% D₂O, a value of R_g similar to that of the Hfq_{Ec65} protein alone ($R_g \sim 21.0 \pm 0.7$ $\text{\AA}$) would be expected, since the scattering contribution of the RNA is matched by the solvent, in accordance with experimental matchpoint determination (see above).

Fig.3.12 shows a plot of the experimental determination of matchpoints for the protein (Hfq_{Ec65}) and RNA (DsrA₃₄). By relating the intrinsic scattering (I_0) to the sample concentration (c), sample transmission relative to that of H₂O (T) and the pathlength (l), and plotting their square root as a function of the D₂O content; the intersection with the value-axis will denote the point (D₂O-content) where I_0 equals zero, and the specific component has been matched by the solvent scattering. For Hfq_{Ec65} we find the matchpoint to be 42% D₂O, and for DsrA₃₄ the matchpoint is 71% D₂O.

Table II. SANS Guinier-analysis				
	R_g Guinier (Å)	I_0 Guinier	R_g Extended (Å)	I_0 Extended
Hfq _{Ec65} 0%	22.5 ± 1.2	0.113 ± 0.01	22.0 ± 0.3	0.111 ± 0.01
Hfq _{Ec65} 75%	21.0 ± 0.7	0.105 ± 0.01	20.7 ± 0.5	0.104 ± 0.01
Hfq _{Ec65} 100%	21.3 ± 0.2	0.313 ± 0.02	21.2 ± 0.2	0.312 ± 0.02
Hfq _{DsrA} 0%	34.1 ± 0.7	0.117 ± 0.02	29.8 ± 0.3	0.104 ± 0.01
Hfq _{DsrA} 75%	31.0 ± 1.3	0.028 ± 0.01	23.1 ± 0.7	0.019 ± 0.01
Hfq _{DsrA} 100%	31.6 ± 0.4	0.055 ± 0.01	28.8 ± 0.2	0.050 ± 0.01

Parameters derived from Guinier analyses based on conventional ($s \times R_g < 1.3$) and extended range of momentum transfer ($1.0 < s \times R_g < 2.0$).

The neutron transmission values measured at 75% D₂O indicated that the experiment indeed was acquired at the desired contrast level. Taking the small size of the RNA into account, with subsequent small contribution to the overall scattering, and considering the experimental error, it is reasonable to assume that the DsrA₃₄ RNA-fragment was matched at 75% D₂O.

The low concentration and poor sample quality, due to increased precipitation caused by the kinetic isotope effect, resulted in low intensity data acquired for the complex in the presence of D₂O. Assuming a Guinier-range at slightly higher momentum transfer (Guinier_{Extended}: $1.0 < s \times R_g < 2.0$), the derived parameters display the expected tendency of indicating an $R_g \sim 23$ Å for Hfq_{DsrA} at 75% D₂O, where the RNA is matched. Using the same extended Guinier-region for Hfq_{DsrA} at 100% D₂O returns a value of $R_g \sim 28.8 \pm 0.2$ Å, which is similar to that of Hfq_{DsrA} at 0% D₂O $R_g \sim 29.8 \pm 0.3$ Å (Table II), as expected. However, Guinier's law is by definition shape-independent and applicable only in the low s -range, where the particle form-factor does not affect the scattering intensity. Therefore the parameters derived by assuming an extended Guinier-range can serve only for speculation.

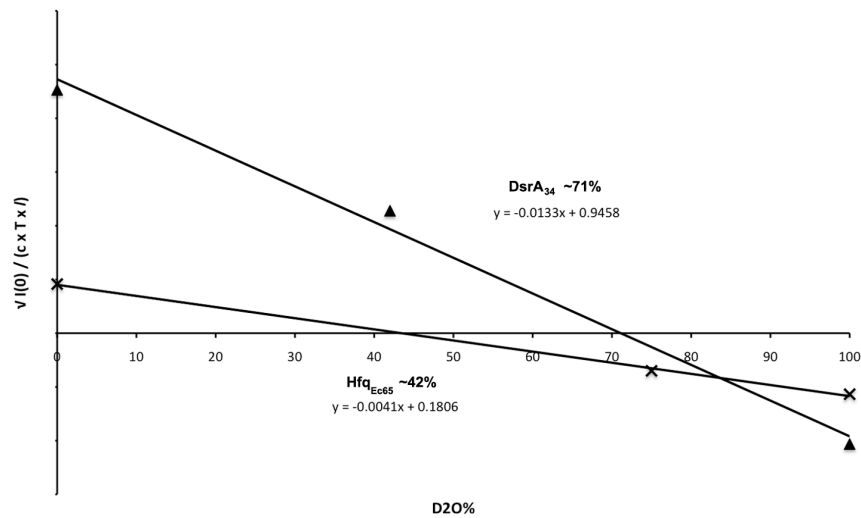


Fig.3.12: Experimental matchpoint determination. For Hfq_{Ec65} we find ~42% D₂O and for DsrA₃₄ ~71% D₂O.

3.3.4.1 Molecular mass estimate

The molecular mass (M_w) of the sample molecule can be derived from contrast variation data by comparison to the scattering of H_2O . When recording data at different contrast, the proportion of the scattering occurring from the different components of the complex changes as a function of their scattering length, governed by its chemical composition (see above). By normalizing the I_0 to concentration (c) and to the amount of neutrons transmitted relative to that of H_2O (T), and combining with an instrument specific expression for the difference scattering density (which will change at different contrast levels), it is possible to estimate the molecular mass of the sample particle, relative to that of H_2O ¹⁴⁴.

Retrieving estimates of molecular mass from this approach for the Hfq_{Ec65} protein alone at 0%, 75% and 100% D_2O , resulted in a reasonably stable value of 38 ± 5 kDa for the Hfq_{Ec65} protein alone (Hfq_{Ec65} $M_{w(Calc)} = 43.2$ kDa), again demonstrating the protein to exist as a hexamer in solution and establishing a proof of concept for the sample under investigation. For Hfq_{DsrA} no stable value for the molecular mass could be found for the three experiments in the contrast series. Based on the standard Guinier-analysis we calculate the molecular mass at 0% $D_2O \sim 73 \pm 7$ kDa, 75% $D_2O \sim 243 \pm 35$ kDa and 100% $\sim 105 \pm 13$ kDa.

First, the problem of selecting the correct value of extinction coefficient for the correct stoichiometry of the complex could have resulted in significant deviations of measuring the sample concentration. Reasoned by the stable value of R_g across the contrast series, the stoichiometry of the complex appeared to be the same in all experiments, and the extinction coefficient was therefore considered to be constant. By rationale of Lambert-Beer's law ($A = \epsilon \times c \times l$) the numerical value of absorption was then used directly as a measure for concentration when normalizing the intrinsic scattering. The concentration of the samples was assessed by direct measurement of the absorption at 260 nm and 280 nm on a Thermo-Scientific Nanodrop-1000 spectrophotometer, which has a very large dynamical range for linear absorption measurements, due to its very short light path (l). Absorption measurements were found in the range of 15 absorption units (a.u.) (max. 20 a.u.) when normalized to a 1 cm light path, which is possible to measure accurately with this type of spectrophotometer. Despite the generous dynamic range of this instrument, it is possible that high content of nucleic acid in the samples has led to absorption measurements outside the linear range. This would lead to faulty assessment of low sample concentration, which would lead to increased molecular mass. Assuming a 1:1 stoichiometry for the three samples of Hfq_{DsrA} the concentration was found in the range of $\sim 1-3$ mg/ml, where the expected value was ~ 7 mg/ml.

Secondly, the Guinier-analyses using different ranges of momentum transfer indicate a possible presence of precipitate in the complex sample at 75% D₂O. This would lead to increased scattering intensity in the low s -region, resulting in higher R_g and I_0 for this particular sample (Table II). The sample properties and the preparation procedures taken into consideration, it is possible that similar precipitates could have also been present in the other samples of Hfq_{DsrA}, which would be at base of the low data quality for SAS-experiments.

3.3.4.2 Interpreting scattering curves using perceptual criteria

Modeling of scattering data is essentially an ill-posed problem – as there will exist multiple models fitting the data equally well. Therefore, the approach is to make *a priori* assumptions about the sample particle. As proposed by Dmitri Svergun in 1992, using a few perceptual criteria it is possible to derive real space distribution function from the relative intensities, representing the scattering particle, “...even with poor experimental data.” (line 4, in ¹⁵¹).

For a monodisperse sample this, ‘distance’ or ‘pair’ distribution function ($P(R)$), will represent the probability (P) of finding two scattering bodies separated by given distance (R) inside the particle. The method has been implemented in the program GNOM¹⁵¹ that, under assumption of a correct estimate of the particle largest dimension (D_{MAX}), will optimize parameters for and carry out indirect Fourier transform, allowing for real-space evaluation of scattering data.

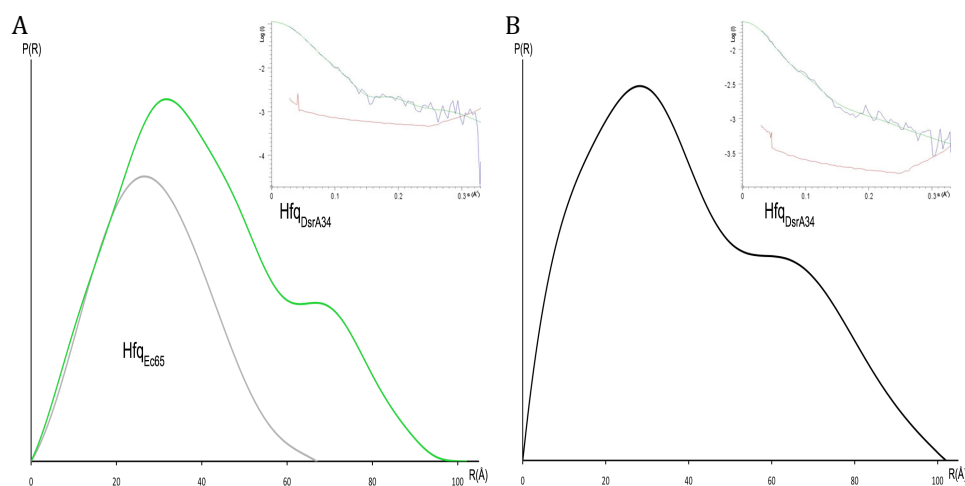


Fig.3.13: Real-space pair-distribution functions $P(R)$ calculated for SANS-curves of Hfq_{Ec65} and Hfq_{DsrA} using the program GNOM, restraining the function to zero at $D_{MIN} = 0$ Å and $D_{MAX} = 102$ Å. Inserted is the fit to the data in reciprocal space. **(A)** Overlay of $P(R)$ of Hfq_{Ec65} (grey) and Hfq_{DsrA} (green) at 0% D₂O. **(B)** $P(R)$ of Hfq_{DsrA} calculated from data acquired at 75% D₂O.

Notable difference can be found between the $P(R)$ based on the SAXS-data for Hfq_{Ec65} and Hfq_{DsrA} , as can be seen from Fig.3.7A (above). Here, in Fig.3.13A the $P(R)$ based on the SANS-data at 0% D_2O for Hfq_{Ec65} and Hfq_{DsrA} are shown, indicating a similar increase of the particle dimensions. The $P(R)$ was restrained to zero at $D_{MIN} = 0 \text{ \AA}$ and $D_{MAX} = 102 \text{ \AA}$, and the small insert display how the calculated real-space $P(R)$ fit the recorded data in reciprocal space. In Fig.3.13B the $P(R)$ based on Hfq_{DsrA} -data at 75% D_2O is shown, implying the same restraints. The shape of the curve resembles that of the complex at 0% D_2O , and together with the calculated particle $R_g \sim 34.1 \text{ \AA}$ that is increased relative to Hfq_{Ec65} (compare to Table II), this could be another indication that the RNA was not matched at 75% D_2O . However, despite the good fit to the data in reciprocal space (from a visual inspection) it is important to take into account that the calculated $P(R)$ for Hfq_{DsrA} at 75% D_2O was restrained to zero at 0 and 102 $\text{\AA}$. When releasing these restraints this $P(R)$ -function for Hfq_{DsrA} at 75% becomes un-physical (i.e. $P(R)$ deviates from zero at D_{MIN} and D_{MAX}), and no stable solution for can be found.

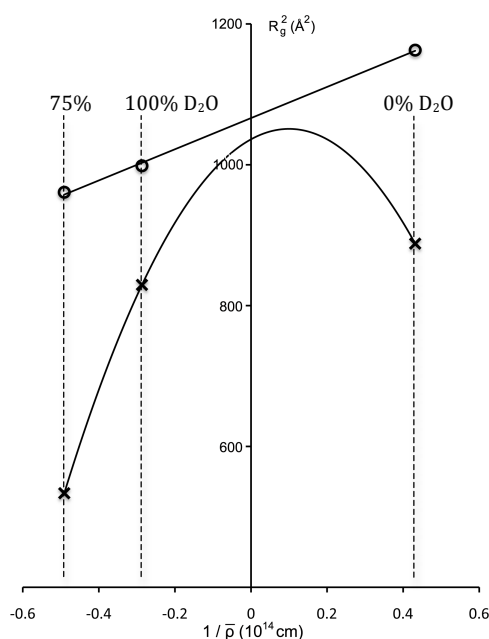


Fig.3.14: Stuhrmann plot of Hfq_{DsrA} data, based on parameters derived from standard Guinier-analysis ($\circ$), and on parameters assuming extended Guinier-region ($\times$).

Another approach to evaluate contrast variation data from a given complex structure was proposed by Heinrich Stuhrmann¹⁵². By plotting the square of R_g obtained at different contrast levels as a function of the designated contrast ($\bar{\rho}$), a graphic representation for the effect of each component on the overall particle R_g is achieved¹⁵³. In Fig.3.14

Stuhrmann plots of $Hf_{q_{DsrA}}$ assuming the two Guinier-ranges are shown. The parameters derived from a normal Guinier-analysis gives rise to an almost perfect linear behavior of the R_g inside the contrast series. Under the assumption that we match the RNA at 75% D_2O , this is an indication that the center of gravity of the particle remains stable throughout the contrast series. This interpretation of the data suggests that the RNA does not have an effect of the overall particle R_g (and shape), and is almost completely encased in the protein phase. Using the parameters from the extended Guinier-region ($1 < s \times R_g < 2$) the plot takes on a quadratic shape, indicating the RNA to affect the R_g (and shape) of the particle. However, in order to confirm this effect data must be acquired at more than three contrast levels.

The direct analysis achieved by the Stuhrmann-plot benefits from being based on a relatively simple Guinier-analysis, estimating the particle R_g . The standard Guinier-analysis for $Hf_{q_{DsrA}}$, by virtue of the linear fit of the Stuhrmann-plot, would indicate that the protein alone gives rise to the increased R_g in the complex. Considering the heat-stable quaternary structure of $Hf_{q_{Ec65}}$, such increase of R_g in this system would only be possible in a complex comprising minimum two hexamers of Hfq. Proposing the complex to comprise more than one hexamer of Hfq leads back to the question regarding the stoichiometry in the particle, and would not be in agreement with the increase in molecular mass and volume observed by SAXS, or the relaxation properties measured by NMR.

Furthermore, a $P(R)$ calculated for a monodispersed sample where two domains of similar size (i.e. two hexamers) are connected by a narrow linker like RNA, would take the shape of two peaks of equal height. The first would represent the *intra*-domain scattering from a single domain, the second the *inter*-domain scattering. These two peaks would be of similar height because the domains are of similar size¹²⁶. In the present $P(R)$ for $Hf_{q_{DsrA}}$ (Fig.3.13A+B) we observe a large peak with a big shoulder, which, in following a similar line of arguments, represents a globular domain with a long appendix extending away. The shoulder has its own maximum, which could indicate a smaller globular domain to exist at the end of the appendix. This interpretation of the $P(R)$ leads back to the initial study of Brescia *et al.*⁹², which indicated Hfq to prefer binding a fragment of the DsrA sRNA comprising a U-rich stretch followed by a hairpin. This initially led to the selection of the DsrA₃₄ RNA-fragment for study, which comprises the regulatory domain 2 hairpin preceeded by a short U-rich stretch.

In Fig.3.15 secondary structure predictions for the DsrA₃₄ RNA-fragment is shown, as calculated by the mFold web-server¹⁵⁴. Of the two proposed folds of DsrA₃₄, which are about equally stable, only the one on the left has the 5'-end U-rich stretch accessible. In conjunction with the P(R) suggesting a second, smaller globular domain to exist in the complex between Hfq_{Ec65} and DrsA₃₄, it is very satisfying to consider this fold of the RNA. The U-rich stretch comprises the Hfq binding site, and in complex with Hfq this fold of DsrA₃₄ is therefore most likely to be the dominant species in solution.

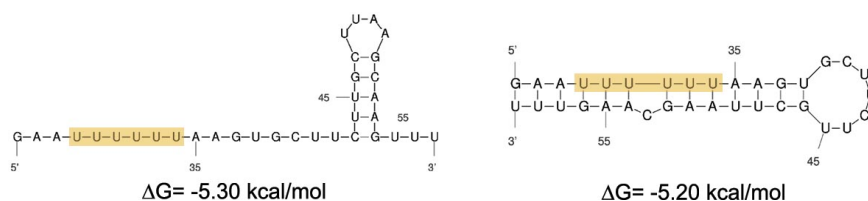


Fig.3.15: Secondary structure predictions of the DsrA₃₄ RNA-fragment of equal stability (ΔG), calculated by the mFold web-server¹⁵⁴. The 5'-uridine-rich Hfq binding site is marked in yellow.

3.3.4.3 Shape-reconstruction assuming monodispersity

Shape-reconstruction assumes the sample to be perfectly monodisperse, and the scattering-curve so to represent the density of the sample particle. Under the assumption of monodispersity, a proposed shape of same density can be reconstructed. If we were to assume the samples of the Hfq_{DsrA} complex to be monodisperse, a hypothetical shape of the sample particle can then be generated. However, it must be emphasised that due to the inconsistencies in deriving sample parameters, this can be regarded only as speculations on the overall model for the solution structure of the Hfq_{DsrA}-complex.

In Fig.3.16A a hypothetical model of Hfq_{DsrA} and its fit to the contrast variation data is shown. The model was generated ab-initio assuming overall P1-symmetry. The model incorporates the full six curves acquired for the protein and the complex at 0, 75 and 100% D₂O, with the D_{MAX} = 102 Å. It is important to emphasize the hypothetical nature of this model, which is reflected in the clearly speculative fit to the data for Hfq_{DsrA} at 75% D₂O.

The model fit the SANS-data for Hfq_{Ec65} with $X^2 \sim 1.38$ at 0%, $X^2 \sim 1.83$ at 75% and $X^2 \sim 1.44$ at 100% D₂O. For Hfq_{DsrA} the fit display higher discrepancy, with $X^2 \sim 1.28$ at 0%, $X^2 \sim 3.53$ at 75% and $X^2 \sim 2.52$ at 100% D₂O, reflecting the samples increased complexity. In Fig.3.16B (from bottom to top) the development of the model can be followed. In the lower panel (Fig.3.16B.1) the crystal structure of Hfq_{Ec65} is shown, with the N-termini modeled by SAXS. In Fig.3.16B.2, Hfq_{Ec65} is superposed onto the protein-phase from SANS shown in blue spheres, as it appears when modeling the complex combining all six curves. The hexameric shape of the conserved Hfq-core emerges clear in this model. In Fig.3.16B.3, green spheres on top of the protein-phase represent the RNA extending away, satisfying D_{MAX} of the particle. The modeled volume of protein and RNA was of equal size. In Fig.3.16B.4 the resulting rigid-body model of Hfq_{DsrA} from SAXS is superposed onto the ab-initio model from SANS. Reasonable agreement indicates some stability of the overall shape of these two models, despite their independently speculative fit to respective data sets.

The SANS-models were calculated leaving the volumes parameters variable, keeping only restraints on the particle largest dimension D_{MAX} and the R_g of the protein-phase.

In the outcome model the two phases of the particle are about same size, which is not consistent with a 1:1 complex. As estimated by tabulated values for the partial specific volume ($\bar{v}$) for globular protein ($7.40 \cdot 10^{23}$ Å³ g⁻¹) and nucleic acid ($5.40 \cdot 10^{23}$ Å³ g⁻¹), in a 1:1 complex between Hfq_{Ec65} (43.200 g/mol) and DsrA₃₄ (10.800 g/mol), the protein should make out about 85% of the total volume. As with the shape-reconstruction of the SAXS-data, the volume of the RNA-phase in

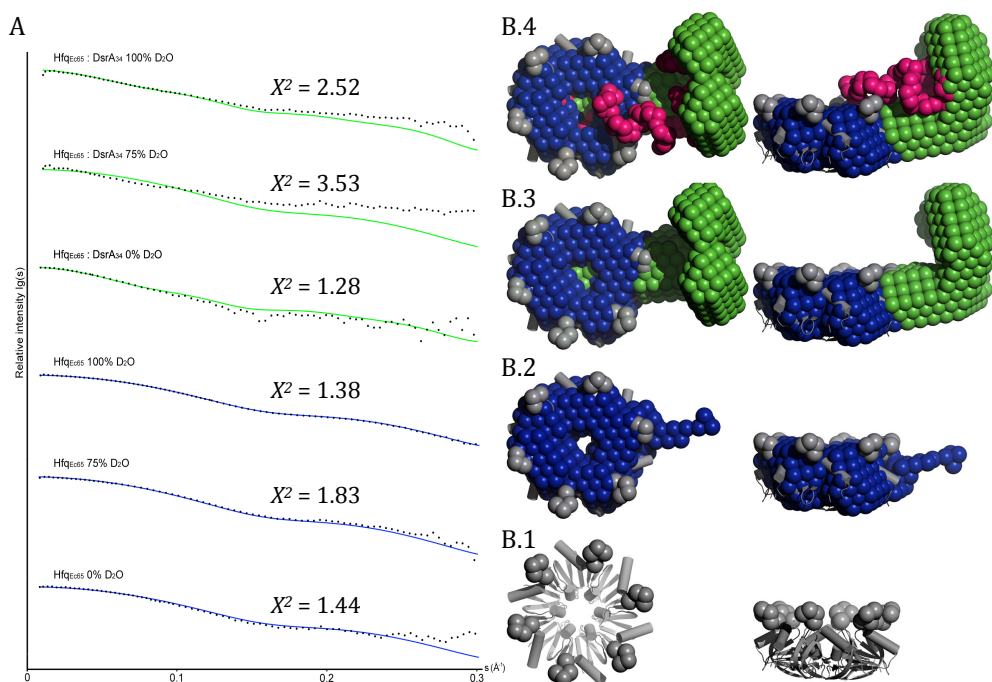


Fig.3.16: Neutron small angle scattering acquired for Hfq_{Ec65} and Hfq_{DsrA}. **(A)** The data shown in black dots acquired at three contrast levels 0%, 75% and 100% D₂O, overlaid with curves (blue: Hfq_{Ec65}; green: Hfq_{DsrA}) representing hypothetical models from shape-reconstruction. The curves were shifted along the intensity scale. **(B.1-4)** Shape-reconstruction of Hfq_{DsrA} assuming monodisperse samples, by the program Monsa. *B.1*: The crystal structure of Hfq_{Ec65} with the N-termini reconstructed from SAXS. *B.2*: Crystal structure overlaid the protein phase from SANS in blue spheres. *B.3*: Model of Hfq_{DsrA} with RNA represented by green spheres. *B.4*: superposition of the resulting models of the complex from SAXS and SANS.

the SANS-model appears rather large in comparison to the protein. When modeling Hfq_{DsrA} by SAXS-data the protein is represented by the known crystal structure, as a rigid-body. The RNA is represented by a number of dummy residues, which we increased to a maximum to improve the fit in the low s -range. Similarly, in the present hypothetical ab-initio P1-model based on the SANS-data, the protein-phase is restrained by its known R_g , and without restraints on the volumes of the phases, the volume of the RNA-phase is increased in the ab-initio model to compensate a poor fit in the low s -range.

3.3.4.4 Volume-restrained shape-reconstruction

In Fig.3.17 shape-reconstructions assuming monodispersity of Hfq_{DsrA} enforcing volume-restraints in accordance with a 1:1 stoichiometry in the complex are shown. The top panel shows the rigid-body model based on SAXS-data with the RNA represented by a total of 68 dummy residues, two per nucleotide instead of three as above. To the right, iterative modeling (10x) displays the RNA extending away from the

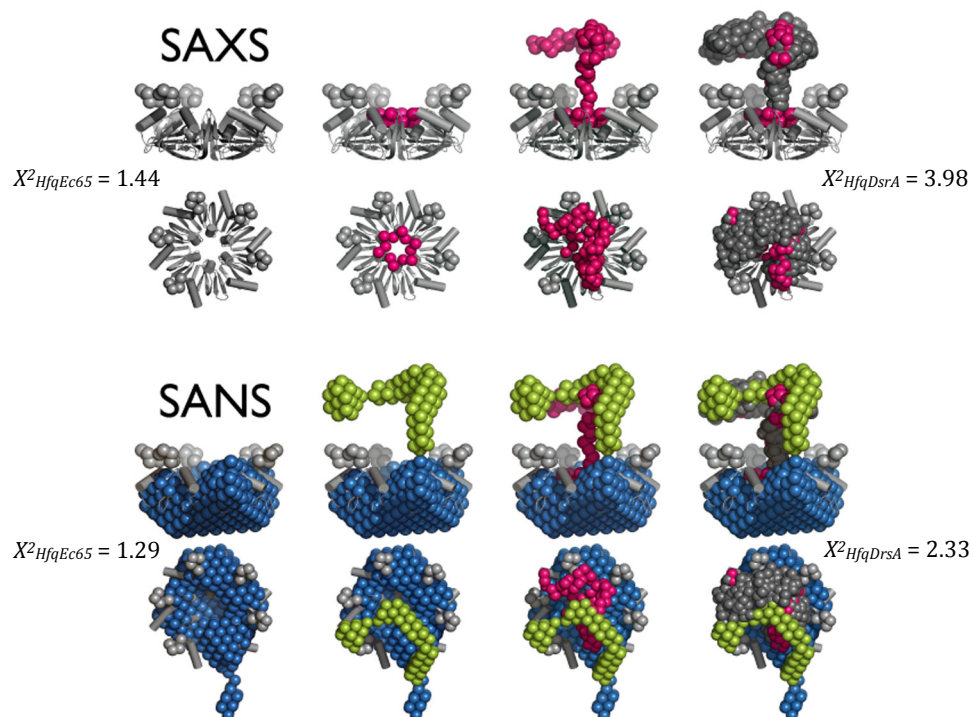


Fig.3.17: Shape-reconstruction of Hfq_{DsrA} implying volume restraints of a 1:1-complex. X^2 -values for the data in H₂O have been noted. *Top-panel*: Rigid-body model based on SAXS-data. In grey is shown the crystal structure of Hfq_{Ec65} with the N-termini modeled by SAXS, with the RNA represented by 68 dummy residues (pink). To the right is shown how iterative modeling leads to a 'horseshoe-shape' of the RNA (black). *Bottom-panel*: Super-position of ab-initio model based on SANS, onto the SAXS-models. Shown in blue spheres the protein-phase resembles the crystal structure of hexameric Hfq_{R65}, and in green spheres the RNA-phase agrees with that of the SAXS-model.

protein, adopting a shape reminiscent of a horseshoe. The SAXS-data for Hfq_{Ec65} is fit with $X^2 \sim 1.44$ and Hfq_{DsrA} $X^2 \sim 3.98$. The lower panel display an ab-initio model based on SANS-data restrained by the expected volume-fraction of the protein (85%) and RNA (15%), $D_{MAX} \sim 102$ Å and the known R_g of Hfq_{Ec65} ~ 23 Å. In agreement with the model based on the SAXS-data, the RNA can be seen emerging from the proximal face of the hexameric protein. This model fit the SANS-data for Hfq_{Ec65} with $X^2 \sim 1.29$ at 0%, $X^2 \sim 2.11$ at 75% and $X^2 \sim 1.74$ at 100% D₂O; and for Hfq_{DsrA} with $X^2 \sim 2.33$ at 0%, $X^2 \sim 3.48$ at 75% and $X^2 \sim 4.32$ at 100% D₂O. In Fig.3.18 the fit to the SANS-data of the two models with and without volume-restraints are compared. The upper panel shows the fit of the model restrained by volume, displaying clear discrepancy in the fit to the data. In the lower panel the fit of the model generated without volume-restraints is shown, displaying a better fit to the experimental data. However, none of the models fit the data at 75% D₂O. Whereas

from a visual inspection, the two models display about equal discrepancy for the data at higher s -range ($s > 1.5$), especially in the low s -range ($s < 0.1$), the fit is improved with increased volume of RNA. This demonstrates how the increased density of RNA improves the fit.

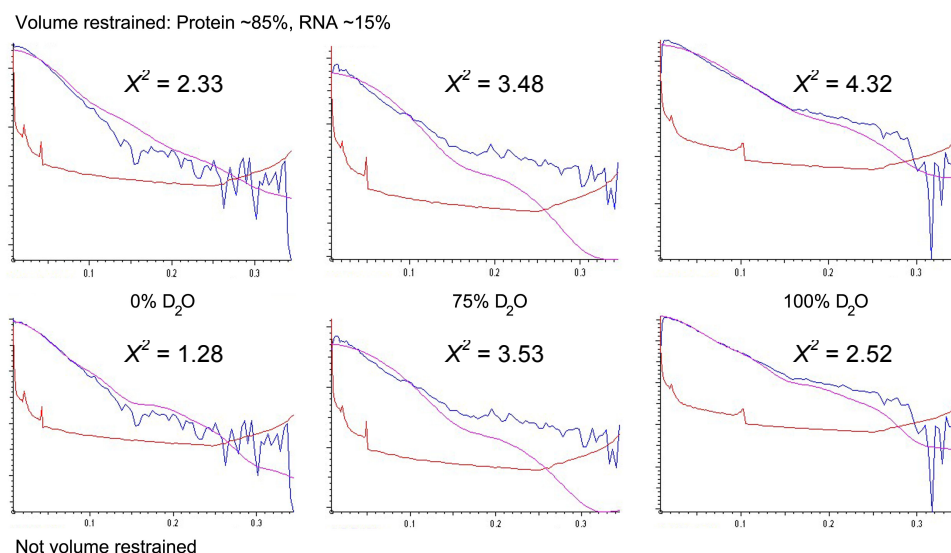


Fig.3.18: Comparison between the SANS-data for Hfq_{DsrA} (in blue) and theoretical scattering-curves (in pink) calculated for the models generated with volume-restraints of Hfq_{DsrA} (*upper*) and without (*lower*). The axis denotes momentum-transfer ($\text{\AA}^{-1}$) plotted against logarithm of scattering intensity. The χ^2 -values for Hfq_{DsrA} have been noted.

3.3.5 Summary & Concluding remarks

From the complementarity of the results presented above, it was sought to meet the goal of establishing a shape-model of RNA, binding on the C-terminally truncated *E.coli* Hfq protein.

The presented SAXS-data support the expected ~25% increase in molecular mass and ~15% increase in particle excluded volume, in a complex of 1:1 stoichiometry between hexameric Hfq_{Ec65} and DsrA₃₄. However, the shape-model based on the SAXS-data does not satisfy the scattering at low angle, despite implying a generous volume for the RNA.

Transverse ^{15}N -relaxation experiments by NMR indicate a ~25% increase in hydrodynamic volume of the particle in the presence of RNA, which is in agreement with a 1:1 complex. Comparative analysis of ^{15}N -HSQC NMR-spectra for Hfq_{Ec65} and Hfq_{DsrA}, as expected, maps the interaction of RNA primarily to the proximal face of Hfq_{Ec65}.

The SANS-data recorded from contrast variation experiments lends further speculations to the validity of the proposed 1:1 model. The

SANS-data is of low quality, but interpretation of the data under assumption of established perceptual criteria sheds further light on the solution behavior in the Hfq_{DsrA}-complex.

In SAS-experiments the recorded scattering is a synthesis of the scattering of all solute components in the sample, subtracted the solvent. The intensity of the scattering increases roughly with the power of six of the radius of the solute, and even a very small amount of aggregation can therefore potentially have a huge effect on the result. In the NMR-experiments such aggregation will not be detected, and in consideration of the sample properties and preparation procedures, the NMR-result is probably more reliable for evaluating the complex stoichiometry.

Based on the SAS-data two sets of shape-models have been generated that are mutually in agreement independent of volume-restraints. However, the shape-model best fitting the SAS-data does not agree well with the NMR relaxation experiments. Both sets of shape-models could in principle agree with the chemical shift pattern from NMR, from which the most interesting observation undoubtedly is the structuring of the N-terminus in complex with RNA.

The N-terminus of Hfq is widely conserved in bacterial species, and previous functional studies have suggested it to be involved in the function of Hfq⁴⁶. The DsrA₃₄ RNA carries in its 3'-end a self-complementary region that forms the domain 2 hairpin of the full-length sRNA. The DsrA₃₄-fragment was selected for studies based on the assumption that the hairpin would be involved in the binding to Hfq, in following of the findings by Brescia *et al.*⁹². However, the mFold-server¹⁵⁴ predicted an alternative secondary structure for the DsrA₃₄-fragment, with a shorter hairpin at its 3'-end. At distance to the expected U-rich Hfq binding site, no steric hindrance would appear to be obstructing the formation of this 3'-hairpin while DsrA₃₄ is in complex with Hfq from the U-rich 5'-end. Yet no imino-signals was observed for the Hfq_{DsrA} complex by NMR, which could indicate a dynamic process in the RNA that lead to line broadening of the NMR-signals. A process that could prevent detection of imino-signals could be the reversible opening and closing of the RNA hairpin. concomitantly with the disordered N-terminus of Hfq becoming structured, the hairpin of DsrA₃₄ unfolds. Such a model would be in agreement with the 'entropy transfer model', previously proposed by Tompa *et al.*⁹⁸ to be of particular application for RNA-chaperones, and in a process of isothermal entropy/enthalpy-compensation this is seamlessly reversible.

Furthermore, that the RNA is self-complementary could in principle lead to RNA oligomerization and nucleate the aggregation, which may potentially have compromised the quality of the all the SAS-data recorded for the complex. SAXS experiments on the DsrA₃₄ RNA alone have previously demonstrated DsrA₃₄ to form large aggregates (data not shown), and recent experiments by DLS have demonstrated the RNA to most likely exist as a dimer at lower temperature, which dissociate as temperature increases (Dr. Euripides Almeida de Rebeiro; personal communication).

From the present study it has been established that binding of the DsrA₃₄ RNA-fragment leads to a structuring of the conserved N-terminus of Hfq. The dynamic nature of the proposed protein:RNA complex can explain why attempts to crystallize the Hfq_{DsrA} complex so far have failed.

4 Structure of Hfq_{Ec65} complex with ATP

Synopsis

The crystal structure of C-terminally truncated *E.coli* Hfq, have been solved in complex with ATP. As expected ATP is coordinated in the distal R-site, similar to the published structure of a complex between Hfq and poly(A)-RNA. Two unique RNA-binding sites have been identified on each face of the Hfq hexamer. On the proximal face resides the uridine specific binding site that has been speculated to also accommodate adenine and cytosine bases, and on the distal face the tripartite A-R-N-site has been shown to accommodate more variability in the sequence. Intricate analysis of the present ATP-complex structure and comparison with other published complex structures, together indicate degrees of freedom in the R-site binding crevice and suggests possible alternative modes of RNA-coordination to be conserved on the distal face of *E.coli* Hfq.

4.1 Preamble

Hfq is classified as an Sm-like protein, adopting the classic OB-fold with a N-terminal helix, followed by five anti-parallel β -strands forming a distinctly bent β -sheet in an SH3-like domain. It adopts a highly stable quaternary structure as a doughnut-shaped homo-hexamer around a central pore. A modified Walker-A motif (-GXXXXGKT-)¹⁵⁵ was proposed identified to the sequence -Gly₂₉IKLQGQI₃₆-, residing on the distal face of the Hfq hexamer⁷⁵. Further work by Arluison *et al.*⁵⁰ supported the location of the nucleotide binding pocket by mutational studies, demonstrating the importance of a π -stacking interaction between Tyr-25 and the nucleotide base. In their proposed model of ATP-binding, the ribose-phosphate is pointing towards the rim of the hexamer, with no possibility for interaction with the modified Walker-A motif.

Here we report the crystal structure of adenosine-tri-phosphate (ATP) binding in the purine-pocket (R-site) of the tripartite binding site (A-R-N site), on the distal face of Hfq protein, as identified by Link and co-workers¹¹¹ in the crystal structure by of *Escherichia coli* Hfq (Hfq_{Ec}) in complex with poly-A₁₅ RNA oligonucleotide. Baba *et al.*¹¹² further established the location of the bindingpocket, in the crystal structure of *Bacillus subtilis* Hfq (Hfq_{Bac}) in complex with a heptameric RNA oligonucleotide (5'-AGAGAGA-3'). The R-site binding pocket is composed of residues from β -strands of two adjacent subunits of the Hfq hexamer (Fig.4.1 insert). It is comprises a crevice formed of hydrophobic and aromatic residues on its sides (Tyr-25, Leu-26', Ile-

30', and Leu-32'; prime' denoting residues from the adjacent subunit) and, with H-bond partners at the bottom and cervixes (Asn-28', Gly-29, Gln-52', Ser-60, Thr-61).

In complex with *E. coli* Hfq, every third nucleotide of the poly-A₁₅ oligonucleotide is inserted into the bindingpocket. In complex with *Bacillus subtilis* every second nucleotide of the (5'-AGAGAGA-3') oligonucleotide is inserted in the crevice (see Fig.4.1 insert), suggesting a similar bindingmode possible for *E. coli* Hfq.

4.2 Results & Discussion

4.2.1 R-site coordination of ATP

Crystals were obtained from a C-terminally truncated fragment, comprising residues 1-65 of *E.coli* Hfq (Hfq_{Ec65}) in complex with ATP. Data collection and refinement statistics are summarized in Table III. Model coordinates and structure factors have been deposited in the PDB under the code 3Q03.

The crystal structure of Hfq_{Ec65} in complex with ATP presented here does not support the model proposed by Arluison *et al.*⁷⁵. In our crystal data, the adenine moiety is turned around 180° with the ribose-phosphate pointing towards the central pore of the hexamer. Taken together our data, those of Link *et al.*¹¹¹ and Baba *et al.*¹¹², confirm this mode of binding and rule out the alternative modeled binding mode.

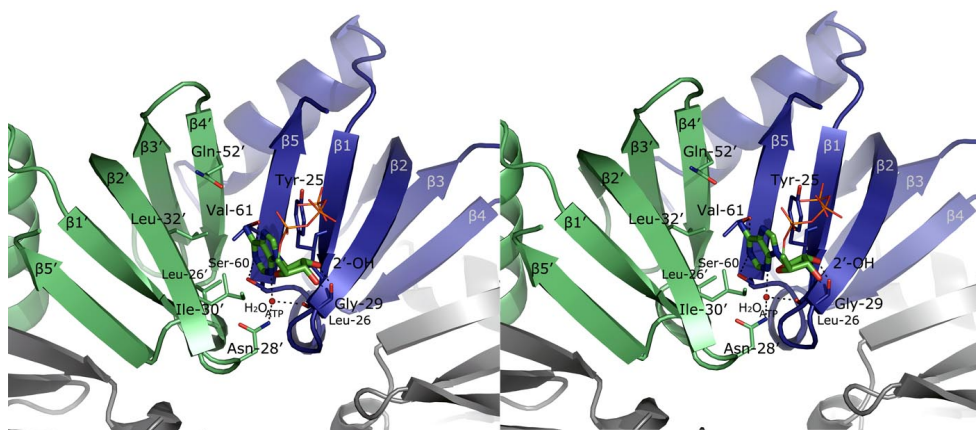


Fig.4.1: A stereoview of the R-site purine binding site between adjacent monomers, on the distal face of *E.coli* Hfq. Side-chains of binding site residues are shown as sticks. A single ATP-moiety is colored in green, with the triple-phosphate protruding in the favored conformation (orange lines). Inserted in the upper right corner is the Hfq hexamer shown from distal face, with the two known binding-modes of RNA overlaid, adopted from the structures of Hfq_{Ec}-Poly-A₁₅ (magenta; PDBid: 3GIB) and Hfq_{Bac}-AGAGAGA- (yellow; PDBid: 3HSB).

Superposition of Hfq_{Ec65} hexamer with the apo-structure (PDBid:1HK9)¹⁰⁵ exhibits an overall r.m.s.d of 0.53 Å over 360 equivalent Cα atoms, thus suggesting no conformational changes upon ATP binding. Fig.4.1 displays a single ATP-unit inserted into the purine binding-site (R-site), on the distal face of the *E.coli* Hfq hexamer.

In the present structure, density corresponding to ATP was observed in four of six binding sites. In three of the four occupied binding sites the ATP is bound similarly as in complex with poly-A₁₅, as reported by Link *et al.*¹¹¹. In the R-site binding pocket the adenine nitrogens are engaged by two H-bonds to side chains (exocyclic N6:O_ε Gln-52' and N1:O_γ Thr-61) and one water-mediated H-bond to N3, coordinated by the backbone carbonyl of Leu-26 and the side chain of Asn-28' (Fig.4.2). In addition, aromatic stacking interaction with Tyr-25 is supported by favorable hydrophobic contacts with the side chains of Leu-26', Ile-30', and Leu-32' of the adjacent subunit (Fig.4.2).

One of the adenosine-units (one out of four occupied binding pockets) is tilted further towards the central pore formed by the hexamer, when compared to the poly-A₁₅ structure. This gives rise to a different hydrogen-bonding pattern (Fig.4.2). In this particular ATP binding site, the adenine exocyclic N6 is H-bonded by Thr-61 O_γ, while N1 is H-bonded to Ser-60 O_γ. Additionally, in the poly-A₁₅ structure a tightly bound water molecule, coordinated by the carbonyl oxygen atom of Leu-26 and side chain of Asn-28', mediates a hydrogen bond to N3 of the bound nucleotide (Fig.4.2; H₂O_{Poly-A}). In the ATP binding site shown in Fig.4.2, this tightly bound water moves in concession with the tilt of the ATP moiety, apparently disrupting the hydrogen-bonding network between the solvent molecule and N3 atom of adenine (Fig.4.2; H₂O_{ATP}). In the region of Gln-52' (towards the rim of the hexamer) two new water molecules are observed, mediating contact to the exocyclic N6 of the tilted ATP moiety (Fig.4.2; W1+W2).

Common to all the binding sites is the formation of a hydrogen bond between the adenosine ribose 2'-OH and carbonyl-group of Gly-29 (Fig.4.2). This interaction was reported to likely convey selectivity for RNA over DNA for this binding-site¹¹¹. Our results corroborate the importance of binding the 2'-OH of the ribose sugar to Gly-29 carbonyl oxygen atom. With its φ/ψ-angles falling in the region around ~85° and ~-10° respectively, Gly-29 adopts main-chain conformation allowed only for glycine residues, which do not bare side-chains. This suggests that any kind of substitution of the glycine residue at this position would dramatically reduce Hfq's binding capacity to ATP.

In all ATP units bound to Hfq_{Ec65} the ribose is modeled in a staggered conformation. The tri-phosphates have in all units been modeled with full occupancy, and several possible conformations were observed for the phosphates of each ligand. The conformation of seemingly highest

electron density was chosen for all ligands, resulting in the phosphates of two ATPs to be involved in crystal contacts, binding to Arg-19 of symmetry related protomers. The general interpretation should though be that the tri-phosphates are present (ie. not hydrolyzed) with variable conformations, and are not involved in specific interactions with Hfq_{Ec65}.

4.2.2 Comparison to coordination of RNA

The structure of the Hfq_{Bac} in complex with a short RNA-oligonucleotide (5'-AGAGAGA-3')¹¹² has demonstrated new properties of the R-site in binding of adenosine. With a sequence homology of 45% and similarity of 73%, the overall r.m.s.d. between *E.coli* and *B. subtilis* (calculated for C $_{\alpha}$ -atoms aa 8-65 of PDBid: 3HSB and *E.coli* Hfq in complex with ATP over 348 equivalent C $_{\alpha}$ atoms) is 0.54 Å. This suggests a high degree of structural conservation of Sm-like fold across strains, as well as no structural changes upon nucleotide binding.

Comparing the structures of *E.coli* and *B. subtilis* Hfq (Fig.4.2), two amino acid substitutions are found in the distal R-site binding site (Tyr-25 $\rightarrow$ Phe-24 and Ile-30' $\rightarrow$ Phe-29'), which upgrades the binding site π -stacking properties to a π -sandwich. Namely, while in *E. coli* Hfq the aromatic adenosine ring is flanked by Tyr-25 and Ile-30', in *B. subtilis* Hfq the nucleotide aromatic moiety is stacked between two aromatic side-chains (Phe-24 and Phe-29' as numbered for Hfq_{Bac}). A calculated r.m.s.d. of 0.53 Å / 0.53 Å for C $_{\alpha}$ / C $_{\beta}$ for R-site residues (*E.coli*: Tyr-25,

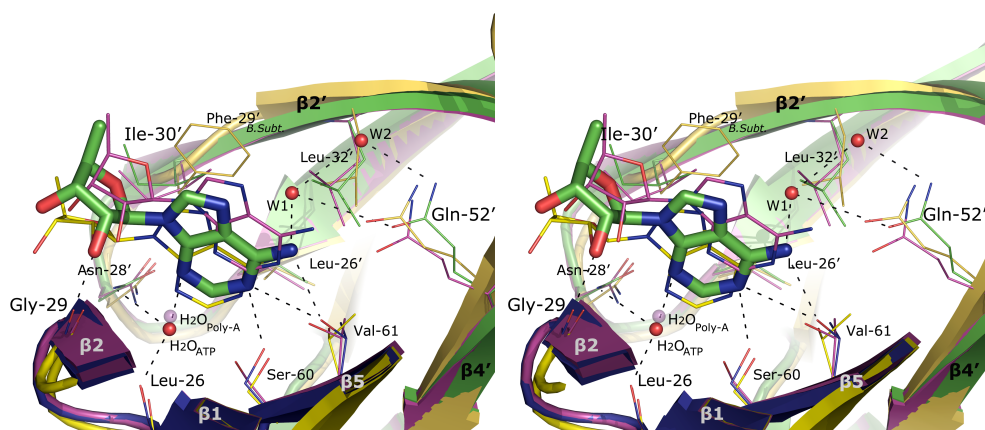


Fig.4.2: A stereo view into the R-site purine binding pocket from the perspective of Phe-25 (removed for clarity). ATP is represented as green sticks (phosphates removed), and the adenosines of Hfq_{Ec}:Poly-A₁₅ (magenta) and Hfq_{Bac}:AGAGAGA- (yellow) as thin sticks. Binding site residues are shown as wire-frame in respective color. Residues coordinating ATP is noted by stipulations. The H-bond mediating water in the Hfq_{Ec65}:ATP structure is located between Leu-26 and Asn-28', with the equivalent water from Hfq_{E.coli}:Poly-A₁₅ in the background. Behind the nucleotide moieties the Phe-29' of Hfq_{Bac} (yellow stick) overlaps with Ile-30' of Hfq_{Ec}.

Leu-26', Asn-28', Gly-29, Ile-30', Leu-32', Gln-52', Ser60, Thr61 and *B. subtilis*: Phe-24, Leu-25', Asn-27', Gly-28, Phe-29', Leu-31', Gln-52', Ser-60, Thr-61) indicate little to none structural changes in the binding site, upon introduction of an additional bulky aromatic group.

Comparing the Hfq_{Ec65}:ATP-complex structure to the Hfq_{Bac}:RNA oligonucleotide complex, a single ATP-unit is coordinated in a manner similar to adenines in *B. subtilis*, which are all moved further towards the central pore of the hexamer when compared to the Hfq_{Ec}:poly-A₁₅ structure. This is possibly an effect of inter-protomer crystal contacts in the *B. subtilis* structure mediated by the RNA oligonucleotide. The slight translation of the adenine disrupts the interaction between O_ε atom of Gln-52' and the adenine exocyclic N6, similar to ATP. The tightly bound water between Leu-26 and the side chain of Asn28' (Fig.4.2; H₂O_{ATP}) is not observed in the *B. subtilis* structure.

Overall comparison between the published structures of Hfq_{Ec}:Poly-A₁₅ (PDBid: 3GIB) and Hfq_{Bac}:AGAGAGA- (PDBid: 3HSB) in combination with the here presented Hfq_{Ec65}:ATP co-crystal structure, indicates that *E.coli* Hfq possibly could bind RNA in a mode similar to that observed in the *B. subtilis* structure, with every second purine-base inserted into the binding-site crevice.

Table III

Data collection and refinement statistics	
Beamline	ID14.1 (ESRF)
Wavelength (Å)	0.934
Resolution (Å)	98.67 - 2.15 (2.21 - 2.15)
Space group	C2
Unit cell (Å,°)	a = 104.30 b = 40.62 c = 100.81 β = 101.8
Molecules / a.u.	6
Unique reflections	21689 (1484)
Completeness (%)	99.81 (88.9)
R _{meas} ^b	0.079 (0.469)
R _{merge} ^c	0.111 (0.432)
Multiplicity	5.61 (5.62)
I/sig(I)	17.96 (3.70)
R ^d / R _{free} ^e	0.2294 / 0.2573
R.m.s.d. bonds (Å)	0.008
R.m.s.d. angles (°)	1.22

$$R_{\text{cryst}} = \sum |F_o - F_c| / \sum F_o$$

^eR_{free} is the cross-validation R_{factor} computed for the test set of reflections (5%) which were omitted in the refinement process.

where I is the intensity of an individual reflection and <I> is the mean intensity of same reflection.

$$\hat{I}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} I_{hi}$$

$$R_{\text{merge}} = \frac{\sum_h \sum_i |I_{hi} - \hat{I}_h|}{\sum_h \sum_i I_{hi}} ; \text{ where } \hat{I}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} I_{hi}$$

^bValues in parentheses are for the highest resolution shell.

4.3 Methods & Materials

4.3.1 Protein expression & purification

The C-terminally truncated Hfq protein from *Escherichia coli*, encompassing residues 1-65 (Hfq_{Ec65}) out of 102 was cloned as described in⁹⁹. The construct containing the *hfq65* gene was obtained by means of PCR using the forward primer (5'-GCTCTAGAAATATAATAGTTTAACTTTAAGAAGGAGATATACATATGGCT AAGGGGCAATCTTTACAAGATCCGTTCCCT-3'), containing a XbaI site (*italics*), and the reverse primer (5'-TTTTTTGAATTCTTACTA **AGACGGGACAACAGTAGAAAT** CG-3'), which contains two stop codons (*underlined*) after the triplet encoding Ser65 (**bold**) as well as an EcoRI site (*italics*). The PCR product was cleaved with XbaI and EcoRI and ligated into the corresponding sites of plasmid pUC19 (New England Biolabs Ltd., USA). From the resulting plasmid pUHfq65, the *hfq65* gene was re-isolated by cleavage with PvuII and then ligated into the EcoRV and NruI sites of pACYC184, yielding plasmid pAHfq65. The resulting plasmid was transformed in the expression strain *E.coli* BL21 (DE3) (Invitrogen) by the heat-shock method.

Protein expression was done by inoculating 1 l Luria-Bertani broth, replenished with 100 µg/ml ampicillin (Sigma-Aldrich) for selection, with 10 ml over-night culture. The expression culture was grown to an OD₆₀₀ ~ 0.6 - 0.8 and over-expression was induced with addition of 0.5 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) for about 3.5-4 hrs, at 37°C. Cells were harvested by centrifuging at 4000 x G for 15 min.

Cells were lysed by french-press at 4°C, into 25 ml / 1 l culture of 50 mM tris-HCl pH ~ 7.4, 1.5 M NaCl, 250 mM MgCl₂, 1 mM phenylmethanesulfonylfluoride (PMSF), 0.5 mM β-mercaptoethanol (BME). After lysis 25 µg/ml of DNase I (Sigma-Aldrich) was added on ice for 20 min, and the lysis-solution was cleared by centrifugation at 45.000 x G for 15 min.

Further purification steps involved initial heat-fractionation followed by a series of FPLC-based chromatographic steps (hydrophobic interaction chromatography, immobilized metal ion chromatography) with a final step of size-exclusion-chromatography, as described in Beich-Frandsen M. *et al.* 2010¹¹⁸. The purified protein was concentrated in Centricon spin-filters (Amicon) to final 15 mg/ml in a buffer containing 50 mM tris-HCl pH 7.4, 200 mM NaCl.

4.3.2 Crystallization & Data collection

Prior to crystallization the protein was mixed with adenosine-triphosphate (ATP; Sigma-Aldrich) in a 1:1 molar-ratio with respect to one subunit of Hfq_{Ec65}. The complex solution was centrifuged at 44.000 x G for 1 hr, to remove and macro-scale impurities and protein aggregates. Nano-litre crystallization trials with JCSG+-screen (Qiagen¹⁵⁶) were set by robot in 96-well vapor diffusion sitting-drop format. Diffraction quality crystals were obtained directly from the JCSG+-screen, and data was collected from a crystal grown in 0.1 M hepes pH ~7.5, 10% (w/v) PEG-8000 and 8% (v/v) ethylene glycol. Presence of ATP in the nano-litre crystallization experiments had a dramatic effect on the success-rate of crystallization, which increased roughly 20-fold. No other single nucleotide displayed the same effect, but did also not antagonize crystallization when compared to the apo-protein. Crystals were also obtained in presence of GTP, CTP and UTP, but no electron density for the ligand could be observed in any of the structures.

Crystals were harvested using cryo-loops of suitable size (Hampton) and flash-cooled in liquid nitrogen without usage of any additional cryo-protectant. A complete dataset was collected at the beamline ID14.1 at the European Synchrotron Facility (ESRF), and processed with the XDS-suite¹⁵⁷ in space group C2, to a maximum resolution of 2.15 Å. Data-collection and refinement statistics can be found in Table III.

4.3.3 Phasing, Model building & Refinement

The structure was solved by molecular replacement, using Molrep¹⁵⁸ from the CCP4-suite¹⁵⁹, with one Hfq protomer subunit from the Hfq hexamer as search-model (PDBid: 1HK9, aa 5-65)¹⁰⁵. The rotation function had six distinct solutions, corresponding each to one of the subunits in the hexamer.

Refinement was performed by phenix.refine in the Phenix-suite¹⁶⁰ employing simulating annealing and energy minimization cycles. Initially tight geometry restraints were applied, releasing them gradually as the R and R_{free} factors diverged. After several rounds of iterative refinement and model building with COOT¹⁶¹, clear density for ATP-molecules was found in four of six binding-sites. Addition of ATP to the model (with occupancy fixed at 1) caused the R and R_{free} factors to converge, and refinement was continued with increasing restraints, until no significant change in R-factors was obtained. A final step of refinement by Refmac^{162,163} in the CCP4-suite¹⁵⁹, was done to polish the weighting between X-ray and stereochemical terms for minimizing the residual. Final validation of the models was performed with MOLPROBITY¹⁶⁴.

4.3.4 Structural analysis and Superpositions

Structural comparison and superposition was done in Pymol¹¹³, using algorithms herein. Pymol was also used for generation of figures. Analysis of the R-site binding pocket was performed using the Protorp server (<http://www.bioinformatics.sussex.ac.uk/protorp/>)¹⁶⁵. Homology between *E. coli* and *B. subtilis* Hfq sequences was quantified using EMBOSS¹⁶⁶.

5 Discussion & Perspectives

Due to its versatile nature the *E. coli* RNA-chaperone Hfq is characterized as a pleiotropic regulator, a quality that likely resides in its partially disordered and flexible C-terminal domain. In contrast to a completely disordered protein, Hfq has a structured Sm-like core as scaffold to support its ubiquitous function. The very stable oligomeric assembly provides an overall doughnut shape that is conserved in many bacteria without significant terminal extensions, and is *per-se* alone sufficient for certain riboregulatory events. With high entropy disposition and the ability to adopt multiple conformations, the terminal domain allows the *E.coli*-type of Hfq additional abilities for interaction.

The work presented in this thesis has shown how *E. coli* Hfq can change shape, demonstrating its rheomorphic nature. The SAXS-model of full-length Hfq displays the C-termini extending laterally away from the structured Sm-core¹¹⁸. The crystal structure of Hfq_{Ec} indicates the C-termini to reside on top of the proximal face¹⁶⁷, in agreement with previous EM-results¹¹⁵. Together with NMR-results that showed the C-terminus to interact with specific residues in the hexameric core, this suggests a possible dynamic transition of the whole C-terminal domain, from lateral to proximal position, with possible implications for function. Interestingly the C-terminal domain, which is required for *E. coli* Hfq to bind RNAs above a certain length, only attains structure when interacting with such long RNAs, as demonstrated by the presented SR-CD studies¹¹⁸. So far no structural information is available on the C-terminus alone or in complex with RNA. However, NMR-studies presented here indicated that the C-terminus switches between a partially structured and disordered state, which demonstrates how the amino acid composition fine-tunes the structural propensity of the C-terminal domain. The structural and functional studies of Hfq_{Ec} showed the structural variability in the latter half of the C-terminus to correlate with function⁹⁹, indicating higher entropic potential residing in the end of the C-terminus is important for function¹¹⁸.

The N-terminus of Hfq was reconstructed by SAXS, demonstrating how the shorter and highly conserved, intrinsically disordered N-termini of Hfq_{Ec65}, which are important for the proper function of Hfq, also have an effect on the overall shape of the protein. Furthermore, small angle scattering studies of the complex between Hfq_{Ec65} and DsrA₃₄ (Hfq_{DsrA}) showed the RNA interacts mainly by the proximal uridine specific binding site, and indicates the RNA is flexible and extends away from Hfq, excluding the possibility of additional interactions with the Sm-core. The initial purpose of this study was to expand our knowledge of the interactions between *E. coli* Hfq and RNA by crystallographic

methods. The regulatory domain 2 of the sRNA DsrA (DsrA₃₄) was selected for this study due to its secondary structure and sequence, as previous studies had indicated that Hfq prefers to bind to a uridine rich stretch of RNA adjacent to a hairpin-loop⁹². The reconstruction of shape-models from small angle scattering studies of the Hfq_{DsrA} complex was probably compromised by low sample quality and polydispersity. The models do not display the desired precise fit to the SAXS-data at low scattering angle, and the SANS-data for the complex is further compromised by a low signal to noise ratio. When fitting the scattering at low angle the excluded volume of the resulting model raises questions about the precise stoichiometry of the complex, which was determined to be the expected 1:1 by NMR relaxation experiments and supported by Guinier analysis of the SAXS-data. However, the two resulting sets of shape-models are in mutual agreement independent of whether volume-restraints are applied or not. This suggests some degree of consistency for the models, both of which could agree with the NMR data, although interpretation of such models should be done with caution given the ill-posed nature of reconstructing 3D-models from 2D-data. Together the data indicate that in the complex between Hfq_{Ec65} and DsrA₃₄, the RNA is protruding from the proximal face of Hfq. Intricate analysis applying perceptual criteria to the SAS-data suggest a small hairpin, that is subject to a dynamic transition between structure and disorder, to form in the 3'-terminal of the DsrA₃₄ RNA in complex with Hfq. The transition from structured to disordered state of the RNA happens in conjunction with the structuring of the N-terminus of Hfq – a feature that was clearly demonstrated by the presented NMR-studies.

The mechanism of action of RNA-chaperones has been proposed to be explained by the entropy-transfer model, in which inherent entropy of a disordered chaperone infers partial unfolding of its target upon binding⁹⁸. Correspondingly, the N-termini of Hfq_{Ec65} becomes structured when binding to DsrA₃₄. The entropic gain from the interaction allows the target to overcome an enthalpy barrier to unfolding, and to refold upon release from the chaperone. Similarly, DsrA₃₄ is tethered to Hfq_{Ec65} by tight binding in the proximal uridine specific binding site, and the 3'-terminal hairpin of the RNA changes between two states: partially folded as a hairpin-loop, and unfolded in complex with the N-terminus.

Also the C-terminus of Hfq_{Ec} becomes more structured in complex with RNA, as described above. Due to its greater length the disordered C-terminus possess an larger entropic potential, as compared to the short N-terminus of Hfq. Interestingly, as shown by SR-CD, the full entropic potential must be satisfied by the unfolding interaction with RNA before the C-terminus can attain structure, which only occurs in complex with long RNAs. The binding and unfolding of the long RNA

continues until equilibrium between the entropic potential and the enthalpy cost of unfolding is reached, an equilibrium that also must take account of the enthalpy contribution from the interaction between RNA and protein. However, no specific RNA-interaction could be identified by mutational analysis (Branislav Vecerek/Udo Bläsi - personal communication). The first half of the C-terminus is mostly positively charged due to pair-wise distribution of histidine residues, whereas the latter half that is flexible and necessary for Hfq to accelerate annealing is negatively charged. Interaction between the acidic and flexible latter half and the RNA-bases will infer an enthalpy contribution, which compensates for the loss of entropy upon binding and structuring. Such isothermal enthalpy/entropy-compensation has been proposed as a method to moderate the affinity of tight interactions, effectively rendering them reversible¹⁶⁸. When multiple diverse binding partners are available, which is the case for the pleiotropic regulator and promiscuous RNA-chaperone Hfq, their interaction is required to be both unspecific and reversible.

The tethering of RNA in the proximal or distal binding sites could be regarded as secondary to the unfolding interaction, but upon unfolding, tethering provides a specific and tight interaction, which is necessary to maintain a stable complex between Hfq and unfolded RNA – similar to the tethering of DsrA₃₄ in the proximal binding site of Hfq_{Ec65}. By transfer of entropy the C-termini of Hfq are able to unfold the RNA, and by tethering the single stranded RNA in the proximal or distal binding site, Hfq obtains thermodynamic control over the kinetic unfolding interaction. The unfolded RNA is suspended between the C-termini of the hexamer in a tight interaction, and tethered in a specific binding site at either face of the Sm-core, without the ability to fold up on itself. Release of the RNA from this complex requires another investment of enthalpy or entropy – cue the concept of an sRNA. The presence of a suitably complementary RNA binding partner, be it mRNA or sRNA, offers a new interaction that takes precedence over Hfq's thermodynamic control, and upon annealing the RNA-duplex is released from Hfq in a zipper-like manner.

One might speculate if the ability of the C-termini to move from lateral to proximal orientation relates to the RNA protruding from the proximal face. Another possibility would be an interaction between the N- and the C-termini, however such an interaction should have been observed in the crystal structure of Hfq_{Ec}. It is very interesting to observe the altered crystal packing in the presence of the C-termini, where the 'honeycomb' double-layers of Hfq hexamers are poised perfectly on top of each other, a phenomenon that has not been observed in other Hfq structures. Between the poised layers a void of only ~30 Å contains both the N- and C-termini, making it necessary for

the C-terminus to pack up on itself. Due to a lattice translocation defect of the crystal, little density was observed in this void. However, as no periodicity can be possible without a degree of order, this could suggest a residual fold in the C-termini to mediate the poised positioning of the hexamers. The SAXS-experiments reveal the C-termini to be in lateral position in solution, suggesting this to be the energetically most favoured confirmation in solution for Hfq. The extending conformation is in agreement with the NMR-relaxation experiments that indicate Hfq_{Ec} to be 'puffed up', appearing larger than expected for a globular protein of that molecular mass. The dynamic transition by the C-termini, from a possible partially folded state in proximal position to a laterally extending conformation, could function as the mechanism to unfold RNA, in a manner similar to opening an umbrella. Suspending an RNA-molecule between the core binding site and the tip of the C-termini might also serve to explain how a short RNA, tethered in the Sm-core, cannot reach to bind and invoke structure on the acidic C-terminal tip. A similar mechanism might work for RNA that is tethered in the binding site on the distal face.

The distal face binding site has mainly been implicated in binding of poly(A)-RNA and adenine-rich adjacent sequence, like the AAYAA-motif upstream in RpoS^{53,111}. However, this does not rule out the possibility of a cooperative function of the distal face with the C-termini. Crystallographic data presented here showed that the distal face purine specific 'R'-site binding pocket allows a degree of freedom for interaction, which suggest the distal binding site to be capable of accommodating variable sequences. This would mean the proximal uridine binding site provides specificity to Hfq, while the distal face binding site provides for a secondary less specific interaction, in a manner similar to the AAYAA-motif in the RpoS-leader. On the other hand, it has also been proposed that RNA binding partners compete for the proximal binding site, this however from a study that employed only very short oligonucleotides⁵². Furthermore, based on the structural studies of the uridine complex, the proximal binding site has been proposed to be able to accommodate adenine and cytosine nucleotides^{108,110}. This taken together diminishes the specificity of the proximal face, and provides Hfq with yet another degree of freedom for its ubiquitous nature.

Why has no structural work been published with both the proximal and distal binding site of Hfq occupied? Binding studies have shown Hfq able to occupy both binding sites simultaneously⁹⁵, and visualization of such a structure would provide very interesting insights to the function of the Sm-core of Hfq. Furthermore, applying small angle scattering studies to such a ternary complex, and gradually extending the RNA, could shed some light on the behavior of the C-termini. Finally, the solution behavior of the isolated C-terminus could be studied by NMR.

The long C-termini provide *E. coli* and other gram-negative pathogens with additional degrees of freedom, by the increased entropic potential that makes more RNA-targets available for unfolding. Short Hfq homologs carry only the N-termini, which are sufficient for certain aspects of riboregulation. Considering that most information-flow in the cell requires getting the message through to translation, it is easy to imagine a player like Hfq to have a pleiotropic influence. Hfq gets the job done by keeping RNA unfolded. Interplay between the structured and disordered domains of Hfq composes its versatile nature, which act in cooperation with RNA to provide circumstances for interaction following the dogma: *Binding promotes unfolding – unfolding promotes annealing – annealing promotes release of Hfq!*

6 Bibliography

1. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561-563 (1970).
2. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
3. Waters, L. S. & Storz, G. Regulatory RNAs in bacteria. *Cell* **136**, 615-628 (2009).
4. Wassarman, K. M. 6S RNA: a small RNA regulator of transcription. *Curr Opin Microbiol* **10**, 164-168 (2007).
5. Babitzke, P. & Romeo, T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr Opin Microbiol* **10**, 156-163 (2007).
6. Kazantsev, A. V. & Pace, N. R. Bacterial RNase P: a new view of an ancient enzyme. *Nat Rev Microbiol* **4**, 729-740 (2006).
7. Esakova, O. & Krasilnikov, A. S. Of proteins and RNA: the RNase P/MRP family. *RNA* **16**, 1725-1747 (2010).
8. Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR-a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**, 181-186 (2008).
9. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845 (2008).
10. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712 (2007).
11. Wilderman, P. J. *et al.* Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc Natl Acad Sci U S A* **101**, 9792-9797 (2004).
12. Brantl, S. Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol* **10**, 102-109 (2007).
13. Lorenz, C. *et al.* Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Res* **38**, 3794-3808 (2010).
14. Lorenz, C., von Pelchrzim, F. & Schroeder, R. Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat Protoc* **1**, 2204-2212 (2006).
15. Toledo-Arana, A. *et al.* The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* **459**, 950-956 (2009).
16. Giangrossi, M. *et al.* A novel antisense RNA regulates at transcriptional level the virulence gene *icsA* of *Shigella flexneri*. *Nucleic Acids Res* **38**, 3362-3375 (2010).

-
17. Repoila, F. & Darfeuille, F. Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biol Cell* **101**, 117-131 (2009).
 18. Gorke, B. & Stulke, J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* **6**, 613-624 (2008).
 19. Urbanowski, M. L., Stauffer, L. T. & Stauffer, G. V. The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol Microbiol* **37**, 856-868 (2000).
 20. Gottesman, S. *et al.* Small RNA regulators and the bacterial response to stress. *Cold Spring Harb Symp Quant Biol* **71**, 1-11 (2006).
 21. Johansen, J., Rasmussen, A. A., Overgaard, M. & Valentin-Hansen, P. Conserved small non-coding RNAs that belong to the sigmaE regulon: role in down-regulation of outer membrane proteins. *J Mol Biol* **364**, 1-8 (2006).
 22. Gorke, B. & Vogel, J. Noncoding RNA control of the making and breaking of sugars. *Genes Dev* **22**, 2914-2925 (2008).
 23. De Lay, N. & Gottesman, S. The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. *J Bacteriol* **191**, 461-476 (2009).
 24. Kaberdin, V. R. & Bläsi, U. Translation initiation and the fate of bacterial mRNAs. *FEMS Microbiol Rev* **30**, 967-979 (2006).
 25. Kawamoto, H., Koide, Y., Morita, T. & Aiba, H. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol Microbiol* **61**, 1013-1022 (2006).
 26. Morfeldt, E., Taylor, D., von Gabain, A. & Arvidson, S. Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense RNA, RNAIII. *EMBO J* **14**, 4569-4577 (1995).
 27. Repoila, F., Majdalani, N. & Gottesman, S. Small non-coding RNAs, co-ordinators of adaptation processes in *Escherichia coli*: the RpoS paradigm. *Mol Microbiol* **48**, 855-861 (2003).
 28. Urban, J. H. & Vogel, J. Two seemingly homologous noncoding RNAs act hierarchically to activate glmS mRNA translation. *PLoS Biol* **6**, e64 (2008).
 29. Papenfort, K. & Vogel, J. Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. *Res Microbiol* **160**, 278-287 (2009).
 30. Majdalani, N., Cunnig, C., Sledjeski, D., Elliott, T. & Gottesman, S. DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc Natl Acad Sci U S A* **95**, 12462-12467 (1998).

-
31. Lease, R. A. & Belfort, M. Riboregulation by DsrA RNA: transactions for global economy. *Mol Microbiol* **38**, 667-672 (2000).
 32. Franze de Fernandez, M. T., Eoyang, L. & August, J. T. Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature* **219**, 588-590 (1968).
 33. Franze de Fernandez, M. T., Hayward, W. S. & August, J. T. Bacterial proteins required for replication of phage Q ribonucleic acid. Purification and properties of host factor I, a ribonucleic acid-binding protein. *J Biol Chem* **247**, 824-831 (1972).
 34. Kajitani, M., Kato, A., Wada, A., Inokuchi, Y. & Ishihama, A. Regulation of the *Escherichia coli* *hfq* gene encoding the host factor for phage Q beta. *J Bacteriol* **176**, 531-534 (1994).
 35. Azam, T. A. & Ishihama, A. Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J Biol Chem* **274**, 33105-33113 (1999).
 36. Sonnleitner, E., Schuster, M., Sorger-Domenigg, T., Greenberg, E. P. & Bläsi, U. Hfq-dependent alterations of the transcriptome profile and effects on quorum sensing in *Pseudomonas aeruginosa*. *Mol Microbiol* **59**, 1542-1558 (2006).
 37. Tsui, H. C., Feng, G. & Winkler, M. E. Negative regulation of *mutS* and *mutH* repair gene expression by the Hfq and RpoS global regulators of *Escherichia coli* K-12. *J Bacteriol* **179**, 7476-7487 (1997).
 38. Vecerek, B., Moll, I. & Bläsi, U. Translational autocontrol of the *Escherichia coli* *hfq* RNA chaperone gene. *RNA* **11**, 976-984 (2005).
 39. Vytvytska, O. *et al.* Host factor I, Hfq, binds to *Escherichia coli* *ompA* mRNA in a growth rate-dependent fashion and regulates its stability. *Proc Natl Acad Sci U S A* **95**, 14118-14123 (1998).
 40. Tsui, H. C., Leung, H. C. & Winkler, M. E. Characterization of broadly pleiotropic phenotypes caused by an *hfq* insertion mutation in *Escherichia coli* K-12. *Mol Microbiol* **13**, 35-49 (1994).
 41. Ding, Y., Davis, B. M. & Waldor, M. K. Hfq is essential for *Vibrio cholerae* virulence and downregulates sigma expression. *Mol Microbiol* **53**, 345-354 (2004).
 42. Robertson, G. T. & Roop, R. M. J. The *Brucella abortus* host factor I (HF-I) protein contributes to stress resistance during stationary phase and is a major determinant of virulence in mice. *Mol Microbiol* **34**, 690-700 (1999).
 43. Sonnleitner, E. *et al.* Reduced virulence of a *hfq* mutant of *Pseudomonas aeruginosa* O1. *Microb Pathog* **35**, 217-228 (2003).
 44. Muffler, A., Fischer, D. & Hengge-Aronis, R. The RNA-binding protein HF-I, known as a host factor for phage Qbeta RNA replication, is essential for *rpoS* translation in *Escherichia coli*. *Genes Dev* **10**, 1143-1151 (1996).

-
45. Brown, L. & Elliott, T. Efficient translation of the RpoS sigma factor in *Salmonella typhimurium* requires host factor I, an RNA-binding protein encoded by the *hfq* gene. *J Bacteriol* **178**, 3763-3770 (1996).
 46. Brown, L. & Elliott, T. Mutations that increase expression of the *rpoS* gene and decrease its dependence on *hfq* function in *Salmonella typhimurium*. *J Bacteriol* **179**, 656-662 (1997).
 47. Cunnig, C., Brown, L. & Elliott, T. Promoter substitution and deletion analysis of upstream region required for *rpoS* translational regulation. *J Bacteriol* **180**, 4564-4570 (1998).
 48. Zhang, A. *et al.* The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. *EMBO J* **17**, 6061-6068 (1998).
 49. Sittka, A. *et al.* Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* **4**, e1000163 (2008).
 50. Arluison, V. *et al.* Spectroscopic observation of RNA chaperone activities of Hfq in post-transcriptional regulation by a small non-coding RNA. *Nucleic Acids Res* **35**, 999-1006 (2007).
 51. Hopkins, J. F., Panja, S. & Woodson, S. A. Rapid binding and release of Hfq from ternary complexes during RNA annealing. *Nucleic Acids Res* (2011).
 52. Hwang, W., Arluison, V. & Hohng, S. Dynamic competition of DsrA and *rpoS* fragments for the proximal binding site of Hfq as a means for efficient annealing. *Nucleic Acids Res* (2011).
 53. Soper, T. J. & Woodson, S. A. The *rpoS* mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA* **14**, 1907-1917 (2008).
 54. Moller, T., Franch, T., Udesen, C., Gerdes, K. & Valentin-Hansen, P. Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes Dev* **16**, 1696-1706 (2002).
 55. Zhang, A., Wassarman, K. M., Ortega, J., Steven, A. C. & Storz, G. The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell* **9**, 11-22 (2002).
 56. Papenfort, K. *et al.* SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global *omp* mRNA decay. *Mol Microbiol* **62**, 1674-1688 (2006).
 57. Afonyushkin, T., Vecerek, B., Moll, I., Bläsi, U. & Kaberdin, V. R. Both RNase E and RNase III control the stability of *sodB* mRNA upon translational inhibition by the small regulatory RNA RyhB. *Nucleic Acids Res* **33**, 1678-1689 (2005).
 58. Sharma, C. M., Darfeuille, F., Plantinga, T. H. & Vogel, J. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev* **21**, 2804-2817 (2007).

-
59. Hussein, R. & Lim, H. N. Disruption of small RNA signaling caused by competition for Hfq. *Proc Natl Acad Sci U S A* **108**, 1110-1115 (2011).
 60. Busi, F. *et al.* Auto-assembly as a new regulatory mechanism of noncoding RNA. *Cell Cycle* **8**, 952-954 (2009).
 61. Cayrol, B. *et al.* Auto-assembly of *E. coli* DsrA small noncoding RNA: Molecular characteristics and functional consequences. *RNA Biol* **6**, 434-445 (2009).
 62. Updegrove, T. B., Correia, J. J., Galletto, R., Bujalowski, W. & Wartell, R. M. *E. coli* DNA associated with isolated Hfq interacts with Hfq's distal surface and C-terminal domain. *Biochim Biophys Acta* **1799**, 588-596 (2010).
 63. Herschlag, D. RNA chaperones and the RNA folding problem. *J Biol Chem* **270**, 20871-20874 (1995).
 64. Herschlag, D., Khosla, M., Tsuchihashi, Z. & Karpel, R. L. An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. *EMBO J* **13**, 2913-2924 (1994).
 65. Clodi, E., Semrad, K. & Schroeder, R. Assaying RNA chaperone activity in vivo using a novel RNA folding trap. *EMBO J* **18**, 3776-3782 (1999).
 66. Moll, I., Leitsch, D., Steinhauser, T. & Bläsi, U. RNA chaperone activity of the Sm-like Hfq protein. *EMBO Rep* **4**, 284-289 (2003).
 67. Rajkowitsch, L. & Schroeder, R. Dissecting RNA chaperone activity. *RNA* **13**, 2053-2060 (2007).
 68. Rudiger, S., Schneider-Mergener, J. & Bukau, B. Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. *EMBO J* **20**, 1042-1050 (2001).
 69. Maki, K., Uno, K., Morita, T. & Aiba, H. RNA, but not protein partners, is directly responsible for translational silencing by a bacterial Hfq-binding small RNA. *Proc Natl Acad Sci U S A* **105**, 10332-10337 (2008).
 70. Lee, E. J. & Groisman, E. A. An antisense RNA that governs the expression kinetics of a multifunctional virulence gene. *Mol Microbiol* **76**, 1020-1033 (2010).
 71. Nielsen, J. S. *et al.* Defining a role for Hfq in Gram-positive bacteria: evidence for Hfq-dependent antisense regulation in *Listeria monocytogenes*. *Nucleic Acids Res* **38**, 907-919 (2010).
 72. Aiba, H. Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr Opin Microbiol* **10**, 134-139 (2007).
 73. Brennan, R. G. & Link, T. M. Hfq structure, function and ligand binding. *Curr Opin Microbiol* **10**, 125-133 (2007).
 74. Valentin-Hansen, P., Eriksen, M. & Udesen, C. The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol Microbiol* **51**, 1525-1533 (2004).

-
75. Sukhodolets, M. V. & Garges, S. Interaction of *Escherichia coli* RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry* **42**, 8022-8034 (2003).
 76. Arluison, V., Mutyam, S. K., Mura, C., Marco, S. & Sukhodolets, M. V. Sm-like protein Hfq: location of the ATP-binding site and the effect of ATP on Hfq-RNA complexes. *Protein Sci* **16**, 1830-1841 (2007).
 77. Resch, A., Vecerek, B., Palavra, K. & Bläsi, U. Requirement of the CsdA DEAD-box helicase for low temperature riboregulation of rpoS mRNA. *RNA Biol* **7**, 796-802 (2010).
 78. Olejniczak, M. Despite Similar Binding to the Hfq Protein Regulatory RNAs Widely Differ in Their Competition Performance. *Biochemistry* **50**, 4427-4440 (2011).
 79. Lease, R. A. & Woodson, S. A. Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J Mol Biol* **344**, 1211-1223 (2004).
 80. Hajnsdorf, E. & Regnier, P. Host factor Hfq of *Escherichia coli* stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proc Natl Acad Sci U S A* **97**, 1501-1505 (2000).
 81. Folichon, M. *et al.* The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res* **31**, 7302-7310 (2003).
 82. Folichon, M., Allemand, F., Regnier, P. & Hajnsdorf, E. Stimulation of poly(A) synthesis by *Escherichia coli* poly(A)polymerase I is correlated with Hfq binding to poly(A) tails. *FEBS J* **272**, 454-463 (2005).
 83. Mohanty, B. K., Maples, V. F. & Kushner, S. R. The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in *Escherichia coli*. *Mol Microbiol* **54**, 905-920 (2004).
 84. Morita, T., Maki, K. & Aiba, H. RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev* **19**, 2176-2186 (2005).
 85. Butland, G. *et al.* Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531-537 (2005).
 86. Haynes, C. *et al.* Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol* **2**, e100 (2006).
 87. Worrall, J. A. *et al.* Reconstitution and analysis of the multienzyme *Escherichia coli* RNA degradosome. *J Mol Biol* **382**, 870-883 (2008).
 88. Vecerek, B., Beich-Frandsen, M., Resch, A. & Bläsi, U. Translational activation of rpoS mRNA by the non-coding RNA DsrA and Hfq does not require ribosome binding. *Nucleic Acids Res* **38**, 1284-1293 (2010).

-
89. Masse, E., Escorcia, F. E. & Gottesman, S. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev* **17**, 2374-2383 (2003).
 90. Moll, I., Afonyushkin, T., Vytvytska, O., Kaberdin, V. R. & Bläsi, U. Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. *RNA* **9**, 1308-1314 (2003).
 91. Moller, T. *et al.* Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell* **9**, 23-30 (2002).
 92. Brescia, C. C., Mikulecky, P. J., Feig, A. L. & Sledjeski, D. D. Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure. *RNA* **9**, 33-43 (2003).
 93. Battiste, J. L. *et al.* Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* **273**, 1547-1551 (1996).
 94. Geissmann, T. A. & Touati, D. Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J* **23**, 396-405 (2004).
 95. Mikulecky, P. J. *et al.* *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat Struct Mol Biol* **11**, 1206-1214 (2004).
 96. Sonnleitner, E. *et al.* Functional effects of variants of the RNA chaperone Hfq. *Biochem Biophys Res Commun* **323**, 1017-1023 (2004).
 97. Dunker, A. K. *et al.* Intrinsically disordered protein. *J Mol Graph Model* **19**, 26-59 (2001).
 98. Tompa, P. & Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J* **18**, 1169-1175 (2004).
 99. Vecerek, B., Rajkowitsch, L., Sonnleitner, E., Schroeder, R. & Bläsi, U. The C-terminal domain of *Escherichia coli* Hfq is required for regulation. *Nucleic Acids Res* **36**, 133-143 (2008).
 100. Soper, T., Mandin, P., Majdalani, N., Gottesman, S. & Woodson, S. A. Positive regulation by small RNAs and the role of Hfq. *Proc Natl Acad Sci U S A* **107**, 9602-9607 (2010).
 101. Stark, H., Dube, P., Luhrmann, R. & Kastner, B. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539-542 (2001).
 102. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K., Li, J. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475-480 (2009).
 103. Moskaleva, O. *et al.* The structures of mutant forms of Hfq from *Pseudomonas aeruginosa* reveal the importance of the conserved His57 for the protein hexamer organization. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**, 760-764 (2010).

-
104. Beggs, J. D. Lsm proteins and RNA processing. *Biochem Soc Trans* **33**, 433-438 (2005).
 105. Sauter, C., Basquin, J. & Suck, D. Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic Acids Res* **31**, 4091-4098 (2003).
 106. Sun, X., Zhulin, I. & Wartell, R. M. Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res* **30**, 3662-3671 (2002).
 107. Newman, A. J. & Nagai, K. Structural studies of the spliceosome: blind men and an elephant. *Curr Opin Struct Biol* **20**, 82-89 (2010).
 108. Schumacher, M. A., Pearson, R. F., Moller, T., Valentin-Hansen, P. & Brennan, R. G. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J* **21**, 3546-3556 (2002).
 109. Nikulin, A. *et al.* Structure of *Pseudomonas aeruginosa* Hfq protein. *Acta Crystallogr D Biol Crystallogr* **61**, 141-146 (2005).
 110. Sauer, E. & Weichenrieder, O. Structural basis for RNA 3'-end recognition by Hfq. *Proc Natl Acad Sci U S A* (2011).
 111. Link, T. M., Valentin-Hansen, P. & Brennan, R. G. Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci U S A* **106**, 19292-19297 (2009).
 112. Baba, S., Someya, T., Kawai, G., Nakamura, K. & Kumasaka, T. Expression, crystallization and preliminary crystallographic analysis of RNA-binding protein Hfq (*YmaH*) from *Bacillus subtilis* in complex with an RNA aptamer. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**, 563-566 (2010).
 113. DeLano, W. L. MacPyMOL: a PyMOL-based molecular graphics application for MacOS X. *DeLano Scientific LLC, Palo Alto, CA* (2007).
 114. Jousset, A., Metzinger, L. & Felden, B. On the facultative requirement of the bacterial RNA chaperone, Hfq. *Trends Microbiol* **17**, 399-405 (2009).
 115. Arluison, V. *et al.* The C-terminal domain of *Escherichia coli* Hfq increases the stability of the hexamer. *Eur J Biochem* **271**, 1258-1265 (2004).
 116. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-405 (2000).
 117. Mittag, T., Kay, L. E. & Forman-Kay, J. D. Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* **23**, 105-116 (2010).
 118. Beich-Frandsen, M. *et al.* Structural insights into the dynamics and function of the C-terminus of the *E. coli* RNA chaperone Hfq. *Nucleic Acids Res* **39**, 4900-4915 (2011).

-
119. Urlaub, H., Raker, V. A., Kostka, S. & Luhrmann, R. Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J* **20**, 187-196 (2001).
 120. Rambo, R. P. & Tainer, J. A. Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Current opinion in structural biology* **20**, 128-137 (2010).
 121. Vecerek, B., Moll, I., Afonyushkin, T., Kaberdin, V. & Bläsi, U. Interaction of the RNA chaperone Hfq with mRNAs: direct and indirect roles of Hfq in iron metabolism of *Escherichia coli*. *Mol Microbiol* **50**, 897-909 (2003).
 122. Resch, A. *et al.* Translational activation by the noncoding RNA DsrA involves alternative RNase III processing in the rpoS 5'-leader. *RNA* **14**, 454-459 (2008).
 123. Jacrot, B. The study of biological structures by neutron scattering from solution. *Reports on Progress in Physics* **39**, 911 (1976).
 124. Nierhaus, K. H. *et al.* Structure of the elongating ribosome: arrangement of the two tRNAs before and after translocation. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 945 (1998).
 125. Timmins, P. A., Langowski, J. & Brown, R. S. An elongated model of the *Xenopus laevis* transcription factor IIIA-5S ribosomal RNA complex derived from neutron scattering and hydrodynamic measurements. *Nucleic Acids Res* **16**, 8633-8644 (1988).
 126. Svergun, D. I. & Koch, M. H. J. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics* **66**, 1735 (2003).
 127. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **40**, 191-285 (2007).
 128. Svergun, D. I. *et al.* Solution scattering structural analysis of the 70 S *Escherichia coli* ribosome by contrast variation. II. A model of the ribosome and its RNA at 3.5 nm resolution. *J Mol Biol* **271**, 602-618 (1997).
 129. Svergun, D. I. *et al.* Solution scattering structural analysis of the 70 S *Escherichia coli* ribosome by contrast variation. I. Invariants and validation of electron microscopy models. *J Mol Biol* **271**, 588-601 (1997).
 130. Roessle, M. W. *et al.* Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *Journal of Applied Crystallography* **40**, s190-s194 (2007).

-
131. Round, A. R. *et al.* Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33. *Journal of Applied Crystallography* **41**, 913-917 (2008).
 132. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. ATSAS 2.1-towards automated and web-supported small-angle scattering data analysis. *Journal of Applied Crystallography* **40**, 223-228 (2007).
 133. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**, 1277-1282 (2003).
 134. Semenyuk, A. V. & Svergun, D. I. GNOM-a program package for small-angle scattering data processing. *Journal of Applied Crystallography* **24**, 537-540 (1991).
 135. Porod, G. in *Small-angle X-ray scattering* (ed Kratky, O. G. a. O.) 17-51 (Academic Press, London, 1982).
 136. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **42**, 342-346 (2009).
 137. Petoukhov, M. V. & Svergun, D. I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* **89**, 1237-1250 (2005).
 138. Svergun, D., Barberato, C. & Koch, M. H. J. CRY SOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* **28**, 768-773 (1995).
 139. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* **80**, 2946-2953 (2001).
 140. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293 (1995).
 141. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **89**, 496-514 (1990).
 142. Kay, L. E., Ikura, M. & Bax, A. Proton-proton correlation via carbon-carbon couplings: A three-dimensional NMR approach for the assignment of aliphatic resonances in proteins labeled with carbon-13. *Journal of the American Chemical Society* **112**, 888-889 (1990).
 143. Mikkelsen, K. & Nielsen, S. O. Acidity measurements with the glass electrode in H₂O-D₂O mixtures. *The Journal of Physical Chemistry* **64**, 632-637 (1960).
 144. Jacrot, B. & Zaccai, G. Determination of molecular weight by neutron scattering. *Biopolymers* **20**, 2413-2426 (1981).

-
145. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* **76**, 2879-2886 (1999).
 146. Petoukhov, M. V. & Svergun, D. I. Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution. *Eur Biophys J* **35**, 567-576 (2006).
 147. Petoukhov, M. V. *et al.* Conformation of polypyrimidine tract binding protein in solution. *Structure* **14**, 1021-1027 (2006).
 148. Akiyama, S. Quality control of protein standards for molecular mass determinations by small-angle X-ray scattering. *Journal of Applied Crystallography* **43**, (2010).
 149. Cheng, A. C., Chen, W. W., Fuhrmann, C. N. & Frankel, A. D. Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains. *J Mol Biol* **327**, 781-796 (2003).
 150. Rajkowitsch, L. *et al.* RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* **4**, 118-130 (2007).
 151. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **25**, 495-503 (1992).
 152. Stuhrmann, H. B. Unique aspects of neutron scattering for the study of biological systems. *Reports on Progress in Physics* **67**, 1073 (2004).
 153. Ibel, K. & Stuhrmann, H. B. Comparison of neutron and X-ray scattering of dilute myoglobin solutions. *Journal of molecular biology* **93**, 255-265 (1975).
 154. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415 (2003).
 155. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the alpha-and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO journal* **1**, 945 (1982).
 156. Newman, J. *et al.* Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr D Biol Crystallogr* **61**, 1426-1431 (2005).
 157. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J.Appl.Cryst.* **26**, 795-800 (1993).
 158. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *Journal of Applied Crystallography* **30**, 1022-1025 (1997).
 159. Nr.4, C. C. P. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763 (1994).

-
160. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
 161. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
 162. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255 (1997).
 163. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184-2195 (2004).
 164. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-83 (2007).
 165. Reynolds, C., Damerell, D. & Jones, S. ProtorP: a protein-protein interaction analysis server. *Bioinformatics* **25**, 413-414 (2009).
 166. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276-277 (2000).
 167. Beich-Frandsen, M., Vecerek, B., Sjoblom, B., Bläsi, U. & Djinovic-Carugo, K. Structural analysis of full-length Hfq from *Escherichia coli*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**, 536-540 (2011).
 168. Jen-Jacobson, L., Engler, L. E. & Jacobson, L. A. Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure* **8**, 1015-1023 (2000).

7 Abbreviations

D _{MAX}	Maximum dimension
D _{MIN}	Minimum dimension
DLS	Dynamic Light Scattering
FRET	Fluorescence Resonance Energy Transfer
FT-IR	Fourier Transform Infrared Spectroscopy
Hfq	Host factor Q
HSQC	Heteronuclear Single Quantum Coherence
NMR	Nuclear Magnetic Resonance
P(R)	Pair distribution function
R _g	Radius of Gyration
Rbs	Ribosomal Binding Site
r.m.s.d.	Root Mean Square Deviation
SAS	Small Angle Scattering
sRNAs	Small regulatory Ribo Nucleic Acid
SR-CD	Synchrotron Radiation Circular Dichroism
Utr	Untranslated region

8 Paper I

8.1 Supplementary material

8.1.1 Purification of Hfq proteins

Lysis was achieved by disrupting the cells in a French-press (1g cells / 5ml) after resuspension in 50mM Tris-HCl pH 8, 1.5 M NaCl, 250 mM MgCl₂, 1 mM EDTA, 1 mM β -mercapto-ethanol (BME), 1 mM phenyl-methyl-sulphonyl-flouride (PMSF) and 10 μ g/ml DNase I. Subsequently cell-debris was removed by centrifugation (45.000 x G / 20 min). After centrifugation the supernatant was heated to 85°C for 45 min, with subsequent removal of precipitates by centrifugation (45.000 x G / 20 min). Ni²⁺-affinity purification was performed for Hfq_{Ec} and Hfq_{Ec85} without recombinant addition of histidines to the protein, but utilizing that the protein harbors four histidines at the C-terminus. In the hexameric protein this totals to 24 histidines, providing the high affinity to the Ni²⁺-affinity column. The protein fractions were loaded onto a 1 ml His-Trap FF column (GE Healthcare), charged with 0.1M NiSO₄ and equilibrated with buffer containing 50mM Tris-HCl pH 8, 1.5 M NaCl, 1M urea, 0.5 mM EDTA, 0.5 mM BME, 0.1 mM PMSF, washed extensively with 50 column volumes with the same buffer and eluted by addition of 500 mM imidazole to the buffer.

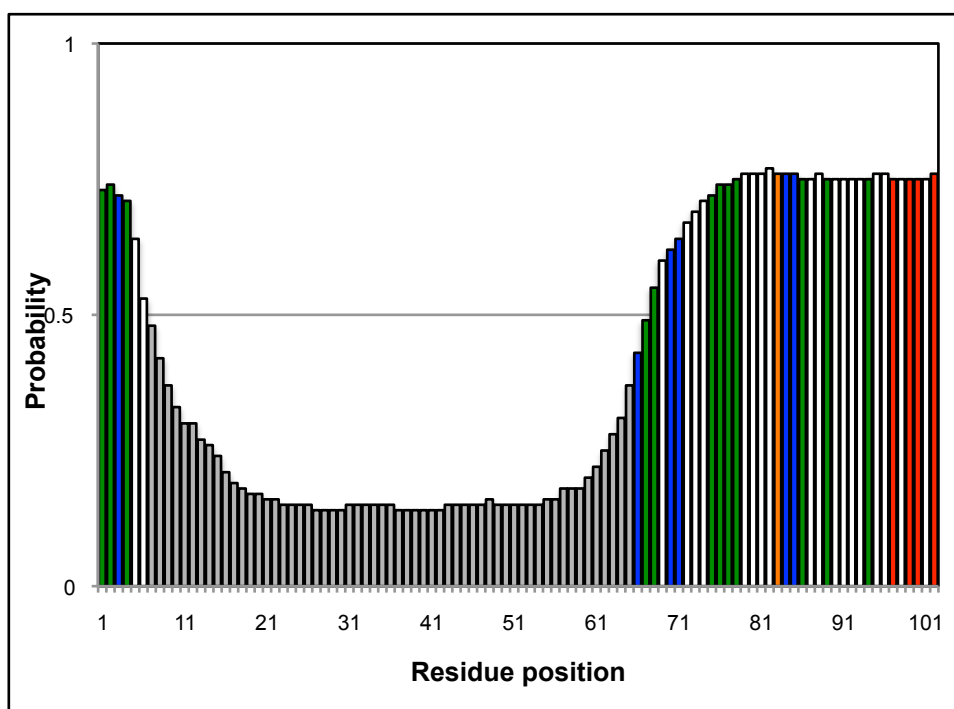
Hydrophobic Interaction Chromatography (HIC) was applied to Hfq_{Ec65}, Hfq_{Ec75} as well as to Hfq_{Sa} and Hfq_{Bs} which did not bind to the Ni²⁺-affinity column in their native form. After heating and removal of the precipitates, (NH₄)₂SO₄ was added to a final concentration of 1 M. The proteins were loaded onto a 1 ml HP Butyl-FF column, previously equilibrated with buffer containing 50 mM Tris-HCl pH 8, 1.5 M NaCl, 1.5 M (NH₄)₂SO₄, 0.5 mM EDTA, 0.5 mM BME, 0.1 mM PMSF, and washed extensively for ~30 column volumes with the same buffer. Elution was achieved by lowering the salt-concentration by switching to a buffer containing 50mM Tris-HCl pH 8, 200 mM NaCl, 0.5 mM EDTA, 0.5 mM BME, 0.1 mM PMSF.

Anion-exchange chromatography (aIEX) was employed to remove remaining nucleic-acid impurities, which co-purify with the protein. The aIEX-procedure utilizes that the isoelectric point of nucleic-acid is several orders of magnitude below that of the protein. The eluted protein from the washing-step was diluted to a lower salt-concentration, which would still effectively shield the vague negative charge on the protein and keep it from binding to the anion-matrix. The washed protein was diluted with 50 mM Tris-HCl pH~8 to a salt concentration below 200 mM NaCl (or a conductivity below 140 μ S x m⁻¹), and loaded on to a 6 mL ReSource Q column, which had been equilibrated with 50mM Tris-HCl pH 8, 200 mM NaCl, 0.5 mM EDTA, 0.5 mM BME, 0.1 mM PMSF. After loading, the protein was collected in the flow-through. The RNA-impurities could be eluted from the column with a high concentration of salt. After aIEX the proteins were concentrated

by one additional processing by the respective type chromatography washing-step. For Ni²⁺-affinity, the flow-through from aIEX was diluted with 50 mM Tris-HCl pH~8 to lower the imidazole concentration (<50 mM), and the solution was loaded onto a 1 ml His-Trap FF equilibrated with 50mM Tris-HCl pH 8, 1.5 M NaCl, 1M urea, 0.5 mM EDTA, 0.5 mM BME, 0.1 mM PMSF, and eluted by addition of 500 mM imidazole to the buffer. For HIC, the flow-through from aIEX was diluted with 50 mM Tris-HCl pH~8, 4 M NaCl to increase the ionic strength (>1.5M NaCl, or +400 $\mu\text{S} \times \text{m}^{-1}$). The solution was loaded on a 1 ml HP Butyl-FF column, previously equilibrated with buffer containing 50 mM Tris-HCl pH 8, 1.5 M NaCl, 1.5 M (NH₄)₂SO₄, 0.5 mM EDTA, 0.5 mM BME, 0.1 mM PMSF, and eluted by switching to a buffer containing 50mM Tris-HCl pH 8, 200 mM NaCl, 0.5 mM EDTA, 0.5 mM BME, 0.1 mM PMSF. The protein was concentrated further using a centrifugal filter unit (Amicon Ultracell 10 kDa cut-off) to a volume applicable to size-exclusion chromatography (below 5 ml). Size exclusion chromatography was performed on a GE Healthcare, Tricorn 16/600 Superdex 200 column, prior equilibrated with buffer 50 mM Tris-HCl pH~7.2, 200 mM NaCl, 0.5 mM EDTA, 0.1 mM PMSF. For VUV-CD samples, size-exclusion chromatography was performed on a GE Healthcare, Tricorn 10/300 Superdex 200 column, prior equilibrated with buffer 50 mM Na-PO₄ pH~7.2, 200 mM NaI. For NMR samples, size-exclusion chromatography was performed on a GE Healthcare, Tricorn 16/600 Superdex 200 column, prior equilibrated with buffer 50 mM Na-PO₄ pH~7.2, 200 mM NaCl. Fractions collected from size-exclusion were concentrated under stirring using a 10 ml Amicon nitrogen pressure cell equipped with a Millipore Ultrafiltration membrane (10 kDa cut-off) to ~20 mg/ml in same buffer.

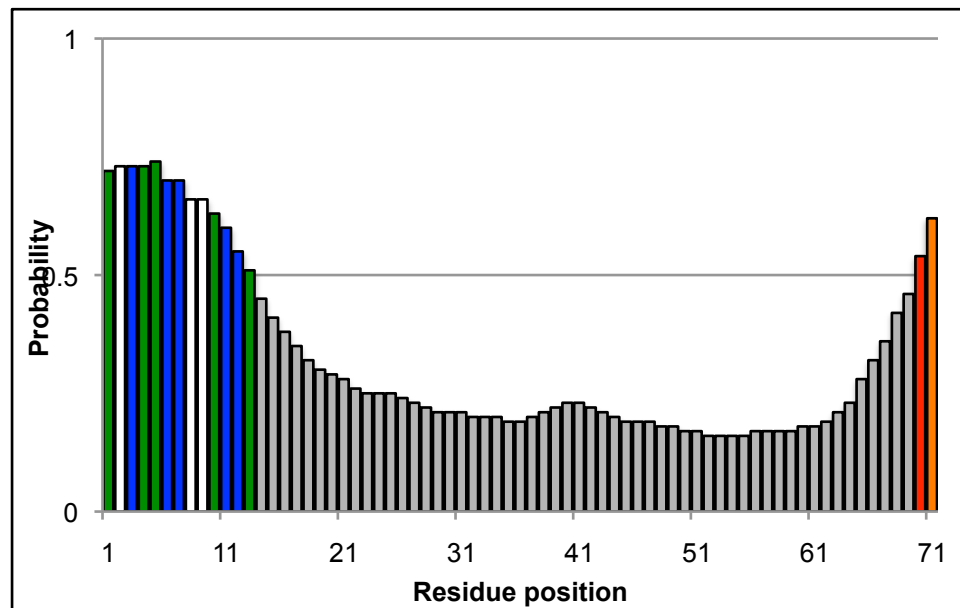
8.1.2 Supplementary Figure S1

Protein disorder prediction of full-length Hfq_{Ec} primary sequence, generated by consensus prediction (<http://prdos.hgc.jp/meta/>), using prediction-servers PrDOS, DISOPRED2, DisEMBL, DISPROT (VSL2P), DISprot, IUPred. A probability value higher than 0.5 indicates disordered residues, colored accordingly: RED: Acidic, BLUE: Basic, WHITE: Polar, GREEN: Apolar, ORANGE: Aromatic. Structured residues are colored in grey.



8.1.3 Supplementary Figure S2

Protein disorder prediction of full-length Hfq_{Mj} primary sequence, generated by consensus prediction (<http://prdos.hgc.jp/meta/>), using prediction-servers PrDOS, DISOPRED2, DisEMBL, DISPROT (VSL2P), DISprot, IUPred. Probability values higher than 0.5 indicates disordered residues, colored accordingly: RED: Acidic, BLUE: Basic, WHITE: Polar, GREEN: Apolar, ORANGE: Aromatic. Structured residues are colored in grey.



9 Paper II