Structural Insights into Noncanonical Mechanisms of Translation

Nathan Rhys James

MRC Laboratory of Molecular Biology

Clare Hall
University of Cambridge

Dissertation submitted for the degree of Doctor of Philosophy
April 2017
This dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration except where declared in the main text and acknowledgements.

This work does not exceed the limit of 60,000 words, excluding the figures, tables, and bibliographies.

Nathan Rhys James
April 2017
SUMMARY

Structural Insights into Noncanonical Mechanisms of Translation

Nathan R. James

Translation is the process by which proteins are synthesized from the instructions in the genetic code. Translation is mediated by the ribosome, a large ribonucleoprotein complex, in concert with messenger RNA (mRNA), transfer RNA (tRNA), and a variety of proteins. The canonical mechanism of translation, introduced in Part I of my thesis, is divided into four distinct phases: initiation, elongation, termination, and recycling. Under unusual circumstances, each phase of translation can also proceed via a number of noncanonical mechanisms, many of which are vitally important for cellular growth or viral infectivity. My thesis describes structural insights into two such noncanonical mechanisms.

The aim of the first project, described in Part II, was to structurally characterize a noncanonical mechanism of translational termination in bacteria. In the absence of a stop codon, ribosomes arrest at the 3′ end of an mRNA and are unable to terminate. In bacteria, the primary mechanism for rescuing such nonstop complexes is known as trans-translation. In the absence of a functional trans-translation system, however, the small protein ArfA recognizes the empty mRNA channel and recruits the release factor RF2 to the ribosome, enabling termination to occur. Using single-particle electron cryomicroscopy (cryo-EM), I obtained four high-resolution structures of nonstop complexes that reveal the mechanism of ArfA-mediated ribosome rescue and have wider implications for understanding canonical termination in bacteria.

The aim of the second project, described in Part III, was to gain structural insights into a noncanonical mechanism of translational initiation in eukaryotes known as internal ribosome entry. Instead of a 5′ cap, many viruses contain intricately structured, cis-acting internal-ribosome-entry sites (IRESs) within their genomes that direct end-independent initiation. The IRES of hepatitis-C virus (HCV), for example, interacts directly with the mammalian ribosome and functionally replaces many of the canonical initiation factors. However, the mechanism by which the HCV IRES coordinates assembly of an initiation complex and progresses through the initiation phase remains poorly understood. I developed a method for purifying native ribosomal complexes from cell lysate that enabled me to obtain multiple cryo-EM maps of the HCV IRES in complex with the 80S ribosome, including a previously unseen conformation of the IRES induced by rotation of the ribosomal small subunit, and to make progress towards capturing earlier steps in the initiation pathway.
ACKNOWLEDGEMENTS

First, I want to thank Venki for giving me the opportunity to follow my curiosities and work in such an amazing environment as the MRC LMB. Although he has been busy in his new presidential role, Venki has encouraged me and trusted in my ability to do science even in the midst of setbacks and extreme competition in the ribosome field. My second supervisor, Kiyoshi Nagai, has been friendly and supportive, and my university supervisor, Richard Jackson, has been a great resource for knowledge of translational initiation. In addition, Ian Brierley has kindly helped me with experiments, Manu Hegde has been a keen listener and an essential contributor to all my projects, and my examiners, Alan Warren and Christian Spahn, made my viva a fairly pleasant experience and provided constructive feedback for the thesis.

I should also thank the members of the Ramakrishnan group, past and present, who have made life in the lab interesting and fun over the last few years. Alexey, Tanweer, Israel, and Ann helped me settle into the lab as a new starter; Jason, Brian, Nirupa, and Chris have kept the lab lively and fun; Yuliya has provided practical help with experiments; Song, an honorary ribosomologist, has been inspirational and a great help with preparing presentations; José has been a great teacher and proofreader; and the new postdocs Sebastian and Vish have shown me that it is possible for projects to be even more ambitious than mine have been. Special thanks should also go to Alan, who has been absolutely indispensable in all aspects of my scientific life (except those that involve pipettes). I owe Alan a tremendous debt for helping me get my first primary-author publication.

Other people at the LMB have also helped me in various ways. Christos and Shaoxia taught me how to use the microscopes; Sjors and Kai have helped with software issues; and Jake and Toby made all the data processing possible. Anthony, Wilson, Xiao-chen, Julian, Lisa, Suyang, and Beatrice have all been supportive and friendly, and the other members of the Nagai group have provided day-to-day help in the lab.

Lastly, I should thank my family for supporting me over the last few years, and Siân for being a great encouragement and occasionally extracting me from the workbench.

Soli Deo gloria.
## CONTENTS

Summary .................................................................................................................. iii

Acknowledgements ................................................................................................. iv

Contents ................................................................................................................... v

List of abbreviations ................................................................................................. viii

List of figures ........................................................................................................... x

List of tables ............................................................................................................. xi

Part I: A general introduction to translation ............................................................. 1

  1.1 Introduction ........................................................................................................ 1

    1.1.1 The translation pathway ........................................................................... 1

    1.1.2 Translational termination in bacteria ...................................................... 3

    1.1.3 Translational initiation in eukaryotes ...................................................... 6

  1.2 Bibliography ...................................................................................................... 11

Part II: The mechanism of ArfA-mediated ribosome rescue ................................... 17

  2.1 Introduction ....................................................................................................... 17

    2.1.1 Ribosome-associated quality control .................................................. 17

    2.1.2 Generating a nonstop mRNA ............................................................. 18

    2.1.3 Nonstop decay in bacteria ..................................................................... 19

      2.1.3.1 Trans-translation ................................................................. 20

      2.1.3.2 ArfA-mediated ribosome rescue .............................................. 22

      2.1.3.3 ArfB-mediated ribosome rescue .............................................. 23

  2.2 Results ............................................................................................................... 25

    2.2.1 Scientific aims ....................................................................................... 25

    2.2.2 Sample preparation and data processing .............................................. 25

    2.2.3 Interpretation ......................................................................................... 31

      2.2.3.1 Specificity for nonstop complexes ............................................ 31

      2.2.3.2 Specificity for RF2 .................................................................. 37

      2.2.3.3 The mechanism of accommodation ......................................... 37

  2.2.4 Discussion ................................................................................................... 43
LIST OF ABBREVIATIONS

2D  two dimensional
3D  three dimensional
ATP  adenosine 5′-triphosphate
CrPV  cricket paralysis virus
cryo-EM  electron cryomicroscopy
CSFV  classical-swine-fever virus
cT2  concentrated T2
CTP  cytidine 5′-triphosphate
DNase  deoxyribonuclease
EDTA  2,2′,2″,2″″-(ethane-1,2-diyl)dinitrilo)tetraacetic acid
EGTA  ethylene glycol-bis(2-aminoethyl)ether)-N,N,N′,N′-tetraacetic acid
EMCV  encephalomyocarditis virus
EMDB  Electron Microscopy Data Bank
FCwSS  focused classification with signal subtraction
FSC  Fourier shell correlation
GDPCP  guanosine 5′-(β,γ-methylene)triphosphate
GDPNP  guanosine 5′-(β,γ-imido)triphosphate
GTP  guanosine 5′-triphosphate
GTPase  guanosine 5′-triphosphatase
HalV  halastavi-árvá virus
HCV  hepatitis-C virus
HEK  human embryonic kidney
HEPES  2-(4-[2-hydroxyethyl]piperazin-1-yl)ethylsulphonic acid
IPTG  isopropyl β-D-1-thiogalactopyranoside
IRES  internal-ribosome-entry site
MLR  mRNA-like region
mRNA  messenger RNA
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTP</td>
<td>nucleoside 5′-triphosphate</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PKR</td>
<td>protein kinase R</td>
</tr>
<tr>
<td>ppGpp</td>
<td>guanosine 5′-diphosphate 3′-diphosphate</td>
</tr>
<tr>
<td>pppGpp</td>
<td>guanosine 5′-triphosphate 3′-diphosphate</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>RQC</td>
<td>ribosome-associated quality control</td>
</tr>
<tr>
<td>RRL</td>
<td>rabbit reticulocyte lysate</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>SAXS</td>
<td>small-angle X-ray scattering</td>
</tr>
<tr>
<td>TBE</td>
<td>tris–borate–EDTA</td>
</tr>
<tr>
<td>TLD</td>
<td>tRNA-like domain</td>
</tr>
<tr>
<td>tmRNA</td>
<td>transfer-messenger RNA</td>
</tr>
<tr>
<td>tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TtRF2</td>
<td><em>Thermus thermophilus</em> RF2</td>
</tr>
<tr>
<td>UTP</td>
<td>uridine 5′-triphosphate</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

I.1 | The elongation cycle in bacteria ................................................................. 2
I.2 | Canonical termination in bacteria .................................................................. 4

II.1 | The fate of the ribosome during the stringent response .................................... 18
II.2 | The trans-translation pathway ....................................................................... 21
II.3 | ArfB-mediated ribosome rescue ...................................................................... 23
II.4 | Data collection ................................................................................................. 25
II.5 | In-silico 3D classification ................................................................................ 27
II.6 | Quality of the maps and models ..................................................................... 29
II.7 | Structures of nonstop complexes recognized by ArfA and ArfA(A18T) .............. 30
II.8 | Conformations of the decoding nucleotides .................................................. 31
II.9 | Interactions between ArfA, RF2, and the ribosome ........................................ 32
II.10 | Sequence alignments ....................................................................................... 34
II.11 | ArfA, SmpB, and ArfB contact the mRNA entry channel with distinct C-termini ... 38
II.12 | Switch-loop stabilization and RF2 accommodation ....................................... 40
II.13 | Release-factor domain movements ................................................................ 42
II.14 | Canonical termination in eukaryotes ............................................................. 44
II.15 | ArfA-mediated ribosome rescue ................................................................... 45

III.1 | Secondary-structure representations of several IRESs ..................................... 62
III.2 | Secondary-structure representation of the HCV genome .................................. 67
III.3 | Secondary-structure representation of the HCV IRES .................................... 68
III.4 | Purification and analysis of mRNA .................................................................. 77
III.5 | Initial attempts at the StreptoTag method ...................................................... 80
III.6 | Purification of 80S initiation complexes using biotinylated IRES mRNA .......... 83
III.7 | Data collection ................................................................................................ 84
III.8 | In-silico 3D classification of 80S ribosomes in complex with the HCV IRES ...... 86
III.9 | Comparison of class III with Christian Spahn’s structure .............................. 87
LIST OF TABLES

I.1 | Mechanisms of translational initiation in eukaryotes ........................................ 7

II.1 | Components of each sample .................................................................................. 26
II.2 | Data-collection and model statistics ........................................................................ 28

III.1 | The current classification system for IRESs found in animal viruses ......................... 64
III.2 | Phenotypes of HCV-IRES mutants ........................................................................... 72
III.3 | Statistics for the 80S cryo-EM dataset ..................................................................... 85
III.4 | Candidate proteins for the density visible over the exit tunnel ................................ 96
III.5 | Buffers and reagents with set pH values .................................................................. 108
I.1 INTRODUCTION

I.1.1 THE TRANSLATION PATHWAY

Translation is the process by which proteins are synthesized from the instructions in the genetic code. Translation is mediated by the ribosome, a large ribonucleoprotein complex (Palade, 1955), in concert with messenger RNA (mRNA), transfer RNA (tRNA), and a variety of proteins. The mRNA contains several functionally important regions including a site for ribosome recruitment and a sequence of codons known as a coding region, open reading frame, or cistron. Each tRNA contains an anticodon that is complementary to a particular codon within the mRNA, while the 3′ end of the tRNA is charged with a specific aminoacyl moiety. Thus, each aminoacyl-tRNA functions as an adapter between the genetic code and the primary structure of the synthesized protein (Hoagland et al., 1958; Crick, 1958). The ribosome itself functions as a molecular ratchet, with the directionality of translation conferred by the action of guanosine 5′-triphosphatases (GTPases) which function as pawls that irreversibly commit the ribosome to certain conformational changes.

The canonical translation pathway is divided into four phases: initiation, elongation, termination, and recycling. During the first phase, initiation, the ribosomal subunits are recruited to an mRNA and assembled into a mature ribosomal complex. This is followed by elongation, the most highly conserved phase of translation, which is directed by orthologous proteins in bacteria, archaea, eukaryotes, and organelles (Voorhees & Ramakrishnan, 2013). During the elongation phase, each codon within an mRNA is decoded by an aminoacyl-tRNA (figure I.1). This occurs in the A site of the ribosomal small subunit and results in accommodation of the aminoacyl moiety into the peptidyltransferase centre of the large subunit. The ribosome then catalyses a condensation reaction between the peptidyl-tRNA in the P site and the newly acquired aminoacyl-tRNA in the A site, thereby transferring the nascent polypeptide to a new tRNA and extending it by one residue. Upon peptidyl transfer, the ribosomal small subunit rotates relative to the large subunit, causing the deacylated tRNA and the peptidyl-tRNA to adopt P/E and A/P hybrid states, respectively, before they are translocated fully into the E and P sites via the action of a translational GTPase (EF-G in bacteria; eEF2 in eukaryotes). Once the GTPase has dissociated, the ribosome is
then ready to start another round of elongation or enter the termination phase of translation in which the nascent polypeptide is released. The ribosomal subunits are then split apart and recycled for another round of translational initiation.

**Figure 1.1** | The elongation cycle in bacteria. The ribosomal large (50S) and small (30S) subunits are shown in blue and yellow, respectively. Elongation is the most highly conserved phase of translation, remaining essentially the same in bacteria, archaea, eukaryotes, and organelles. An aminoacyl-tRNA is initially recruited to the A site with its cofactor EF-Tu. Upon GTP hydrolysis, EF-Tu dissociates and the aminoacyl-tRNA accommodates into the peptidyltransferase centre of the 50S subunit. This is followed by peptidyl transfer and rotation of the 30S subunit, resulting in dissociation of a deacetylated tRNA from the E site. The peptidyl-tRNA in its hybrid conformation is recognized by EF-G, which catalyses translocation of the remaining tRNAs and mRNA through the ribosome.

Under unusual circumstances, however, each phase of translation can also proceed via a diverse array of noncanonical mechanisms. For example, initiation can be
manipulated or even bypassed by cis-acting motifs within the mRNA (Johnson et al., 2017), elongation can be interrupted by frameshifting (Dunkle & Dunham, 2015), recoding (Fischer et al., 2016), or stalling (Richter et al., 2015; Keiler, 2015), termination can proceed even in the absence of a stop codon (Gagnon et al., 2012; James et al., 2016), and recycling can be bypassed in favour of reinitiation on the same mRNA (Skabkin et al., 2013; Zinoviev et al., 2015). These regulatory mechanisms are often vitally important for cellular growth or viral infectivity.

My thesis describes the structural characterization of two such noncanonical mechanisms of translation. In order to understand these mechanisms, it is necessary to set the context by first describing the canonical mechanisms of bacterial termination and eukaryotic initiation.

I.1.2 TRANSLATIONAL TERMINATION IN BACTERIA

Translational termination occurs when a stop codon enters the A site of the ribosome as a result of translocation during the elongation phase. Although the genetic code varies between species, with stop codons being particularly susceptible to recoding in mitochondrial genomes (Jukes & Osawa, 1990), the triplets amber (UAG), ochre (UAA), and umber (UGA) function as stop codons in the standard genetic code utilized by most species (Brenner et al., 1965; Brenner et al., 1967). These codons generally lack cognate tRNAs and are instead recognized by proteins known as release factors. The release factors of bacteria, mitochondria, and plastids (Zhou et al., 2012a) are unrelated to those of archaea and the eukaryotic cytosol (Jackson et al., 2012), yet these machineries fulfil the same functions as a result of convergent evolution. Here I will describe the canonical mechanism of translational termination in bacteria (figure I.2), which provides the background for the noncanonical mechanism of ArfA-mediated ribosome rescue described in Part II.

When the bacterial ribosome encounters a stop codon, a class-I release factor is recruited to the A site (Capecchi, 1967; Zhou et al., 2012a). There are two class-I release factors involved in stop-codon recognition, both of which recognize the ochre codon. RF1 also recognizes amber, whereas RF2 recognizes umber (Scolnick et al., 1968). These specificities are conferred by a recognition loop between the β4- and β5-
strands (figure II.10B,C) containing a motif (PXT in RF1; SPF in RF2) that recognizes a particular conformation of the decoding centre (Ito et al., 2000). Stop-codon recognition is thought to induce allosteric conformational changes in the release factor, leading to stabilization of the switch loop between the α7-helix and the β9-strand (figure II.10B,C), which in turn causes the far end of domain 3 to rise into the peptidyltransferase centre (Laurberg et al., 2008; Korostelev et al., 2008; Weixlbaumer et al., 2008; Korostelev et al., 2010). The conserved GGQ motif in domain 3 then catalyses hydrolysis of the ester bond between the P-site tRNA and the nascent polypeptide (Jin et al., 2010).

Crystal structures of RF1 (Shin et al., 2004) and RF2 (Vestergaard et al., 2001; Zoldák et al., 2007) in isolation have shown that these proteins adopt a compact conformation that is incompatible with peptidyl-tRNA hydrolysis. The same conformation has also been observed in solution by small-angle X-ray scattering (SAXS) (Zoldák et al., 2007), although another SAXS study observed only the
extended conformation (Vestergaard et al., 2005). To rationalize these data, it has been hypothesized that class-I release factors adopt an initial codon-sampling state upon binding the ribosome (Rawat et al., 2003; Klaholz et al., 2003). Stop-codon recognition would then induce a conformational change and accommodation into the A site. Support for this hypothesis comes from hydroxyl-radical-probing (He & Green, 2010) and Förster-resonance-energy-transfer (Trapp & Joseph, 2016) experiments that show discrete release-factor conformations on the ribosome depending on the identity of the A-site codon. However, due to its presumably transient nature, structural data for a codon-sampling state have been lacking.

Upon peptidyl-tRNA hydrolysis, the class-I release factor remains bound in the A site of the ribosome and must be dislodged by the class-II release factor RF3 (Freistroffer et al., 1997), a translational GTPase most closely related to the translocase EF-G (Margus et al., 2007; Atkinson, 2015). Previously, it was thought that RF3 initially bound the post-release complex in its apo state, with the ribosome acting as a guanosine exchange factor that would enhance the binding of guanosine 5′-triphosphate (GTP) to RF3 (Zavialov et al., 2001; Zavialov et al., 2002; Pallesen et al., 2013). Subsequent kinetic studies have shown this model to be incorrect (Koutmou et al., 2014; Peske et al., 2014). Instead, RF3 behaves much like EF-G, primarily existing in a GTP-bound state in solution. Upon binding the post-release complex, RF3 is thought to induce rotation of the ribosomal small subunit relative to the large subunit (Jin et al., 2011; Zhou et al., 2012b), destabilizing the interaction between the class-I release factor and the ribosome. RF3 then hydrolyses GTP and dissociates, either shortly before or concurrently with the class-I release factor (Shi & Joseph, 2016). The ribosome then enters the recycling phase of translation in which the ribosomal subunits are split apart by the concerted action of RRF and EF-G.

A complication to this model is the fact that many bacterial species lack orthologues of RF3 (Margus et al., 2007). Furthermore, Escherichia coli can survive in the absence of a functional RF3, although the mutant strain has impaired growth (Mikuni et al., 1994; Grentzmann et al., 1994). This implies that class-I release factors have an intrinsic ability to dissociate from the ribosome after peptidyl-tRNA hydrolysis, with the efficiency of dissociation varying between species. There may also be alternative release factors that functionally replace RF3 in some species.
Part II of the thesis describes the mechanism of ArfA-mediated ribosome rescue, which is a noncanonical mechanism of translational termination in bacteria. In the absence of a stop codon, ribosomes arrest at the 3′ end of an mRNA and fail to terminate. Such nonstop complexes are usually rescued by a mechanism known as trans-translation (Keiler, 2015). However, in the absence of a functional trans-translation system, the small protein ArfA recognizes the empty mRNA channel and substitutes for the absent stop codon by recruiting the release factor RF2 to the A site of the ribosome (James et al., 2016). Using single-particle electron cryomicroscopy (cryo-EM), I obtained four high-resolution structures of nonstop complexes that reveal the mechanism of ArfA-mediated ribosome rescue and have wider implications for understanding canonical termination in bacteria.

I.1.3 TRANSLATIONAL INITIATION IN EUKARYOTES

Translational initiation is the process by which an elongation-competent ribosome is assembled on the start codon of an mRNA. The later steps of the initiation phase, particularly start-codon recognition and subunit joining, are conserved between bacteria, archaea, and eukaryotes. The eukaryotic pathway, however, involves major mechanistic additions and a greatly expanded repertoire of initiation factors.

Although there are many exceptions and complications, it is possible to sketch a general outline of translational initiation in eukaryotes that holds more or less true for most scenarios (Jackson et al., 2010; Hinnebusch & Lorsch, 2012; Aylett & Ban, 2017). First, the initiation factors eIF1, eIF1A, eIF2, eIF3, and eIF5, together with the initiator Met-tRNAi^{AUG}, bind the ribosomal small subunit to assemble the 43S preinitiation complex. This complex is then recruited to the mRNA, either to the 5′ end or an internal position, before the mRNA is loaded into the channel between the head and body of the small subunit. Following this, the initiation complex often scans along the mRNA in the 5′-to-3′ direction until it reaches the start codon (Kozak, 2002). Recognition of the start codon induces substantial rearrangement of the initiation complex, involving tighter closure of the mRNA channel and displacement of several initiation factors (Hashem et al., 2013; Hussain et al., 2014; Llácer et al., 2015; des Georges et al., 2015). Lastly, eIF5B recruits the ribosomal large subunit to
form the mature 80S initiation complex (Fernández et al., 2013). Upon dissociation of eIF5B, this complex is ready to enter the elongation phase of translation.

It is possible to define five mechanisms of translational initiation in eukaryotes (table I.1). These mechanisms are distinct from each other and may be further broken down into a variety of subcategories that depend on the particular properties of the mRNA and other factors involved in the initiation pathway. Canonical initiation is the most recognizable mechanism, though still poorly understood. This mechanism is defined by its dependence on a 7-methylguanosine cap covalently linked to the 5′ end of the mRNA via a 5′-to-5′ triphosphate bridge (Shatkin, 1976). The 5′ cap is usually recognized by eIF4E (Sonenberg et al., 1978), but other cap-binding proteins such as NCBP2 (Ishigaki et al., 2001) or eIF3d (Lee et al., 2016) can also direct canonical initiation. The cap-binding protein then recruits the preinitiation complex via interactions with other initiation factors, notably the scaffold protein eIF4G, and initiation proceeds largely as described above.

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Cap dependent?</th>
<th>End dependent?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canonical initiation</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>VPg-directed initiation</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>CITE-directed initiation</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Leaderless initiation</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Internal ribosome entry</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

Table I.1 | Mechanisms of translational initiation in eukaryotes, showing their dependence on the 5′ cap or 5′ end of an mRNA.

The remaining four mechanisms all represent noncanonical, cap-independent strategies for directing translational initiation. For example, instead of a cap, some viruses attach a protein to the 5′ end of their genomic RNA. This protein, termed the VPg (for viral protein, genome linked) has important functions in the viral infection cycle, most notably in replication and translation (Goodfellow, 2011; Jiang & Laliberté, 2011). Several VPg proteins have been shown to bind initiation factors, in particular eIF4E (Goodfellow et al., 2005; Hosmillo et al., 2014), and are therefore hypothesized to function as cap mimics, directing a mechanism similar to canonical
Part I: A general introduction to translation

initiation. However, a more complicated mechanism seems to be directed by the noroviral VPg. Although it binds eIF4E, this interaction is not required for translational activity (Chaudhry et al., 2006). Instead, the noroviral VPg interacts directly with eIF4G (Leen et al., 2016) and eIF3d (Daughenbaugh et al., 2003). More than just a cap mimic, the noroviral VPg makes novel contacts with initiation factors and directs initiation via a mechanism that is quite different to the canonical mechanism (Daughenbaugh et al., 2006; Chung et al., 2014).

Translational initiation can also be directed by cis-acting motifs within the mRNA known as cap-independent translational enhancers (CITEs) that directly recruit the preinitiation complex to internal positions (Terenin et al., 2013). Although there is no requirement for a 5′ cap or VPg, CITE-directed initiation is strictly dependent on the mRNA having an unobstructed 5′ end. This end dependence is the main distinction between CITE-directed initiation and internal ribosome entry (Terenin et al., 2017). Some of the best-characterized CITEs are located towards the 3′ ends of plant viruses and interact with their respective 5′ ends through long-range kissing-loop interactions and protein contacts (Simon & Miller, 2013). However, CITEs appear to be widely distributed among viral and cellular species and may be located anywhere within an mRNA. Indeed, any motif that interacts with the ribosome or initiation factors can function as a CITE (Jünemann et al., 2007; Terenin et al., 2013). It is also possible for capped mRNAs to contain translational enhancers that reduce the dependence of translational initiation on the 5′ cap and eIF4E. This may result in composite mechanisms that show aspects of both canonical and CITE-directed initiation, as seems to be the case for the mRNA of histone H4 (Martin et al., 2011) and the genomic RNA of flaviviruses (Edgil et al., 2006). Similarly, polyadenosine tails function as cis-acting translational enhancers that recruit initiation factors to internal positions and direct initiation in an end-dependent manner (Munroe et al., 1990; Wells et al., 1998), even in the absence of a 5′ cap (Preiss & Hentze, 1998).

Another mechanism, leaderless initiation, is somewhat of a catch-all for several different mechanisms that occur on leaderless or near-leaderless mRNA. The near or total absence of a 5′ untranslated region (UTR) precludes scanning and may affect the mechanism of mRNA loading. Remarkably, the best-characterized version of leaderless initiation, conserved between bacteria, archaea, and eukaryotes, has been shown to require only the ribosome and fMet- or Met-tRNA_{\text{AUG}} (Balakin et al., 1992;
Jones III et al., 1992; Moll et al., 2004; Udagawa et al., 2004; Andreev et al., 2006). Because it seems to depend on intrinsic properties of the ribosome rather than trans-acting initiation factors, this version of leaderless initiation is thought to represent the ancestral mechanism utilized by the last universal common ancestor of all cellular species (Londei, 2005), and may even be a molecular fossil of the RNA world (Allen & Frank, 2007). Alternatively, leaderless initiation can follow a similar pathway to canonical initiation, mediated by eIF2, although this is antagonized by eIF1 (Pestova & Kolupaeva, 2002), or it can rely on the alternative initiation factor eIF2D to recruit the Met-tRNA\textsubscript{i}\textsuperscript{AUG} (Dmitriev et al., 2010). The Met-tRNA\textsubscript{i}\textsuperscript{AUG} may also be recruited by eIF5B in a bacterial-like mechanism of translational initiation (Akulich et al., 2016). It seems that, in the absence of scanning, the initiation complex has a reduced dependence on eIF2 and may instead follow a variety of pathways that culminate in recruitment of a Met-tRNA\textsubscript{i}\textsuperscript{AUG} and subunit joining.

Perhaps the most distinctive mechanism of translational initiation is internal ribosome entry (Kieft, 2008; Filbin & Kieft, 2009; Balvay et al., 2009; Johnson et al., 2017; Lee et al., 2017; Yamamoto et al., 2017), which is directed by cis-acting motifs within the mRNA known as internal-ribosome-entry sites (IRESs). These motifs recruit the ribosome and initiation factors directly to internal positions, but tend to be larger and more intricately structured than CITEs. Most importantly, however, internal ribosome entry is the only mechanism that is end independent, having no requirement for an accessible 5′ end (Terenin et al., 2017). Because it is the most widely known noncanonical mechanism of initiation, internal ribosome entry has often been invoked to explain a variety of instances in viruses and cells where translation seems to be independent of a 5′ cap. However, end independence is more difficult to demonstrate than cap independence, and many of these instances are better accounted for by CITE-directed initiation or other end-dependent mechanisms. The structures and mechanisms of well-characterized IRESs are described in the Introduction to Part III.

The focus of Part III is on a particular example of internal ribosome entry directed by the IRES of hepatitis-C virus (HCV), which interacts directly with the mammalian ribosome and functionally replaces many of the canonical initiation factors (Johnson et al., 2017). However, the mechanism by which the HCV IRES directs translational initiation remains poorly understood. I developed a method for purifying native
ribosomal complexes from cell lysate that enabled me to obtain four cryo-EM maps of the HCV IRES in complex with the 80S ribosome, including two functional states and a previously unseen conformation of the IRES induced by rotation of the ribosomal subunits, and I have subsequently made progress towards capturing earlier steps in the initiation pathway.
I.2 BIBLIOGRAPHY


Part I: A general introduction to translation


Part I: A general introduction to translation


Part I: A general introduction to translation


Part I: A general introduction to translation


II.1 INTRODUCTION

II.1.1 RIBOSOME-ASSOCIATED QUALITY CONTROL

In all cells, quality-control pathways incessantly monitor the pool of synthesized proteins for signs of damage, misfolding, or mislocalization (Shoemaker & Green, 2012; Inada, 2013; Roy & Jacobson, 2013; Wolff et al., 2014; Shao & Hegde, 2016). Such aberrant proteins may aggregate, have a dominant-negative effect on cellular pathways, or even possess deleterious gain-of-function mutations. Quality control is therefore vital for survival. On the other hand, promiscuous recognition of normal proteins, which vastly outnumber aberrant proteins, would waste resources and could severely impair the ordinary functioning of the cell, meaning that quality-control pathways must recognize a wide variety of possible aberrations with high specificity.

Most quality-control pathways target aberrant proteins after they have been synthesized and released from the ribosome. However, in ribosome-associated quality control (RQC), also known as mRNA surveillance, nascent polypeptides are targeted for degradation (Dimitrova et al., 2009; Bengtson & Joazeiro, 2010), along with the mRNA (Frischmeyer et al., 2002; Doma & Parker, 2006) and sometimes even the ribosome itself (Cole et al., 2009). Because an mRNA must be translated before it can be recognized as defective, this is the only way in which a particular mRNA can be targeted for degradation before the end of its natural lifespan (Brandman & Hegde, 2016). As such, RQC pathways are vitally important for maintaining the quality of both mRNA and protein.

RQC pathways typically recognize ribosomes that have stalled during translational elongation and are unable to terminate (Buskirk & Green, 2017). A variety of situations can lead to stalling. If the problem can be resolved without leading to a build-up of aberrant proteins, trans-acting factors may enable the stalled ribosome to resume translation of the mRNA. We might refer to these ribosomes as paused. For example, ribosomes often pause during synthesis of polypeptide motifs, due to the restricted $\psi$ and $\phi$ angles of proline residues that constrain the geometry of the peptidyl-tRNA, and require the activity of EF-P (in bacteria) (Blaha et al., 2009) or eIF5A (in eukaryotes) (Schmidt et al., 2016) to enhance the peptidyl-transfer reaction (Ude et al., 2013; Doerfel et al., 2013; Peil et al., 2013).
Sometimes, however, translation can arrest indefinitely. Several mechanisms can lead to translational arrest, including amino-acid starvation (Brown et al., 2016), structures within the mRNA (Doma & Parker, 2006), or loss of an in-frame stop codon (Keiler, 2015). As they cannot resume translation, the ribosomal subunits must be rescued to prevent a lethal build-up of arrested complexes (Ramadoss et al., 2013). The aberrant mRNA and nascent polypeptide are also targeted for degradation.

II.1.2 Generating a Nonstop mRNA

All cells have therefore evolved a number of RQC pathways that recognize and rescue arrested ribosomes. In bacteria, many of these pathways lead to activation of a ribonuclease (RNase) and subsequent endonucleolytic cleavage of the mRNA. These pathways therefore converge on a single mechanism, trans-translation (described in section II.1.3.1), the bacterial nonstop-decay pathway, which rescues ribosomes that are stalled at the 3′ end of an mRNA. These arrested ribosomes are known as nonstop complexes and are unable to terminate due to the absence of a stop codon.

![Image](image_url)

**Figure II.1** | The fate of the ribosome during the stringent response to amino-acid starvation. RelA recognizes an uncharged tRNA in the A site and synthesizes the alarmones guanosine 5′-diphosphate 3′-diphosphate (ppGpp) and guanosine 5′-triphosphate 3′-diphosphate (pppGpp) (Brown et al., 2016). What governs the eventual dissociation of RelA is unclear. The alarmones set off the stringent response, leading to activation of RelE which cleaves mRNA in the A site (Neubauer et al., 2009). RelE is promiscuous, targeting all actively translating ribosomes in the cell and leaving them arrested on 3′-truncated mRNA.

Nonstop complexes may occur spontaneously during translational elongation as a result of stop-codon readthrough by near-cognate tRNA or ribosomal frameshifting (Giudice & Gillet, 2013; Keiler, 2015). Alternatively, the mRNA itself may be defective. It is estimated that between 2 and 4% of bacterial mRNAs lack an in-frame stop codon due to aberrant transcription or endonucleolytic cleavage (Ito et al., 2011).
Part II: The mechanism of ArfA-mediated ribosome rescue

The latter often occurs as a programmed response to specific stimuli. For example, the stringent response to amino-acid starvation is initially stimulated by RelA, which recognizes ribosomes that have arrested with an uncharged tRNA in the A site (Brown et al., 2016), and ultimately leads to activation of the endoRNase RelE (figure II.1). Although it has some specificity, preferring the sequence YRG (where Y is a pyrimidine and R is a purine), RelE is promiscuous enough to cleave any mRNA between the second and third residues of the A-site codon (Pedersen et al., 2003; Neubauer et al., 2009). This leaves the ribosome arrested on a 3′-truncated mRNA, which is then targeted for nonstop decay (Christensen & Gerdes, 2003).

Other problems that lead to translational arrest include rare codons, structures within the mRNA (Doma & Parker, 2006), damage to the mRNA (Simms et al., 2014) or ribosomal RNA (rRNA) (LaRivière et al., 2006; Cole et al., 2009), and polyadenosine tails (Juszkiewicz & Hegde, 2017; Guydosh & Green, 2017). These ribosomes all have one feature in common: they are all arrested in the middle of an mRNA but their A sites are at least transiently unoccupied. The RQC pathway that deals with such complexes is referred to as no-go decay in eukaryotes. An equivalent pathway exists in bacteria and culminates in endonucleolytic cleavage of the A-site codon (Hayes & Sauer, 2003; Sunohara et al., 2004a; Sunohara et al., 2004b; Li et al., 2006; Li et al., 2007). The identity of the endoRNase remains unknown, but a number of candidates (including RelE) have been excluded. However, endonucleolytic cleavage has been shown to depend on the prior activity of RNase II, which may act in synergy with polynucleotide phosphorylase, a component of the proteobacterial degradosome (Garza-Sánchez et al., 2009; Janssen et al., 2013). Both proteins digest the mRNA from the 3′ end, which seems to be a prerequisite for subsequent cleavage in the A site. Perhaps this shortening of the mRNA is necessary to dislodge it from the channel and provide access to an endoRNase. This would explain how arrested complexes are recognized specifically, in contrast to the promiscuous activity of RelE.

II.1.3 NONSTOP DECAY IN BACTERIA

Thus, nonstop decay is arguably the most important RQC pathway in the bacterial cell because it has to cope with high traffic of arrested ribosomes from a variety of sources. In most bacteria, as well as plastids and some mitochondria, the primary
Part II: The mechanism of ArfA-mediated ribosome rescue

mechanism of nonstop decay is trans-translation (Keiler, 2015). This mechanism is totally distinct from the nonstop-decay pathway of the eukaryotic cytosol, with none of the same factors and major mechanistic differences, and yet it leads to essentially the same outcomes (ribosome rescue, degradation of the nascent polypeptide, and degradation of the mRNA). In addition, bacteria have evolved at least two alternative mechanisms of ribosome rescue mediated by the small proteins ArfA and ArfB. I will discuss each of these mechanisms in turn.

II.1.3.1 Trans-translation

Trans-translation involves an unusual RNA molecule known as transfer-messenger RNA (tmRNA), which functions as both a tRNA and an mRNA (Keiler et al., 1996; Atkins & Gesteland, 1996). In some bacteria, tmRNA is split into two chains, but the overall conformation is retained (Keiler et al., 2000). The tRNA-like domain (TLD) of tmRNA resembles the acceptor stem and T arm of an Ala-tRNA and, as a consequence, is alanylated on its 3’ CCA end by the corresponding synthetase (Komine et al., 1994; Ushida et al., 1994). The TLD, along with its cofactor SmpB (Karzai et al., 1999; Gutmann et al., 2003), is then loaded into the A site of the ribosome by EF-Tu (Rudinger-Thirion et al., 1999; Barends et al., 2000; Valle et al., 2003) (figure II.2). SmpB inserts a C-terminal α-helix into the mRNA channel (Sundermeier et al., 2005; Kurita et al., 2010; Neubauer et al., 2012), enabling specific recognition of a nonstop complex (Ivanova et al., 2004; Moore & Sauer, 2005). The nascent polypeptide is then transferred to the TLD, which is translocated through the ribosome by EF-G as if it were an elongator tRNA (Weis et al., 2010; Fu et al., 2010; Ramrath et al., 2012).

In addition to the TLD, tmRNA contains an mRNA-like region (MLR) that is loaded into the mRNA channel of the 30S subunit, allowing translational elongation to continue in trans (Tu et al., 1995; Williams & Bartel, 1996; Felden et al., 1996; Williams et al., 1999). This presents a topological problem for tmRNA, as the channel is closed during the elongation phase and must be opened by tilting the head of the 30S subunit (Weis et al., 2010; Ramrath et al., 2012). The details are not yet understood, but opening of the channel seems to be mediated at least in part by helix 2b, located near the TLD, which is forced underneath the central protuberance of the 60S subunit and acts as a wedge to tilt the 30S head downwards. In addition, up to
four pseudoknots fold into a ring-like structure that loops around the beak of the 30S head and connects the ends of the MLR to the TLD.

The MLR encodes a degradation signal followed by a stop codon that recruits a class-I release factor and allows translational termination to proceed as normal, thereby rescuing the ribosomal subunits and directing the tagged polypeptide for proteolytic degradation (Gottesman et al., 1998; Herman et al., 1998; Flynn et al., 2001; Choy et al., 2007).

Figure II.2 | The trans-translation pathway. EF-Tu delivers tmRNA and SmpB to the nonstop complex. SmpB recognizes the empty mRNA channel, allowing the TLD to accommodate into the peptidyltransferase centre. The nascent polypeptide is then transferred to the TLD, which adopts a hybrid conformation upon rotation of the 30S subunit. SmpB and tmRNA are subsequently translocated into the P site by EF-G in much the same way as an elongator tRNA. At the same time, the first codon of the MLR enters the A site and another tRNA can be recruited. The elongation cycle is then repeated in trans until the degradation tag is synthesized and a stop codon is reached, allowing canonical termination to occur. The tagged polypeptide is targeted for proteolysis, while RNase R degrades the 3′-truncated mRNA.

Trans-translation also leads to degradation of the 3′-truncated mRNA (Yamamoto et al., 2003). In E. coli, this degradation is mediated by RNase R, which is recruited to the nonstop complex (Richards et al., 2006; Ge et al., 2010; Venkataraman et al., 2014a). Mutations in tmRNA at the 3′ end of the coding region disrupt the recruitment of RNase R (Mehta et al., 2006; Venkataraman et al., 2014b), but whether tmRNA interacts directly with it or has an allosteric effect remains unclear. Many bacteria, including E. coli, lack 5′-to-3′ exoRNases and instead rely on 3′-to-5′ exoRNases such as RNase R (Mohanty & Kushner, 2016). This leads to a topological problem in
nonstop decay, as the 3′ end is protected inside the ribosome. Thus, RNase R presumably waits until the 3′ end emerges from the mRNA channel.

The trans-translation pathway is necessary for virulence in many bacterial species (Julio et al., 2000; Okan et al., 2010; Mann et al., 2012; Svetlanov et al., 2012), yet is not found in human cells. Trans-translation is therefore a promising target for antibiotic development (Keiler, 2015). Some progress has already been made in this regard, as a number of compounds have been shown to inhibit trans-translation in vitro and impair the growth of Bacillus anthracis, Shigella flexneri, and Mycobacterium smegmatis (Ramadoss et al., 2013). However, any therapeutic approach would need to circumvent the back-up mechanisms of alternative ribosome-rescue factors ArfA and ArfB, which enable some species to survive in the absence of a functional trans-translation system (Giudice & Gillet, 2013; Keiler, 2015).

II.1.3.2 ArfA-mediated ribosome rescue

ArfA in particular has been shown to act as a fail-safe for trans-translation. In E. coli, ArfA can support continued growth in the absence of tmRNA (Komine et al., 1994; Chadani et al., 2010). The double knockout of ArfA and tmRNA, however, is lethal to the cell. The arfA transcript contains a stem-loop within the coding region that is recognized and cleaved by RNase III. Ribosomes subsequently stall on this 3′-truncated mRNA and are rescued by the trans-translation system, which targets the truncated ArfA protein for degradation (Chadani et al., 2011a; Garza-Sánchez et al., 2011; Schaub et al., 2012). A small population of transcripts are thought to escape this fate, leading to synthesis of full-length ArfA, but the C-terminal tail of the full-length protein is unstable and confers a short half-life, keeping its expression to a minimum (Chadani et al., 2011a). However, when the trans-translation system is impaired or overwhelmed, full-length ArfA may rescue the synthesis of its truncated isoform, leading to a rapid increase in the concentration of truncated, though functional, ArfA proteins.

ArfA is a small protein that binds in the A site of the ribosome and specifically recruits RF2 (but not RF1) to nonstop complexes (Chadani et al., 2012; Shimizu, 2012). In this manner, ArfA functionally mimics a stop codon, but the structural basis of this mechanism has been unclear. RF2 then catalyses release of the nascent polypeptide. In contrast to the trans-translation system, ArfA does not target the
mRNA or nascent polypeptide for degradation and is therefore better suited to functioning as a fail-safe for *trans*-translation rather than a frontline mechanism of ribosome rescue. Nevertheless, the fact that ArfA alone can support growth demonstrates that ribosome rescue is the most important function of nonstop decay.

**II.1.3.3 ArfB-mediated ribosome rescue**

ArfB, by contrast, is homologous to domain 3 of the class-I release factors and is able to directly hydrolyse peptidyl-tRNA within nonstop complexes (Chadani *et al.*, 2011b; Handa *et al.*, 2011). The C-terminal α-helix of ArfB, separated from the globular N domain by an unstructured region, binds in the mRNA channel while the catalytic GGQ motif accommodates into the peptidyltransferase centre (Gagnon *et al.*, 2012). In this way, ArfB mediates ribosome rescue without targeting the mRNA or nascent polypeptide for degradation (figure II.3).

![Figure II.3](image)

*Figure II.3* | ArfB-mediated ribosome rescue. ArfB recognizes the nonstop complex by inserting a C-terminal α-helix into the empty mRNA channel. The N domain of ArfB, homologous to the catalytic domains of RF1 and RF2, accommodates into the peptidyltransferase centre and hydrolyses the peptidyl-tRNA. Although the ribosomal subunits are rescued, the mRNA and nascent polypeptide are not targeted for degradation. The mechanism by which ArfB dissociates from the ribosome has not been determined, but might simply be spontaneous.

The *arfB* gene is found in 34% of sequenced bacterial genomes (Feaga *et al.*, 2014), and an orthologue with a plastidial targeting signal has been identified in plant genomes (Feaga *et al.*, 2016). Another orthologue, ICT1, is also found in mammalian mitochondrial where, in addition to being a soluble protein, it has become an integral component of the mitoribosome (Amunts *et al.*, 2015; Greber *et al.*, 2015) termed mL62 (Greber & Ban, 2016).

In *Caulobacter crescentus*, which lacks an orthologue of ArfA, the double knockout of tmRNA and ArfB is synthetically lethal, demonstrating that ArfB is necessary for
survival in the absence of trans-translation (Feaga et al., 2014). However, ArfB cannot fully compensate for the loss of trans-translation, as the tmRNA knockout strain has a severe growth defect (Keiler & Shapiro, 2003). This effect is even more pronounced in *E. coli*, which cannot survive in the absence of both tmRNA and ArfA unless ArfB is overexpressed (Chadani et al., 2011b). Furthermore, no phenotypic impairment in either species has been observed when ArfB is knocked out. Thus, the function of ArfB in vivo remains contentious. A clue to a possible function emerged when it was found that, although ArfB preferentially rescues nonstop complexes, it can also rescue ribosomes that are stalled in the middle of an mRNA (Shimizu, 2012). In this way, ArfB resembles the eukaryotic protein pelota (Shao et al., 2016). Perhaps ArfB functions primarily as a back-up for the bacterial equivalent of no-go decay, rather than nonstop decay.
II.2 RESULTS

II.2.1 SCIENTIFIC AIMS

ArfA is an important fail-safe that enables bacterial cells to survive and grow when trans-translation is impaired or overwhelmed. The mechanism that governs its expression is well characterized, but the structural basis by which ArfA rescues nonstop complexes was not known. To obtain a detailed mechanistic understanding of ArfA-mediated ribosome rescue, I solved the structures of multiple nonstop complexes by cryo-EM to resolutions between 3.0 and 3.4 Å (figure II.6). The structures reveal two distinct conformations of RF2 on the ribosome and explain how ArfA recognizes nonstop complexes, recruits RF2, and mediates accommodation of the catalytic GGQ motif into the peptidyltransferase centre.

Figure II.4 | Data collection. (A) A representative micrograph. (B) Representative classes from reference-free 2D classification.

II.2.2 SAMPLE PREPARATION AND DATA PROCESSING

I prepared four samples of *E. coli* ribosomes programmed with 3′-truncated mRNA and a nonhydrolysable fMet-NH-tRNA<sub>\text{AUG}</sub> that serves as a peptidyl-tRNA analogue. A similar nonhydrolysable tRNA has been used previously to capture termination complexes in an on-pathway state after accommodation of RF2 but before peptidyl-tRNA hydrolysis (Jin *et al.*, 2010). Each sample also contained ArfA, ArfA and RF2,
ArfA(A18T) and RF2, or ArfA and *Thermus thermophilus* RF2 (TtRF2) (table II.1). The intention was to capture distinct conformational states with each sample, to better characterize the mechanism of ArfA-mediated ribosome rescue. The amino-acid substitution A18T abolishes the ability of ArfA to support peptidyl-tRNA hydrolysis, though the mutant can still bind nonstop complexes and recruit RF2 (Chadani *et al.*, 2010; Shimizu, 2012). Similarly, TtRF2 is recruited to the ribosome by wild-type ArfA but fails to hydrolyse the peptidyl-tRNA (Zeng *et al.*, 2017).

<table>
<thead>
<tr>
<th>Sample</th>
<th>Component</th>
<th>mRNA</th>
<th>fMet-NH$<em>2$tRNA$</em>{AUG}$</th>
<th>ArfA</th>
<th>RF2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>70S ribosome</td>
<td>mRNA</td>
<td>fMet-NH$<em>2$tRNA$</em>{AUG}$</td>
<td>ArfA</td>
<td>–</td>
</tr>
<tr>
<td>II</td>
<td>70S ribosome</td>
<td>mRNA</td>
<td>fMet-NH$<em>2$tRNA$</em>{AUG}$</td>
<td>ArfA</td>
<td>RF2</td>
</tr>
<tr>
<td>III</td>
<td>70S ribosome</td>
<td>mRNA</td>
<td>fMet-NH$<em>2$tRNA$</em>{AUG}$</td>
<td>ArfA(A18T)</td>
<td>RF2</td>
</tr>
<tr>
<td>IV</td>
<td>70S ribosome</td>
<td>mRNA</td>
<td>fMet-NH$<em>2$tRNA$</em>{AUG}$</td>
<td>ArfA</td>
<td>TtRF2</td>
</tr>
</tbody>
</table>

Table II.1 | Components of each sample. Each component was visible in the final maps, with the exception of ArfA in sample I. RF2 seems to be necessary to stabilize ArfA in a well-defined conformation.

Micrographs were collected for each sample and processed in RELION (Fernández-Leiro & Scheres, 2016) (figure II.4; see section II.3.3 in the Materials and methods). After an initial round of three-dimensional (3D) classification without alignment, one major class was isolated from each dataset. For the sample with wild-type ArfA and RF2, this class contained both proteins, but the local resolution was slightly worse than for the surrounding ribosome. This is because the ribosome dominates the signal in the reconstruction, making it difficult to classify minor differences resulting from flexibility or differential occupancy. To improve the occupancy of ArfA and RF2, I made a local mask around RF2 and subtracted the surrounding density from the extracted particles, which were then subjected to another round of 3D classification using the local mask. This procedure, known as focused classification with signal subtraction (FCwSS) (Bai *et al.*, 2015; Brown *et al.*, 2016), enabled me to discard unoccupied ribosomes from the final reconstruction, despite the similarity of their overall conformations. The classified particles were subsequently replaced with the original particles (from before the signal subtraction) and refined. A final round of 3D classification with alignment was used as a final clean-up to improve the angular
accuracy of the reconstruction. The classification scheme is outlined in figure II.5. A similar scheme was used for each of the other datasets.

**Figure II.5** | *In-silico* 3D classification. A single well-defined map for the nonstop complex containing wild-type ArfA and RF2 was isolated after three rounds of *in-silico* classification. 186,210 particles selected from 2D classification were classified while preserving the Euler angles from a prior 3D refinement step. Nonstop complexes were separated from empty 70S ribosomes, isolated 50S subunits, and poorly aligned particles. FCwSS, using a mask over RF2, removed further empty ribosomes. A final 3D classification, this time with angular sampling, removed any remaining particles that poorly aligned. The final reconstruction has a nominal resolution of 2.97 Å. All other datasets were classified similarly, with any changes described in the Materials and methods.
**Part II: The mechanism of ArfA-mediated ribosome rescue**

<table>
<thead>
<tr>
<th></th>
<th>70S (empty A site)</th>
<th>70S•ArfA•RF2</th>
<th>70S•ArfA•(A18T)•RF2</th>
<th>70S•ArfA•TtRF2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Particles</td>
<td>140,027</td>
<td>139,792</td>
<td>141,950</td>
<td>54,465</td>
</tr>
<tr>
<td>Pixel size (Å)</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>Defocus mean (µm)</td>
<td>−2.1</td>
<td>−2.3</td>
<td>−2.4</td>
<td>−2.4</td>
</tr>
<tr>
<td>Defocus range (µm)</td>
<td>−0.7 to −3.9</td>
<td>−0.6 to −4.8</td>
<td>−0.5 to −5.8</td>
<td>−1.1 to −3.9</td>
</tr>
<tr>
<td>Voltage (kV)</td>
<td>300</td>
<td>300</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Electron dose (e⁻ Å⁻²)</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td><strong>Model composition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-hydrogen atoms</td>
<td>146,646</td>
<td>149,895</td>
<td>149,614</td>
<td>149,470</td>
</tr>
<tr>
<td>Protein residues</td>
<td>5,971</td>
<td>6,379</td>
<td>6,344</td>
<td>6,321</td>
</tr>
<tr>
<td>RNA bases</td>
<td>4,628</td>
<td>4,628</td>
<td>4,628</td>
<td>4,628</td>
</tr>
<tr>
<td>Ligands (Zn²⁻/Mg²⁻/H₂O)</td>
<td>2/449/2</td>
<td>2/453/2</td>
<td>2/450/2</td>
<td>2/462/3</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>3.1</td>
<td>3.0</td>
<td>3.1</td>
<td>3.4</td>
</tr>
<tr>
<td>Map sharpening B-factor (Å²)</td>
<td>−79.6</td>
<td>−77.0</td>
<td>−75.8</td>
<td>−91.3</td>
</tr>
<tr>
<td>Average B factor (Å²)</td>
<td>95.8</td>
<td>75.3</td>
<td>99.1</td>
<td>85.2</td>
</tr>
<tr>
<td>FSC_average</td>
<td>0.89</td>
<td>0.90</td>
<td>0.89</td>
<td>0.85</td>
</tr>
<tr>
<td>CRef (Å)</td>
<td>3.1</td>
<td>3.0</td>
<td>3.1</td>
<td>3.4</td>
</tr>
<tr>
<td><strong>Rms deviations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.006</td>
<td>0.006</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.08</td>
<td>1.10</td>
<td>1.16</td>
<td>1.20</td>
</tr>
<tr>
<td><strong>Validation (proteins)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MolProbity score (percentile)</td>
<td>2.0 (99&lt;sup&gt;th&lt;/sup&gt;)</td>
<td>2.1 (99&lt;sup&gt;th&lt;/sup&gt;)</td>
<td>2.2 (99&lt;sup&gt;th&lt;/sup&gt;)</td>
<td>2.4 (98&lt;sup&gt;th&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Clashscore, all atoms (percentile)</td>
<td>2.7 (100&lt;sup&gt;th&lt;/sup&gt;)</td>
<td>2.8 (100&lt;sup&gt;th&lt;/sup&gt;)</td>
<td>3.2 (100&lt;sup&gt;th&lt;/sup&gt;)</td>
<td>4.2 (100&lt;sup&gt;th&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Good rotamers (%)</td>
<td>83.0</td>
<td>82.2</td>
<td>80.9</td>
<td>77.1</td>
</tr>
<tr>
<td><strong>Ramachandran plot</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favoured (%)</td>
<td>94.9</td>
<td>94.7</td>
<td>93.9</td>
<td>93.4</td>
</tr>
<tr>
<td>Outliers (%)</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td><strong>Validation (RNA)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct sugar puckers (%)</td>
<td>97.8</td>
<td>97.9</td>
<td>98.2</td>
<td>97.8</td>
</tr>
<tr>
<td>Good backbone conformations (%)</td>
<td>81.4</td>
<td>81.3</td>
<td>80.8</td>
<td>78.8</td>
</tr>
<tr>
<td><strong>Accession codes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMDB</td>
<td>EMD-3493</td>
<td>EMD-3489</td>
<td>EMD-3490</td>
<td>EMD-3492</td>
</tr>
<tr>
<td>PDB</td>
<td>5MDZ</td>
<td>5MDV</td>
<td>5MDW</td>
<td>5MDY</td>
</tr>
</tbody>
</table>

**Table II.2 | Data-collection and model statistics.**
Figure II.6 | Quality of the maps and models. (A) FSC curves for each cryo-EM reconstruction, shown in black. The resolutions at FSC=0.143 and FSC=0.5 (CRef) are indicated with a red and a grey dashed line, respectively. FSC curves of the fit of the refined model to the final map for each structure are shown in green. The self- and cross-validated correlations are shown in purple and blue, respectively. (B) The cryo-EM map for each complex, coloured according to the local resolution. (C) Isolated density for RF2 and TtRF2, coloured according to the local resolution, for each complex. (D) Density and model for ArfA for each complex. In the complex with wild-type ArfA and RF2, additional density is visible that corresponds to the ordered N-terminus of ArfA.

In three of the datasets, the density for ArfA was well resolved, allowing it to be built de novo (figure II.6D). RF2 was similarly well resolved (figure II.6C). However,
in the dataset collected without RF2, no ArfA was visible in the map even after extensive 3D classification. It is thought that ArfA initially binds nonstop complexes in the absence of RF2, but it may sample many conformational states and require RF2 to fix it in one position. This dataset was used to build a model of a nonstop complex with an unoccupied A site (see section II.3.3 in the Materials and methods). The statistics for the datasets and models are recorded in table II.2.

**Figure II.7 |** Structures of nonstop complexes recognized by ArfA and ArfA(A18T). (A) Overview of the ribosome programmed with a 3′-truncated mRNA in complex with a nonhydrolysable fMet-NH\textsubscript{i}tRNA\textsubscript{AUG}, ArfA, and RF2 (left). RF2 adopts a similar conformation with ArfA as it does with a UAA stop codon (PDB accession code 4V67) (right). There is no structure of *E. coli* RF2 bound to the ribosome, so TtRF2 is shown. The catalytic GGQ motif of domain 3 is accommodated within the peptidyltransferase centre (inset). (B) Overview of the ribosome with ArfA(A18T) (left). The preaccommodated conformation of RF2 resembles the conformation seen in the crystal structure of the isolated protein (PDB accession code 1GQE) (right). The GGQ motif is placed up against the P-site tRNA. The movement of domain 1 results from contacts with the bL12 stalk base.
II.2.3 INTERPRETATION

The structures reveal that RF2 adopts two distinct conformations on the ribosome (figure II.7). The interactions that confer specificity for nonstop complexes and RF2 remain similar between the two states, but the observation of a preaccommodated state helps to distinguish the regions of ArfA that are responsible for initial recognition from those that mediate accommodation of RF2 into the peptidyltransferase centre. Here, each of these interactions is described in turn.

Figure II.8 | Conformations of the decoding nucleotides. (A) Conformation of the decoding centre with an unoccupied A site. A1493 is disordered in this and all ArfA-containing structures. (B) In the preaccommodated state, ArfA(A18T) recognizes a vacant A site without remodelling the decoding centre. (C) Upon accommodation, closer contacts between wild-type ArfA, RF2, and the ribosome cause A1492 to switch from a syn to an anti configuration. (D) In contrast to ArfA, a UAA stop codon remodels the decoding centre during canonical termination (PDB accession code 4V67).

II.2.3.1 Specificity for nonstop complexes

The decoding centre, comprising three conserved nucleotides in the A site of the 30S subunit, is known to make functionally important interactions with the A-site codon in translational termination (Weixlbaumer et al., 2008; Laurberg et al., 2008; Korostelev et al., 2008; Korostelev et al., 2010). As ArfA functionally mimics a stop codon, it is interesting to compare the conformation of the decoding centre in canonical and
Part II: The mechanism of ArfA-mediated ribosome rescue

ArfA-mediated termination. In the three ArfA-containing structures, ArfA contacts the decoding-centre nucleotides G530 and A1492 of the 16S rRNA (figure II.8). A1492 is stacked within helix 44, interacting with ArfA via the rRNA backbone. The third key nucleotide of the decoding centre, A1493, is flipped out and disordered. These nucleotides are unchanged from their positions in an unoccupied A site (figure II.8A). This is in contrast to canonical termination, in which the stop codon remodels the decoding centre by flipping out A1492 and placing A1493 into helix 44 (figure II.8D). ArfA therefore seems to interact specifically with a vacant decoding centre.

**Figure II.9** | Interactions between ArfA, RF2, and the ribosome. (A) The N-terminal half of ArfA binds RF2 while the C-terminal tail occupies the mRNA channel. The path of ArfA as it emerges from the channel entrance, which can be traced in unsharpened maps, is indicated with a green dashed line. The boxed area indicates the region of ArfA shown in (B). (B) Interactions between ArfA and the rRNA lining the mRNA channel for the region highlighted in (A), as viewed from the channel entrance. (C) Both wild-type ArfA and ArfA(A18T) form an antiparallel β-addition motif with the β-sheet of RF2 domain 2. F25 packs against an RF2-specific hydrophobic pocket formed by V198 and F217. With wild-type ArfA, this pocket is also recognized by W319 from the switch loop of accommodated RF2. Single-letter abbreviations for the amino-acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

It has recently been shown that the moderately conserved P23 residue of ArfA (figure II.10A), which contacts A1492, is functionally important, as a P23A substitution severely impairs the rate of peptidyl-tRNA hydrolysis (Zeng et al., 2017). The authors proposed that P23 might help to stabilize the stacking of A1492 within helix 44. However, the observation that A1492 is flipped in by default, when the
mRNA channel is empty (figure II.8A), suggests that P23 may be more important for maintaining the overall conformation of ArfA, as it mediates a 3_10-helical turn around the body of RF2 (figure II.10A). Thus, an E30A substitution, which would alter the contact with G530, has minimal effect on peptidyl-tRNA hydrolysis (Zeng et al., 2017), as expected if ArfA passively recognizes the conformation of the vacant decoding centre.

The C-terminal tail of ArfA protrudes into the mRNA channel downstream of the decoding centre, occupying space that would otherwise contain mRNA, and makes extensive contacts with the rRNA around the inner walls of the channel (figure II.9A,B), in agreement with hydroxyl-radical-probing data (Kurita et al., 2014). Although the C-terminus lacks secondary structure, the conserved KGKGS motif (figure II.10A; residues 34 to 37) forms an expansive helical turn that may help to anchor ArfA within the channel (figure II.9B). Recent mutagenic data have shown that, in the region spanning residues 26 to 41, no single positively charged residue has a significant influence on peptidyl-tRNA hydrolysis (Zeng et al., 2017), implying that the cumulative effect of many electrostatic contacts is important for the function of ArfA. Further downstream, the C-terminus interacts with helix 1 of the 16S rRNA and then contacts helix 18 on the opposite wall of the channel (figure II.9B). The model could be built up to Gly46, near the channel entrance; beyond this, scattered density is visible and has more recently been interpreted as a putative α-helix (Demo et al., 2017), but there are few contacts with the ribosome and it was not possible to place any additional residues. This is consistent with the last 17 residues being truncated in the mature protein and dispensable for ribosome rescue (Garza-Sánchez et al., 2011).

The position of the C-terminal tail would clash with mRNA in an actively translating ribosome and thus confers specificity for nonstop complexes. A similar model has been proposed for ArfB (Gagnon et al., 2012) and SmpB (Neubauer et al., 2012), which both insert their C-termini into the mRNA channel, although these tails and their interactions with the ribosome are distinct from ArfA (figure II.11). A discriminatory function for the ArfA C-terminus is consistent with the dramatic decline of ArfA-dependent peptide release as the mRNA is extended in the 3′ direction (Shimizu, 2012; Zeng & Jin, 2016). However, recent data have shown that ArfA can bind ribosomes, and recruit RF2, even when mRNA occupies the channel (Kurita et al., 2014). Thus it seems that simultaneous occupancy of the channel may
Part II: The mechanism of ArfA-mediated ribosome rescue

![Diagram of ArfA mechanism](image)
Part II: The mechanism of ArfA-mediated ribosome rescue

**Escherichia coli**

- **Actinobacillus pleuropneumoniae**
- **Citrobacter farmacci**
- **Citrobacter koseri**
- **Enterobacter cloacae**
- **Haemophilus influenzae**
- **Haemophilus parasuis**
- **Halomonas chromatireducens**
- **Klebsiella pneumoniae**
- **Neisseria meningitidis**
- **Pantoaea ananatis**
- **Paraglaciecola polaris**
- **Proteus vulgaris**
- **Providencia alcalifaciens**
- **Pseudalteromonas translucida**
- **Pseudomonas aeruginosa**
- **Salmoneilla derby**
- **Salmoneilla gallinarum**
- **Serratia plymuthica**
- **Serratia symbiotica**
- **Shewanella oneidensis**
- **Shigella sonnei**
- **Streptococcus pneumoniae**
- **Vibrio campbellii**
- **Vibrio cholerae**
- **Yersinia aldovae**
- **Yersinia berovieri**

**Thermus thermophilus**
Part II: The mechanism of ArfA-mediated ribosome rescue.
Part II: The mechanism of ArfA-mediated ribosome rescue

**Figure II.10** | Sequence alignments. (A) Alignment of selected ArfA sequences. The numbering and secondary structure correspond to *E. coli* ArfA. (B) Alignment of the recognition and switch loops of RF2 from ArfA-containing species and from *T. thermophilus*, which lacks an ArfA orthologue. The numbering and secondary structure correspond to *E. coli* RF2. Hydrophobic residues that confer specificity for ArfA are indicated with arrowheads. (C) Alignment of the recognition and switch loops of RF1. The numbering corresponds to *E. coli* RF1. All sequences were obtained from UniProt (UniProt Consortium, 2015), aligned with MUSCLE (Edgar, 2004), and visualized in Jalview (Waterhouse et al., 2009). Residues are coloured only when they agree with the consensus sequence and shaded according to the percentage of agreeing residues.

be possible, but ArfA would not be able to adopt the conformation seen here and only in the absence of mRNA would it be able to position RF2 in a productive conformation for peptidyl-tRNA hydrolysis.

**II.2.3.2 Specificity for RF2**

How does ArfA interact with RF2, and what confers the specificity for this particular protein? The basis for RF2 recruitment is provided by ArfA residues 25 to 30, which form an antiparallel β-addition motif with the β5-strand of RF2 domain 2 (figure II.9C). In addition to these backbone interactions, specificity for RF2 seems to be conferred by F25 of ArfA, which protrudes into a conserved hydrophobic pocket of RF2 formed by V198 from the β4-strand and F217 from the β5-strand (figure II.9C). RF1 lacks such a pocket in all known ArfA-containing species (figure II.10B,C), likely explaining the exclusive recruitment of RF2 to ArfA (Chadani et al., 2012; Shimizu, 2012).

The recognition loop of RF2 between the β4- and β5-strands, containing the SPF motif that confers stop-codon specificity (Weixlbaumer et al., 2008), faces solvent and does not interact with ArfA or the decoding centre. This is consistent with the SPF motif being dispensable in ArfA-mediated termination (Chadani et al., 2012).

**II.2.3.3 The mechanism of accommodation**

In the complex with wild-type ArfA, RF2 adopts an extended conformation that resembles the crystal structures of release factors bound to stop codons (Weixlbaumer et al., 2008; Laurberg et al., 2008; Korostelyev et al., 2008; Korostelyev et al., 2010) (figure II.7A). The catalytic GGQ motif of domain 3 is accommodated within the peptidyltransferase centre (figure II.7A, inset), which is consistent with a shared mechanism of catalysis for canonical and ArfA-mediated termination. In the presence
of ArfA(A18T), however, RF2 adopts a compact conformation that resembles the crystal structures of isolated release factors (Vestergaard et al., 2001; Shin et al., 2004; Zoldák et al., 2007). The only difference is the position of domain 1, which has moved away from the other domains due to contacts with the bL12 stalk base. In this preaccommodated state, domain 3 lies across an eight-stranded β-sheet formed by domains 2 and 4, with the GGQ motif facing the anticodon arm of the P-site tRNA, ~60 Å from the peptidyltransferase centre (figure II.7B). This structure validates previous predictions that class-I release factors can adopt a compact conformation on the ribosome that likely functions as an initial codon-sampling state (Rawat et al., 2003; Klaholz et al., 2003), as described in Part I, section I.1.2.

![Figure II.11](image)

**Figure II.11** | ArfA, SmpB, and ArfB contact the mRNA entry channel with distinct C-termini. (A) ArfA has an unstructured C-terminal tail that protrudes into the channel. (B) SmpB (PDB accession code 4V8Q) (Neubauer et al., 2012) inserts a C-terminal α-helix deep into the channel, occupying a position near the entrance. The globular N domain binds tmRNA. (C) The C-terminal α-helix of ArfB (PDB accession code 4V95) (Gagnon et al., 2012) occupies a higher position close to the decoding centre. The globular N domain is homologous to domain 3 of RF2 and accommodates into the peptidyltransferase centre.

Here, the preaccommodated conformation results from the A18T substitution, which prevents the interdependent folding of the ArfA N-terminus and RF2 switch loop. The switch loop connects the α7-helix of domain 3 to the β9-strand of domain 4 and has been proposed to mediate accommodation during canonical termination by
undergoing a disorder-to-order transition (Laurberg et al., 2008). In agreement with this, the RF2 switch loop and the first 14 residues of ArfA(A18T) are disordered. A18T is located within the α-helix of ArfA that packs against the central β-sheet of RF2, but faces away from this interface (figure II.12A). However, in the structure with wild-type ArfA, the N-terminus of ArfA is ordered and turns tightly to run antiparallel to this α-helix (figure II.12B). Compared to alanine, a polar threonine would be unable to pack against I11 in this conformation. As expected, an I11N substitution also severely impairs ArfA activity (Zeng et al., 2017), but a hydrophobic A18C substitution has little effect (Kurita et al., 2014). The curvature of this region, which packs tightly between the switch loop and helix 69 of the 23S rRNA, is also important for accommodation, as a G9V substitution severely impairs peptidyl-tRNA hydrolysis (Zeng et al., 2017).

Comparing the wild-type and mutant structures of ArfA therefore makes it possible to confidently identify the region of ArfA that mediates accommodation of RF2, after its initial recruitment, and to better understand the mechanism by which this accommodation occurs. In the wild-type structure, the ordered N-terminus of ArfA induces conformational changes in the RF2 switch loop. As also occurs in stop-codon recognition (Weixlbaumer et al., 2008; Laurberg et al., 2008; Korostelev et al., 2008; Korostelev et al., 2010), the switch loop becomes partially α-helical, extending the α7-helix (figure II.12B). This is followed by a ~100° kink as the conserved W319 is drawn into the RF2-specific hydrophobic pocket where it contacts ArfA (figure II.9C). An additional α-helical turn leads into the β9-strand. Thus, compression of the switch loop draws the α7-helix, together with the rest of domain 3, from its interface with domain 2 such that the GGQ motif rises into the peptidyltransferase centre (figure II.12C).

Upon accommodation, there is also a ~10° rotation of RF2 domains 2 and 4 (figure II.12C) which increases the interface area with the ribosome by 25%. Together with a movement of the ArfA α-helix, these changes result in a tighter fit between ArfA, RF2, and the decoding centre. The movement of domains 2 and 4 also appears to pull domain 1, together with the bL12 stalk base, into closer contact with the ribosomal small subunit (figure II.13A). All these conformational changes likely function to stabilize RF2 in its accommodated state.
Part II: The mechanism of ArfA-mediated ribosome rescue
Part II: The mechanism of ArfA-mediated ribosome rescue

Figure II.12 | Switch-loop stabilization and RF2 accommodation. (A) In the structure of the nonstop complex with ArfA(A18T), the first 14 residues of ArfA(A18T) and the switch loop between domains 3 and 4 of RF2 are disordered. (B) In the structure with wild-type ArfA, the ordered N-terminus of ArfA helps to stabilize the switch loop of RF2, which extends the α7 helix. A1913 from helix 69 of the 23S rRNA stacks with A1492 from helix 44 of the 16S rRNA, while C1914 from helix 69 stabilizes residues 10 to 14 of ArfA. (C) RF2 recruited by ArfA(A18T) adopts a compact conformation with the GGQ loop disordered and facing the P-site tRNA. Superposition with accommodated RF2 (outlined) reveals that all four domains move during accommodation. (D) TtRF2 has a switch-loop composition (top) incompatible with ArfA and adopts a preaccommodated conformation on the ribosome (bottom), similar to E. coli RF2 recruited by ArfA(A18T). Red arrowheads denote residues of TtRF2 with long side chains that would clash with ArfA. (E) During stop-codon recognition, the switch loop of TtRF2 is stabilized by interactions that are dependent on the remodelled decoding centre.

At the same time, the decoding nucleotide A1492 of the 16S rRNA switches from a syn to an anti configuration within helix 44 (figure II.8C) and stacks with A1913 at the apex of helix 69 of the 23S rRNA. This may help to reposition helix 69, drawing it forwards and allowing the backbone of ArfA (residues Q10, K12, and N14) to mimic a base-pairing interaction with C1914, which coordinates the turn adopted by residues 11 to 14 (figure II.12B). The phosphate of C1914 also forms hydrogen bonds with the side chains of K8 and H21. The former interaction is functionally important, as a K8A substitution impairs ArfA activity (Zeng et al., 2017).

To further explore the interaction between ArfA and the switch loop, and to test the model for the mechanism of accommodation, I also determined the structure of a nonstop complex with wild-type ArfA and TtRF2, which retains the conserved hydrophobic pocket but has a distinct switch-loop composition (figure II.12D). Thermus thermophilus lacks an ArfA orthologue, and it may therefore be predicted that TtRF2 would fail to make key interactions with the ArfA protein of E. coli. Indeed, it has recently been shown that TtRF2 can be recruited to the ribosome by E. coli ArfA, but cannot hydrolyse the peptidyl-tRNA (Zeng et al., 2017). In agreement with these data, the structure shows that TtRF2 is recruited to the nonstop complex, interacting with ArfA through the hydrophobic pocket, but it adopts a compact conformation similar to that of E. coli RF2 recruited by ArfA(A18T) (figure II.12D) and the N-terminus of ArfA remains disordered. The inability to accommodate presumably results from clashes between ArfA and the longer side chains of the TtRF2 switch loop. Thus, taken together, the structures demonstrate that
accommodation of RF2 is dependent on switch-loop stabilization mediated by the ArfA N-terminus.

**Figure II.13** | Release-factor domain movements. (A) ArfA-mediated accommodation of RF2 involves an inward movement of domain 1, together with the bL12 stalk base (23S rRNA helices 42 to 44 and the ribosomal proteins uL10 and uL11). This movement brings the three-stranded coiled coil of domain 1 into closer contact with the rRNA of the 30S beak, which may help to stabilize the accommodated state. (B) A superposition of preaccommodated RF2 with accommodated RF2 (outlined) shows that all four domains move during accommodation. The GGQ loop of domain 3 rises by ~60 Å to accommodate into the peptidyltransferase centre. This movement is concurrent with a rotation of domains 2 and 4 and an inward movement of domain 1. (C) The conformational changes of RF2 during ArfA-mediated accommodation are analogous to those that the unrelated eRF1 undergoes on a stop codon. The eukaryotic protein is delivered to the ribosome in a preaccommodated conformation by eRF3 (PDB accession code 5LZT) (Shao et al., 2016). GTP hydrolysis and subsequent dissociation of eRF3 provide an additional checkpoint before the GGQ motif in the M domain of eRF1 can accommodate into the peptidyltransferase centre (PDB accession code 5LZU).

These direct contacts between ArfA and the switch loop appear to emulate the interactions between RF2 and the ribosome that result from stop-codon-dependent rearrangement of the decoding centre (figure II.12E). During canonical termination, the conserved tryptophan of the switch loop stacks with flipped-out A1492 while the extension of the α7-helix interacts directly with helix 69, which adopts a different conformation when A1492 is flipped out (Weixlbaumer et al., 2008; Korostelev et al.,...
Thus, in the absence of a remodelled decoding centre, ArfA provides alternative interactions that mediate accommodation into the A site.

**II.2.4 DISCUSSION**

The delivery of RF2 in a preaccommodated state followed by a conformational switch that depends on specific recognition of the decoding centre shares many parallels with the universal elongation (Voorhees & Ramakrishnan, 2013) (figure I.1), eukaryotic termination (figures II.13B,C and II.14), and eukaryotic RQC pathways (Shao *et al.*, 2016). Aminoacyl-tRNAs, eRF1, and pelota all adopt preaccommodated conformations during initial codon sampling that prevent premature commitment to peptidyl transfer, termination, or ribosome rescue, respectively. As a tRNA mimic, the heterodimer of tmRNA and SmpB also adopts a preaccommodated conformation in which SmpB, like ArfA, probes the mRNA channel before committing to nonstop decay (Valle *et al.*, 2003) (figure II.2). A similar mechanism for bacterial class-I release factors during canonical termination is consistent with the 4,000-fold difference in dissociation rates between stop and sense codons, despite similar association rates (Hetrick *et al.*, 2009), and may explain the remarkable accuracy of termination in bacteria (~10^{-5}), which is comparable to the fidelity of aminoacyl-tRNA selection (Freistroffer *et al.*, 2000).

A key difference, however, is that the aforementioned mechanisms are all gated by a translational GTPase (EF-Tu or eEF1A, eRF3, and HBS1L, respectively), with hydrolysis of GTP irreversibly separating the initial selection step from the subsequent accommodation step (Blanchard *et al.*, 2004). This may be because ribosome-bound aminoacyl-tRNAs, eRF1, and pelota are all to some extent spring loaded, being held in a high-energy state by their respective cofactors. In translational elongation, the aminoacyl-tRNA is held in a bent conformation on the ribosome by EF-Tu or eEF1A (Schmeing *et al.*, 2009; Fischer *et al.*, 2015; Shao *et al.*, 2016) and accommodates by relaxing into the A site. Similarly, in the eukaryotic termination and RQC pathways, the M domains of eRF1 and pelota are initially held away from the A site, but dissociation of eRF3 or HBS1L allows the α8- and α9-helices to join up, forming a single, long α-helix that directs the M domain towards the peptidyltransferase centre (Becker *et al.*, 2011; Brown *et al.*, 2015b; Shao *et al.*, 2016).
Part II: The mechanism of ArfA-mediated ribosome rescue

(figure II.13C). The bacterial class-I release factors, by contrast, seem to adopt compact low-energy conformations in isolation (Vestergaard et al., 2001; Shin et al., 2004; Zoldák et al., 2007) and require specific interactions with a stop codon or ArfA to stabilize the high-energy accommodated state in which domain 3 rises up into the peptidyltransferase centre (figure II.13B). In other words, rather than being springs, the class-I release factors function as mechanical levers.

**Figure II.14** | Canonical termination in eukaryotes. As in bacteria, eRF1 is initially recruited to the ribosome in a compact, preaccommodated conformation. Stop-codon recognition induces allosteric conformational changes, leading to GTP hydrolysis, dissociation of eRF3, and accommodation of eRF1 into the A site. The M domain of eRF1 flips outwards from its eRF3-bound position, bringing the catalytic GGQ motif into the peptidyltransferase centre where it hydrolyses the peptidyl-tRNA. ABCE1 is then recruited to the ribosome and splits the ribosomal subunits. Eukaryotes therefore lack a clear demarcation between the termination and recycling phases of translation. The eukaryotic RQC pathway follows a similar mechanism, with pelota and HBS1L replacing eRF1 and eRF3, respectively.

In summary, the structures described here reveal the mechanism of ArfA-mediated ribosome rescue (figure II.15) and have wider implications for canonical termination. The discovery of a predicted but previously unseen preaccommodated state of RF2 demonstrates that termination involves a conformational switch and helps to explain the accuracy of canonical termination in bacteria.

Shortly after this work was published (James et al., 2016), four other research groups released their own papers detailing the mechanism of ArfA-mediated ribosome rescue in *E. coli* (Huter et al., 2017; Ma et al., 2017; Zeng et al., 2017;
Demo et al., 2017). All four papers describe the fully accommodated state containing wild-type ArfA and RF2. The maps are of similar quality and the interpretations are largely identical. The paper from Hong Jin’s group contains additional mutagenic data in support of the model that help to pick apart the functional significance of individual residues (Zeng et al., 2017). I have incorporated these data into the description of my structures in section II.2.3.

**Figure II.15 | ArfA-mediated ribosome rescue.** ArfA recognizes the unoccupied decoding centre of a nonstop complex. I was not able to capture a state with ArfA before RF2 binds, presumably because ArfA samples many conformations. RF2 is initially recruited in a compact, preaccommodated state. This is followed by accommodation into the A site, as in canonical termination. RF3 is not required in this pathway, but the mechanism by which ArfA and RF2 dissociate remains unclear.

Only one other paper, from Andrei Korostelev’s group, also describes the preaccommodated state of RF2 (Demo et al., 2017). Interestingly, and in contrast to my results, both states emerged with similar abundance from the same dataset, which was classified using FREALIGN (Grigorieff, 2016). The authors propose that this discrepancy may be due to differences in sample preparation or data processing. The latter explanation is unlikely, as I subjected the wild-type dataset to three rounds of 3D classification, including a round of FCwSS, yet only one dominant class emerged from each round (figure II.5). Furthermore, the protocols for sample preparation were similar, both involving stepwise addition of components and long incubations at 37
°C (see section II.3.2 in the Materials and methods). In both cases, this should have provided ample time for the nonstop complex to assemble and for RF2 to accommodate fully into the A site. The mRNA constructs were also similar, as both ended immediately after the P site (without a 3′ overhang).

The most likely explanation is that Demo et al. (2017) used a recombinant ArfA protein that retained an N-terminal hexahistidine tag. This may have destabilized the interactions between ArfA and the switch loop, biasing the equilibrium towards the preaccommodated state. In all four other studies, including my own, the affinity tags on ArfA were removed by endoproteolysis (see section II.3.1 in the Materials and methods) (Huter et al., 2017; Ma et al., 2017; Zeng et al., 2017).

Another important difference is that Demo et al. (2017) used an uncharged tRNAi_AUG. I used a nonhydrolysable fMet-NH-tRNAi_AUG, while Huter et al. (2017) used a mutant of RF2 in which the catalytic GGQ motif had been substituted with nonfunctional GAQ. The structures from Demo et al. (2017) are therefore more reminiscent of post-hydrolysis states. It may be that, after peptidyl-tRNA hydrolysis, RF2 retains some flexibility and is able to return to a preaccommodated-like state, possibly as a precursor to subsequent dissociation. Nevertheless, despite the appeal of this explanation, the structures from Ma et al. (2017) and Zeng et al. (2017) also used uncharged tRNAi_AUG, yet these groups did not observe such a state.

This relates to a puzzling property of ArfA that has yet to be explained in mechanistic detail. In canonical termination, the GTPase activity of RF3 is generally required to dislodge RF2 (or RF1) from its stabilized position in the A site (Freistroffer et al., 1997). In ArfA-mediated termination, however, RF3 has no effect on release of the nascent polypeptide when RF2 is present in substoichiometric amounts, implying that RF2 and ArfA can dissociate freely without requiring RF3 for multiple turnovers (Zeng & Jin, 2016). Indeed, when RF2 is in excess, RF3 actually inhibits the single-turnover reaction, presumably through competitive interactions. When bound to GTP and in the absence of a class-I release factor, RF3 prefers to interact with the ribosome in its rotated state (Jin et al., 2011; Zhou et al., 2012), and it has been hypothesized that ArfA locks the ribosome in a nonrotated state (Zeng & Jin, 2016). Another explanation might be that ArfA mimics RF3 insofar as it enables RF2 to return to a loosely bound preaccommodated-like state after peptidyl-tRNA
Part II: The mechanism of ArfA-mediated ribosome rescue

hydrolysis (Demo et al., 2017). RF3 would presumably induce a similar conformational change, given that RF2 seems to adopt a compact conformation in isolation (Vestergaard et al., 2001; Shin et al., 2004; Zoldák et al., 2007). Perhaps ArfA and RF3 competitively interact with the same surface on RF2. However, in the absence of high-resolution structures containing both RF2 and RF3, this hypothesis remains speculative.

In addition, a number of other questions remain that relate to the wider biology of RQC pathways in bacteria. First and foremost, the environmental conditions that overwhelm trans-translation and trigger the ArfA response have not been determined. The function of ArfB is also unclear, as it is able to rescue nonstop complexes (Chadani et al., 2011b; Handa et al., 2011) yet, in E. coli, is unable to compensate for the simultaneous loss of tmRNA and ArfA. Does ArfB primarily recognize other types of arrested ribosome, stalled in the middle of an mRNA (Shimizu, 2012)? There may also be other alternative ribosome-rescue factors, explaining how some species, such as Bacillus subtilis, can survive without any apparent nonstop-decay pathway (Shin & Price, 2007). A potential candidate is RF-H, a class-I release factor that retains the catalytic GGQ motif but lacks domain 1 (Baranov et al., 2006).

Much progress has recently been made towards understanding how nascent polypeptides are targeted for degradation in bacteria and eukaryotes. The fate of the mRNA, however, is less well understood. Apart from RelE, the RNases involved in bacterial RQC pathways are poorly characterized. RNase R (Richards et al., 2006), RNase II, and the proteobacterial degradosome (Garza-Sánchez et al., 2009) have all been shown to function in RQC but their mechanisms and interactions with the ribosome have not been elucidated, while the endoRNase that cleaves mRNA in the bacterial equivalent of no-go decay has yet to be identified (Keiler, 2015). Understanding the mechanisms of ribosome-dependent mRNA decay in bacteria, as well as eukaryotes, remains an important goal.
II.3 MATERIALS AND METHODS

II.3.1 PURIFICATION OF RIBOSOMES AND RECOMBINANT ARFA

Ribosomes were purified by Dr. Y. Gordiyenko (MRC LMB) from 50 g of *E. coli* MRE 600 cells that had been harvested in mid-log phase. Cells were sonicated in buffer R200 (20 mM tris–HCl (pH 7.5), 200 mM NH₄Cl, 10 mM Mg(OAc)₂, 6 mM 2-mercaptoethanol, 0.1 mM benzamidine, and 0.1 mM phenylmethylsulphonyl fluoride). The lysate was centrifuged twice at 30,000 ×g for 30 min to remove cell debris. The supernatant was layered onto cushions of 1.1 M sucrose in buffer R200 and ribosomes were pelleted through the cushions by ultracentrifugation in a 45Ti rotor (Beckman Coulter) at 205,000 ×g for 18 h at 4 °C. This ultracentrifugation step was repeated twice more in buffer R200 and once in buffer R500 (as R200, but containing 500 mM NH₄Cl). After each ultracentrifugation, the pellets were resuspended in buffer by gentle shaking on ice for 2 h. The final pellets were resuspended in buffer R60 (20 mM tris–HCl (pH 7.5), 60 mM NH₄Cl, 10 mM Mg(OAc)₂, 6 mM 2-mercaptoethanol).

Six sucrose gradients (15 to 30%) were prepared in buffer R60. Each gradient was loaded with ~208 *A₂₆₀* units of the resuspension and ultracentrifuged in an SW-28 rotor (Beckman Coulter) at 58,000 ×g for 18 h at 4 °C. The fractions containing 70S ribosomes were pooled, diluted to 180 ml with buffer G (5 mM HEPES–KOH (pH 7.4), 50 mM KCl, 10 mM NH₄Cl, 10 mM Mg(OAc)₂, and 6 mM 2-mercaptoethanol), and ultracentrifuged again in a 45Ti rotor (Beckman Coulter) at 205,000 ×g for 18 h at 4 °C. The pellets were washed with buffer G and resuspended in buffer G to a final concentration of ~6.8 µM. Aliquots of 20 µl were flash-frozen in liquid nitrogen and stored at −80 °C.

A pET16b vector encoding *E. coli* ArfA (residues 1 to 60) with an N-terminal decahistidine tag followed by a Factor Xa cleavage site was a gift from Dr. Y. Shimizu (Riken). ArfA was purified by Dr. A. Brown (MRC LMB) from *E. coli* BL21 (DE3) cells (Novagen). The cells were induced at an OD₆₀₀ of 0.5 with a final concentration of 1 mM IPTG and grown for another 4 h at 37 °C. The cells were then sonicated in lysis buffer (50 mM HEPES–KOH (pH 7.5), 1 M NH₄Cl, 5 mM MgCl₂,
40 mM imidazole, and 6 mM 2-mercaptoethanol) and applied to a HisTrap column (GE Healthcare). ArfA was eluted over a gradient up to 1 M imidazole and dialysed overnight against dialysis buffer (50 mM HEPES–KOH (pH 7.5), 50 mM KCl, 5 mM MgCl₂, 2 mM CaCl₂, and 6 mM 2-mercaptoethanol) containing Factor Xa (NEB). The dialysed solution was again passed through a HisTrap column to isolate untagged ArfA, which was further purified by gel filtration using a Superdex 75 16/600 column (GE Healthcare) equilibrated with buffer G. The same procedure was used to purify ArfA(A18T).

A 3′-truncated version of Z4C mRNA (Selmer et al., 2006) that ends immediately after the start codon was chemically synthesized (GE Dharmacon). E. coli fMet-NH-tRNAᵢ₄₈₅ containing a nonhydrolysable amide bond between A76 and the N-formylmethionyl residue was prepared as described (Voorhees et al., 2009). TtRF2 was purified as described (Weixlbaumer et al., 2008). Dr. R.S. Hegde (MRC LMB) provided His-tagged E. coli RF2 from the PURE system (Shimizu et al., 2001).

II.3.2 SAMPLE PREPARATION AND DATA COLLECTION

Reaction mixtures of 500 µl containing 130 nM E. coli 70S ribosomes were prepared in buffer G. The mixtures were incubated for 5 min at 37 °C after addition of each of the following: 520 nM mRNA, 1.3 µM fMet-NH-tRNAᵢ₄₈₅, 1.3 µM wild-type ArfA, and either 1.3 µM RF2 or TtRF2. Additional reactions were prepared with 3.2 µM ArfA(A18T) and 1.3 µM RF2, and with 1.3 µM wild-type ArfA but without RF2. The mixtures were held on ice before being frozen onto grids.

UltrAuFoil R 1.2/1.3 grids (for the mixture with TtRF2) and QUANTIFOIL R 2/2 grids (for the other mixtures) were coated with a continuous carbon film (~60 Å in thickness) and glow-discharged for 30 s at 6 V. 3 µl of each mixture were incubated on the grids for 30 s at 4 °C and 100% humidity. The grids were then blotted for 4 to 5 s and flash-frozen in liquid ethane using a Vitrobot Mark III (FEI).

Micrographs were collected using a Titan Krios at 300 keV with a Falcon II detector and EPU software (all FEI). The dose rate was 40 e⁻Å⁻²⁻s⁻¹, the defocus values
ranged from –2.0 to –3.0 µm, the exposures were 1.2 s with 20 movie frames, and the calibrated magnification of 134,615× gave a pixel size of 1.04 Å.

II.3.3 DATA PROCESSING

The movie frames of each image were aligned using MotionCorr (Li et al., 2013). Parameters for the contrast transfer function were obtained using Gctf (Zhang, 2016) for each motion-corrected micrograph. All subsequent processing was performed in RELION-2.0 (Fernández-Leiro & Scheres, 2016), with the exception of the TtRF2 dataset, which was processed in RELION-1.4. Semiautonomous particle picking (Scheres, 2015) was used to select ribosomes from the micrographs. Incorrectly selected nonribosomal particles were discarded after reference-free two-dimensional (2D) classification. A cryo-EM map of the 70S–RelA complex (EMD-8107) (Brown et al., 2016), low-pass-filtered to 40 Å, was used as an initial reference for the first 3D refinement. An initial round of 3D classification without further aligning the particles was used to discard free 50S subunits, empty 70S ribosomes, and poorly aligned particles (figure II.5). After another refinement, motion correction was performed for each particle using a running average of three movie frames and a standard deviation of one pixel for the translational alignment. Each frame was weighted with a B-factor to compensate for radiation damage.

After this “particle-polishing” step, FCwSS (Bai et al., 2015) was used to isolate particles containing ArfA and RF2 (or an empty A site for the dataset without RF2). An additional 3D classification without alignments and two rounds of FCwSS were performed on the dataset with TtRF2 to improve the quality of the map. For the other three datasets, a final 3D classification with local angular searching was performed as a final clean-up of the data. Despite three rounds of classification, no ArfA was visible in the dataset without RF2. This dataset was used to model a nonstop complex with an unoccupied A site.

The final statistics for each dataset are provided in table II.2. High-resolution noise substitution was used to correct for the effects of applying a mask during the Fourier-shell-correlation (FSC) calculations (Chen et al., 2013). Reported resolutions are based on the FSC=0.143 criterion (Rosenthal & Henderson, 2003) (figure II.6A).
Part II: The mechanism of ArfA-mediated ribosome rescue

Before visualization, density maps were corrected for the modulation transfer function of the Falcon II detector and sharpened by applying a negative B-factor that was estimated using automated procedures (Rosenthal & Henderson, 2003). Local resolution was estimated using ResMap (Kucukelbir et al., 2014) (figure II.6B,C).

II.3.4 MODEL BUILDING

Models of the *E. coli* 70S ribosome and fMet-tRNA$_{\text{AUG}}$ (Brown et al., 2016) were docked into the maps using Chimera (Pettersen et al., 2004). Each chain was subsequently rigid-body-fitted in Coot (Emsley et al., 2010) and all chains were manually adjusted to fit the density. The crystal structures of *E. coli* RF2 (PDB accession code 1GQE) (Vestergaard et al., 2001) and TtRF2 (PDB accession code 2IHR) (Zoldák et al., 2007) were rigid-body-fitted into the A site and partially rebuilt to fit the densities. ArfA was built *de novo*.

II.3.5 MODEL REFINEMENT AND VALIDATION

Reciprocal-space refinement was performed in REFMAC v5.8 optimized for cryo-EM maps using external restraints generated by ProSMART and LIBG (Brown et al., 2015a). The fit of the model to the map density was quantified using FSC$_{\text{average}}$ (Brown et al., 2015a) and CRef (Rosenthal & Henderson, 2003). CRef is a measure of the resolution when the FSC between the refined model and map is 0.5, and should agree closely with resolution of the map estimated from the FSC=0.143 criterion (Rosenthal & Henderson, 2003). Model statistics were obtained using MolProbity (Chen et al., 2010) and provided in table II.2. Cross-validation against overfitting was performed as described (Brown et al., 2015a) (figure II.6A).

Figures were prepared using PyMOL (DeLano, 2002), Chimera (Pettersen et al., 2004), and Adobe Illustrator.
II.4 BIBLIOGRAPHY


Part II: The mechanism of ArfA-mediated ribosome rescue


Christensen S.K., Gerdes K. (2003) RelE toxins from Bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA. Mol. Microbiol. 48: 12389–1400


Part II: The mechanism of ArfA-mediated ribosome rescue


Part II: The mechanism of ArfA-mediated ribosome rescue


Part II: The mechanism of ArfA-mediated ribosome rescue


Part II: The mechanism of ArfA-mediated ribosome rescue


Sunohara T., Jojima K., Yamamoto Y., Inada T., Aiba H. (2004b) Nascent-peptide-mediated ribosome stalling at a stop codon induces mRNA cleavage resulting in nonstop mRNA that is recognized by tmRNA. *RNA* 10: 378–386


Part II: The mechanism of ArfA-mediated ribosome rescue


III.1 INTRODUCTION

III.1.1 INTERNAL RIBOSOME ENTRY

A number of eukaryotic viruses utilize an alternative mechanism of translational initiation termed internal ribosome entry (Kieft, 2008; Filbin & Kieft, 2009; Balvay et al., 2009; Johnson et al., 2017; Lee et al., 2017; Yamamoto et al., 2017). This mechanism is directed by cis-acting IRESs, highly structured motifs within the mRNA that interact with the ribosome and a subset of initiation factors. Many IRESs also require IRES trans-acting factors (ITAFs) that are specific to certain cell types or encoded within the viral genome. The key definition of an IRES, as opposed to any other cap-independent motif, is that it directs end-independent initiation (Terenin et al., 2017). In other words, IRESs recruit the ribosome directly to internal positions within an mRNA, without needing an accessible 5' end (table I.1), and at least in principle can direct translational initiation from a circular mRNA (Chen & Sarnow, 1995; Chen & Sarnow, 1998). As such, IRESs primarily affect the ribosome-recruitment and mRNA-loading steps of the initiation phase, but there is mounting evidence that they also mediate functionally important conformational changes throughout initiation and even into the initial round of the elongation cycle.

IRESs have diverse structures and mechanisms (figure III.1), having adapted to the particular requirements of each virus, but they all ensure the efficient translation of viral genes even when the translation of cellular genes is downregulated. Indeed, some viruses actively disrupt cellular translation yet are able to translate their own genes thanks to the peculiar properties of their IRESs.

Many putative IRESs have been identified in diverse species, including viruses and cells, but only a few have convincingly been shown to direct initiation in an end-independent manner. The existence of cellular IRESs in particular is disputed, as many positive results have been shown to result from artefacts (Kozak, 2003; Kozak, 2005; Gilbert, 2010; Jackson, 2013; Terenin et al., 2017). A few candidates have been shown to bind initiation factors or the ribosome with at least some specificity, resulting in reduced cap dependence, but they nevertheless require an accessible 5' end and presumably function as CITEs or other types of cis-acting regulatory motif (Shatsky et al., 2010). In contrast, the best-characterized IRESs are all found in RNA
viruses and are classified into four types (Jackson et al., 2010), although a growing number of newly discovered IREs have yet to be classified (table III.1).

**Figure III.1** | Secondary-structure representations of several IREs, not drawn to scale. Bacterial tmRNA is shown for comparison. The start and stop codons are labelled. In many viruses, including EMCV, CSFV, and HalV, not every domain in the 5' UTR is required for minimal IRES activity, but the additional domains may enhance translational initiation or otherwise function as translational regulators. Most IREs manipulate initiation complexes, but type-IV IREs (such as the CrPV IGR IRES) and bacterial tmRNA bypass the initiation phase by mimicking canonical elongation complexes when bound to the ribosome.
The largest IRESs, belonging to types I and II, are found in picornaviruses. These related IRESs were the first to be discovered, concurrently in poliovirus (Pelletier & Sonenberg, 1988) and encephalomyocarditis virus (EMCV) (Jang et al., 1988), respectively, and typically require all the standard initiation factors except the cap-binding protein eIF4E (Jackson et al., 2010). They also tend to associate only weakly with the 40S subunit, requiring proteins such as eIF4G to stabilize the initiation complex. The key mechanistic difference is that type-I IRESs recruit the 40S subunit to an upstream position from which it scans (Kuge et al., 1989) or shunts (Hellen et al., 1994) to the start codon, whereas type-II IRESs place the start codon directly into the P site without scanning (Kaminski et al., 1990; Kaminski et al., 1994). They are also found in different genera, with type I being characteristic of enteroviruses and type II of aphthoviruses and cardioviruses. Related IRESs with distinct mechanisms have also been identified in other picornaviruses (Glass et al., 1993; Yu et al., 2011).

The IRESs of type III are smaller than most others and comprise only two essential domains, although a number of peripheral domains may also influence translational initiation. These IRESs are found in pestiviruses such as classical-swine-fever virus (CSFV) (Rijnbrand et al., 1997) and hepaciviruses such as HCV (Tsukiyama-Kohara et al., 1992; Wang et al., 1993), both of which belong to the Flaviviridae. The HCV IRES is the best characterized and is described in detail in section III.1.2. Several type-III IRESs have also been identified in picornaviruses (Kaku et al., 2002; Pisarev et al., 2004; Chard et al., 2006; Bakhshesh et al., 2008; Liu et al., 2011; Hellen & de Breyne, 2007; Asnani et al., 2015), presumably as a result of recombination with an ancestral member of the Flaviviridae. This demonstrates that, much like other genetic elements, IRESs are discrete modules that can be exchanged between species and provide selective advantages to their new carriers.

The smallest IRESs belong to type IV and are generally found in the intercistronic or intergenic region (IGR) of dicistroviruses such as cricket paralysis virus (CrPV). The IGR contains three pseudoknots (Kanamori & Nakashima, 2001) that interact directly with the ribosomal subunits and are able to initiate translation without using any initiation factors, Met-tRNA<sub>i AUG</sub>, or even an AUG codon (Deniz et al., 2009). Instead, the first pseudoknot (PKI) mimics the anticodon stem-loop of a tRNA. It was originally thought that PKI is positioned in the P site, like Met-tRNA<sub>i AUG</sub> (Wilson et al., 2000a; Costantino et al., 2008), but recent high-resolution structures have shown...
that it initially binds in the A site and requires eEF2-mediated translocation before the first aminoacyl-tRNA can be recruited (Fernández et al., 2014; Koh et al., 2014). The IRES-bound ribosome thus resembles an elongation complex, allowing the IRES to bypass the initiation phase altogether. Subsequent structures have shown that the IRES moves like an inchworm as it translocates through the ribosome, with the gap between PKI and the other pseudoknots contracting before expanding again (Murray et al., 2016; Abeyrathne et al., 2016; Johnson et al., 2017; Yamamoto et al., 2017).

<table>
<thead>
<tr>
<th>Type</th>
<th>Families</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Picornaviridae</td>
<td>Enterovirus</td>
</tr>
<tr>
<td>II</td>
<td>Picornaviridae</td>
<td>Aphthovirus, Cardiovirus</td>
</tr>
<tr>
<td>III</td>
<td>Flaviviridae</td>
<td>Hepacivirus, Pestivirus</td>
</tr>
<tr>
<td></td>
<td>Picornaviridae</td>
<td>Aalivirus, Anativirus, Aquamavirus, Avihepatovirus, Colbovirus, Kobuvirus, Kunsagivirus, Limnipivirus, Megrivirus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Parescovirus, Pasivirus, Phacovirus, Rafivirus, Sakobuvirus, Sapelovirus, Senecavirus, Teschovirus, Tremovirus</td>
</tr>
<tr>
<td>IV</td>
<td>Dicistroviridae</td>
<td>Aparavirus (IGR), Cripavirus (IGR), Triatovirus (IGR)</td>
</tr>
<tr>
<td></td>
<td>Unassigned</td>
<td>Halastavi-árvá virus (IGR)</td>
</tr>
<tr>
<td></td>
<td>Dicistroviridae</td>
<td>Aparavirus (leader), Cripavirus (leader), Triatovirus (leader)</td>
</tr>
<tr>
<td></td>
<td>Hepeviridae</td>
<td>Orthohepevirus</td>
</tr>
<tr>
<td></td>
<td>Picornaviridae</td>
<td>Hepatovirus, Kobuvirus</td>
</tr>
<tr>
<td></td>
<td>Unassigned</td>
<td>Halastavi-árvá virus (leader)</td>
</tr>
</tbody>
</table>

Table III.1 | The current classification system for IRESs found in animal viruses with sense, single-stranded RNA genomes, used by most researchers. As is apparent from the number of unassigned IRESs, this system fails to account for the diversity of IRESs and the potential homology between different types, but it is not possible to improve upon it significantly until many of these IRESs are better characterized.

The Picornaviridae and Dicistroviridae, which infect vertebrates and arthropods, respectively, are thought to be related and are classified together in the order Picornavirales. As such, dicistoviruses (Wilson et al., 2000b) and the recently
discovered halastavi-årva virus (HalV) (Abaeva et al., 2016) also contain IRESs in their 5′ leaders that superficially resemble those of picornaviruses and may have diverged from an ancient ancestor that existed during the early evolution of animals. There are important mechanistic differences, however, as these leader IRESs share no sequence identity with picornaviral IRESs and often have distinct initiation-factor requirements. Translational initiation from the HalV leader IRES, for example, involves reverse scanning from the initial recruitment site and does not require the eIF4F complex (Abaeva et al., 2016).

A number of putative IRESs have also been identified in other groups of eukaryotic viruses such as the Totiviridae, which includes leishmaniavirus (Maga et al., 1995), totivirus (Huang & Ghabrial, 1996), and giardiavirus (Garlapati & Wang, 2004). The latter is the best characterized, with the entire 5′ UTR of the giardiaviral sense strand and a large upstream portion of the coding region being necessary for efficient translation in transfected cells (Yu & Wang, 1996). These regions each contain a pseudoknot and multiple stem-loops, disruption of which results in a total loss of translational activity (Garlapati et al., 2001; Garlapati & Wang, 2002; Garlapati & Wang, 2005). The host of giardiavirus, *Giardia lamblia*, is a unicellular metamonad that has a number of highly divergent features such as 70S-like ribosomes (Sogin et al., 1989), fewer initiation factors (Rezende et al., 2014), and exclusively leaderless or near-leaderless mRNAs (Li & Wang, 2004). The ribosomes of *G. lamblia* therefore lack the ability to scan through the 5′ UTR of the giardiaviral sense strand, implying that giardiavirus must contain a true IRES quite unlike those of other eukaryotic viruses (Garlapati & Wang, 2009; Garlapati et al., 2011).

It is also worth noting that, despite their comparative simplicity, the Shine–Dalgarno sequences of bacteria, archaea, and organelles are functionally equivalent to eukaryotic IRESs. By base-pairing with the 16S rRNA, a Shine–Dalgarno sequence is able to direct translational initiation on a circular mRNA (Perriman & Ares Jr., 1998; Abe et al., 2013) or downstream of an obstruction (Wilson & Beckmann, 2011). This implies that bacterial, archaeal, and organellar ribosomes are more versatile than their eukaryotic counterparts, which cannot direct internal initiation unless they are manipulated by a large, highly structured motif within the mRNA. However, bacteria and some organelles do contain a highly structured IRES-like molecule in the form of tmRNA (described in Part II, section II.1.3.1). Although it mediates the continuation
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

in trans of the elongation phase, rather than starting a new initiation phase, tmRNA nevertheless directly recruits the ribosome to an internal position in an end-independent manner and directs translation of a new cistron (Keiler, 2015). Indeed, tmRNA resembles the viral type-IV IRESs insofar as it bypasses the initiation phase, mimics an elongator tRNA, initially binds in the A site, and is loaded into the mRNA channel in a stepwise manner by translocation through the ribosome (Weis et al., 2010; Fu et al., 2010; Ramrath et al., 2012). It even requires an ITAF-like protein in the form of SmpB. It may be that structural studies of tmRNA will provide insights into the general properties of IRES-directed translational initiation.

III.1.2 The HCV IRES

Hepatitis C is an infectious disease that primarily affects the liver (Gravitz, 2011). The disease is often asymptomatic, but chronic infection can lead to severe complications such as cirrhosis and hepatocellular carcinoma (di Bisceglie, 2000; Hoofnagle, 2002). The causative agent, HCV, is the type species of the genus Hepacivirus in the Flaviviridae. The virus is primarily blood borne (Alter, 2011) and is classified into seven distinct genotypes with only ~70% sequence identity (Nakano et al., 2012). The genotypes are further divided into hundreds of subtypes.

The genomic single-stranded RNA of HCV is ~9,600 nt in length and encodes a long precursor polyprotein that is endoproteolytically cleaved into ten structural and catalytic proteins (Kato, 2000), as well as a short protein in the +1 frame (Vassilaki & Mavromara, 2009). The most highly conserved parts of the HCV genome are the 5′ and 3′ regions on either side of the long cistron, which contain a number of structural motifs that are essential for the viral infection cycle (Romero-López & Berzal-Herranz, 2014).

HCV primarily infects hepatocytes, as well as peripheral blood mononuclear cells such as lymphocytes and monocytes (Wang et al., 1992). Upon entering a host cell, the genomic RNA directs translational initiation from a type-III IRES in its 5′ UTR. This IRES recruits a subset of initiation factors and is regulated by cis-acting motifs within the HCV genome (Romero-López & Berzal-Herranz, 2014; Fricke et al., 2015) (figure III.2). Later in the infection cycle, translational initiation is downregulated in
favour of genome replication from the 3’ end. At high copy numbers the HCV RNA begins to dimerize, before being packaged into capsids and exocytosed from the host cell (Cristofari et al., 2004).

III.1.2.1 Structure of the HCV IRES

When first sequenced, the 5’ UTR of HCV was found to be highly conserved between isolates (Han et al., 1991), more so than the main coding region of the genome. The 5’ UTR was also found to bear homology to the 5’ UTRs of pestiviruses, and yet was distinct from those of other flaviviruses. The putative IRES was first identified in 1992 using a bicistronic mRNA in mammalian cell lysate (Tsukiyama-Kohara et al., 1992) and later confirmed by mRNA transfection (Wang et al., 1993). This was the first IRES to be discovered outside the Picornaviridae.

Figure III.2 | Secondary-structure representation of the HCV genome (sense strand), showing experimentally verified long-range base-pairing interactions as dashed lines (Romero-López & Berzal-Herranz, 2014). Interactions that repress translational initiation are shown in red. Interactions that enhance or are otherwise necessary for IRES activity are shown in green. At least twelve additional long-range base-pairing interactions have been predicted (Fricke et al., 2015), and other long-range interactions may be mediated by protein contacts. Regions corresponding to the IRES, cis-acting replication element (CRE), variable region, and X tail are labelled. Domains VII and VIII lie towards the 5’ end of the genome, in the region that encodes the core protein.

The structural organization of the HCV IRES was originally determined by a combination of phylogenetic comparisons with related viruses (Honda et al., 1999; Rijnbrand et al., 2000), nucleolytic and chemical probing (Brown et al., 1992; Wang...
et al., 1995), and mutagenesis (Wang et al., 1994; Zhao & Wimmer, 2001). The 5’ UTR of HCV contains four domains (Brown et al., 1992), including the essential pseudoknot IIIf (Wang et al., 1995), with only domains II and III being required for IRES activity (Pestova et al., 1998; Honda et al., 1999) (figure III.3).

![Figure III.3](image_url)

**Figure III.3** | Secondary-structure representation of the HCV IRES (domains II to IV), in context with domains I, V, and VI, showing the proposed conformational change between open and closed states. The long-range interaction between domain VI and the 5’ UTR is disrupted by miR-122, favouring the open state. The apical loop of subdomain IIId base-pairs with the 18S rRNA. The apical loop of subdomain IIb interacts with domain IV and the Met-tRNA\(_{\text{AUG}}\) and contains a conserved GCC triplet that affects the first translocation step of translational elongation.

The structural organization of the HCV IRES was originally determined by a combination of phylogenetic comparisons with related viruses (Honda et al., 1999; Rijnbrand et al., 2000), nucleolytic and chemical probing (Brown et al., 1992; Wang et al., 1995), and mutagenesis (Wang et al., 1994; Zhao & Wimmer, 2001). The 5’ UTR of HCV contains four domains (Brown et al., 1992), including the essential pseudoknot IIIf (Wang et al., 1995), with only domains II and III being required for IRES activity (Pestova et al., 1998; Honda et al., 1999) (figure III.3).

In the intervening years, the HCV IRES has become a model for the study of internal ribosome entry and has been extensively characterized (Fraser & Doudna, 2007; Lukavsky, 2009; Johnson et al., 2017; Yamamoto et al., 2017). Structures of
subdomains in isolation have been determined by X-ray crystallography (Kieft et al., 2002; Dibrov et al., 2007; Zhao et al., 2008; Berry et al., 2011) and nuclear-magnetic-resonance spectroscopy (Lukavsky et al., 2000; Collier et al., 2002; Lukavsky et al., 2003), and the whole IRES (domains II to IV) has been characterized by small-angle X-ray scattering (Pérard et al., 2013). There are also low-resolution cryo-EM reconstructions of the HCV IRES bound to the ribosome (Spahn et al., 2001; Boehringer et al., 2005) or eIF3 (Siridechadilok et al., 2005; Sun et al., 2013), as well as subnanometre-resolution structures of the HCV IRES in complex with the 80S ribosome and eIF5B (Yamamoto et al., 2014) and a partial CSFV IRES as part of a 48S-like (or 48S*) initiation complex (Hashem et al., 2013a). Most recently, high-resolution structures have been published, showing the HCV IRES bound to the ribosome in off-pathway states (Quade et al., 2015; Yamamoto et al., 2015).

III.1.2.2 Functions of IRES domains

Although the domains of the HCV IRES have been subjected to extensive mutagenic and biochemical studies, a number of important questions remain that relate to their functions in translational initiation. For example, despite its conservation, domain I appears to have no function at all, while domain IV, which contains the start codon within its apical loop, is unwound upon loading of the mRNA (Reynolds et al., 1996), is not found in some other type-III IRESs (Fletcher & Jackson, 2002; Chard et al., 2006; Asnani et al., 2015), and seems to function only to slow the rate of initiation (Honda et al., 1996) (table III.2). Nevertheless, the first 20 nt or so of the coding region (as opposed to an artificial sequence) are necessary for efficient translational initiation (Reynolds et al., 1995; Hwang et al., 1998; Fletcher et al., 2002), but whether the sequence itself is important or simply the absence of a very stable secondary structure is debatable (Rijnbrand et al., 2001).

The best-characterized domain is the large domain III, which primarily functions as a binding platform for the initiation complex. Domain III binds in an extended conformation along the trailing edge of the 40S subunit (Spahn et al., 2001; Boehringer et al., 2005). These interactions are mediated by subdomains IIIa, c, d, e, and f (Kieft et al., 2001), which bind the 40S subunit with nanomolar affinity (Kieft et al., 2001; Otto & Puglisi, 2004) and position the start codon directly into the P site.
without scanning (Berry et al., 2010; Berry et al., 2011). In addition, subdomain IIIb and the IIIabc junction bind eIF3 (Kieft et al., 2001; Hashem et al., 2013a).

Domain II, on the other hand, is conformationally dynamic (Fuchs et al., 2015) and only weakly associates with components of the initiation complex (Kieft et al., 2001). This domain seems to have a variety of functions throughout the initiation phase that depend on its mobility and curvature. These functions are difficult to distinguish, however, as disrupting the conformation of domain II, or deleting it entirely, stalls the initiation complex in a very early state before loading of the mRNA and displacement of eIF3j from the channel (Fraser et al., 2009) (table III.2). The wild-type IRES has been shown to manipulate the 40S subunit, tilting the head backwards to partially open the channel (Spahn et al., 2001). Domain II is therefore necessary for a series of conformational changes including GTP hydrolysis by eIF2γ, dissociation of eIF2 and eIF5, and subunit joining (Locker et al., 2007), but whether it plays an active role in these processes or is simply a bystander after loading the mRNA remains unclear.

Domain II is therefore essential for translational initiation. However, when the apical loop of subdomain IIb is disrupted, the HCV IRES is able to progress through initiation until it stalls during the elongation phase, after peptidyl transfer but before the first translocation step (Filbin et al., 2013) (table III.2). This implies that domain II has a specific function in the context of the 80S ribosome, but the observed phenotype could result from gain-of-function interactions that inhibit translocation. The same mutations also affect mRNA loading (Filbin et al., 2011), and a low-resolution cryo-EM reconstruction of one such mutant bound to the 40S subunit shows that the whole IRES, including domain III, adopts a different conformation that is rotated and extended relative to its usual position (Filbin et al., 2013). It is not clear how these data relate to the mechanism of the HCV IRES.

Four conserved stem-loops, domains V to VIII, lie downstream of the HCV IRES in the region that encodes the core protein (Ina et al., 1994; Smith & Simmonds, 1997; Tuplin et al., 2002; Tuplin et al., 2004) (figure III.2). The first two of these domains are important for viral infectivity and translational activity (McMullan et al., 2007; Vassilaki et al., 2008), and have been implicated in translation of the +1 frame (Choi et al., 2003). Domain VI has been shown to repress translational initiation by partially unwinding and base-pairing with the upstream region between domains I and II.
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

(Wang et al., 2000; Kim et al., 2003; Beguiristain et al., 2005), resulting in a closed conformation (figure III.3). This repression is not affected by the synthesized core protein but can be relieved by mutations that disrupt the long-range interaction.

The microRNA miR-122, expressed only in hepatocytes, is essential for infectivity and complementary to the region between domains I and II (Jopling et al., 2005). The base-pairing interactions with miR-122 (in trans) and domain VI (in cis) are mutually exclusive. Binding of miR-122 therefore drives a conformational switch between the closed and open states in vitro (Díaz-Toledano et al., 2009) (figure III.3). It is hypothesized that the closed state represses translational initiation by inhibiting recruitment of the 40S subunit to the IRES, thereby explaining how miR-122 is able to enhance translational initiation (Henke et al., 2008). Despite the attractiveness of this model, the relevance of the conformational switch to translational initiation in vivo remains uncertain. For example, the closed state has also been shown to require a high, nonphysiological concentration of magnesium (4 to 6 mM) (García-Sacristán et al., 2015). Furthermore, the aminoglycoside Geneticin stabilizes the open state and inhibits translation from the IRES (Ariza-Mateos et al., 2016), although the open state may be a side effect rather than a cause of the inhibition. In another study, translational enhancement by miR-122 was shown to depend on the 3′ UTR and Argonaute proteins rather than a conformational switch (Roberts et al., 2011).

III.1.2.3 Recruitment of the Met-tRNA\textsubscript{\textsc{AUG}}

Despite being extensively studied, the mechanism by which the HCV IRES directs translational initiation is poorly understood. Even the initiation-factor requirements of the HCV IRES are uncertain. It was originally shown, using purified initiation factors and ribosomes, that the IRES requires only Met-tRNA\textsubscript{\textsc{AUG}}, eIF2, and eIF3 to assemble a 48S* initiation complex (Pestova et al., 1998; Hellen & Pestova, 1999), though eIF5 and eIF5B were also necessary for maturation into the 80S complex. However, recent data have greatly complicated this model.

The first complication relates to eIF2. Not long after its initial characterization in vitro, data started to emerge showing that the HCV IRES can direct translational initiation in cells even when eIF2α is inhibited by phosphorylation (Koev et al., 2002; Robert et al., 2006). It was then shown that, remarkably, the HCV IRES and the related CSFV IRES can use eIF5B instead of eIF2 to recruit the Met-tRNA\textsubscript{\textsc{AUG}}
(Terenin et al., 2008; Pestova et al., 2008). This bacterial-like mechanism was the first demonstration that eIF5B retains its ancestral function of Met-tRNA_{i}^{AUG} recruitment. A similar mechanism was subsequently observed on leaderless mRNA (Akulich et al., 2016), implying that eIF2 is dispensable when the start codon is loaded directly into the P site without scanning. Further studies revealed a function for the alternative initiation factor eIF2D, which delivers the Met-tRNA_{i}^{AUG} to the initiation complex in a GTP-independent fashion (Dmitriev et al., 2010). A similar function was also found for eIF2A, which contacts the apical loop of subdomain IIId, mediates initiation when eIF2α is phosphorylated, and is even relocalized from the nucleus to the cytosol upon HCV infection (Kim et al., 2011). Nevertheless, the relevance of these mechanisms to translational initiation in vivo is questionable, as the hepaciviral proteins NS5A (Gale Jr. et al., 1997) and E2 (Taylor et al., 1999), and even the IRES itself (Vyas et al., 2003), have been shown to inhibit protein kinase R (PKR), the kinase responsible for phosphorylating eIF2α during viral infection. However, other studies have found that HCV activates PKR upon infection (Kang et al., 2009; Garaigorta & Chisari, 2009).

<table>
<thead>
<tr>
<th>Region</th>
<th>Mutant phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subdomain IIId</td>
<td>Prevents binding to the 40S subunit</td>
</tr>
<tr>
<td>Subdomain IIIb</td>
<td>Prevents binding of eIF3, eIF2, and Met-tRNA_{i}^{AUG}</td>
</tr>
<tr>
<td>Domain II</td>
<td>Prevents mRNA loading and displacement of eIF3j</td>
</tr>
<tr>
<td>Apical loop of subdomain IIb</td>
<td>Prevents translocation</td>
</tr>
<tr>
<td>Start codon</td>
<td>Prevents subunit joining</td>
</tr>
<tr>
<td>Domain IV</td>
<td>Stabilization represses initiation; disruption enhances initiation</td>
</tr>
</tbody>
</table>

**Table III.2** | Phenotypes of HCV-IRES mutants. Generally, disrupting the integrity of a domain has the same effect as deleting it entirely. The precise step affected by the mutation is often unclear, as only downstream consequences can be detected.

Other initiation factors also seem to affect the stability and conformation of the 48S* initiation complex when included in toeprinting assays. For example, eIF1 has been shown to destabilize binding of the Met-tRNA_{i}^{AUG} to type-III IRESs, presumably by keeping the mRNA channel in a fairly open conformation (Pestova et al., 2008; de Breyne et al., 2008). On the other hand, eIF1A counteracts the effect of eIF1 and
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

stabilizes the Met-tRNA$_i^{AUG}$ when eIF2 is absent. Interestingly, deleting domain II of the CSFV IRES eliminates the sensitivity to eIF1-mediated destabilization, but the mutant complexes cannot progress to subunit joining. These effects probably result from a failure to load the mRNA.

A recent study sought to resolve these discrepancies and determine how the Met-tRNA$_i^{AUG}$ is recruited to the HCV IRES (Jaafar et al., 2016). In contrast to previous studies, knocking down eIF2A and eIF2D had no effect on translation from the IRES, even in stressed cells (in which eIF2\(\alpha\) is phosphorylated). Instead, translation in the absence of eIF2 was shown to depend on eIF1A, which stabilizes the binding of Met-tRNA$_i^{AUG}$ to the initiation complex (Schreier et al., 1977; Trachsel et al., 1977) and enhances subunit joining via interactions with eIF5B (Goumans et al., 1980; Acker et al., 2006). Initiation complexes assembled on the IRES and purified from human cell lysate were found to contain eIF1A (Jaafar et al., 2016). Further experiments confirmed that the function of eIF1A in this context depends on its ability to interact with mRNA in the channel. When the start codon was mutuated to CUG, however, eIF1A actually inhibited translation, with the inhibition being most severe when eIF2 was absent. Thus, although not essential for translational initiation on the HCV IRES, eIF1A is a potent start-codon-dependent enhancer that helps the IRES to pick up free Met-tRNA$_i^{AUG}$ without a chaperone.

Another important consideration is that the HCV IRES is unlikely to encounter naked 40S subunits in vivo (Asano et al., 2001). Instead, it will recruit a 43S preinitiation complex and remodel it to suit the particular requirements of internal ribosome entry. Thus, when eIF2 is present, the IRES must displace it from the initiation complex while retaining the Met-tRNA$_i^{AUG}$. In agreement with this prediction, mixing the IRES with 43S complexes results in loss of eIF2 and increased retention of eIF1A and eIF5B (Jaafar et al., 2016). This remodelling depends on the integrity of domain II, as expected. The mechanism of remodelling remains to be determined, but the first step presumably involves resolution of the steric clash between subdomain IIb and eIF2\(\alpha\) in the E site of the 40S subunit (Hussain et al., 2014; Llácer et al., 2015). I hypothesize that eIF2\(\alpha\) rotates around the shoulder of the Met-tRNA$_i^{AUG}$, adopting a conformation similar to that seen in the crystal structure of isolated eIF2 (with Met-tRNA$_i^{AUG}$) (Schmitt et al., 2012). This movement may occur concurrently with mRNA loading. As an interesting consequence, the HCV IRES
might be able to recruit eIF2 even when eIF2α is phosphorylated, though this has yet to be tested directly.

Having displaced eIF2α, the apex of subdomain IIb would mimic the anticodon stem-loop of a tRNA in the E site (Quade et al., 2015; Yamamoto et al., 2015). If the E site were empty, the Met-tRNA_{i}^{AUG} would be free to move laterally when the head of the 40S subunit swivels around its long axis. However, the presence of a tRNA mimic in the E site (whether eIF2α or subdomain IIb) would sterically block this movement and force the Met-tRNA_{i}^{AUG} to squeeze eIF1 as the mRNA channel closes, leading to its dissociation and replacement by the N-terminal domain of eIF5 (Llácer et al., unpublished).

### III.1.2.4 Interactions with eIF3

Another complication relates to eIF3. Biochemical studies have shown that eIF3 (particularly subunits a, b, and c) has specific affinity for subdomains IIIb and IIIC, including junction IIIabc, of the HCV IRES (Buratti et al., 1998; Sizova et al., 1998; Kieft et al., 2001; Pérard et al., 2009; Sun et al., 2013). This interaction is crucial for translational initiation, as disrupting subdomain IIIb leads to dissociation of both eIF3 and eIF2 from the initiation complex (Ji et al., 2004) (table III.2). Similarly, omitting eIF3 from an in-vitro-reconstituted translation system results in loss of the Met-tRNA_{i}^{AUG} (Pestova et al., 1998). It can therefore be inferred that eIF3 and, by extension, domain III of the HCV IRES are required to stabilize the binding of eIF2 and the Met-tRNA_{i}^{AUG}.

However, as discussed in section III.1.2.3, initiation from the HCV IRES can proceed without eIF2 (Jafar et al., 2016), calling into question the necessity for eIF3. Furthermore, the recent structure of a partial CSFV IRES in a 48S* complex revealed eIF3 to be bound solely to domain III, with no direct contacts between eIF3 and the 40S subunit (Hashem et al., 2013a). This surprising observation led to the hypothesis that type-III IRESs sequester eIF3, preventing its use by the host cell, while the IRESs themselves have no use for it. This explanation remains speculative, as there are no data supporting the prediction that host translational activity is downregulated during infection by these viruses and it fails to explain how disrupting subdomain IIIb affects the binding of eIF2 and the Met-tRNA_{i}^{AUG}.
Regardless of its function, eIF3 must be remodelled by the HCV IRES. Domain III of the IRES (Quade et al., 2015; Yamamoto et al., 2015) would clash with the PCI domains of eIF3a and eIF3c (Hashem et al., 2013b; Hussain et al., 2014; Llácer et al., 2015; des Georges et al., 2015) at the trailing edge of the 40S subunit. These subunits must be displaced from their positions in a canonical initiation complex and have both been shown to bind the IRES in the absence of the 40S subunit (Sun et al., 2013). Mutagenesis of eIF3a has a similar effect to mutating subdomain IIId of the IRES (Ji et al., 2004) or omitting eIF3 altogether (Pestova et al., 1998). Mutating eIF3c, however, has a milder effect and seems to inhibit a later step in the initiation phase, recruitment of eIF5B. There is no obvious way to reconcile these mutational data with known structures.

Another subunit, eIF3j, occupies the channel between the head and body of the 40S subunit and must be displaced for mRNA loading (Benne & Hershey, 1978; Unbehaun et al., 2004; Fraser et al., 2007). This displacement is mediated by domain II of the IRES (Fraser et al., 2009; Filbin & Kieft, 2011), but whether it results from direct or allosteric interactions remains to be determined. It might, however, involve synergy with domain III, which potentially competes with eIF3j for binding to eIF3b (ElAntak et al., 2007; Pérard et al., 2009). Domain III binds eIF3b with a $K_d$ of 2 µM, while eIF3j binds with a $K_d$ of 20 µM. The binding sites also overlap, and it is therefore hypothesized that domain III displaces eIF3j from this contact.

Other conformational changes in eIF3 must also occur, apparently leading to its complete detachment from the 40S subunit, as observed with the partial CSFV IRES (Hashem et al., 2013a). However, this structure contradicts previous low-resolution cryo-EM data of the HCV IRES bound to eIF3 without the 40S subunit (Sun et al., 2013). Furthermore, the partial CSFV IRES lacked domain II and should therefore have captured an earlier state before mRNA loading and eIF3j displacement. Whether this structure represents an artefact, a hitherto-unpredicted on-pathway state, or a conformation specific to CSFV remains to be determined.
III.2 Results

III.2.1 Scientific aims

The HCV IRES is an important model for studying internal ribosome entry (Fraser & Doudna, 2007; Lukavsky, 2009; Johnson et al., 2017; Yamamoto et al., 2017), is a potential therapeutic target for antiviral drugs (Davis & Seth, 2011; Hoffman & Liu, 2011), and provides a means of stabilizing initiation factors in complex with the ribosome (Hashem et al., 2013a; Yamamoto et al., 2014). The aim of this project was to capture the HCV IRES in functional, on-pathway complexes with the mammalian ribosome and other factors. This would enable me to determine the structure of the IRES, describe its interactions with the ribosome, and characterize the mechanism by which it progresses through the initiation phase of translation.

In order to achieve these goals, I have developed a reproducible method for purifying natively assembled initiation complexes, bound to the HCV IRES, from mammalian cell lysate. Native purification has the potential to be faster than in-vitro reconstitution, adaptable to a variety of different complexes, and more realistic because the conditions in the lysate more closely resemble those inside a cell. Such a method would never provide the yield or purity necessary for crystallization and carries an increased risk of sample degradation, but recent advancements in cryo-EM have made it possible to obtain high-resolution structures of heterogeneous samples with only picomoles of material. Using this method, I have obtained four cryo-EM maps of the HCV IRES in complex with the 80S ribosome, including two on-pathway states and a previously unseen conformation of the IRES, and I have subsequently made progress towards capturing earlier steps in the initiation pathway.

III.2.2 Purification of IRES mRNA

I initially prepared the IRES mRNA following the protocol that has traditionally been used in the laboratory (described in section III.3.2 of the Materials and methods). The mRNA was transcribed in vitro and purified by electrophoresis through denaturing urea–polyacrylamide gels, electroelution, dialysis, and centrifugal concentration. This was followed by a renaturation reaction in which the mRNA was
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

heated in the presence of magnesium cations. As a result, the protocol took almost a week to prepare a single batch of mRNA and carried the risk of misfolding and nucleolytic degradation.

Figure III.4 | Purification and analysis of mRNA. (A) Purification of in-vitro-transcribed mRNA, showing intermediates in the protocol. The lanes are labelled T (in-vitro-transcription mixture), P1 (phenol extract 1), P2 (phenol extract 2), P3 (phenol extract 3), C (chloroform extract), F (final stock). (B) Gel filtration of the phenol-extracted mRNA. The first peak appears at the void volume and contains the mRNA, which is further purified by ethanol precipitation. The second peak occurs at one column volume and contains free ribonucleotides. The third peak occurs reproducibly with each prep after more than a full column volume, regardless of the mRNA sequence, but its identity remains unclear. No precipitate appears when these fractions are mixed with ethanol. (C) In-vitro-translation reactions in cT2 with 35S-L-methionine, demonstrating that the IRES mRNAs prepared by this method are translationally active. The HCV construct includes domains I to VII. Addition of miR-122 (250 nM, in a fourfold excess relative to the mRNA), which is essential for the hepaciviral infection cycle (Jopling et al., 2005; Wilson & Huys, 2013) and reported to enhance translation of the full-length HCV genome (Henke et al., 2008), has no effect on translational activity in cT2. The EMCV construct includes domains D to O.

As I had to prepare a variety of different mRNA constructs over the course of the project, it was necessary to develop a faster purification method. The first step was to remove the proteins (T7 RNA polymerase and inorganic pyrophosphatase) and linearized plasmid using acidic phenol (figure III.4A). Under acidic conditions, the negative charge of DNA is neutralized and it partitions into the organic layer. The mRNA, however, remains in the aqueous layer. This was followed by gel filtration to remove the free ribonucleotides (figure III.4B). The purified mRNA was then concentrated by ethanol precipitation. By following this protocol, four batches of mRNA, totalling up to ~80 mg depending on the sequence, can be prepared in five
days, starting from the linearized plasmids. Furthermore, at no point is the mRNA denatured or subjected to elevated temperatures. The mRNA constructs prepared in this way are translationally active (figure III.4C) and were used for subsequent biochemical and structural analyses.

### III.2.3 Purification of native ribosomal complexes

A number of my colleagues have recently published a series of papers that describe high-resolution cryo-EM structures of native ribosomal complexes, stalled at various points in the translation pathway (Shao et al., 2015; Brown et al., 2015; Voorhees & Hegde, 2016; Shao et al., 2016). These complexes were purified from rabbit reticulocyte lysate (RRL) via an N-terminal FLAG tag on the nascent polypeptide. The FLAG tag is encoded within the mRNA construct and must be synthesized before the complexes can be purified. As a result, although this is a powerful method for capturing elongation, termination, and RQC complexes, it cannot be applied to translational initiation. A different approach is needed.

#### III.2.3.1 Purification of StreptoTagged complexes

In the absence of a nascent polypeptide, an alternative strategy is to tag the mRNA, which should be present in the target complex and is the only component added exogenously to the lysate. A variety of methods have been developed for tagging mRNA, but these tags generally suffer from poor efficiency. This is because RNA tags tend to bind weakly to affinity resin, or, in the case of biotin, bind so tightly that they can only be eluted under denaturing conditions. In an attempt to overcome these limitations, a streptomycin-binding aptamer, known as a StreptoTag, has been developed (Bachler et al., 1999; Windbichler & Schroeder, 2006). This structured motif is incorporated into the mRNA construct and is designed to have high affinity for dihydrostreptomycin-coupled resin yet elute efficiently in the presence of free streptomycin in the elution buffer. The StreptoTag has reportedly been used to purify 48S* initiation complexes from RRL (Locker et al., 2006; Locker & Lukavsky, 2007; Locker et al., 2007).

I initially set out to reproduce this purification method. The full protocol is described in section III.3.3 of the Materials and methods. In brief, a construct
containing a 3’ StreptoTag downstream of the HCV IRES (figure III.5A) was prepared by *in-vitro* transcription and incubated in RRL (Hunt & Jackson, 1974; Pelham & Jackson, 1976) in the presence of 2 mM guanosine 5′-(β,γ-imido)triphosphate (GDPNP). This nonhydrolysable analogue of GTP is thought to stall the 48S* complex by inhibiting the GTPase activity of eIF2γ (Merrick, 1979). Although a low concentration (~1 µM) is sufficient when using purified proteins *in vitro* (Shao *et al.*, 2016), higher concentrations (2 to 10 mM) are generally used in lysates due to the presence of endogenous GTP and the high concentrations of other GTPases that would compete for binding (Locker *et al.*, 2006; Locker & Lukavsky, 2007; Simonetti *et al.*, 2016).

The mixture was then passed through a column of dihydrostreptomycin-coupled Sepharose 6B. Bound complexes were eluted by competition with free streptomycin in the elution buffer (figure III.5B). The eluted complexes were further purified by pelleting and sucrose-gradient fractionation (figure III.5C). The plots shown in figure III.5 are taken from my first attempt and are similar to those of Locker *et al.* (2006), suggesting that the method worked, although the absence of a peak corresponding to 80S ribosomes is suspicious given the high concentration of magnesium cations in the reaction (see section III.2.6.2). In addition, the 48S* and 40S peaks from the sucrose gradient migrated with their expected relative sizes in a native agarose gel (figure III.5D). However, the yield was not high enough to also run a denaturing gel or send the sample for mass spectrometry, so the identity and composition of each peak remain uncertain.

Unfortunately, I was unable to reproduce these initial results. While some other research groups have used the StreptoTag for small-scale analytical experiments (Garlapati & Wang, 2009; Nicholson *et al.*, 2010), none have used the method for preparative-scale purification and a few, such as the groups of Ian Brierley and Kiyoshi Nagai (personal communications), have struggled with reproducibility. Although it was not possible to continue with the StreptoTag approach, I was able to identify a number of problems that helped to inform further work.

Firstly, in my hands, the StreptoTag construct bound to the column with much higher affinity than that reported in the literature. Locker *et al.* (2006) used a streptomycin concentration of 10 µM in their elution buffer, but this was not
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

sufficient to elute my construct. A linear elution from 0 to 100 mM demonstrated that a streptomycin concentration of at least 20 mM was required for efficient elution (figure III.5B). There is no obvious explanation for this difference in affinity, as the sequence of the StreptoTag in my construct is identical to that described previously (Windbichler & Schroeder, 2006). Regarding the streptomycin concentration, there is some inconsistency in the literature, with another group using a concentration of 1 mM for elution (Nicholson et al., 2010). I used a concentration of 50 mM in subsequent attempts to ensure maximal elution, which allowed me to obtain the results shown in figure III.5.

**Figure III.5** | Initial attempts at the StreptoTag method. (A) Design of the construct. The StreptoTag is a streptomycin-binding aptamer that should enable gentle elution from
A number of other problems with this method were apparent in hindsight. For example, the method includes two concentration steps, the first by pelleting and the second using a filter. Initiation factors are known to be transiently associated with the 40S subunit, resulting in heterogeneous samples even when reconstituted in vitro, crosslinked, and purified further (José Llácer, personal communication). Pelleting has the potential to disrupt the initiation complex and dislodge factors. The 40S subunit also has a tendency to aggregate on cryo-EM grids, which may be exacerbated by pelleting. Furthermore, there is anecdotal evidence that mammalian ribosomes stick to filters during centrifugal concentration, resulting in severe loss of material. This may not be a problem for large-scale preps, but pull-downs of native ribosomal complexes tend to have a low yield, providing only enough material for a handful of grids.

How then can native ribosomal complexes be concentrated after purification? The safest strategy is to elute the complexes at the concentration needed to prepare grids for cryo-EM, or preferably at a slightly higher concentration, thus avoiding the concentration step entirely. This would require very efficient elution. For 80S ribosomes, pelleting is also a valid strategy, as these complexes are more robust and less prone to aggregation.

The buffers used throughout the purification also included a high concentration of potassium cations (up to 300 mM), which, over the course of a two-day protocol with two concentration steps and multiple dilutions, may have resulted in the loss of weakly associated proteins from the complexes (for example, see section III.2.5.4).

If the StreptoTag method had been reproducible enough to prepare cryo-EM grids, I would have encountered another problem. Sucrose was included in all the buffers at a concentration of 8% up until the pelleting step, presumably to act as a cushion when pelleting the complexes, and was also present in the final sample as a result of the sucrose gradient. Sucrose increases the viscosity of aqueous solution and, as a
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

consequence, greatly increases the thickness of amorphous ice on grids. This can be compensated for by increasing the blot time, but only when the sucrose concentration is low. Sucrose also increases the level of background noise in micrographs, reducing the contrast of the particles (de Carlo et al., 1999; Thompson et al., 2016). Any method for purifying complexes for cryo-EM should include a step in which residual sucrose is removed, or simply avoid using sucrose altogether.

III.2.3.2 Purification of biotinylated complexes

Having gained experience from the StreptoTag method, I decided to try a simpler approach. Rather than aiming straight for the 48S* initiation complex, which may be difficult to purify intact, I developed a protocol for purifying 80S complexes containing the HCV IRES. The 80S ribosome can be pelleted without disrupting its integrity or resulting in severe aggregation on cryo-EM grids (Brown et al., 2015). It can also be stalled more easily, as a variety of antibiotics specifically target the elongation phase of translation and can be preincubated in the lysate to stall the ribosome before it can translocate away from the start codon. I therefore incubated the in-vitro-translation mixtures with cycloheximide, which is soluble to at least 50 mM in water and effectively stalls translation in cell lysates. Cycloheximide binds in the E site of the 60S subunit with a $K_d$ of $\sim$0.14 µM and stalls the ribosome in a pre-translocation state (Obrig et al., 1971; Schneider-Poetsch et al., 2010; Garreau de Loubresse et al., 2014; Lareau et al., 2014).

I also redesigned the IRES mRNA, replacing the StreptoTag with an unstructured region at the 3′ end that is complementary to a biotinylated oligo (figure III.6A). The interaction between biotin and streptavidin is one of the strongest in nature, with a $K_d$ of $\sim$10$^{-14}$ M, and is routinely used to purify nucleic acids. Ian Brierley kindly helped me to get started with this method, which was based on a similar approach from his research group (Namy et al., 2006).

Because the interaction between biotin and streptavidin-coated beads is essentially unbreakable under nondenaturing conditions, another method of elution is required. One such method involves endonucleolytic cleavage of the mRNA. This can be achieved by annealing a DNA oligo upstream of the biotinylated oligo but downstream of the HCV IRES (Namy et al., 2006). The beads can then be incubated with RNase H, which cleaves hybrid double helices of DNA and RNA. Another
potentially gentler method involves competitive annealing of an oligo that is complementary to the entire length (40 nt) of the biotinylated oligo. The fully annealed oligos have a predicted melting temperature of ~66 °C. The overlap between the biotinylated oligo and the IRES mRNA is only 20 nt (with a melting temperature of ~46 °C) and can therefore be outcompeted by the new oligo. This method has previously been used to purify spliceosomal complexes (Wojciech Galej, personal communication). The results of these methods are shown in figure III.6B. Although the digestion with RNase H gave slightly better elution, I proceeded with the competitive-annealing method as this seemed to elute well enough, involved fewer components, and could be done on ice.

**Figure III.6** | Purification of 80S initiation complexes using biotinylated IRES mRNA. (A) Design of the construct. The biotinylated oligo anneals to a region at the 3′ end of the mRNA. The overhang allows competitive annealing of a complementary oligo. (B) Trial elutions, using an upstream DNA oligo and RNase H or competitive annealing of an oligo complementary to the biotinylated oligo. The incubation with RNase H seemed to result in slightly better elution, but the competitive-annealing method was chosen because it is gentler and faster. The thick band of ~12 kDa represents streptavidin. The lanes are labelled P (pellet), M (markers), W1 (wash 1), E (eluate), W2 (wash 2), B (beads after elution). Wash 1
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

was the final wash before elution. Wash 2 was an additional wash after elution. Equivalent volumes were loaded into lanes W1, E, W2, and B. (C) Purification of 80S initiation complexes. The complexes were pelleted, resuspended, and incubated with streptavidin-coated beads. After washing, the complexes were eluted, pelleted again, and resuspended in a small volume (~40 µl). The lanes are labelled M (markers), P1 (pellet 1), F (flowthrough), W1 (wash 1), B (beads after elution), S (supernatant), P2 (pellet 2). Equivalent volumes were loaded into lanes P1 and F, and into lanes W1 and B.

I also started using a version of nuclease-treated RRL, known as concentrated T2 (cT2), developed by Ramanujan Hegde and prepared in house (see section III.3.4 in the Materials and methods) (Sharma et al., 2010). This mixture is optimized for high yields of synthesized protein and enabled me to control the salt and buffer conditions of the in-vitro-translation reactions, which was helpful for identifying some later methodological problems. In-vitro-translation reactions in cT2 using the HCV IRES and the EMCV IRES are shown in figure III.4C.

![Figure III.7](image_url)

**Figure III.7** Data collection. (A) A representative micrograph of sample P2. (B) Representative classes from reference-free 2D classification. Some free 40S subunits are visible; these would have been discarded after 3D classification.

The biotinylated oligo was annealed to the IRES mRNA and incubated in cT2 that had been premixed with cycloheximide. Complexes were pelleted through a sucrose cushion, bound to streptavidin-coated beads, washed, eluted with the competitive oligo, and pelleted again (figure III.6C). The final resuspension was used to prepare grids for cryo-EM. The second pelleting step was introduced to further concentrate the complexes, as eluting in a small volume did not produce a high enough
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

concentration for grids. However, despite repeated attempts with ever higher $A_{260}$ values, it was not possible to improve the concentration on the grids beyond $\sim50$ particles per image (at a calibrated magnification of 104,478×). I was not able to determine the cause of this problem, and it ultimately limited the sizes of the cryo-EM datasets. Nevertheless, from just one dataset I was able to determine high-resolution reconstructions of several distinct complexes.

### III.2.4 Data collection and processing

Micrographs were collected and processed in RELION (Fernández-Leiro & Scheres, 2016) (figure III.7). Three major classes containing density that corresponds to the HCV IRES were isolated after an initial round of 3D classification without alignment. Two of these represent on-pathway states containing tRNA$_{AUG}$, whereas the other represents an off-pathway state with eEF2. To improve the occupancy of the HCV IRES, FCwSS was undertaken on each of these classes using a local mask on domains II and III of the IRES. This revealed a fourth class, containing only 3,363 particles, in which the 40S subunit has adopted a rolled conformation. In the two classes that represent on-pathway states, the local resolution of the P/P- or P/E-site tRNA$_{AUG}$ was further improved by another round of FCwSS. The classification scheme is outlined in figure III.8 and the statistics for the dataset are recorded in table III.3.

<table>
<thead>
<tr>
<th></th>
<th>Class I</th>
<th>Class II</th>
<th>Class III</th>
<th>Class IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>eEF2</td>
<td>Rolled</td>
<td>Nonrotated</td>
<td>Rotated</td>
</tr>
<tr>
<td>Particles</td>
<td>11,313</td>
<td>3,363</td>
<td>12,425</td>
<td>16,362</td>
</tr>
<tr>
<td>Pixel size (Å)</td>
<td>1.34</td>
<td>1.34</td>
<td>1.34</td>
<td>1.34</td>
</tr>
<tr>
<td>Defocus mean (µm)</td>
<td>−2.3</td>
<td>−2.3</td>
<td>−2.2</td>
<td>−2.3</td>
</tr>
<tr>
<td>Defocus range (µm)</td>
<td>−0.4 to −4.8</td>
<td>−0.4 to −4.8</td>
<td>−0.4 to −4.8</td>
<td>−0.4 to −4.8</td>
</tr>
<tr>
<td>Voltage (kV)</td>
<td>300</td>
<td>300</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Electron dose ($e\cdot Å^{-2}$)</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>4.1</td>
<td>6.6</td>
<td>3.9</td>
<td>3.8</td>
</tr>
<tr>
<td>Angular accuracy (°)</td>
<td>0.513</td>
<td>0.834</td>
<td>0.460</td>
<td>0.415</td>
</tr>
</tbody>
</table>

Table III.3 | Statistics for the 80S cryo-EM dataset.
**III.2.5 Interpretation of cryo-EM maps**

As I was processing this dataset and continuing to optimize the purification method, Nenad Ban’s group published a paper detailing the structure of the HCV IRES bound to the 80S ribosome (though only the 40S subunit was modelled) at a resolution of 3.9 Å (Quade et al., 2015). Not long afterwards, a second paper was published from Christian Spahn’s group, which described the same structure, also at 3.9 Å, as well as a novel conformation of the IRES bound to the 80S ribosome in the presence of a P-site tRNA (Yamamoto et al., 2015). These structures generally agree well with my maps (figure III.9A). I was not ready to write up a paper at the time, however, as I still needed to improve the maps and build the models. I therefore decided to focus on other projects.

**Figure III.8** | *In-silico* 3D classification of 80S ribosomes in complex with the HCV IRES. Poorly aligned particles, empty 80S ribosomes, and free 60S subunits are grouped together as junk particles for simplicity. Three IRES-containing classes were isolated after a single round of classification without simultaneous alignments. Subsequent FCwSS with a mask around the HCV IRES revealed a fourth class of complexes with a rolled conformation. The classes representing on-pathway states were further improved by FCwSS with a mask around the P- or A/P-site tRNA.

Here, I will describe the cryo-EM maps that I obtained of the HCV IRES in complex with the 80S ribosome, including details of the molecular interactions. Given the
agreement between my maps and the recently published structures, I will use the published models as a basis to discuss various aspects of the biological interpretation.

### III.2.5.1 General features of the maps

In all four classes, most of domain III is clearly visible bound to the trailing edge of the 40S subunit (figure III.10A). The exception is the apical region of subdomain IIIb, which appears to be flexible in the absence of eIF3 (Hashem et al., 2013a). With the exception of class IV (see section III.2.5.2), the overall conformation of the IRES resembles that of previous low-resolution reconstructions in which domain II adopts a bent conformation as it reaches into the E site (Spahn et al., 2001; Boehringer et al., 2005). Subdomain IIb contacts uS7, uS11, eS25, and the 18S rRNA (Quade et al., 2015; Yamamoto et al., 2015) (figure III.10B). In so doing, it mimics some of the interactions that the anticodon stem-loop of a tRNA would make with the E site. Similar contacts are also made by eIF2α (Hussain et al., 2014; Llácer et al., 2015) and type-IV IRESs (Fernández et al., 2014; Koh et al., 2014). An interaction with eS25 is essential for translational initiation from various unrelated IRESs (Landry et al., 2009; Hertz et al., 2013). Unlike a tRNA, however, subdomain IIb does not base-pair with the E-site codon. Instead, the conserved G82 in the apical loop of subdomain IIb protrudes into the mRNA channel and contacts A339, which is located at the −3 position relative to the start codon and stacks against G961 of the 18S rRNA. This interaction presumably stabilizes the conformation of the mRNA in the channel and helps to fix the start codon in the P site (Filbin & Kieft, 2011).

![Figure III.9](image.png)

**Figure III.9** | Comparison of class III from my dataset with the published structure from Christian Spahn’s group, shown in grey (PDB accession code 5FLX) (Yamamoto et al.,...
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

2015). (A) Fit of the model into class III, showing good agreement for the conformations of the 40S body and HCV IRES. (B) Side view of the ribosome, showing the different degrees of head tilting. The mRNA channel has a fully closed conformation in class III, but is more open in Spahn’s model. The published structure from Nenad Ban’s group has an intermediate degree of head tilting (Quade et al., 2015).

The recently published structures from Yamamoto et al. (2015) show that, in the absence of a P-site tRNA, the mRNA channel partially opens as a result of the 40S head tilting ~17°. Quade et al., (2015) also observed tilting, though to a lesser degree. Only one of my maps, the low-abundance class II, shows evidence of tilting (in addition to rolling; see section III.2.5.2), presumably due to the absence of tRNA and eEF2 in this class. The mRNA channel is closed in the other three classes (figure III.9B). Domain II of the HCV IRES interacts primarily with the head of the 40S subunit and moves as a result of tilting, but the overall conformation of the IRES remains similar between the open and closed states.

It is debatable whether subdomain IIb of the HCV IRES contacts the bottom of the E site at any point during translational initiation; at least until it reaches the first step of elongation represented by class III (see section III.2.5.2). On the one hand, the IRES can function in vitro by binding the 40S subunit and subsequently recruiting initiation factors and the Met-tRNAi^AUG. This conformation would therefore represent the initial binding state. However, this state is unlikely to occur in vivo, as the IRES would primarily interact with 43S preinitiation complexes (see section III.1.2.3 in the Introduction) (Asano et al., 2001). It is possible that subdomain IIb may rest at the bottom of the E site at other points in the initiation pathway, perhaps after the clash with eIF2α has been resolved, but other conformations have also been observed (Yamamoto et al., 2014; Yamamoto et al., 2015).

A previous low-resolution reconstruction of the HCV IRES in a cycloheximide-stalled complex revealed density that connects subdomain IIa to the L1 stalk (Boehringer et al., 2005). No such density can be seen in my maps or those from Quade et al. (2015) and Yamamoto et al. (2015), even at low threshold. Quade et al. (2015) suggest that this density might be attributable to domain I or the aptamer used to purify the complex (Boehringer et al., 2005). My mRNA construct also included domain I, so the latter explanation is more likely.
Biochemical studies have shown that subdomain IIId and junction IIIef are responsible for the nanomolar affinity between the IRES and the 40S subunit (Kieft et al., 2001; Otto & Puglisi, 2004). These regions are anchored to expansion segment ES7 (or helix 26) of the 18S rRNA, which has reduced flexibility compared to structures lacking the IRES (Voorhees et al., 2014). In particular, the apical loop of subdomain IIId forms a kissing-loop interaction in which the conserved GGG triplet base-pairs with residues 1116 to 1118 of ES7 (figure III.10B). This validates previous data showing that the GGG triplet is essential for binding and that mutants can be rescued by compensatory mutations in ES7 (Matsuda & Mauro, 2014).

The apical loops of subdomains IIIa and IIIc contact ribosomal protein eS27 (figure III.10B). These interactions depend on the four-way IIIabc junction connecting these subdomains, which adopts an antiparallel conformation that brings them into close proximity (Quade et al., 2015; Yamamoto et al., 2015). This is in contrast to the isolated crystal structure of this region, in which the junction adopts a parallel conformation (Kieft et al., 2002). Fluorescence data have shown that, in the context of the full IRES and in the presence of magnesium cations, the junction exists in a dynamic equilibrium with roughly equal populations of parallel and antiparallel conformations (Melcher et al., 2003). Whether the 40S subunit induces a conformational change upon binding or selectively interacts with the antiparallel conformation has not yet been determined.

In canonical initiation, eS27 is an important anchor point for eIF3 (see section III.1.2.4) (Hashem et al., 2013b; Hussain et al., 2014; Llácer et al., 2015; des Georges et al., 2015). As a result, the HCV IRES would exclude eIF3 from this interface, potentially leading to the complete detachment of eIF3 from the 40S subunit (Hashem et al., 2013a), although this begs the question as to how the IRES displaces eIF3 when recruited to a 43S preinitiation complex in which eIF3 would already be bound.

Subdomain IIIf, which folds into a pseudoknot, is located at the exit of the mRNA channel where it contacts eS28 (figure III.10B). In this position, the subdomain functions as an anchor point that fixes the position of the start codon in the P site (Kieft et al., 2001; Berry et al., 2010; Berry et al., 2011). Indeed, this region of the IRES resembles the structure formed by a bacterial Shine-Dalgarno sequence with its complementary sequence in the 16S rRNA (though the IRES acts in trans relative to
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

the ribosome) and seems to dictate the position of the start codon in much the same way. Upon entering the elongation phase, translocation would add an additional three residues between the E site and subdomain IIIf, presumably resulting in displacement of the latter, but whether a single translocation event is sufficient for displacement, and whether this is sufficient to disrupt the other contacts with domain III, remains to be determined.

Figure III.10 | The model of the HCV IRES (pink) bound to the 40S subunit from Christian Spahn’s group (PDB accession code 5FLX) (Yamamoto et al., 2015). (A) Overview of the
structure, showing the arrangement of subdomains and the position of the IRES on the trailing edge of the 40S subunit (yellow). Subdomain IIb cannot be seen in cryo-EM maps due to its flexibility. (B) Contacts between the IRES and the 40S subunit, showing the proteins at the binding interface and helices 23 (h23) and 26 (h26) of the 18S rRNA.

### III.2.5.2 Differences between the classes

While the conformation of the HCV IRES generally remains the same, there are a number of differences between the classes in my dataset (figure III.11). The first of these, class I, has a resolution of 4.1 Å and contains the HCV IRES and eEF2 but no tRNA. The 40S subunit is also partially rotated relative to the 60S subunit. This class represents an off-pathway state in which the HCV IRES has bound directly to an 80S ribosome without going through the initiation pathway. Similar off-pathway classes containing eEF2 in the A site have been observed in other datasets of natively purified 80S ribosomes (Voorhees et al., 2014). In the absence of tRNA, the HCV IRES adopts the conformation described in section III.2.5.1, with subdomain IIb resting at the bottom of the E site and mimicking an anticodon stem-loop.

![Figure III.11](image)

**Figure III.11** | Overview of the four classes that contain the HCV IRES. Class I represents an off-pathway state containing eEF2 but no tRNA. Class II contains neither tRNA nor eEF2, but is distinguished by rolling of the 40S subunit. Class III represents an on-pathway state at the transition between the initiation and elongation phases of translation, containing a Met-tRNA\textsubscript{i}\textsuperscript{AUG} in the P site. Class IV represents a later step in the first round of elongation, after peptidyl transfer and rotation of the 40S subunit, and contains a deacylated tRNA\textsubscript{i}\textsuperscript{AUG} and a peptidyl-tRNA\textsuperscript{AGC} in hybrid conformations.

Despite the low number of particles, class II has a resolution of 6.6 Å. This class also lacks tRNA but, in contrast to class I, has very poor occupancy for eEF2. Although there is some weak density in the A site, this class was not further classified.
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

because of its small size. The major difference between classes I and II, however, is that class II shows evidence of subunit rolling. This movement, a $\sim 6^\circ$ rotation of the 40S subunit around its long axis, has been observed in previous structures of mammalian cytosolic and mitochondrial ribosomes (Budkevich et al., 2014; Yamamoto et al., 2014; Amunts et al., 2015). Yamamoto et al. (2015) also identified a rolled class with the HCV IRES. As the IRES makes no contacts with the 60S subunit, this movement seems to result from the intrinsic properties of the 80S ribosome. My data suggest that off-pathway binding of eEF2 fixes the ribosome in an unrolled, partially rotated state.

The remaining two classes represent on-pathway states with tRNA, neither of which has been observed previously. Class III has a nonrotated conformation and contains a Met-tRNA$^\text{AUG}$ in the P site. This is the mature 80S initiation complex, ready to receive the ternary complex of Ser-tRNA$^\text{AGC}$, eEF1A, and GTP. In this class, the IRES adopts the same conformation as described in section III.2.5.1.

Yamamoto et al. (2015) also describe a map containing density for a P-site tRNA. The presence of a tRNA was unexpected and probably resulted from residual contamination of the ribosome prep. The published map has an overall resolution of 5.8 Å, whereas my class III averages 3.9 Å. At these resolutions, the P-site tRNA cannot be definitively identified as an initiator tRNA$^\text{AUG}_i$ (with the most likely alternative being an elongator tRNA$^\text{AUG}$), but this seems a fair assumption. In my maps, the lack of density in the exit tunnel of the 60S subunit also supports the conclusion that multiple rounds of elongation have not occurred. The major difference, however, is that, in the published map, subdomain IIb of the HCV IRES reaches upwards to contact the E site of the 60S subunit, $\sim 55$ Å from its position on the 40S subunit. This movement is even more pronounced than the previously observed movement induced by eIF5B, which pushes the Met-tRNA$^\text{AUG}_i$ towards the E site, enabling subdomain IIb to contact the opposite shoulder (Yamamoto et al., 2014). The function of these movements is unclear. The most pronounced movement in the recent map may represent a state in which subdomain IIb mimics the contacts that an E-site tRNA would make with the 60S subunit, as it also mimics an anticodon stem-loop in its conventional position. The movements may also relate to mutagenic data showing that the apical loop of subdomain IIb affects the first translocation event of the elongation phase (described in section III.1.2.2 of the Introduction) (Filbin et
Neither of these explanations provides a definitive function, however, and additional structures will be required to better characterize the conformational dynamics of the HCV IRES in the context of the 80S ribosome.

Class IV, which represents the most abundant state and the only conformation of the IRES not previously described in the scientific literature, reaches the highest resolution of 3.8 Å. In this class, the 40S subunit adopts a rotated conformation relative to the 60S subunit, with a deacylated tRNA<sup>i</sup><sub>AUG</sub> and probably a peptidyl-tRNA<sup>AGC</sup> in hybrid conformations. This state occurs after peptidyl transfer but before recruitment of eEF2 and subsequent translocation. The P/E-site tRNA<sup>i</sup><sub>AUG</sub> and the anticodon stem-loop of the A/P-site peptidyl-tRNA<sup>AGC</sup> are well resolved, but the upper regions of the peptidyl-tRNA<sup>AGC</sup> appear to be flexible, with only scattered density visible despite FCwSS. This is expected, as a short dipeptidyl moiety would be insufficient to anchor the tRNA<sup>AGC</sup> to the peptidyltransferase centre.

The ribosome remains stalled over the start codon with the HCV IRES bound to the trailing edge, as in the other classes. However, as a result of subunit rotation, the L1 stalk has moved down to contact the tRNA<sup>i</sup><sub>AUG</sub> and stabilize its P/E hybrid conformation. Together, the tRNA<sup>i</sup><sub>AUG</sub> and L1 stalk occlude the E site and prevent it being occupied by subdomain IIb of the IRES. As a consequence, domain II has flipped out and become disordered, leaving only domain III visible in the map.

The HCV IRES is known to bind the 40S subunit through a number of stable interactions. These are presumably broken after initiation is completed, but the mechanism by which the IRES dissociates from the ribosome is not known. This structure provides a partial answer by showing that dissociation occurs in a stepwise manner, with domain II being displaced from the E site before domain III detaches from the 40S subunit. A similar hypothesis was proposed a number of years ago, based on low-resolution data (Boehringer et al., 2005).

It is also worth mentioning that cycloheximide inhibits translocation by blocking a step before rotation of the ribosomal subunits (Schneider-Poetsch et al., 2010; Lareau et al., 2014). The presence of a rotated class in my dataset presumably results from the fact that cycloheximide was not included in any of the buffers after the sucrose cushion. Cycloheximide is a reversible noncompetitive inhibitor, meaning that the degree of inhibition depends solely on its concentration. Bound cycloheximide would
therefore dissociate during the purification, allowing the ribosomes to adopt a rotated conformation if the P-site tRNA$_{i}^{\text{AUG}}$ has been deacylated.

### III.2.5.3 Capturing 80S complexes with anisomycin

I also tried to capture a post-translocation state using anisomycin (Grollman, 1967), an elongation inhibitor that binds in the peptidyltransferase centre (Hansen et al., 2003). In contrast to cycloheximide, anisomycin preferentially binds post-translocation ribosomes (Barbacid & Vázquez, 1974; Barbacid & Vázquez, 1975), resulting in a distinct footprint (Lareau et al., 2014). The aim was to trap the HCV IRES after a single translocation event to determine whether domain III was still bound to the ribosome and, if so, to identify its conformation.

However, I was not able to identify a distinct post-translocation state in this dataset (data not shown). Most of the particles lacked the IRES, and the few classes with bound IRES were identical to the cycloheximide-stalled pre-translocation states described above. This may be because the IRES completely detaches from the ribosome upon translocation. An alternative explanation, however, is that anisomycin may have failed to stall the complexes after just one translocation event. Anisomycin is known to be a leaky inhibitor, allowing multiple translocations before eventually stalling the ribosome (Anthony & Merrick, 1992; Kozak, 1998; Dmitriev et al., 2003).

A more precise method of capturing a post-translocation state would be to mutate the third codon to a stop codon and incubate the in-vitro-translation reaction with a dominant-negative mutant of eRF1 that is unable to hydrolyse the peptidyl-tRNA (Brown et al., 2015). This would stall the ribosome after a single translocation event with the mutant eRF1 bound to the stop codon in the A site. As a consequence, the ribosome would stall in a nonrotated conformation with the deacylated tRNA$_{i}^{\text{AUG}}$ in the E site. Such a structure would answer whether the IRES dissociates from the ribosome upon translocation or remains at least partially bound.

### III.2.5.4 Density over the exit tunnel

During initial optimization of the protocol, small preliminary datasets were collected using samples that contained potassium acetate (rather than chloride, which improved the particle distribution on grids) and had not undergone the second pelleting step. Presumably as a result of these gentler conditions, the cryo-EM maps contained
additional density over the exit tunnel corresponding to expansion segment ES27 and a protein of ~50 kDa (figure III.12). The fold identifies this protein as belonging to the family of methionyl aminopeptidases. One such protein, Arx1, has previously been observed in the same location on the *S. cerevisiae* 60S subunit (Bradatsch *et al.*, 2012; Greber *et al.*, 2012). BackPhyre (Kelley *et al.*, 2015) was used to identify structural homologues of Ebp1 (Monie *et al.*, 2007; Kowalinski *et al.*, 2007), the mammalian orthologue of Arx1, in the human genome (which is better annotated than the rabbit genome). Nine proteins were identified with 100% confidence, listed in table III.4. Four of these candidates (Ebp1, MetAP1, MetAP2, and prolidase) have the correct mass and subcellular localization.

![Figure III.12](image)

*Figure III.12* | Density over the exit tunnel corresponding to Ebp1 and ES27 (grey), present in preliminary datasets. A model of Ebp1 (PDB accession code 2Q8K) (Kowalinski *et al.*, 2007) was fitted to the density in Chimera. The overall conformation is very similar to that of Arx1 bound to the yeast ribosome (Bradatsch *et al.*, 2012; Greber *et al.*, 2012).

Of the four candidates, Ebp1 itself fits best in the observed density. This is a little surprising, as MetAP1 or MetAP2 might be considered more likely to remain constitutively bound to the 60S subunit where they would be ready to remove the N-terminal methionyl residue as soon as it emerges from the exit tunnel. Ebp1, by contrast, is a moonlighting protein that is known to inhibit phosphorylation of eIF2α (Squatrito *et al.*, 2006), function as an ITAF for some type-II IRESs (Pilipenko *et al.*, 2000), play a role in ribosome biogenesis (Nissan *et al.*, 2002; Squatrito *et al.*, 2004;
Hung & Johnson, 2006), and regulate transcription in response to ErbB3 activation (Yoo et al., 2000; Xia et al., 2001; Zhang et al., 2003; Zhang & Hamburger, 2005; Liu et al., 2006). There is no evidence that Ebp1 affects translation from the HCV IRES, so its function in these complexes is unclear. Perhaps it functions more generally in translation by keeping the ribosome in a particular conformation until the nascent polypeptide emerges from the exit tunnel during the elongation phase.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Mass (kDa)</th>
<th>Localization</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ebp1</td>
<td>44</td>
<td>Cytosol</td>
<td>Implicated in many pathways</td>
</tr>
<tr>
<td>MetAP1</td>
<td>43</td>
<td>Cytosol</td>
<td>Removes N-terminal methionyl from nascent polypeptides</td>
</tr>
<tr>
<td>MetAP2</td>
<td>53</td>
<td>Cytosol</td>
<td>Removes N-terminal methionyl from nascent polypeptides</td>
</tr>
<tr>
<td>MetAP1D</td>
<td>37</td>
<td>Mitochondria</td>
<td>Removes N-terminal methionyl from nascent polypeptides</td>
</tr>
<tr>
<td>Prolidase</td>
<td>55</td>
<td>Cytosol</td>
<td>Cleaves dipeptides with a C-terminal prolyl</td>
</tr>
<tr>
<td>XProAP1</td>
<td>70</td>
<td>Cytosol</td>
<td>Removes any N-terminal residue that is linked to prolyl</td>
</tr>
<tr>
<td>XProAP2</td>
<td>76</td>
<td>Cytosol</td>
<td>Removes any N-terminal residue that is linked to prolyl</td>
</tr>
<tr>
<td>XProAP3</td>
<td>57</td>
<td>Mitochondria</td>
<td>Removes any N-terminal residue that is linked to prolyl</td>
</tr>
<tr>
<td>SPT16</td>
<td>120</td>
<td>Nucleus</td>
<td>Remodels chromatin</td>
</tr>
</tbody>
</table>

Table III.4 | Candidate proteins for the density visible over the exit tunnel in preliminary datasets that were prepared under gentler conditions.

### III.2.6 Purification of Native 48S* Initiation Complexes

In light of the recent papers from the research groups of Nenad Ban and Christian Spahn, I decided to focus my efforts on the 48S* initiation complex. This would allow me to build upon the methods that I had developed, with the aim of obtaining a high-resolution structure of the HCV IRES in a functional complex with initiation factors and Met-tRNAi^{AUG}.

#### III.2.6.1 Problems with GDPCP and magnesium

To capture 48S* complexes, I initially tried to use the same protocol as for 80S complexes, simply replacing cycloheximide with another nonhydrolysable analogue of GTP, guanosine 5’- (β,γ-methylene)triphosphate (GDPCP), at a concentration of 2 mM. According to the literature, this should be sufficient to stall eIF2 bound to the
48S* complex in RRL and prevent recruitment of the 60S subunit (Locker et al., 2006; Locker & Lukavsky, 2007; Locker et al., 2007; Simonetti et al., 2016). Although the final yield of complexes was low, the protocol seemed to work. The purified complexes are shown in figure III.13B. However, at the same time, I wanted to see if it were possible to capture elongation factors bound to 80S ribosomes by adding GDPCP after stalling the complexes with cycloheximide. When complexes were prepared with cycloheximide alone, 80S ribosomes were visible in the micrographs (figure III.13C), but subsequent addition of GDPCP resulted in a mess of particles mostly resembling 40S subunits (figure III.13D). The protocols were otherwise identical, so the only explanation is that GDPCP disrupted the integrity of the stalled 80S ribosomes.

![Figure III.13](image_url)

**Figure III.13** | Micrographs showing the effect of excess GDPCP on ribosome integrity. (A) Purified human 40S subunits with IRES mRNA, showing the normal appearance of 40S subunits by cryo-EM. (B) Native complexes prepared with 2 mM GDPCP. The particles
appear to be mostly 40S subunits, as expected, with some 80S contaminants. (C) Native complexes prepared with 1 mM cycloheximide. All the particles are 80S ribosomes, as expected. (D) Native complexes prepared with 1 mM cycloheximide followed by 2 mM GDPCP. Surprisingly, most of the particles appear to be 40S subunits, with only a few 60S or 80S particles. Cycloheximide would have stalled the 80S complexes, but the subsequent addition of GDPCP disrupted the integrity of the ribosomes.

All nucleoside 5′-triphosphates (NTPs) are fairly strong acids that exist predominantly as $4^-$ anions in a neutral solution. The presence of multiple negatively charged groups with lone pairs of electrons means that NTPs are chelating agents that sequester metal cations (Storer & Cornish-Bowden, 1976). Although they chelate a number of metals, by far the strongest interaction is with magnesium (Wilson & Chin, 1991). As a result, intracellular NTPs exist mostly in a magnesium-bound form (figure III.14).

![Figure III.14](image)

**Figure III.14** | Chelation of a magnesium cation by GTP. The overall charge of the complex is $2^-$. The lone-pair donors bind the magnesium cation with roughly tetrahedral geometry.

Many ribonucleoprotein complexes require magnesium cations to maintain their tertiary and quaternary structures (Klein *et al.*, 2004; Shenvi *et al.*, 2005; Petrov *et al.*, 2012). Ribosomes themselves are a major reservoir of magnesium, such that knocking down the expression of rRNA causes a significant rise in the concentration of free cytosolic cations (Akanuma *et al.*, 2014). Indeed, magnesium is so important for ribosome integrity that a low concentration instigates a stress response in bacteria.
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

whereby transcription of rRNA is suppressed and residual magnesium is conserved to maintain protein synthesis (Pontes et al., 2016). Adding a high concentration of a chelating agent to lysate therefore has the unintended, though predictable, consequence of disrupting ribosome integrity, leading to subunit separation and even partial denaturation of ribosomal subunits (Allen & Wong, 1986; Bonincontro et al., 1993; Yamamoto et al., 2010). A similar effect can be produced by greatly raising the concentration of monovalent cations such as potassium (Blobel & Sabatini, 1971). Having torn apart the ribosomes, the newly freed 40S subunits may form off-pathway interactions with other components of the translational machinery. Proteins are largely unaffected by magnesium, but their interactions with ribosomal subunits may be disrupted and the properties of enzymes could be altered.

The salt sensitivity of in-vitro-translation systems has wider implications, casting doubt upon a number of published biochemical and structural studies that were conducted without controlling the magnesium concentration. For example, in a recent cryo-EM study, partial 48S complexes were purified from RRL that had been incubated with 10 mM GDPNP, apparently without controlling the salt concentrations (Simonetti et al., 2016). Ribosome integrity would have been disrupted, but exactly how this would affect the assembly and purification of initiation complexes is unclear. Other structural studies of the ribosome have used high NTP concentrations without adverse effects, but the buffers used in these studies typically contain a high concentration of magnesium (Schmeing et al., 2009; Ramrath et al., 2012; Tourigny et al., 2013). It is important to consider the ratio between NTPs and divalent cations.

III.2.6.2 Off-pathway subunit joining

In light of these data, I repeated the in-vitro-translation reaction with 2 mM GDPCP premixed with an equimolar concentration of magnesium chloride. Somewhat unexpectedly, when the purified complexes were observed by cryo-EM, they turned out to be 80S ribosomes (figure III.15). It was therefore not possible to stall 48S* initiation complexes under these conditions, presumably because the salt conditions (with a high concentration of divalent cations and a low concentration of monovalent cations) favoured spontaneous, off-pathway subunit joining. The HCV IRES also has a tendency to interact directly with 80S ribosomes, especially when the concentration
of divalent cations is high (Lancaster et al., 2006; Quade et al., 2015; Yamamoto et al., 2015).

Another possible explanation is that the added GDPCP was not able to stall the initiation pathway. To account for this possibility, I had used a batch of cT2 that had been prepared without addition of GTP and adenosine 5′-triphosphate (ATP). Nevertheless, it is still possible that the endogenous GTP already present in the lysate may have been sufficient to outcompete the exogenous GDPCP.

IIII.2.6.3 Ribosome pelleting and add-back of 40S subunits

These problems demanded a new strategy for capturing 48S* initiation complexes. I therefore decided to take a different approach, removing all the ribosomes from cT2 by pelleting and then adding back purified 40S subunits. By incubating the 40S subunits in ribosome-depleted cT2, it should be possible to assemble 43S preinitiation complexes before adding the IRES mRNA. Ribosomal subunits were purified from human embryonic kidney (HEK) cells through multiple sucrose cushions and sucrose gradients. The conditions for subunit separation were based on those described by Blobel & Sabatini (1971) (see section III.3.8 in the Materials and methods).
Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome

To simplify the purification method, I also decided to presaturate the beads with IRES mRNA and wash away the excess before incubating the beads directly in the *in-vitro*-translation mixture. This avoided the initial pelleting step, which could be damaging for 48S* complexes, and maximized the efficiency of the pull-down by preventing the formation of excess complexes that would not have enough space to bind the beads.

**Figure III.16** | Purification of 48S* initiation complexes using biotinylated IRES mRNA. (A) Pull-down assay for 40S subunits and initiation factors, using ribosome-depleted cT2. The streptavidin-coated beads were presaturated with IRES mRNA, incubated in the reaction mixtures, washed, resuspended in loading dye and loading onto the gel. The IRES mRNA specifically pulls down 40S subunits and initiation factors. The thick band of ~12 kDa represents streptavidin, and some higher bands result from crosslinked streptavidin. (B) Elution assay for IRES mRNA using DNase I, showing efficient elution under all conditions tested. The beads were resuspended in loading dye and loaded onto the gel. (C) Elution assay for 40S subunits. Presaturated beads were incubated with purified 40S subunits, washed, and incubated with DNase I on ice for 1 h. The lanes are labelled M (markers), D (DNase I), F (flowthrough), W1 (wash 1), B1 (beads before elution), E (eluate), W2 (wash 2), B2 (beads
after elution). Wash 1 was the final wash before elution. Wash 2 was an additional wash after elution. Equivalent volumes were loaded into lanes F, W1, B1, E, W2, and B2. (D) Purification of 48S* complexes from ribosome-depleted cT2. The beads were presaturated with IRES mRNA, washed, incubated in the in-vitro-translation mixture, washed, incubated with DNase I to elute the complexes, and washed again. The lanes are labelled M (markers), D (DNase I), W1 (wash 1), B1 (beads before elution), E (eluate), W2 (wash 2), B2 (beads after elution). Equivalent volumes were loaded into lanes W1, B1, E, W2, and B2.

The IRES mRNA used in these experiments and the preliminary data collection (described in section III.2.7) was also simpler, as it lacked domains I and II. Deletion of these domains has no effect on ribosome binding but stalls the initiation complex before subunit joining with eIF3 still bound (Pestova et al., 1998; Kieft et al., 2001; Ji et al., 2004; Otto & Puglisi 2004; Locker et al., 2007) (see table III.2). An initial pull-down experiment with this mRNA construct showed good specificity for 40S subunits and initiation factors (figure III.16A).

Due to the difficulties of obtaining a high yield with the competitive oligo, I also started using deoxyribonuclease (DNase) I to digest the biotinylated oligo and elute the complexes. An initial elution assay with IRES mRNA demonstrated very efficient elution even when the beads were incubated on ice (figure III.16B). Elution of 40S subunits was also efficient (figure III.16C). Elution was less efficient in the context of 48S* complexes (figure III.16D), but this method is still the most efficient that I have tried so far.

I used this updated method to purify 48S* complexes. The beads were presaturated with IRES mRNA and incubated directly in the in-vitro-translation mixture. After washing, the complexes were eluted with DNase I in a small volume (figure III.16D). The eluate was used immediately to prepare grids for cryo-EM.

### III.2.7 Preliminary data collection

As before, micrographs were collected using an FEI Titan Krios and processed in RELION (Fernández-Leiro & Scheres, 2016). After 2D classification, it became apparent that the dataset has severe orientation bias (figure III.17A,B), resulting in a very streaky map after an initial round of 3D refinement (figure III.17C). This map could not be improved by 3D classification. The HCV IRES is visible, bound to the
trailing edge of the 40S subunit. There is also scattered density above and below the 40S subunit, roughly in the expected locations of eIF3 and other initiation factors, but no details are discernable.

**Figure III.17** | The initial cryo-EM dataset of purified 48S* initiation complexes. (A) Representative classes from reference-free 2D classification. All classes show the same view. (B) Orientational bias in the initial reconstruction. The dataset is mostly populated by only two similar views. (C) The initial reconstruction after 3D refinement, showing severe orientation bias. An mRNA construct lacking domains I and II of the HCV IRES was used in all the 48S* experiments and the data collection, explaining why only domain III is visible in the map. Scattered density appears above and below the 40S subunit around the expected locations of eIF2 and eIF3, but no features are identifiable.

This was merely the first preliminary dataset for the 48S* complexes and there is much room for optimization, but the initial results are promising. It will be necessary to improve the sample preparation, both in terms of the purification method (to increase the yield; discussed in section III.2.8.3) and the grid preparation (to overcome the orientation bias). It is possible that simply improving the sample purity may alleviate the orientation bias. Alternatively, detergents, graphene oxide (Pantelic et al., 2010), and different glow-discharging conditions can be tried.

**III.2.8 DISCUSSION**

In summary, I have developed a method for purification of native ribosomal complexes that has enabled me to obtain high-resolution cryo-EM reconstructions of the HCV IRES in complex with the 80S ribosome, including a previously unseen conformation of the IRES in an on-pathway state. This conformation demonstrates that, upon entering the elongation phase of translation, the IRES dissociates from the
ribosome in a stepwise manner, with domain II detaching before domain III. Although my data agree with the recently published structures from the research groups of Nenad Ban and Christian Spahn, I was several months behind these groups in terms of progress towards publication and therefore decided to focus on the 48S* initiation complex while also working on the ArfA project (described in Part II). I have made some initial progress towards obtaining a structure of a 48S* complex, but much optimization remains to be done.

In working towards these goals, I have also developed a simple method for rapid, nondenaturing purification of \textit{in-vitro}-transcribed RNA on a large scale, and have gained experience working with mammalian cell lysates. I hope to apply what I have learnt to develop a human \textit{in-vitro}-translation system and an improved method for purifying native initiation complexes.

**III.2.8.1 Remaining questions regarding the HCV IRES**

Although the structure of the HCV IRES has now been determined at high resolution, there are still a number of important questions relating to its structure and mechanism. Many of these relate to its function in the 48S* initiation complex. How does domain II of the IRES load the mRNA into the channel and displace eIF3j and eIF1 (Fraser \textit{et al.}, 2009; Filbin \textit{et al.}, 2011)? How is the clash between eIF2α (Hussain \textit{et al.}, 2014; Llácer \textit{et al.}, 2015) and subdomain IIb (Quade \textit{et al.}, 2015; Yamamoto \textit{et al.}, 2015) resolved? Does eIF3 totally detach from the 40S subunit and, if so, how does this occur (Hashem \textit{et al.}, 2013a)? And how do eIF1A and eIF5B interact with the initiation complex in the context of the IRES (Jaafar \textit{et al.}, 2016)?

Furthermore, the HCV IRES remains useful as a tool to capture several of the mammalian initiation factors in a functional complex, with eIF3 in particular still lacking a complete high-resolution structure. The 48S* complex is therefore an important target for structural analysis. Furthermore, \textit{in-vitro}-translation reactions in ribosome-depleted lysate, with 40S subunits added back, would presumably result in assembly of a mature initiation complex containing eIF5B, which has only ever been observed in isolation (Kuhle & Ficner, 2014a; Kuhle & Ficner, 2014b; Zheng \textit{et al.}, 2014) or in the context of the 80S ribosome (Fernández \textit{et al.}, 2013; Yamamoto \textit{et al.}, 2014). It may also be possible to capture earlier steps in the initiation pathway using mutants of the HCV IRES (see table III.2).
Little is known about how the HCV IRES functions in the context of the 80S ribosome, as most mutations tend to stall initiation at the earlier 48S* stage (Fraser et al., 2009). The only paper on the IRES that contains biochemical data pertaining to the 80S ribosome shows that subdomain IIb affects the first translocation step of the elongation phase (see section III.1.2.2 in the Introduction) (Filbin et al., 2013), but nothing is known beyond this. In my maps and in published structures, domain II remains conformationally dynamic due to interactions with the Met-tRNA$_{\text{AUG}}$, but my data seem to conflict with those of Yamamoto et al. (2015). It may therefore be worth improving the resolution of the two on-pathway states that I have captured, and also obtaining structures of 80S complexes with a mutant IRES in which the apical loop of subdomain IIb has been disrupted.

**III.2.8.2 Developing an *in-vitro*-translation system**

In order to tackle these and other projects, it would be extremely useful to have an *in-vitro*-translation system that accurately recapitulates translational initiation. This is especially important because cell lysates in general, and RRL in particular, are poor systems for studying the molecular interactions of translational initiation. There are a number of reasons for this. First and foremost, translation in nuclease-treated lysates shows little to no dependence on the presence of a 5′ cap or 3′ polyadenosine tail, regardless of the mRNA being tested (Terenin et al., 2017). Inhibition of endogenous cap-dependent initiation is known to cause an increase in translational activity from uncapped mRNA, probably due to relieved competition for the eIF4F complex and other initiation factors. Thus, removing the background protein synthesis by nuclease treatment has the unintended side effect of enhancing cap-independent initiation (Soto Rifo et al., 2007). *In-vitro*-translation assays in nuclease-treated lysates cannot therefore be used to demonstrate cap or end independence, as every mRNA shows IRES-like activity to some extent (Terenin et al., 2017). Furthermore, reticulocytes are highly specialized cells that lack nuclei and synthesize vast quantities of only a few proteins (primarily globins). None of their endogenous mRNAs have especially long or structured 5′ UTRs, and exogenous constructs with such characteristics are poorly translated (Andreev et al., 2009; Dmitriev et al., 2009). Indeed, scanning is impaired in various lysates (Vassilenko et al., 2011), though most severely in RRL.
Another problem relates to the concentrations of monovalent and divalent cations. The intracellular concentration of magnesium cations is generally estimated to be between 0.1 and 0.9 mM (Garfinkel et al., 1986; Zhang et al., 2011; Tashiro et al., 2014), depending on the species, cell type, and other conditions. The concentration in RRL from commercial suppliers is unknown, but the final concentration in my in-vitro-translation reactions prepared with cT2 is estimated to be ~3 mM (though the addition of ~1 mM each of ATP and GTP during preparation would reduce the concentration of free cations; see section III.3.5 in the Materials and methods). When combined with the low concentration of potassium cations (~80 mM), this presumably favours spontaneous association of ribosomal subunits and off-pathway interactions between the mRNA and ribosomes (Terenin et al., 2017), as implied by the presence of off-pathway states in my cryo-EM datasets. High concentrations of magnesium (above ~4 mM) actually inhibit both the canonical and IRES-directed mechanisms of translational initiation by restricting the conformational dynamics of the ribosome (Shenvi et al., 2005). The buffer conditions of cT2 have been optimized for protein yield, but a large proportion of synthesized proteins may therefore result from noncanonical initiation.

Human cytoplasmic extracts are apparently less susceptible to some of these problems and are a better choice for recapitulating proper translational initiation (Terenin et al., 2017). HEK cells, for example, have minimal specialization and synthesize a range of different proteins. It should be possible to preserve cap-dependent initiation in HEK-cell lysate by keeping the salt concentrations at a more physiological level and skipping the nuclease treatment (Soto Rifo et al., 2007). The resultant in-vitro-translation system would be less efficient for protein synthesis due to competition with endogenous mRNAs, but might be a better choice for biochemical and structural analysis of natively assembled initiation complexes.

III.2.8.3 Purifying native ribosomal complexes

Having set up a highly controlled in-vitro-translation system with HEK-cell lysate, I intend to develop new methods for purifying native ribosomal complexes bound to the HCV IRES. To ensure that the complexes assemble in a stepwise manner, I plan to use ribosome-depleted HEK-cell lysate preincubated with 40S subunits before addition of the IRES mRNA (in excess relative to the 40S subunits). The resultant
48S* complexes could then be isolated as they are or incubated with 60S subunits in the presence of cycloheximide to assemble 80S complexes. Samples would then be purified by polyethylene-glycol precipitation or pelleting through a sucrose cushion. This method should be sufficient to enrich for the desired complexes and would provide a much higher yield than affinity purification. The presence of IRES mRNA in the resuspended pellet can be tested by phenol extraction, and I can use nuclease-treated lysate if competition with endogenous mRNA is a problem.

Pelleting is a risky strategy for purifying 48S* complexes due to their fragility. If this turns out to be a problem, it will be necessary to develop a more efficient affinity method. A strategy that can be tried immediately is to use an oligo with a desthiobiotin tag, which binds Strep-Tactin resin and can be eluted with free biotin in the elution buffer (Kohler et al., 2017). Alternatively, it may be possible to attach a protein-based tag to the mRNA. The aim would be to elute the 48S* complexes at a concentration high enough to prepare grids for cryo-EM.
III.3 MATERIALS AND METHODS

III.3.1 BUFFERS AND REAGENTS

The pH value of each buffer was adjusted using the appropriate acid or alkali. For an overview of the buffer stocks, see table III.5.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Acid / alkali</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDTA</td>
<td>NaOH</td>
<td>8.0</td>
</tr>
<tr>
<td>EGTA</td>
<td>NaOH</td>
<td>8.0</td>
</tr>
<tr>
<td>HEPES</td>
<td>KOH</td>
<td>7.4</td>
</tr>
<tr>
<td>KOAc</td>
<td>HOAc</td>
<td>7.5</td>
</tr>
<tr>
<td>Mg(OAc)_2</td>
<td>HOAc</td>
<td>7.5</td>
</tr>
<tr>
<td>NTP</td>
<td>NaOH</td>
<td>7.0</td>
</tr>
<tr>
<td>tris</td>
<td>HCl</td>
<td>7.6 / 8.0</td>
</tr>
<tr>
<td>Zn(OAc)_2</td>
<td>HOAc</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Table III.5 | Buffers and reagents with set pH values. Zinc cations were kept in an aqueous solution of Zn(OAc)_2 at pH 5.0 because Zn(OH)_2 is insoluble at neutral pH and the tetrahydrate ZnCl_2(H_2O)_4 precipitates from aqueous solutions of ZnCl_2.

III.3.2 PURIFICATION OF mRNA CONSTRUCTS

Plasmids were synthesized (Integrated DNA Technologies) and transformed into *E. coli* DH5α. Cultures were grown at 37 °C overnight in 3 l of LB medium with 50 µg·ml⁻¹ kanamycin. Each plasmid was prepared using a QIAGEN Plasmid Giga Kit, linearized at 37 °C overnight using EcoRV, and purified from the reaction mixture by phenol extraction and ethanol precipitation.

Stocks of 10× transcription buffer (400 mM tris–HCl (pH 8.0), 10 mM spermidine, 0.1% Triton X-100, 50 mM dithiothreitol) and 100 mM of each NTP–NaOH (pH 7.0) were stored at −80 °C. *In-vitro*-transcription reactions (1× transcription buffer, 30 µg·ml⁻¹ linearized plasmid, 6 mM ATP–NaOH, 6 mM CTP–NaOH, 6 mM GTP–NaOH, 6 mM UTP–NaOH, 30 mM MgCl_2, 15 µg·ml⁻¹ inorganic pyrophosphatase, 20
µg·ml⁻¹ T7 RNA polymerase) were prepared in volumes of 5 ml and incubated at 37 °C for 4 h before being flash-frozen in liquid nitrogen and stored at −80 °C overnight. Each reaction mixture was defrosted in a water bath at 20 °C and centrifuged in an A-4-81 rotor (Eppendorf) at 3,250 ×g for 10 min to remove precipitated pyrophosphate.

In the original protocol, the supernatant was mixed with loading dye (80% methanamide, 10 mM EDTA–NaOH (pH 8.0), 1 mg·ml⁻¹ bromophenol blue, 1 mg·ml⁻¹ xylene cyanol) and loaded into four large urea–polyacrylamide gels. Each gel was electrophoresed at 37 W for 6 h. Ultraviolet light was used to visualize the band containing the transcripts. The band was excised, cut into small pieces, and stored at −20 °C overnight. The pieces were then placed in an Elutrap (GE Healthcare) filled with TBE buffer (90 mM tris, 90 mM boric acid, 2 mM EDTA–NaOH) and electrophoresed at 200 V for 6 h at 4 °C. Each hour, the current was reversed for 30 s, eluate was transferred from the trap into a fresh tube, and the trap was refilled with TBE buffer. The eluate was dialysed overnight against dialysis buffer (10 mM NH₄OAc–HOAc (pH 5.0), 50 mM KCl) and concentrated using a Centricon tube (Merck Millipore) with a cutoff of 30 kDa. The eluate was then incubated in a water bath at 90 °C for 1 min, 60 °C for 15 min, and slowly cooled to 30 °C. MgCl₂ was then added to a final concentration of 5 mM and the eluate incubated at 30 °C for another 15 min. Aliquots were flash-frozen in liquid nitrogen and stored at −80 °C.

In the new protocol, the transcripts were purified by acidic phenol extraction, gel filtration using a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated with gel-filtration buffer (10 mM NH₄OAc–HOAc (pH 5.0), 150 mM KCl, 5 mM MgCl₂), and ethanol precipitation. Each transcript was resuspended in storage buffer (10 mM NH₄OAc–HOAc (pH 5.0), 50 mM KCl, 5 mM MgCl₂) and diluted to a final concentration of 25 µM. Aliquots were flash-frozen in liquid nitrogen and stored at −80 °C.

**III.3.3 STREPTOTAG PURIFICATION OF 48S* COMPLEXES**

Dihydrostreptomycin was coupled to epoxy-activated Sepharose 6B (GE Healthcare) as described (Bachler et al., 1999; Windbichler & Schroeder, 2006).
Purification of 48S* complexes was performed as described (Locker et al., 2006; Locker & Lukavsky, 2007), with a few minor changes. 1 ml of untreated RRL (Promega) was diluted with 3 ml of binding buffer (20 mM tris–HCl (pH 7.6), 120 mM KCl, 10 mM MgCl₂, 8% sucrose), 2.5 µl of RNasin Plus (Promega), and half a tablet of cOmplete EDTA-free protease-inhibitor cocktail (Roche). The mixture was incubated at 37 °C for 10 min. Puromycin dihydrochloride was then added to a final concentration of 1 mM and the mixture was incubated on ice for 10 min and at 37 °C for 10 min. GDPNP was added to a final concentration of 2 mM and the mixture was incubated at 37 °C for 5 min. Lastly, IRES mRNA was added to a final concentration of 1 µM and the mixture was incubated at 37 °C for another 10 min.

1 g of dihydrostreptomycin-coupled Sepharose 6B (GE Healthcare) was loaded into a column and attached to an ÄKTApurifier (GE Healthcare) in a cold room. The column was equilibrated with binding buffer. The in-vitro-translation mixture was loaded onto the column and washed through until a stable baseline was reached. Elution buffer (the same as binding buffer, plus 50 mM streptomycin) was then passed through the column. Fractions corresponding to the eluate peak were combined and ultracentrifuged in a 90Ti rotor (Beckman Coulter) at 173,000 ×g for 16 h at 4 °C. The pellet was resuspended in resuspension buffer (20 mM tris–HCl (pH 7.6), 100 mM KOAc, 200 mM KCl, 2.5 mM MgCl₂, 2 mM dithiothreitol) and loaded into a sucrose gradient (15 to 40%) prepared in resuspension buffer. The gradient was ultracentrifuged in an SW-28 rotor (Beckman Coulter) at 87,000 ×g for 16 h at 4 °C. The fractions containing putative 48S* complexes were pooled and concentrated using a Centricon tube (Merck Millipore) with a cutoff of 50 kDa to a final concentration of 10 A₂₆₀ units per ml. Aliquots were flash-frozen in liquid nitrogen and stored at −80 °C.

III.3.4 PREPARATION OF CT2

Dr. R.S. Hegde prepared a stock of mammalian tRNA from liver and an amino-acid mixture containing nineteen of the standard amino acids (except l-methionine), each at a concentration of 1 mM. Subsequently, 50 ml of RRL (Green Hectares) was incubated with 0.5 ml of 15 U·µl⁻¹ micrococcal nuclease and 0.5 ml of 100 mM
CaCl\textsubscript{2} for 12 min at 25 °C in a water bath. The reaction was stopped with 0.5 ml of 200 mM EGTA–NaOH. 21 ml of this mixture was added to 1.2 ml of 1 M HEPES–KOH (pH 7.4), 600 µl of 5 M KOAc–HOAc, 120 µl of 1 M MgCl\textsubscript{2} in a fresh tube. The following reagents were then added in sequential order: 2.4 ml of amino-acid mixture, 0.6 ml of 1.2 M creatine phosphate, 0.6 ml of 100 mM ATP–NaOH, 0.6 ml of 100 mM GTP–NaOH, 0.24 ml of 250 mM glutathione, 0.18 ml of 100 mM spermidine, 0.12 ml of 20 mg·ml\textsuperscript{-1} creatine kinase, and 0.3 ml of 20 mg·ml\textsuperscript{-1} tRNA, to a final volume of 30 ml.

The final mixture of cT2 is used as a 2× stock and contains ~70% nuclease-treated RRL, 40 mM HEPES–KOH (pH 7.4), 100 mM KOAc–HOAc, 4 mM MgCl\textsubscript{2}, 80 µM amino acids (except L-methionine), 24 mM creatine phosphate, 2 mM ATP–NaOH, 2 mM GTP–NaOH, 2 mM glutathione, 600 µM spermidine, 80 µg·ml\textsuperscript{-1} creatine kinase, and 200 µg·ml\textsuperscript{-1} tRNA. Aliquots were flash-frozen in liquid nitrogen and stored at −80 °C.

In an in-vitro-translation reaction using cT2, the final concentrations of potassium and magnesium cations are estimated to be ~80 mM and ~2.3 mM, respectively, including endogenous cations from the RRL. Assuming that each NTP chelates one magnesium cation, the concentration of free magnesium cations will be ~0.3 mM, although additional MgCl\textsubscript{2} was added for the reactions (see section III.3.5).

**III.3.5 Biotin purification of 80S complexes**

A biotin-tagged oligo, fully 2’-O-methylated to protect against nucleolytic degradation, was synthesized (Integrated DNA Technologies). The 5’ end was connected to a biotinyl moiety via a triethylene-glycol linker to more easily access the biotin-binding site of streptavidin. Annealing mixtures (60 µM oligo, 15 µM IRES mRNA, 20 mM tris–HCl (pH 7.6), 100 mM KCl, 5 mM MgCl\textsubscript{2}) were incubated at 70 °C for 2 min, cooled to 37 °C at 0.1 °C·s\textsuperscript{-1}, held at 37 °C for 2 min, and cooled to 4 °C at 0.1 °C·s\textsuperscript{-1}.

*In-vitro*-translation mixtures (50% cT2, 1 mM cycloheximide, 1 mM dithiothreitol, 20 U·µl\textsuperscript{-1} SUPERase-In (Thermo Fisher Scientific)) were supplemented with an
additional 0.75 mM MgCl$_2$ and 0.1 mM spermidine, to a final volume of 3.9 ml. This mixture was incubated with 100 µl of annealing mixture at 32 °C for 10 min and on ice for 10 min before being layered onto 8 ml of sucrose cushion (1 M sucrose, 30 mM HEPES–KOH (pH 7.4), 100 mM KCl, 5 mM MgCl$_2$, 0.2 mM spermidine, 2 mM dithiothreitol, 1 mM cycloheximide, cOmplete EDTA-free protease-inhibitor tablets (Roche)). The cushion was ultracentrifuged at 200,000 ×g for 16 h at 4 °C and the pellet resuspended in HKMD buffer (30 mM HEPES–KOH (pH 7.4), 100 mM KCl, 5 mM MgCl$_2$, 2 mM dithiothreitol).

Streptavidin-coated MagneSphere paramagnetic beads (Promega) were washed with 100 mM sodium phosphate (pH 7.0) and HKMD buffer and incubated with the resuspended pellet for 1 h on ice. The beads were washed again in HKMD buffer and incubated with 400 µl of 10 µM competitive oligo in HKMD buffer for 5 min at 37 °C, 5 min at 20 °C and 10 min on ice. The eluate was then diluted to 1 ml in HKMD buffer and ultracentrifuged at 100,000 ×g for 40 min at 4 °C. The pellet was resuspended to a final concentration of ~15 $A_{260}$ units per ml.

**III.3.6 DATA COLLECTION**

QUANTIFOIL R 2/2 grids were coated with a continuous carbon film (~60 Å in thickness) and glow-discharged for 30 s at 6 V. Aliquots of 3 µl of the resuspended pellet were incubated on the grids for 30 s at 4 °C and 100% humidity. The grids were then blotted for 2 to 3 s and flash-frozen in liquid ethane using a Vitrobot Mark III (FEI).

Micrographs were collected using a Titan Krios at 300 keV with a Falcon II detector and EPU software (all FEI). The calibrated magnification of 104,478× gave a pixel size of 1.34 Å. Details for the data collection are presented in table III.3.

**III.3.7 DATA PROCESSING**

The movie frames of each image were aligned using MotionCorr (Li et al., 2013). Parameters for the contrast transfer function were obtained using Gctf (Zhang, 2016)
for each motion-corrected micrograph. All subsequent processing was performed in RELION-2.0 (Fernández-Leiro & Scheres, 2016). Semiautonomous particle picking (Scheres, 2015) was used to select ribosomes from the micrographs. Nonribosomal particles were discarded after reference-free 2D classification. A cryo-EM map of the 80S ribosome with aminoacyl-tRNA and eEF1A in the A site (EMD-4130), low-pass-filtered to 40 Å, was used as an initial reference for the first 3D refinement. An initial round of 3D classification without alignment was used to isolate three IRES-containing classes (eEF2-bound, nonrotated, and rotated) and discard free 60S subunits, empty 80S ribosomes, and poorly aligned particles (figure III.8). After another refinement, particle polishing was performed for each IRES-containing class (as described in section II.3.3). Two rounds of FCwSS (Bai et al., 2015) were then performed to discard empty ribosomes (figure III.8). The first round was focused on the HCV IRES and led to the isolation of a fourth class with the ribosome in a rolled conformation. The second round was focused on the P/P- or P/E-site tRNA_AUG. No further classification was done after this step, as the resolution of each class was limited by the number of particles, rather than conformational heterogeneity.

The final statistics for each dataset are provided in table III.3. Reported resolutions are based on the FSC=0.143 criterion (Rosenthal & Henderson, 2003). High-resolution noise substitution was used to correct for the effects of applying a mask during the FSC calculations (Chen et al., 2013). Before visualization, density maps were corrected for the modulation transfer function of the Falcon II detector and sharpened by applying a negative B-factor that was estimated using automated procedures (Rosenthal & Henderson, 2003).

### III.3.8 Purification of Human Cytosolic Ribosomal Subunits

Previous and ongoing projects in the laboratory have used a tetracycline-resistant strain of HEK cells (HEK293S TetR GnT1−) that lacks N-acetyl-β-D-glucosaminyltransferase I (Brown et al., 2014; Amunts et al., 2015). This cell line is optimized for large-scale suspension growth in Wave bioreactors (GE Healthcare), making it a good source of human mitochondria as well as cytosolic ribosomes.
HEK293S cells were grown in 10 l of FreeStyle 293 Expression Medium (Life Technologies) supplemented with 5% fetal bovine serum (Life Technologies). Cultures were centrifuged in a JLA-8.1000 rotor (Beckman Coulter) at 1,000 ×g for 10 min. Cell pellets were resuspended in phosphate-buffered saline, centrifuged again at 1,000 ×g for 10 min, resuspended in lysis buffer (50 mM HEPES–KOH (pH 7.4), 10 mM KCl, 1.5 mM MgCl₂, 1 mM dithiothreitol, 1 mM EDTA–NaOH, 1 mM EGTA–NaOH, cOmplete EDTA-free protease-inhibitor tablets (Roche)), and lysed using a glass homogenizer. The lysate was then stirred for 1 h at 4 °C and centrifuged in a JLA-8.1000 rotor (Beckman Coulter) at 800 ×g for 15 min at 4 °C. The supernatant was then subjected to three centrifugations in an SLA-3000 rotor (Thermo Fisher Scientific) at 4 °C, first at 1,000 ×g for 15 min, then at 10,000 ×g for 15 min, and finally at 20,000 ×g for 30 min. The supernatant was transferred to a fresh bottle after each centrifugation.

The final pellet contained mitochondria, which were further purified by Dr. N. Desai, while I purified cytosolic ribosomes from the supernatant. Fourteen sucrose cushions (1 M sucrose, 40 mM HEPES–KOH (pH 7.4), 150 mM KCl, 10 mM MgCl₂, 2 mM dithiothreitol, 8.5% D-mannitol, 500 µM EDTA–NaOH, cOmplete EDTA-free protease-inhibitor tablets (Roche)), each with a volume of 10 ml, were prepared and 50 ml of supernatant was layered onto each cushion. The cushions were ultracentrifuged at 230,000 ×g for 16 h at 4 °C and the pellets resuspended in gradient buffer (40 mM HEPES–KOH (pH 7.4), 150 mM KCl, 10 mM MgCl₂, 2 mM dithiothreitol) and diluted to ~300 U·ml⁻¹. The amount of material was too great to purify in one batch, so aliquots of ~600 A260 units were flash-frozen in liquid nitrogen and stored at −80 °C.

Three aliquots were defrosted, loaded onto six sucrose gradients (15 to 40%) prepared in gradient buffer, and ultracentrifuged in an SW-28 rotor (Beckman Coulter) at 80,000 ×g for 16 h at 4 °C. The fractions containing 80S ribosomes were pooled, diluted to 336 ml with gradient buffer and ultracentrifuged in a 45Ti rotor (Beckman Coulter) at 235,000 ×g for 16 h at 4 °C. The pellets were resuspended in gradient buffer to a final concentration of ~200 U·ml⁻¹.

The conditions for subunit separation were based on described protocols (Blobel & Sabatini, 1971) and further optimized for large-scale purification. A reaction mixture
for subunit separation (100 U·ml⁻¹ ribosomes, 40 mM HEPES–KOH (pH 7.4), 500 mM KCl, 3 mM MgCl₂, 2 mM dithiothreitol, 2 mM puromycin dihydrochloride) was incubated at 32 °C for 10 min and on ice for 10 min before being loaded onto six sucrose gradients (10 to 35%) prepared in separation buffer (40 mM HEPES–KOH (pH 7.4), 500 mM KCl, 3 mM MgCl₂, 2 mM dithiothreitol) and ultracentrifuged in an SW-28 rotor (Beckman Coulter) at 110,000 ×g for 16 h at 4 °C.

The fractions containing 40S and 60S subunits were pooled separately into 45Ti tubes, topped up with subunit buffer (20 mM HEPES–KOH (pH 7.4), 50 mM KCl, 5 mM MgCl₂, 2 mM dithiothreitol), and ultracentrifuged in a 45Ti rotor (Beckman Coulter) at 235,000 ×g for 16 h at 4 °C. Each pellet was resuspended in subunit buffer to a final concentration of 200 U·ml⁻¹. Aliquots were flash-frozen in liquid nitrogen and stored at −80 °C.

### III.3.9 Biotin Purification of 48S* Complexes

A biotin-tagged DNA oligo, with a 5′ triethylene-glycol linker, was synthesized (Integrated DNA Technologies). 100 µl of annealing mixture (10 µM oligo, 10 µM IRES mRNA) in storage buffer was incubated at 70 °C for 2 min, cooled to 37 °C at 0.1 °C·s⁻¹, held at 37 °C for 2 min, and cooled to 4 °C at 0.1 °C·s⁻¹. The mixture was then diluted with 300 µl of HKMZD buffer (40 mM HEPES–KOH (pH 7.4), 50 mM KCl, 3 mM MgCl₂, 10 µM Zn(OAc)₂–HOAc, 2 mM dithiothreitol).

Ribosome-depleted cT2 was prepared by ultracentrifuging 1 ml of cT2 in a TLA-120.2 rotor (Beckman Coulter) at 500,000 ×g for 1 h at 4 °C. The supernatant was transferred to a fresh tube and used to prepare an in-vitro-translation mixture (50% cT2 supernatant, 50 µM Zn(OAc)₂–HOAc, 5 U·ml⁻¹ 40S subunits, 80 µM L-methionine).

600 µg of streptavidin-coated MagneSphere paramagnetic beads (Promega) were washed with HKMZD buffer and incubated with the diluted annealing mixture on ice for 30 min. The beads were washed again with HKMZD buffer, with each supernatant being discarded. The beads were then incubated in the in-vitro-translation mixture at 32 °C for 5 min and on ice for 10 min. The beads were washed again and before being
incubated with 1 U·ml\(^{-1}\) DNase I in HKMZCD buffer (the same as HKMZD buffer, plus 2 mM CaCl\(_2\)) on ice for 1 h. The eluate was transferred to a fresh tube and immediately used to prepare grids for cryo-EM, as described in section III.3.6.
III.4 Bibliography


Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome


Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome


Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome


Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome


Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome


Part III: Structural analysis of the HCV IRES bound to the mammalian ribosome


Zheng A., Yu J., Yamamoto R., Ose T., Tanaka I., Yao M. (2014) X-ray structures of eIF5B and the eIF5B–eIF1A complex: the conformational flexibility of eIF5B is restricted on the ribosome by interaction with eIF1A. Acta Cryst. D70: 3090–3098